

# STATISTIČKE METODE U KINEZILOGIJI

prof.dr.sc. Miljenko Marušić

Kontakt: [miljenko.marusic@math.hr](mailto:miljenko.marusic@math.hr)

# Literatura

## Osnovna literatura:

W.J. Vincent, J.P. Weir, Statistics in kinesiology, 4th edition, Human Kinetics, Champaign, IL, USA, 2012

## Dopunska literatura:

- 1 B. Petz, Osnovne statističke metode za nematematičare, 3. dopunjeno izdanje, Naklada Slap, Jastrebarsko, 1997.
- 2 Pivac, B. Šego, Statistika, Alka Script
- 3 S. Siegel i N.J. Castellan, Jr., Nonparametric Statistics for the behavioral sciences, 2nd edition, McGraw-Hill, 1988.
- 4 N. Viskiće-Štalec. Elementi faktorske analize, Fakultet za fizičku kulturu, Zagreb, 1991.
- 5 A. Fulgosi, Faktorska analiza, 2. dopunjeno izdanje, Školska knjiga, Zagreb, 1984.
- 6 D.W. Hosmer, S. Lemeshow, Applied logistic regression, Wiley, New York, 1989.

## Zadaci statistike:

- prikupljanje podataka,
- uređivanje i sažimanje,
- analiza podataka
- zaključivanje na osnovu prikupljenih podataka

## Ciljevi kolegija

- Sposobnost čitanja i razumijevanja raznih statističkih studija. Za razumijevanje ovakvih izvještaja osoba mora biti upoznata s riječnikom, simbolima, konceptima i postupcima korištenim u izvještaju.
- Sposobnost sažimanja podataka, donošenje generalnih zaključaka na osnovu rezultata istraživanja.
- Cilj doktorskog studija je osposobljavanje za znanstveni rad. Budući da su statistički postupci osnova većine istraživanja u kineziologiji, poznavanje statistike je neophodno.

# Osnovni pojmovi.

**Varijabla** ili **obilježje** je karakteristika (svojstvo, atribut) koja može poprimiti različite vrijednosti.

Primjeri varijable (obilježja):

- visina (osobe)
- težina (osobe)
- indeks tjelesne težine (BMI)
- duljina skoka
- vrijeme (skijaša u slalomu)
- broj pobjeda na turniru
- ocjena na ispitu
- klub natjecatelja

**Podatak** je izmjerena vrijednost varijable.

Skup svih podataka čini **skup podataka** (*engl.* data set)

Primjer:

- Marko je 100 m pretrčao za 11 s.

Obilježje: vrijeme na 100 m

Podatak: 11 s

- Za drugog natjecatelja vrijednost obilježja može biti drugačije.

Ivan je 100 m pretrčao za 10.5 s.

Obilježje: vrijeme na 100 m

Podatak: 10.5 s

- Andrija je na utakmici pretčao 13 km.

Obilježje: Pretrčani put na utakmici

Podatak: 13 km



**Populacija** je skup jedinki koje imaju neko zajedničko svojstvo a koje su predmet proučavanja.

Definicija populacije ovisi o istraživanju.

### **Primjer.**

Zanima nas broj učenika neke škole koji se dodatno bave sportskom aktivnošću.

Rezultat: za svakog učenika provjeriti da li se bavi sportskom aktivnošću. Prebrojiti sve takve učenike.

Populacija: Svi učenici dotične škole.

Obilježje: Bavljenje dodatnom sportskom aktivnošću.

Podatak: da/ne

**Uzorak** je (bilo koji) podskup populacije.

### **Primjer.**

Populacija: Svi učenici neke škole.

Uzorak:

- Svi učenici prvih razreda.
- Svi učenici sedmih razreda.
- Svi učenici 3.a razreda.

**Element skupa** (jedinica, entitet, objekt) je osnovni element populacije.

**Statistički skup** = populacija ili uzorak.

**Opseg skupa** - broj elemenata, jedinica, u skupu.

**Deskriptivna statistika** se sastoji od prikupljanja, organizacije, sažimanja i prikaza podataka.

**Inferencijalna statistika (statističko zaključivanje)** se sastoji od generalizacije s uzorka na populaciju, testiranja hipoteza, određivanja veza između varijabli i predviđanja.

Deskriptivna statistika - Zaključujemo o skupu o kojem imamo podatke.

Inferencijalna statistika - Na temelju podataka o manjem skupu (uzorku) donosimo zaključke o svojstvima šireg skupa (populaciji).

## Deskriptivna statistika

Zaključujemo o skupu o kojem imamo podatke.

### Primjer.

Od 920 učenika Osnovne škole X, njih 230 je uključeno u izvannastavne sportske aktivnosti. Dakle, 25% učenika je uključeno u izvannastavne sportske aktivnosti.

- Populacija: svi učenici Osnovne škole X;
- Za svakog je učenika poznato je li uključen u izvannastavnu aktivnost ili nije.

## Inferencijalna statistika

Na temelju podataka o manjem skupu (uzorku) donosimo zaključke o svojstvima šireg skupa (populaciji).

### Primjer.

Od 30 učenika 2.a razreda Osnovne škole X, njih 12 je uključeno u izvannastavne sportske aktivnosti. Dakle, 40% učenika Osnovne škole X je uključeno u izvannastavne sportske aktivnosti.

- Populacija: svi učenici Osnovne škole X;
- Uzorak: učenici 2.a razreda Osnovne škole X;
- Za svakog je učenika 2.a razreda poznato je li uključen u izvannastavnu aktivnost ili nije.
- Na osnovu podataka o jednom razredu zaključujemo o cijeloj školi.

# Klasifikacija varijabli (podataka)

## Kvalitativne (kategorijske) varijable

- vrijednost varijable se nalazi u točno jednoj kategoriji.

## Kvantitativne varijable

- vrijednost varijable je rezultat mjerenja na numeričkoj skali.

# Primjeri kvalitativnih varijabli:

## **zemlja porijekla natjecatelja:**

Austrija, Italija, Mađarska, Češka, Slovenija, Francuska, ...

## **spol**

muški, ženski

## **klub natjecatelja:**

Mladost, Dubrava, Primorje, ...

## **Atletska disciplina:**

100 m, 1000 m, skok u vis, skok u dalj, bacanje koplja, bacanje kladiva, ...



# Primjeri kvalitativnih varijabli:

## ocjena:

- nedovoljan (1)
- dovoljan (2)
- dobar (3)
- vrlo dobar (4)
- odličan (5)

## stupanj zadovoljstva :

- jako nezadovoljan
- nezadovoljan
- zadovoljan
- jako zadovoljan

## Primjeri kvantitativnih varijabli:

- broj natjecatelja
- kapacitet stadiona
- visina skoka
- duljina skoka
- broj postignutih koševa jednog igrača

## Podjela kvalitativnih varijabli:

**nominalne varijable** - ne postoji uređaj između kategorija

**redoslijedne (ordinalne, uređajne) varijable** - postoji uređaj između kategorija

Nominalne varijable: Nema bolje (gore) odnosno veće (manje) vrijednosti

# Primjeri nominalnih varijabli:

## **zemlja porijekla natjecatelja:**

Austrija, Italija, Mađarska, Češka, Slovenija, Francuska, ...

## **spol**

muški, ženski

## **klub natjecatelja:**

Mladost, Dubrava, Primorje, ...

## **Atletska disciplina:**

100 m, 1000 m, skok u vis, skok u dalj, bacanje koplja, bacanje kladiva, ...

## Redosljedna varijabla:

- Za bilo koje dvije vrijednosti možemo reći koja je bolja (gora) odnosno veća (manja)
- Postoji uređaj

## Primjer:

- stupanj zadovoljstva
- ocjena

# Podjela kvantitativnih varijabli:

**diskretne varijable** - skup vrijednosti je konačan ili prebrojiv

**neprekidne varijable** - poprimaju sve vrijednosti unutar nekih granica

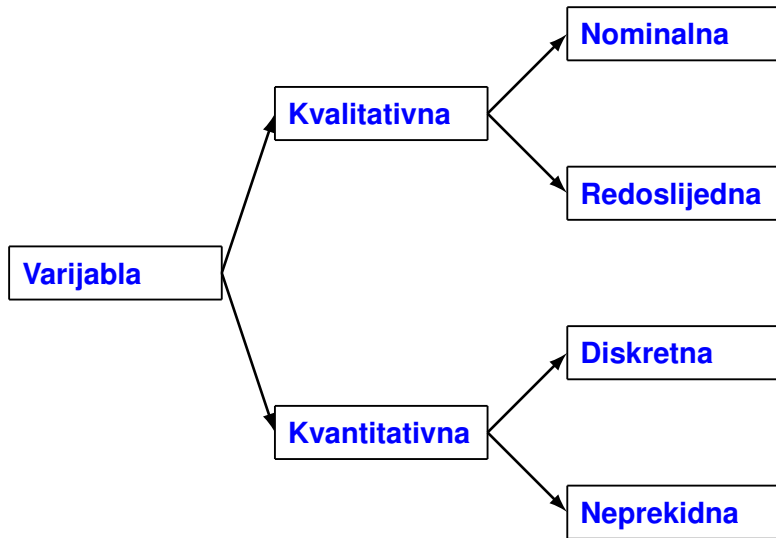
## Diskretne varijable

- Konačan skup vrijednosti  
Broj pogodaka u gađanju puškom
- Prebrojiv skup vrijednosti
  - npr. ukoliko vrijednost varijable može biti bilo koji cijeli broj
  - skup vrijednosti je beskonačan
- rezultat prebrajanja

## Neprekidne varijable

- visina skoka, duljina skoka, vrijeme
- varijabla može poprimiti bilo koju vrijednost unutar nekih granica
- rezultat mjerenja

# Klasifikacija varijabli.





## Primjer.

Marko je dao 23 koša.

Varijabla: Broj postignutih koševa.

Tip varijable: kvantitativna - diskretna

## Primjer.

Ivan je odvozio slalom za 55 s.

Varijabla: Vrijeme vožnje.

Tip varijable: kvantitativna - neprekidna

## Primjer.

Ivica je pretrčao maraton.

Varijabla: Pretrčan maraton.

Tip varijable: kvalitativna - nominalna

## Primjer.

Marica je za svoj nastup ocjenjena s ocjenom 8.

Varijabla: Ocjena nastupa.

Tip varijable: kvalitativna - redoslijedna

**Napomena.** Varijable mogu biti i kombinacija navedenih varijabli. Npr. skijaški skokovi - kombinacija duljine skoka i ocjene sudaca.

## Primjer.

Ivica je na cilj stigao šesti.

Varijabla: Plasman na utrci.

Tip varijable: kvalitativna - redoslijedna

**Napomena.** Iako je vrijednost numerička, radi se o kvalitativnoj varijabli.

# Frekvencije

Na zahtjev vlasnika fitness centra, voditelj je tijekom dana bilježio spol gostiju. Sljedeći dan je rezultat prosljedio svome nadređenom:

M M Ž Ž M Ž M M Ž M Ž M Ž M Ž Ž M Ž Ž M Ž Ž M Ž M Ž M Ž M Ž M Ž  
 M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M M M Ž M  
 Ž M Ž M Ž M Ž Ž M Ž M Ž M Ž Ž Ž M Ž M Ž Ž M Ž M Ž M Ž M Ž M Ž M  
 Ž Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž  
 M Ž M Ž Ž M Ž M Ž M Ž Ž M Ž M Ž M Ž Ž Ž Ž M Ž M Ž M Ž M Ž M Ž M  
 Ž M Ž M Ž M Ž M Ž Ž M Ž Ž M Ž M Ž Ž M Ž M Ž M Ž Ž M Ž Ž M Ž Ž M  
 Ž M Ž M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž  
 Ž M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž Ž M Ž Ž M Ž M Ž  
 M Ž M Ž M Ž Ž M Ž M Ž M Ž M Ž M Ž M Ž Ž M Ž M Ž Ž M Ž Ž M Ž M Ž  
 Ž M Ž M Ž M Ž M Ž M Ž M Ž

## Sirovi podaci

- originalni rezultat mjerenja

**nepregledno!**

## Niz podataka

- podaci posloženi po veličini (redu)





**Niz podataka** je još uvijek nepregledan.

Prebrojimo: U nizu se M pojavljuje 172 puta a Ž 225 puta.

**Frekvencija** je broj pojavljivanja određene vrijednosti varijable (u skupu podataka).

Frekvencija se ponekada naziva još i **apsolutna frekvencija**.

Oznaka:  $f_i$  - frekvencija  $i$ -te vrijednosti

U primjeru: 172 muškaraca i 225 žena

$$f_1 = 172, \quad f_2 = 225$$

opseg skupa = zbroj frekvencija ( $N = \sum f_i$ )

**Distribucija frekvencija** - niz vrijednosti obilježja i pripadajućih frekvencija.

**Primjer.** U prošlom primjeru bilo je 172 muškaraca i 225 žena.

Distribucija frekvencija:

Spol	frekvencija
M	172
Ž	225

**Relativna frekvencija** je omjer frekvencije i opsega (veličine) skupa:

$$r_i = \frac{f_i}{N}$$

Oznaka:  $r_i$  je relativna frekvencija  $i$ -te vrijednosti.

**Kumulativna frekvencija** je broj jedinki s vrijednošću obilježja manjim ili jednakim od pojedine vrijednosti obilježja.

postupno zbrajanje frekvencija:  $c_i = f_1 + f_2 + \dots + f_i$

frekvencije pozitivne  $\Rightarrow$  kumulativne frekvencije su rastuće  
(nepadajuće)

**Relativna kumulativna frekvencija** je omjer kumulativne frekvencije i veličine skupa.

Relativna kumulativna frekvencija je **udio** jedinki s vrijednošću obilježja manjim ili jednakim od pojedine vrijednosti obilježja.

**Primjer.**

Dob	Frekvencija	Relativna frekvencija	Kumulativna frekvencija	Relativna kumulativna frekvencija
19	1	0.05	1	0.05
20	0	0.00	1	0.05
21	5	0.25	6	0.30
22	2	0.10	8	0.40
23	2	0.10	10	0.50
24	1	0.05	11	0.55
25	4	0.20	15	0.75
26	1	0.05	16	0.80
27	0	0.00	16	0.80
28	1	0.05	17	0.85
29	0	0.00	17	0.85
30	1	0.05	18	0.90
31	1	0.05	19	0.95
32	1	0.05	20	1.00

# Prikazivanje podataka

- Tablično prikazivanje
- Grafičko prikazivanje

## Primjer.

### Podaci:

Spol	Dob	Spol	Dob
M	21	M	26
Ž	25	Ž	23
M	19	Ž	22
Ž	32	M	21
Ž	21	Ž	30
M	25	Ž	28
M	23	M	22
M	25	M	31
Ž	21	Ž	25
M	24	M	21

## Tablica.

Spol	f
Muški	11
Ženski	9

ili

Spol	f
Muški	11
Ženski	9
Ukupno	20

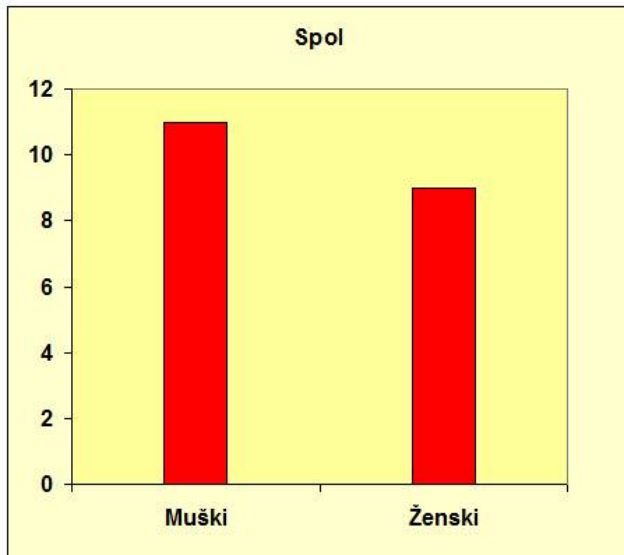


Spol	Frekvencija f	Relativna frekvencija r	Relativna frekvencija (% )
Muški	11	0.55	55
Ženski	9	0.45	45
Ukupno	20	1.00	100

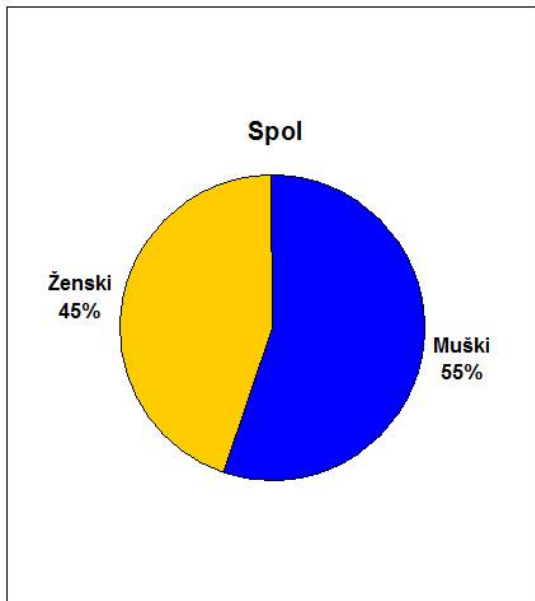
# Grafički prikaz podataka.

- jednostavnost
  - preglednost
  - ali ne i preciznost
- 
- površinski grafikoni
    - grafikon stupaca
    - grafikon krugova i polukrugova
  - linijski grafikoni
  - kartogrami
  - slikovni grafikoni (piktograf,..)

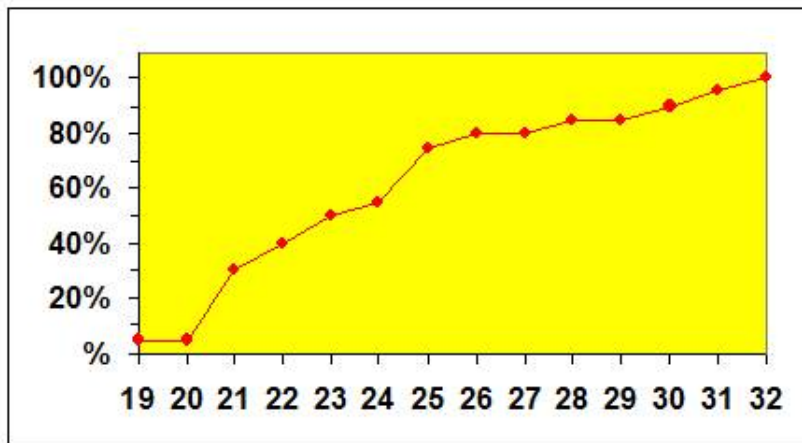
## Grafikon stupaca.



## Grafikon krugova (pita).



## Linijski grafikon.

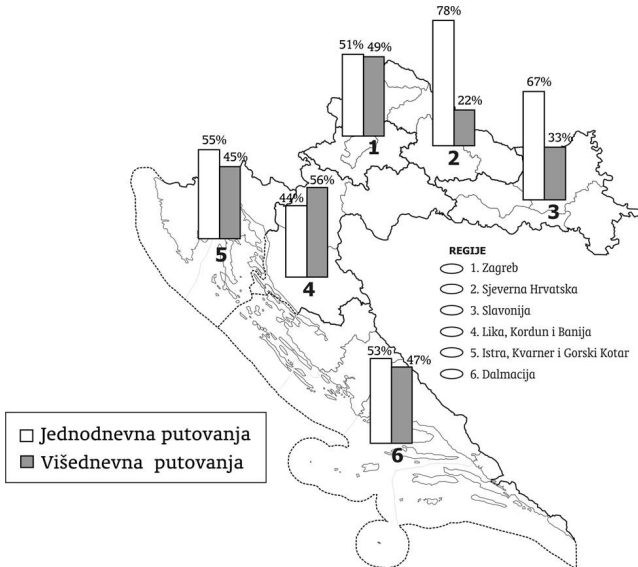


## Piktograf.



predstavlja 10 diplomiranih.

# Kartogram.



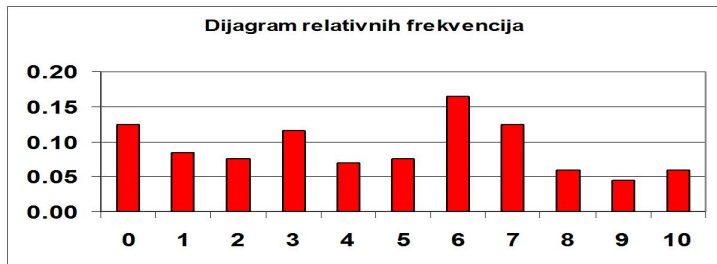
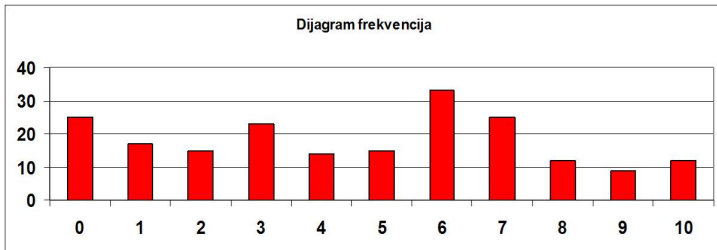
## Distribucija frekvencija prikazuje se:

- tablicama
- histogramima - površinski grafikon distribucije frekvencije
- poligonima frekvencija - linijski grafikon
- stepenasti linijski grafikon

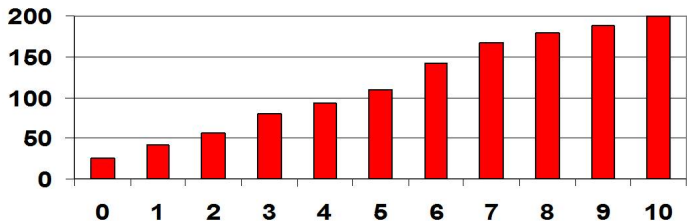


## Primjer.

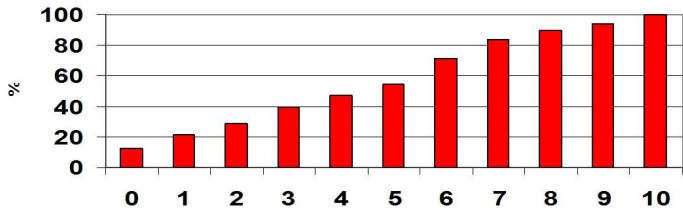
Bodovi na testu	Frekvencija	Kumulativna frekvencija	Relativna frekvencija	Relativna kumulativna frekvencija (%)
0	25	25	0.125	12.5
1	17	42	0.085	21.0
2	15	57	0.075	28.5
3	23	80	0.115	40.0
4	14	94	0.070	47.0
5	15	109	0.075	54.5
6	33	142	0.165	71.0
7	25	167	0.125	83.5
8	12	179	0.060	89.5
9	9	188	0.045	94.0
10	12	200	0.060	100.0
Ukupno	200		1.00	



Dijagram kumulativnih frekvencija



Dijagram relativnih kumulativnih frekvencija



## Primjer.

Bodovi na testu	Kumulativna frekvencija
0	25
1	42
2	57
3	80
4	94
5	109
6	142
7	167
8	179
9	188
10	200

Koliko je studenata dobilo 5 ili 6 bodova?

$$142 - 94 = 48$$

## Primjer.

Bodovi na testu	Relativna kumulativna frekvencija (%)
0	12.5
1	21.0
2	28.5
3	40.0
4	47.0
5	54.5
6	71.0
7	83.5
8	89.5
9	94.0
10	100.0

Koliko je studenata dobilo 7, 8 ili 9 bodova?

$$94.0\% - 71.0\% = 23.0\%$$

# Grupiranje podataka (razredi)

## Primjer.

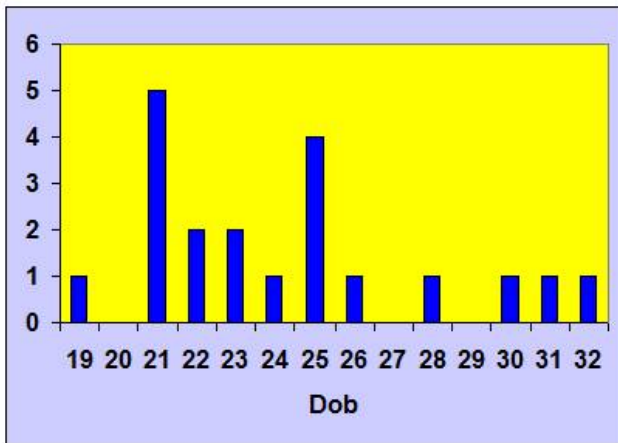
Varijabla DOB

- kvantitativna
- neprekidna

Distribucija frekvencija:

Dob	Frekvencija	Dob	Frekvencija
19	1	26	1
20	0	27	0
21	5	28	1
22	2	29	0
23	2	30	1
24	1	31	1
25	4	32	1

Distribucija frekvencija za varijablu dob:



Ukoliko je broj mogućih vrijednosti velik, podaci se grupiraju u razrede.

**Razredi** su intervali koji:

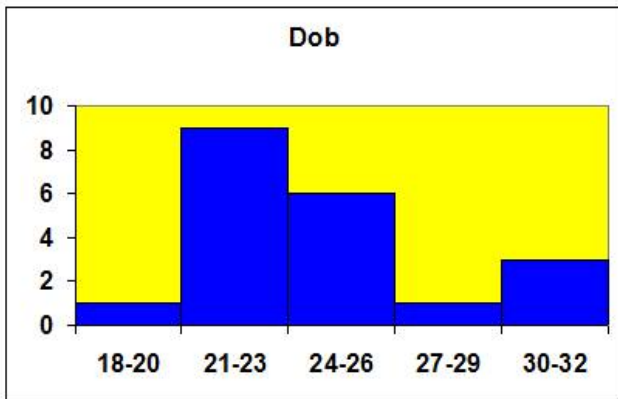
- potpuno pokrivaju cijeli skup mogućih vrijednosti obilježja (**načelo iscrpnosti** - svaki podatak se nalazi u nekom razredu)
- se ne preklapaju (**načelo isključivosti** - svaki podatak se nalazi samo u jednom razredu)

Razredi za varijablu DOB

- 18 – 20
- 21 – 23
- 24 – 26
- 27 – 29
- 30 – 32



Distribucija frekvencija za varijablu dob:



## Frekvencija razreda - broj jedinki u razredu

Razred	Frekvencija razreda
18 – 20	1
21 – 23	9
24 – 26	6
27 – 29	1
30 – 32	3

**Granice razreda** - najmanja i najveća moguća vrijednost koja pripada tom razredu Oznaka:

- $L_i$  donja granica  $i$ -tog razreda
- $U_i$  gornja granica  $i$ -tog razreda

Razred	Donja granica	Gornja granica
18 – 20	18	20
21 – 23	21	23
24 – 26	24	26
27 – 29	27	29
30 – 32	30	32

## Granice razreda

- **se ne preklapaju** kod diskretnog obilježja
- **se preklapaju** kod neprekidnog numeričkog obilježja (gornja granica razreda jednaka je donjoj granici sljedećeg razreda)

Ukoliko je podatak jednak granici razreda, dogovorno se pridružuje jednom razredu (zaokruživanje na dolje ili gore)

## Širina razreda

Ako se granice ne preklapaju, tada je širina  $i$ -tog razreda

$$i_i = L_{i+1} - L_i, i = 1, 2, \dots, k - 1$$

Širina zadnjeg razreda:

$$i_k = U_k - U_{k-1}.$$

ako se granice preklapaju, tada je širina  $i$ -tog razreda

$$i_i = U_i - L_i, i = 1, 2, \dots, k$$

( **Napomena**: formula vrijedi i za slučaj kada se granice ne preklapaju.)

Razredi mogu biti iste ili različite širine.

**Otvoreni razred** - razred bez gornje ili donje granice.

Ponekad zadnji razred nema gornje granice. (Rjeđi je slučaj da prvi razred nema donje granice.)

**Primjer.** Mjesečni prihod.

Razred
0 – 1.000 kn
1.000 – 3.000 kn
3.000 – 5.000 kn
5.000 – 10.000 kn
više od 10.000 kn

U zadnji razred spadaju sve vrijednosti veće od 10.000 kn i gornja granica razreda nije definirana

**Širina razreda** - aritmetička sredina gornje i donje granice razreda:

$$X_i = \frac{L_i + U_i}{2}, \quad i = 1, 2, \dots, k$$

Sredina razreda za otvoreni razred se procjenjuje na temelju poznavanja pojave.

Razred	Donja granica	Gornja granica	Širina razreda	Sredina razreda
18 – 20	18	20	3	19
21 – 23	21	23	3	22
24 – 26	24	26	3	25
27 – 29	27	29	3	28
30 – 32	30	32	3	31

**Korigirana frekvencija** - omjer frekvencije i širine razreda (ili neke druge prikladne veličine proporcionalne širini razreda)

Koriste se u slučaju nejednakih širina razreda.

Kod grupiranja u razrede **poželjno je**:

- koristiti 5 do 20 razreda
- svi razredi su iste širine (osim eventualno prvog i zadnjeg)
- broj razreda je neparan

Za mali skup vrijednosti obilježja:

- Vrijednosti se ne grupiraju
- Uređenjem se ne gubi informacija
- Uvid u sve vrijednosti skupa



## ZADATAK.

Izmjerena je visina učenika jednog razreda. Dobivene vrijednosti su (izražene u cm):

143	156	156	163	167
142	171	170	169	164
138	158	160	162	164
173	157	158	159	160
138	172	166	166	159
120	125	165	136	168

Grupirajte podatke u 6 razreda i odredite distribuciju frekvencija razreda.

Distribuciju frekvencija razreda prikažite tablično (prikažite apsolutne i relativne frekvencije te kumulativne i relativne kumulativne frekvencije) i grafički koristeći histogram, poligon frekvencija i ogivu.

Odredite širinu i sredinu razreda.

# Opis podataka

**CILJ:** sažeto opisati svojstva podataka.

- **Mjere centralne tendencije**
- **Mjere raspršenosti (varijacije)**
- **Mjere asimetrije**
- **Mjere položaja**

# Mjere centralne tendencije

- cilj je odrediti broj oko kojeg se grupiraju podaci  
→ mjere centralne tendencije
- jednim brojem opisujemo skup varijabilnih podataka
- primjeri mjera centralne tendencije: prosječna plaća,

## Mjere centralne tendencije:

- **Potpune** - računaju se na temelju svih podataka
- **Položajne** - određene položajem podataka u nizu

## Mjere centralne tendencije:

### Potpune

- srednja vrijednost (aritmetička sredina)
- geometrijska sredina
- harmonijska sredina

### Položajne

- mod
- medijan

# Srednja vrijednost

**Primjer.** Prosječna plaća u tromesječju: 7500 kn, 7200 kn, 7800 kn.

$$\text{Prosječna plaća} = \frac{7500 + 7200 + 7800}{3} = \frac{22500}{3} = 7500.$$

**Primjer.** Prosječna ocjena. Ocjene: 3, 3, 4, 2, 5, 4.

$$\text{Prosječna ocjena} = \frac{3 + 3 + 4 + 2 + 5 + 4}{6} = \frac{21}{6} = 3.5.$$

Za konačan skup podataka  $x_1, x_2, \dots, x_N$ , **srednja vrijednost** ( $\mu$ ) je aritmetička sredina podataka:

$$\mu = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N} \sum_{i=1}^N x_i.$$

$\sum$  - oznaka za sumiranje. Još se koristi i  $\sum_i x_i$ ,  $\sum x_i$ ,  $\sum x$ .

**Primjer.** Lionel Messi je u Ligi prvaka 2014/15 u jednoj utakmici postigao 3 pogotka, u dvije po 2, u tri po jedan dok u sedam utakmica nije postigao niti jedan zgoditak. Koliko je L. Messi prosječno dao golova po utakmici?

Ukupno je postigao

$$7 \cdot 0 + 3 \cdot 1 + 2 \cdot 2 + 1 \cdot 3 = 0 + 3 + 4 + 3 = 10$$

zgoditaka.

Broj utakmica =  $7 + 3 + 2 + 1 = 13$ .

Prosječan broj zgoditaka =  $\frac{10}{13} = 0.77$ .

Ovdje je zadana distribucija frekvencija:

Broj zgoditaka ( $x_i$ )	Broj utakmica ( $f_i$ )
0	7
1	3
2	2
3	1

Broj zgoditaka:

$$\begin{aligned}
 \mu &= 7 \cdot 0 + 3 \cdot 1 + 2 \cdot 2 + 1 \cdot 3 = \\
 &= f_1 \cdot x_1 + f_2 \cdot x_2 + f_3 \cdot x_3 + f_4 \cdot x_4 = \\
 &= \sum f_i x_i
 \end{aligned}$$

$$\text{Broj utakmica} = 7 + 3 + 2 + 1 = f_1 + f_2 + f_3 + f_4 = \sum f_i$$



# Računanje srednje vrijednosti iz distribucije frekvencija

Neka su  $x_1, x_2, x_3, \dots, x_k$  vrijednosti obilježja i neka su  $f_1, f_2, f_3, \dots, f_k$  pripadne frekvencije. Tada je

$$\mu = \frac{\sum f_i x_i}{\sum f_i} = \frac{\sum f_i x_i}{N}.$$

## Računanje srednje vrijednosti iz relativnih frekvencija

Neka su  $x_1, x_2, x_3, \dots, x_k$  vrijednosti obilježja i neka su  $r_1, r_2, r_3, \dots, r_k$  pripadne relativne frekvencije. Tada je

$$\mu = \sum r_i x_i.$$

Ovaj se izraz lagano dobije iz prethodne formule:

$$\mu = \frac{\sum f_i x_i}{N} = \sum \frac{f_i}{N} x_i = \sum r_i x_i.$$

**Primjer.** Vozilo je 100 km prešlo za 2 sata. Kolika je prosječna brzina?

$$\text{Prosječna brzina} = \frac{100 \text{ km}}{2 \text{ h}} = 50 \text{ km/h.}$$

Brzina ne treba biti konstantna!

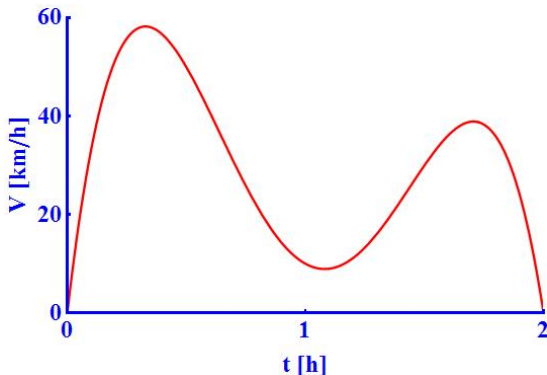
Kako izračunati prosječnu brzinu ukoliko nam je poznata samo brzina tijekom puta?

Tahograf:

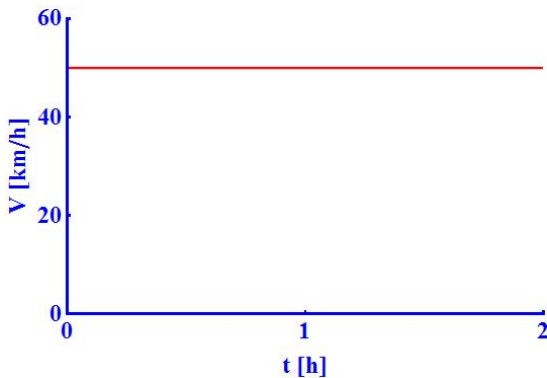


Prosječnu brzinu bismo mogli izračunati ukoliko znamo prijeđeni put.

Koliki je prijeđeni put ukoliko je brzina vozila bila kao na slici?

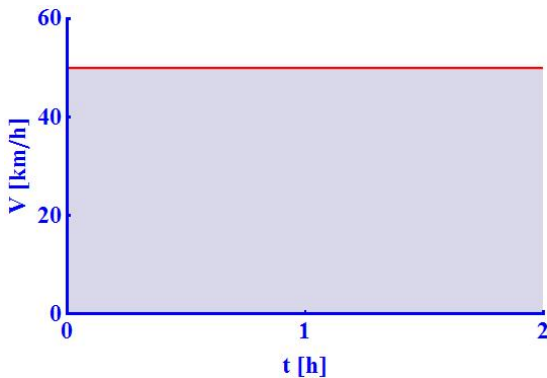


Ukoliko je brzina konstantna:



put je dan s

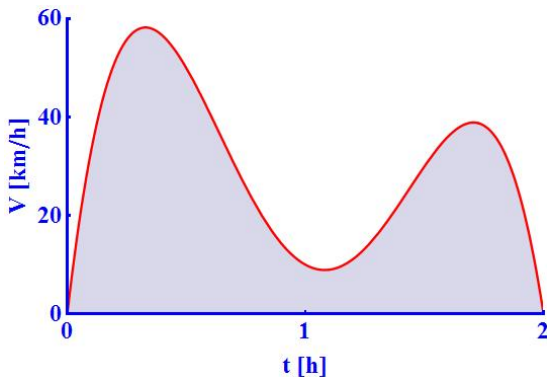
$$s = v \cdot t.$$



$$s = v \cdot t.$$

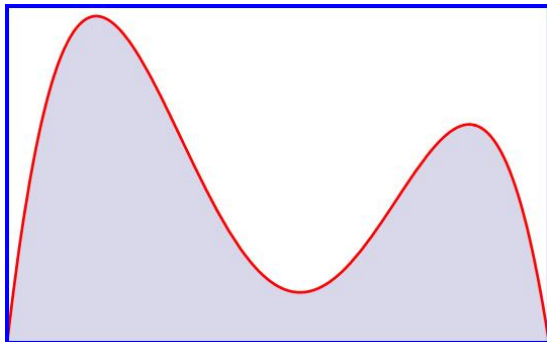
Prijeđeni put je površina pravokutnika!

Prijeđeni put je površina ispod krivulje:



Kako odrediti površinu?

Isprintani graf zaljepimo na čvrstu podlogu (npr. čelični lim) koji površinom odgovara veličini slike:



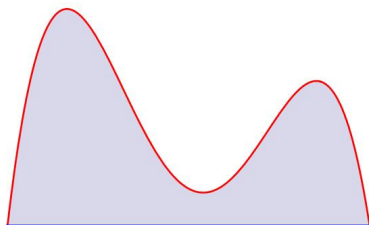
Lim izvažemo.

Zatim lim izrežemo po krivulji i izvažemo dio s traženom površinom.





$$m = 0.125 \text{ kg}$$
$$\text{Put} = P = 120$$



$$m = 0.062 \text{ kg}$$
$$\text{Put} = P = ?$$

Površina i masa su proporcionalni:

$$\text{Put} = \frac{0.062}{0.125} 120 = 59.52$$

$$\text{Prosječna brzina} = \frac{59.52 \text{ km}}{2 \text{ h}} = 29.76 \text{ km/h}$$

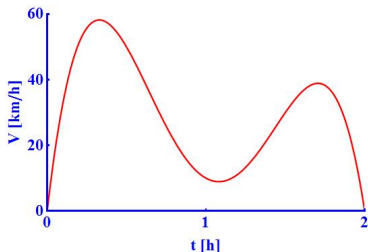
Matematička oznaka za površinu ispod krivulje je:

$$\int_0^2 V(t) dt$$

Znači, prosječna brzina kroz vrijeme  $T$  je dana s:

$$\bar{V} = \int_0^T V(t) dt.$$

**Napomena.** U ovom primjeru je skup podataka bio beskonačan.



**Primjer.** Ekipa A je u 10 min igre posjedovala loptu u sljedećim vremenima:

0:00 – 1:30

2:00 – 3:00

5:00 – 6:00

8:00 – 9:00

Koliki je bio posjed lopte za ekipu A?

Loptu su posjedovali:

$$1.5 + 1 + 1 + 1 = 4.5 \text{ min}$$

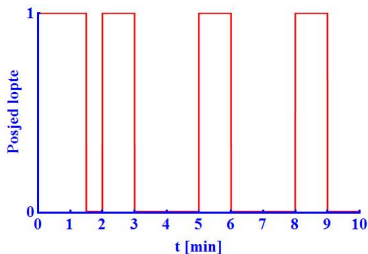
t.j. posjed lopte je bio

$$\frac{4.5}{10} = 0.45 = 45\%$$

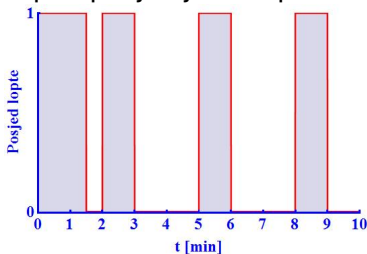
Neka varijabla posjed ima u trenutku  $t$  vrijednost

- 1 ukoliko je ekipa A bila u posjedu lopte
- 0 ukoliko nije.

Tada grafički prikaz posjeda lopte izgleda ovako:



Uočimo da je ukupan posjed jednak površini ispod krivulje:



t.j. jednak je  $\int_0^{10} \textit{posjed}(t) dt$ .

Srednja vrijednost je dana s  $\frac{1}{10} \int_0^{10} \textit{posjed}(t) dt$ .

# Medijan

**Medijan** je vrijednost koja skup podataka dijeli na dva istobrojna dijela tako da je polovica podataka veća a polovica podataka manja od medijana.

Oznaka za medijan: **m, Me**

## Postupak računanja medijana.

Podaci se poredaju prema veličini, od najmanjeg do najvećeg (ili od najvećeg do najmanjeg).

- Ako je broj podataka neparan - medijan je središnji član niza podataka.
- Ako je broj podataka paran - medijan je aritmetička sredina dvaju središnjih članova niza podataka.

**Primjer.** Odredite medijan za podatke

7, 8, 17, 9, 4, 9, 10, 11, 11, 3, 6, 7, 11, 5, 9

Podatke poredamo po veličini:

3	4	5	6	7	7	8	<b>9</b>	9	9	10	11	11	11	17	podaci
1.	2.	3.	4.	5.	6.	7.	<b>8.</b>	9.	10.	11.	12.	13.	14.	15.	redni broj

$$m = 9$$

**Primjer.** Odredite medijan za podatke

4, 2, 3, 5, 2, 1, 3, 4, 2, 1

Podatke poredamo po veličini:

1	1	2	2	<b>2</b>	<b>3</b>	3	4	4	5	podaci
1.	2.	3.	4.	<b>5.</b>	<b>6.</b>	7.	8.	9.	10.	redni broj

$$m = \frac{2 + 3}{2} = 2.5$$

## Određivanje medijana iz distribucije frekvencija.

Ukoliko je zadana distribucija frekvencija, medijan određujemo iz kumulativnih frekvencija.

Odreditmo medijan iz sljedećeg primjera.



Dob	Frekvencija	Relativna frekvencija	Kumulativna frekvencija	Relativna kumulativna frekvencija
19	1	0.05	1	0.05
20	0	0.00	1	0.05
21	5	0.25	6	0.30
22	2	0.10	8	0.40
23	2	0.10	10	0.50
24	1	0.05	11	0.55
25	4	0.20	15	0.75
26	1	0.05	16	0.80
27	0	0.00	16	0.80
28	1	0.05	17	0.85
29	0	0.00	17	0.85
30	1	0.05	18	0.90
31	1	0.05	19	0.95
32	1	0.05	20	1.00

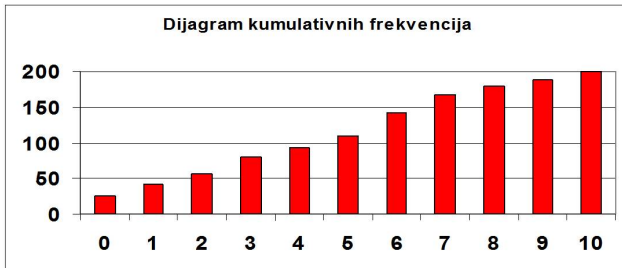
Dob	Frekvencija	Relativna frekvencija	Kumulativna frekvencija	Relativna kumulativna frekvencija
19	1	0.05	1	0.05
20	0	0.00	1	0.05
21	5	0.25	6	0.30
22	2	0.10	8	0.40
<b>23</b>	2	0.10	<b>10</b>	<b>0.50</b>
<b>24</b>	1	0.05	<b>11</b>	<b>0.55</b>
25	4	0.20	15	0.75
26	1	0.05	16	0.80
27	0	0.00	16	0.80
28	1	0.05	17	0.85
29	0	0.00	17	0.85
30	1	0.05	18	0.90
31	1	0.05	19	0.95
32	1	0.05	20	1.00

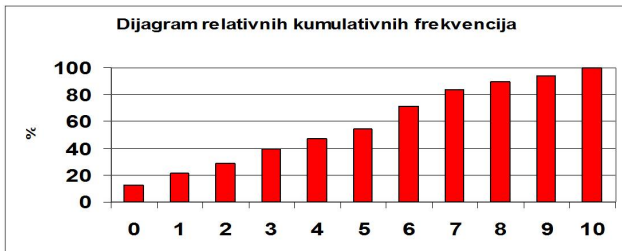
- 10 (50%) podataka je manje ili jednako od 23
- 10 (50%) podataka je veće ili jednako od 24

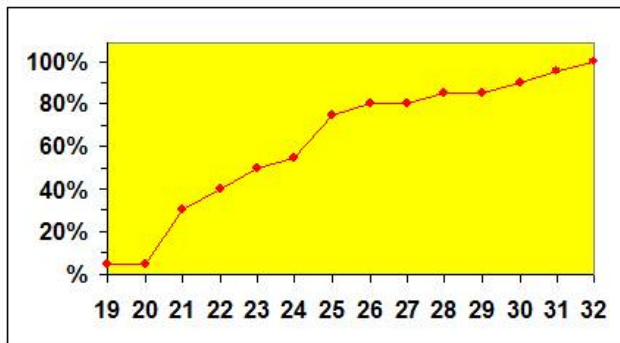
Znači, 23 i 24 su na središnjem mjestu u nizu podataka.

$$m = \frac{23 + 24}{2} = 23.5$$

**Primjer.** Odredite medijan iz sljedećih grafova kumulativne frekvencije:







## Druge mjere centralne tenencije

### Težinska (ponderirana) srednja vrijednost

Za podatke  $x_1, x_2, x_3, \dots, x_N$  dane su težine (ponderi)  
 $w_1, w_2, w_3, \dots, w_N$ .

**Težinska (ponderirana) srednja vrijednost** je dana s

$$\mu = \frac{\sum w_i x_i}{\sum w_i}.$$

**Primjer.** Za podatke 1, 3, 4, 5 i pripadne težine  $1/2, 2/3, 2$  i  $1$ , težinska srednja vrijednost je

$$\frac{\frac{1}{2} \cdot 1 + \frac{2}{3} \cdot 3 + 2 \cdot 4 + 1 \cdot 5}{\frac{1}{2} + \frac{2}{3} + 2 + 1} = \frac{\frac{31}{2}}{\frac{25}{6}} = \frac{93}{25} = 3.72$$

# Geometrijska sredina

Za podatke  $x_1, x_2, x_3, \dots, x_N$  **geometrijska sredina** je definirana s

$$\sqrt[N]{x_1 \cdot x_2 \cdot x_3 \cdot \dots \cdot x_N}$$

**Primjer.** Za podatke 1, 3, 4, 5 geometrijska sredina je

$$\sqrt[4]{1 \cdot 3 \cdot 4 \cdot 5} = \sqrt[4]{60} = 2.78$$



**Primjer.** Kamatna stopa je u 2010. g. bila 12%, u 2011. 17% i u 2013. g. 10%. Kolika je prosječna kamatna stopa za ove tri godine?

Prosječna kamatna stopa je godišnja stopa rasta po kojoj bi konstantni rast kroz 3 godine bio jednak ostvarenom rastu.

Nije

$$\frac{12 + 17 + 10}{3} = 13!$$

$$\text{glavnica} \cdot 1.12 \cdot 1.17 \cdot 1.10 = \text{glavnica} \cdot (1 + p) \cdot (1 + p) \cdot (1 + p)$$

t.j.

$$1.12 \cdot 1.17 \cdot 1.10 = (1 + p)^3$$

odnosno

$$1 + p = \sqrt[3]{1.12 \cdot 1.17 \cdot 1.10} = 1.1296$$

pa je

$$p = 0.1296$$

odnosno, prosječna kamatna stopa je 12.96%.

# Harmonijska sredina

Za podatke  $x_1, x_2, x_3, \dots, x_N$  **harmonijska sredina** je definirana s

$$\frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_N}}.$$

**Primjer.** Za podatke 1, 3, 4, 5 harmonijska sredina je

$$\frac{4}{\frac{1}{1} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5}} = \frac{240}{107} = 2.243.$$

**Primjer.** Vozilo je 100 km vozilo brzinom 40 km/h, sljedećih 100 km brzinom 70 km/h a zadnjih 100 km brzinom 130 km/h. Kolika je prosječna brzina?

Put je  $3 \cdot 100$  km a proteklo vrijeme je:

$$\frac{100}{40} + \frac{100}{70} + \frac{100}{130}.$$

Jer je brzina=put/vrijeme, srednja brzina je

$$\bar{V} = \frac{3 \cdot 100}{\frac{100}{40} + \frac{100}{70} + \frac{100}{130}} = \frac{3}{\frac{1}{40} + \frac{1}{70} + \frac{1}{130}} = 63.86$$

**Harmonijska sredina!**

# Mod

**Mod** je vrijednost varijable s najvećom frekvencijom.

(Najčešća vrijednost.)

**Primjer.** Odredimo mod za podatke

ACCABACABCAABBCBAACCBAAABCCBCAA.

Podatke prvo sortiramo (niz podataka):

AAAAAAAAAAAA BBBB BBBB CCCCCCCCCC

Mod je A jer se pojavljuje najviše puta (12).

Obilježje može imati i više modova. Npr. u nizu

1 1 2 2 2 3 4 4 5 5 5 6 7 7

mod je 2 i 5 (dva moda).

U ovom slučaju govorimo o **bimodalnoj** distribuciji frekvencija.

# Primjena mjera centralne tendencije

Tip obilježja	$\mu$	g	H	Medijan	Mod
Nominalno	NE	NE	NE	NE	DA
Redoslijedno	NE	NE	NE	DA	DA
Kvantitativno	DA	DA	DA	DA	DA

- ako su podaci jednaki nuli ili manji od nule  $\Rightarrow$  ne može se izračunati G ili H
- ako su sve vrijednosti varijable različite  $\Rightarrow$  nema moda
- mjere centralne tendencije mogu biti iste ili različite veličine
- $\min\{x_1, \dots, x_N\} \leq H \leq G \leq \mu \leq \max\{x_1, \dots, x_N\}$

# ZADATAK.

Izračunajte srednju vrijednost, medijan i mod za sljedeći niz podataka:

2	5	5	6	6
8	9	9	11	11
12	13	16	19	20
25	26	26	29	30
30	32	33	34	35
39	39	40	42	43

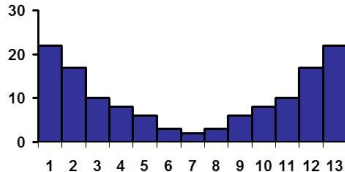
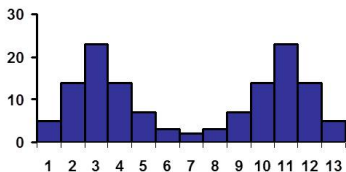
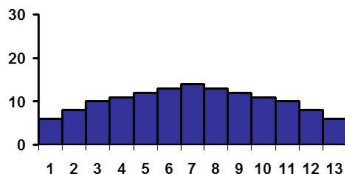
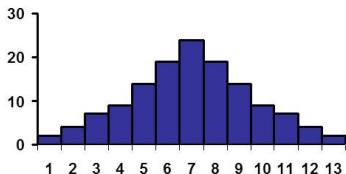
## Mjere raspršenosti (disperzije)

Usporedimo dva niza podataka:

$x_i$	10	60	50	30	40	20	$(N = 6, \mu = 35)$
$y_i$	35	45	30	35	40	25	$(N = 6, \mu = 35)$

Srednje vrijednosti su iste, ali se podaci ipak razlikuju.

Drugi skup podataka je jače grupiran oko srednje vrijednosti.



U svakom histogramu srednja vrijednost je ista (=7).  
Gdje su podaci najviše a gdje najmanje raspršeni?



## Varijanca i standardna devijacija

Za skup podataka  $x_1, x_2, \dots, x_N$  definirajmo **odstupanje**  $i$ -tog podatka od srednje vrijednosti:

$$x_i - \mu.$$

Loša mjera raspršenosti jer

$$\frac{1}{N} \sum_i (x_i - \mu) = 0.$$

tj., prosječno odstupanje je 0, pa nam to ništa ne govori o ukupnoj raspršenosti podataka.

**Kvadratno odstupanje**  $i$ -tog podatka od srednje vrijednosti:

$$(x_i - \mu)^2.$$

- Kvadratno odstupanje je uvijek pozitivno.
- Što je udaljenost podataka od srednje vrijednosti veća to je i kvadratno odstupanje veće.
- Srednje kvadratno odstupanje - mjera odstupanja svih podataka.

**Varijanca** je srednje kvadratno odstupanje podataka od srednje vrijednosti:

$$\sigma^2 = \frac{1}{N} \sum_i (x_i - \mu)^2.$$

$\sigma^2$  - oznaka za varijancu.

Alternativna formula za računanje varijance:

$$\sigma^2 = \frac{1}{N} \sum_i x_i^2 - \mu^2.$$

**Standardna devijacija** je korijen iz varijance:

$$\sigma = \sqrt{\sigma^2}$$
$$\sigma = \sqrt{\frac{1}{N} \sum_i (x_i - \mu)^2}.$$

$\sigma$  - oznaka za standardnu devijaciju.

Mjerna jedinica za standardnu devijaciju je ista kao i za obilježje  $x$ .

**Primjer.** Sandra Perković je na 65. Memorijalu Borisa Hanžekovića (2015) u bacanju diska postigla sljedeće rezultate u pet uspješnih bacanja:

62.30 m, 62.28 m, 69.77 m, 68.95 m, 69.88 m.

Odredite varijancu i standardnu devijaciju ovih bacanja.

Formula za varijancu:

$$\sigma^2 = \frac{1}{N} \sum_i (x_i - \mu)^2.$$

Prvo treba odrediti srednju vrijednost ( $\mu$ ):

$$\mu = \frac{1}{5} (62.30 + 62.28 + 69.77 + 68.95 + 69.88) = \frac{333.18}{5} = 66.636.$$

Varijanca:

$$\begin{aligned}\sigma^2 &= \frac{1}{5} \left[ (62.30 - 63.636)^2 + (62.28 - 63.636)^2 + (69.77 - 63.636)^2 + \right. \\ &\quad \left. + (68.95 - 63.636)^2 + (69.88 - 63.636)^2 \right] = \\ &= \frac{1}{5} \left[ (-4.336)^2 + (-4.356)^2 + 3.134^2 + 2.314^2 + 3.244^2 \right] = \\ &= \frac{1}{5} (18.801 + 18.975 + 9.822 + 5.355 + 10.524) = \\ &= \mathbf{12.695144}\end{aligned}$$

Standardna devijacija:

$$\sigma = \sqrt{\sigma^2} = \sqrt{12.695144} = \mathbf{3.563024558}$$

Preglednije je koristiti tablicu:

$i$	$x_i$	$x_i - \mu$	$(x_i - \mu)^2$
1	62.30	-4.336	18.801
2	62.28	-4.356	18.975
3	69.77	3.134	9.822
4	68.95	2.314	5.355
5	69.88	3.244	10.524
$\sum$	333.18		63.476
$\sum/N$	66.636		12.695

$$\sigma^2 = 12.695$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{12.695144} = 3.563024558$$

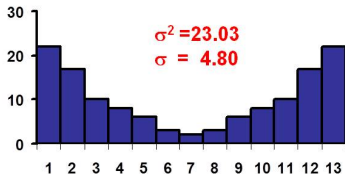
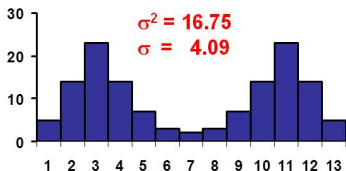
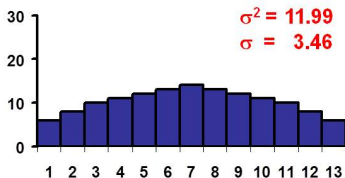
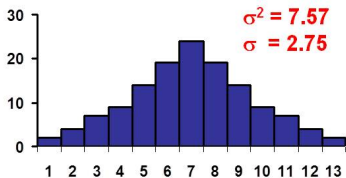
Za alternativnu formulu tablica izgleda ovako:

$i$	$x_i$	$x_i^2$
1	62.30	3881.290
2	62.28	3878.798
3	69.77	4867.853
4	68.95	4754.103
5	69.88	4883.214
$\sum$	333.18	22265.258
$\sum/N$	66.636	4453.052

$$\sigma^2 = \frac{1}{N} \sum_i x_i^2 - \mu^2 = 4453.052 - 66.636^2 = 12.695$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{12.695144} = 3.563024558$$

Varijanca i standardna devijacija za podatke iz uvodnog primjera:





# Računanje varijance iz distribucije frekvencija

Neka su  $x_1, x_2, x_3, \dots, x_k$  vrijednosti obilježja i neka su  $f_1, f_2, f_3, \dots, f_k$  pripadne frekvencije. Tada je

$$\sigma^2 = \frac{1}{N} \sum_i f_i (x_i - \mu)^2.$$

Alternativna formula za računanje varijance:

$$\sigma^2 = \frac{1}{N} \sum_i f_i x_i^2 - \mu^2.$$

## Primjer.

Izračunajte varijancu iz distribucije frekvencija obilježja dob:

Dob	Frekvencija
19	1
20	0
21	5
22	2
23	2
24	1
25	4
26	1
27	0
28	1
29	0
30	1
31	1
32	1

$x_i$ (Dob)	$f_i$ (Frekvencija)	$f_i \cdot x_i$	$f_i \cdot x_i^2$
19	1	19	361
20	0	0	0
21	5	105	2205
22	2	44	968
23	2	46	1058
24	1	24	576
25	4	100	2500
26	1	26	676
27	0	0	0
28	1	28	784
29	0	0	0
30	1	30	900
31	1	31	961
32	1	32	1024
$\sum$	20	485	12013
$\sum/N$		24.25	600.65

$$\begin{aligned}
 \sigma^2 &= \frac{1}{N} \sum_i f_i x_i^2 - \mu^2 = \\
 &= 600.65 - 24.25^2 = \\
 &= \mathbf{12.5875}
 \end{aligned}$$

# Koeficijent varijacije

Varijanca i standardna devijacija ovise o mjernim jedinicama.

- Mjerna jedinica za varijancu je kvadrirana mjerna jedinica varijable.
- Mjerna jedinica za standardnu devijaciju je jednaka mjernoj jedinici varijable.

**Primjer.** Označimo s  $y$  duljinu hitca Sandre Perković izraženog u centimetrima te izračunajmo varijancu i standardnu devijaciju.

- $x$  - duljina hitca u metrima
- $y$  - duljina hitca u centimetrima

$i$	$x_i$	$x_i^2$
1	62.30	3881.290
2	62.28	3878.798
3	69.77	4867.853
4	68.95	4754.103
5	69.88	4883.214
$\sum$	333.18	22265.258
$\sum/N$	66.636	4453.052

$$\begin{aligned}\sigma_x^2 &= 4453.052 - 66.636^2 = \\ &= \mathbf{12.695}\end{aligned}$$

$$\begin{aligned}\sigma_x &= \sqrt{\sigma_x^2} = \sqrt{12.695144} = \\ &= \mathbf{3.563024558}\end{aligned}$$

$i$	$y_i$	$y_i^2$
1	6230	38812900
2	6228	38787984
3	6977	48678529
4	6895	47541025
5	6988	48832144
$\sum$	33318	222652582
$\sum/N$	6663.6	44530516

$$\begin{aligned}\sigma_y^2 &= 44530516 - 6663.6^2 = \\ &= \mathbf{126951.44}\end{aligned}$$

$$\begin{aligned}\sigma_y &= \sqrt{\sigma_y^2} = \sqrt{126951.44} = \\ &= \mathbf{356.3024558}\end{aligned}$$

**Koeficijent varijacije:**

$$CV = \frac{\sigma}{\mu}$$

Koeficijent varijacije nema mjernu jedinicu.

	Duljina hitca	
	m	cm
$\mu$	66.636	6663.6
$\sigma^2$	12.695144	126951.44
$\sigma$	3.56302	356.302
CV	0.053	0.053

Varijabilnost (varianca) duljine hitca u centimetrima je veća, iako se zapravo radi o istim podacima.

Koeficijent varijacije je nepromijenjen.

# Interkvartil

## Kvartili

**Donji (prvi) kvartil ( $Q_1$ )** je vrijednost od koje je 25% podataka manje a 75% podataka veće.

**Gornji (treći) kvartil ( $Q_3$ )** je vrijednost od koje je 75% podataka manje a 25% podataka veće.

Kvartili dijele niz podataka na 4 jednaka dijela.

Medijan = drugi kvartil

### Računanje kvartila

- Niz podataka podijelimo na dva jednaka dijela. Ukoliko je broj podataka neparan, izbacimo medijan.
- Donji (prvi) kvartil je medijan prve polovice niza.
- Gornji (treći) kvartil je medijan druge polovice niza.

**Interkvartil (IQ)** je udaljenost između donjeg i gornjeg kvartila:

$$IQ = Q_3 - Q_1.$$

- Između donjeg i gornjeg kvartila se nalazi 50% podataka.
- Interkvartil je raspon u kojem se nalazi 50% središnjih članova niza podataka.
- Vrijednost prvih i zadnjih 25% podataka ne utječe na vrijednost interkvartila.



# Koeficijent kvartilne devijacije

**Koeficijent kvartilne devijacije (VQ) je:**

$$IQ = Q_3 - Q_1.$$

- Relativna mjera raspršenosti
- $0 \leq VQ < 1$

**Primjer.** Odredite donji i gornji kvartil, interkvartil i koeficijent kvartilne devijacije za sljedeći niz podataka:

1.3 4.1 4.1 4.2 4.4 4.6 5.1 5.2 5.3 5.5 5.5 5.5 5.9 6.1 7.8

**Rješenje.** Prvo određujemo donji i gornji kvartil.

Niz podijelimo na dva jednaka dijela.  $N=15$  je neparan.

1.3 4.1 4.1 4.2 4.4 4.6 5.1

5.2

5.3 5.5 5.5 5.5 5.9 6.1 7.8

Donji kvartil (Q1) je medijan prve polovice niza:

1.3 4.1 4.1 **4.2** 4.4 4.6 5.1

Gornji kvartil (Q3) je medijan druge polovice niza:

5.3 5.5 5.5 **5.5** 5.9 6.1 7.8



# Ostale mjere raspršenosti

## Srednje apsolutno odstupanje

Umjesto kvadratnog odstupanja podataka od srednje vrijednosti:

$$(x_i - \mu)^2$$

mogli smo promatrati apsolutno odstupanje:

$$|x_i - \mu|.$$

Time dobivamo mjeru raspršenosti poznatu kao **srednje apsolutno odstupanje**:

$$\text{s.a.o.} = \frac{1}{N} \sum_i |x_i - \mu|.$$

*(engl. MAD - mean absolute deviation)*

- Može se gledati i srednje apsolutno odstupanje od drugih mjera centralne tendencije.
- Srednje apsolutno odstupanje od srednje vrijednosti.
- Srednje apsolutno odstupanje od medijana.
- Srednje apsolutno odstupanje od moda.
- Umjesto srednje vrijednosti može se koristiti medijan apsolutnih odstupanja → medijalno apsolutno odstupanje.
- I ovdje se može promatrati odstupanje od srednje vrijednosti, medijana ili moda.

**Napomena.** Srednja vrijednost, medijan i standardna devijacija zadovoljavaju:

$$|\mu - m| \leq \sigma.$$

**Napomena.** Srednja vrijednost ( $\mu$ ) je točka s najmanjim srednjim kvadratnim odstupanjem od podataka.

Drugim riječima, za proizvoljan broj  $a$  vrijedi

$$\sigma^2 = \frac{1}{N} \sum_i (x_i - \mu)^2 \leq \frac{1}{N} \sum_i (x_i - a)^2.$$

**Napomena.** Medijan ( $m$ ) je točka s najmanjim srednjim apsolutnim odstupanjem od podataka.

Drugim riječima, za proizvoljan broj  $a$  vrijedi

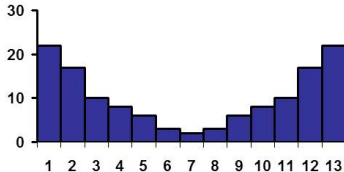
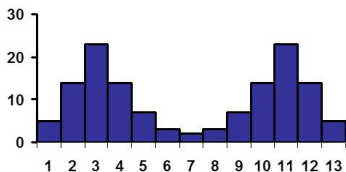
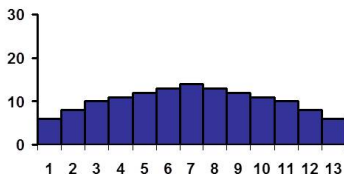
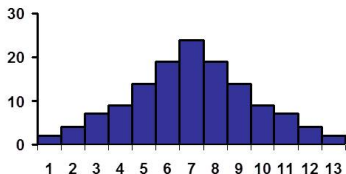
$$\frac{1}{N} \sum_i |x_i - m| \leq \frac{1}{N} \sum_i |x_i - a|.$$

# Raspon varijacije

**Raspon varijacije** je razlika najveće i najmanje vrijednosti u skupu podataka:

$$R_x = x_{\max} - x_{\min}.$$

- najjednostavnija mjera raspršenosti
- temeljena na dva podatka
- iskazan u mjernim jedinicama obilježja
- veći  $R_x$ , veća raspršenost, veći stupanj varijabilnosti



U sva 4 primjera je raspon varijacije isti ( $R = 13 - 1 = 12$ ) iako je varijabilnost podataka naočigled različita!



**Mjere raspršenosti** mogu biti

- u mjernim jedinicama obilježja
- u relativnom iznosu

**Mjere raspršenosti** temeljene na dijelu podataka:

- raspon varijacije
- interkvartil i koeficijent kvartilne devijacije

**Mjere raspršenosti** temeljene na svim podacima:

- varijanca, standardna devijacija, koeficijent varijacije
- srednje apsolutno odstupanje

# Računanje mjera centralne tendencije i raspršenosti za grupirane podatke

## Srednja vrijednost i varijanca

Zadani su razredi i njihove frekvencije ( $f_i$ ).

Za svaki razred izračunamo sredinu razreda  $X_i$  te srednju vrijednost i varijancu računamo kao da svi podaci unutar razreda imaju istu vrijednost ( $X_i$ ).

Srednja vrijednost:

$$\mu = \frac{\sum f_i X_i}{\sum f_i} = \frac{\sum f_i X_i}{N}.$$

Varijanca:

$$\sigma^2 = \frac{1}{N} \sum_i f_i (X_i - \mu)^2.$$

## Medijan i interkvartil

Kod računanja medijana (i kvartila) pretpostavljamo da su unutar svakog razreda frekvencije jednoliko (uniformno) raspoređene (sve vrijednosti imaju istu frekvenciju).

Prvo odredimo **medijalni razred**, t.j. razred u kojem se nalazi medijan.

**Medijan** računamo iz formule

$$m = \frac{\frac{N+1}{2} - (F + 1)}{f_m} w_m + L_m$$

gdje je

- $N$  - opseg (veličina) skupa
- $F$  - zbroj svih frekvencija razreda do, ali ne uključujući, medijalnog razreda (tj. kumulativna frekvencija razreda prije medijalnog)
- $f_m$  - frekvencija medijalnog razreda
- $w_m$  - širina medijalnog razreda
- $L_m$  - donja granica medijalnog razreda ukoliko se granice razreda preklapaju, ili aritmetička sredina donje granice medijalnog razreda i gornje granice prethodnog razreda ukoliko se granice ne preklapaju

# Mod

- U distribuciji frekvencija MOD je vrijednost varijable s najvećom frekvencijom.
- U distribuciji frekvencija s razredima - mod približno određujemo (aproksimiramo ga).
- **Modalni razred** - razred s najvećom korigiranom frekvencijom.
- Mod se ne treba nalaziti u modalnom razredu.
- modalni razred ne treba biti razred s najvećom frekvencijom.

**Napomena.** Mjere centralne tendencije i raspršenosti ne možemo egzaktno izračunati iz grupiranih podataka.  
Računamo ih približno.

# Linearna transformacija varijable

Ako varijabli  $X$  pribrojimo konstantu (broj), što se događa sa srednjom vrijednosti i varijancom?

	$X$	$X + 10$	$X + 20$
	62.30	72.30	82.30
	62.28	72.28	82.28
	69.77	79.77	89.77
	68.95	78.95	88.95
	69.88	79.88	89.88
$\mu$	66.64	76.64	86.64
$\sigma^2$	12.695	12.695	12.695

Varijabla  $Y$  je varijabla  $X$  uvećana za  $b$ :

$$Y = X + b.$$

Tada je

$$\mu_Y = \mu_X + b$$

$$\sigma_Y^2 = \sigma_X^2$$

$$\sigma_Y = \sigma_X$$

gdje je

$\mu_X$  - srednja vrijednost varijable  $X$

$\sigma_X^2$  - varijanca varijable  $X$

$\sigma_X$  - standardna devijacija varijable  $X$

$\mu_Y$  - srednja vrijednost varijable  $Y$

$\sigma_Y^2$  - varijanca varijable  $Y$

$\sigma_Y$  - standardna devijacija varijable  $Y$

Ako varijablu  $X$  pomnožimo s konstantom (brojem), što se događa sa srednjom vrijednosti i varijancom?

	$X$	$10 \cdot X$	$10 \cdot X$
	62.30	623.0	6230
	62.28	622.8	6228
	69.77	697.7	6977
	68.95	689.5	6895
	69.88	698.8	6988
$\mu$	66.64	666.4	6664
$\sigma^2$	12.695	1269.5	126950



Varijabla  $Y$  je varijabla  $X$  pomnožena s  $a$ :

$$Y = aX.$$

Tada je

$$\mu_Y = a\mu_X$$

$$\sigma_Y^2 = a^2\sigma_X^2$$

$$\sigma_Y = |a|\sigma_X$$

gdje je

$\mu_X$  - srednja vrijednost varijable  $X$

$\sigma_X^2$  - varijanca varijable  $X$

$\sigma_X$  - standardna devijacija varijable  $X$

$\mu_Y$  - srednja vrijednost varijable  $Y$

$\sigma_Y^2$  - varijanca varijable  $Y$

$\sigma_Y$  - standardna devijacija varijable  $Y$

Neka je varijabla  $Y$  dana s:

$$Y = aX + b.$$

Tada je

$$\mu_Y = a\mu_X + b$$

$$\sigma_Y^2 = a^2\sigma_X^2$$

$$\sigma_Y = |a|\sigma_X$$

gdje je

$\mu_X$  - srednja vrijednost varijable  $X$

$\sigma_X^2$  - varijanca varijable  $X$

$\sigma_X$  - standardna devijacija varijable  $X$

$\mu_Y$  - srednja vrijednost varijable  $Y$

$\sigma_Y^2$  - varijanca varijable  $Y$

$\sigma_Y$  - standardna devijacija varijable  $Y$

**Primjer.** Na atletskom natjecanju je srednja vrijednost za duljinu skokova bila 7.52 m uz standardnu devijaciju od 0.78 m. Kolika je srednja vrijednost i standardna devijacija duljine skokova izražena u centimetrima?

$X$  - duljina skoka u metrima

$Y$  - duljina skoka u centimetrima

$$Y = 100X$$

$$\mu_Y = 100 \mu_X = 100 \cdot 7.52 = 752$$

$$\sigma_Y = 100 \sigma_X = 100 \cdot 0.78 = 78$$

Srednja vrijednost duljine skoka je 752 cm uz standardnu devijaciju od 78 cm.

Što se događa s srednjom vrijednošću i varijancom zbroja dvije varijable?

**Primjer.** U tablici su prikazani bodovi za dva skoka i ukupni bodovi za deset najboljih skakačica na natjecanju Svjetskog kupa u skijaškim skokovima održanog u Courchevlu (Francuska) 14.8.2015. Usporedite srednju vrijednost, varijancu i standardnu devijaciju za ova tri rezultata.

Skakačica	Bodovi 1. skok (X)	Bodovi 2. skok (Y)	Bodovi Ukupno (Z)
TAKANASHI Sara	122.7	120.7	243.4
ITO Yuki	117.0	122.0	239.0
IRASCHKO-STOLZ Daniela	113.3	110.3	223.6
SEIFRIEDSBERGER Jacqueline	114.6	106.6	221.2
RUPPRECHT Anna	107.7	111.5	219.2
PINKELNIG Eva	109.6	108.8	218.4
ALTHAUS Katharina	112.8	102.1	214.9
ROGELJ Spela	105.6	108.6	214.2
CLAIR Julia	108.3	105.9	214.2
AVVAKUMOVA Irina	105.0	108.4	213.4
Srednja vrijednost ( $\mu$ )	<b>111.66</b>	<b>110.49</b>	<b>222.15</b>
Varijanca ( $\sigma^2$ )	<b>27.53</b>	<b>35.52</b>	<b>101.68</b>
Standardna devijacija ( $\sigma$ )	<b>5.247</b>	<b>5.960</b>	<b>10.084</b>

Iz primjera vidimo da se srednje vrijednosti zbrajaju dok to ne vrijedi za varijancu i standardnu devijaciju:

$X$  - bodovi 1. skoka

$Y$  - bodovi 2. skoka

$Z = X + Y$  - ukupni bodovi

Neka je varijabla  $Z$  zbroj varijabli  $X$  i  $Y$ :

$$Z = X + Y.$$

Tada je

$$\mu_Z = \mu_X + \mu_Y$$

gdje je

$\mu_X$  - srednja vrijednost varijable  $X$

$\mu_Y$  - srednja vrijednost varijable  $Y$

$\mu_Z$  - srednja vrijednost varijable  $Z$

## Primjer.

Na istom natjecanju je srednja vrijednost duljine skoka za muškarce iznosila 86.42 m uz standardnu devijaciju od 5.63 m. Kolika je srednja vrijednost i standardna devijacija bodova za duljinu skoka ukoliko je K-linija bila na 90 m i vrijednost za 1 metar je bila 2 boda?

**Rješenje.** Oznaka

$L$  - duljina skoka

$B$  - broj bodova za duljinu skoka

Bodovi se računaju prema formuli:

$$B = 2 \cdot (L - 90) + 60 = 2 \cdot L - 120.$$

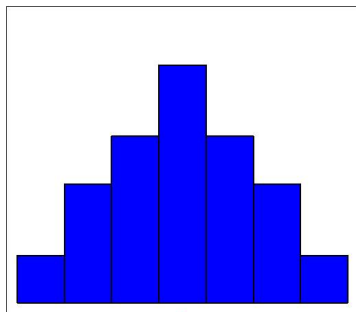
Srednja vrijednost i standardna devijacija su:

$$\mu_B = 2 \cdot \mu_L - 120 = 2 \cdot 86.42 - 120 = 52.84$$

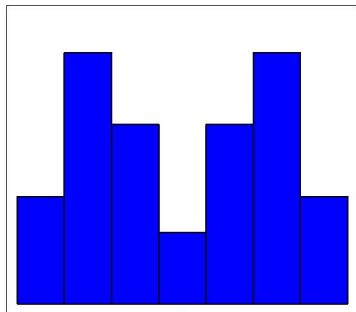
i

$$\sigma_B = 2 \cdot \sigma_L = 2 \cdot 5.63 = 11.26.$$

# Asimetrija



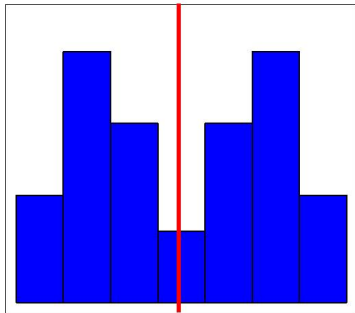
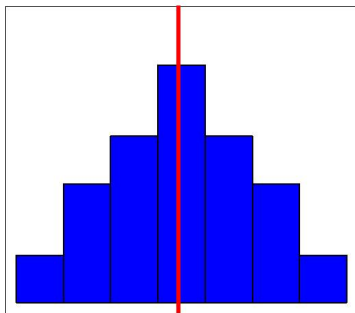
medijan



medijan

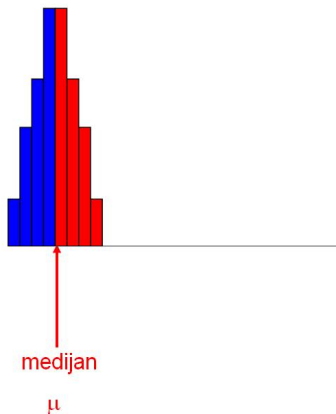


# Mjere asimetrije

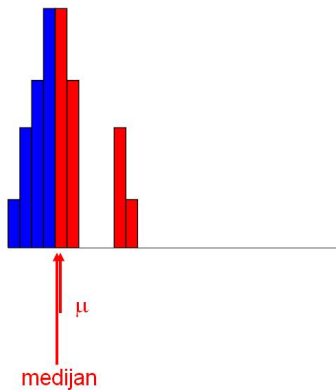


**Simetrična razdioba!**

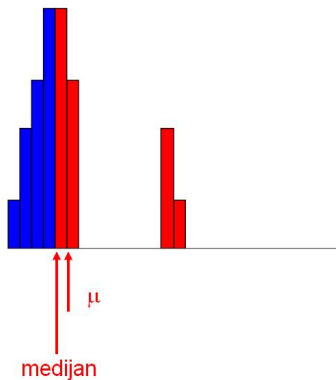
Što se događa sa srednjom vrijednošću i medijanom kada razdioba nije simetrična?



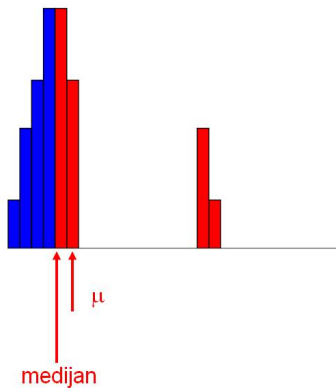
Što se događa sa srednjom vrijednošću i medijanom kada razdioba nije simetrična?



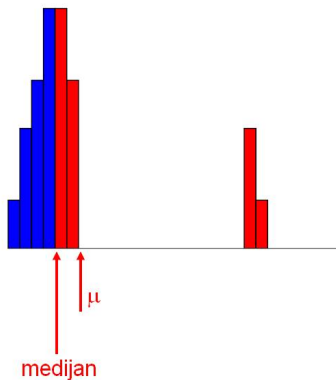
Što se događa sa srednjom vrijednošću i medijanom kada razdioba nije simetrična?



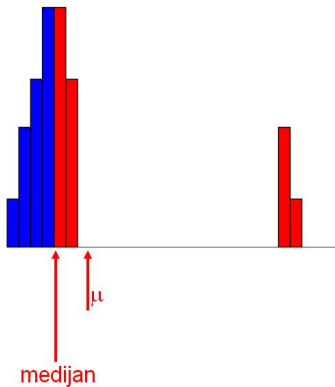
Što se događa sa srednjom vrijednošću i medijanom kada razdioba nije simetrična?



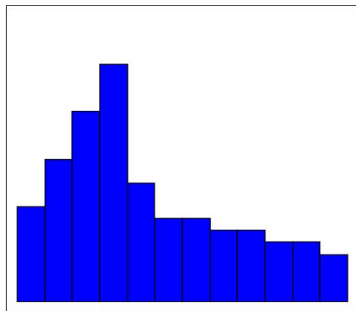
Što se događa sa srednjom vrijednošću i medijanom kada razdioba nije simetrična?



Što se događa sa srednjom vrijednošću i medijanom kada razdioba nije simetrična?

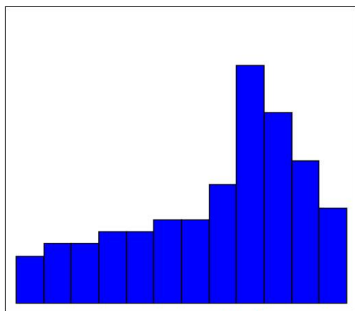


## Asimetrične razdiobe:



medijan

**Pozitivno zakošena**



$\mu$

medijan

**Negativno zakošena**

- Nakošenost (engl. skewness)



# Pearsonov koeficijent zakošenosti

$$Sk_P = 3 \frac{\mu - m}{\sigma},$$

gdje je

$\mu$  - srednja vrijednost

$m$  - medijan

$\sigma$  - standardna devijacija

# Zakošenost

$$\text{skew}(X) = \sum_i \left( \frac{x_i - \mu}{\sigma} \right)^3$$

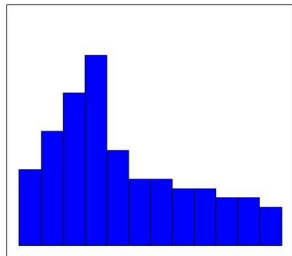
gdje je

$\mu$  - srednja vrijednost

$m$  - medijan

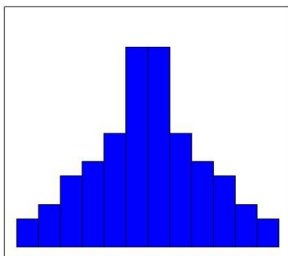
$\sigma$  - standardna devijacija

## Primjer.



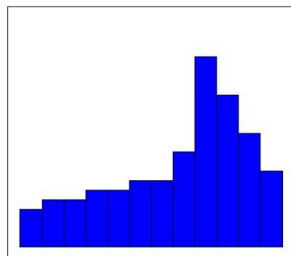
Zakošenost = 0.64

$$Sk_P = 1.22$$



Zakošenost = 0

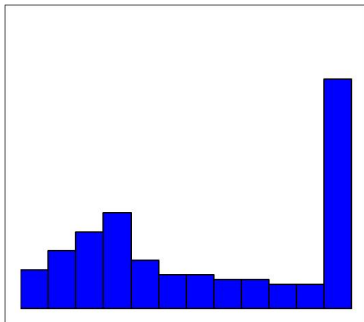
$$Sk_P = 0$$



Zakošenost = -0.64

$$Sk_P = -1.22$$

**Oprez!**

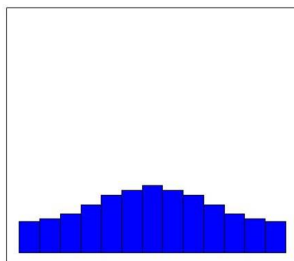
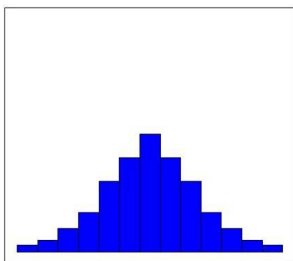
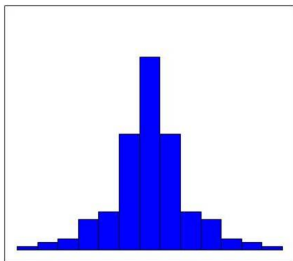


Zakošenost = 0

**Bimodalna distribucija!**

Zakošenost je dobar pokazatelj asimetrije za unimodalne distribucije.

# Spljoštenost



# Koeficijent spljoštenosti

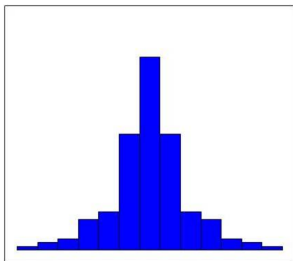
$$kurt(X) = \sum_i \left( \frac{x_i - \mu}{\sigma} \right)^4$$

gdje je

$\mu$  - srednja vrijednost

$m$  - medijan

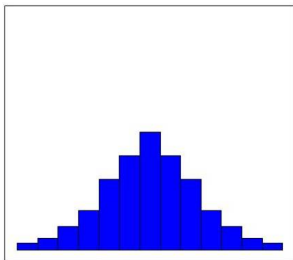
$\sigma$  - standardna devijacija



$$\text{kurt}(X) = 4.31$$

$$\text{kurt}(X) > 3$$

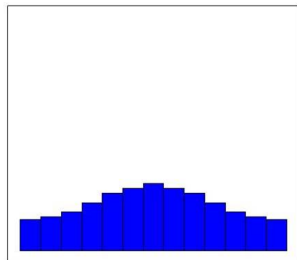
Leptokurtična  
(izbočena)  
distribucija



$$\text{kurt}(X) = 3$$

$$\text{kurt}(X) = 3$$

Mezokurtična  
distribucija



$$\text{kurt}(X) = 2.13$$

$$\text{kurt}(X) < 3$$

Platikurtična  
(spljoštena)  
distribucija

# Položaj podataka

## Rang

Rang je položaj (redni broj) podatka u nizu podataka. Ukoliko dva ili više podataka imaju istu vrijednost, njihov rang je aritmetička sredina rangova koje bi imali da su vrijednosti različite.

**Primjer.** Odredite rangove za zadani skup podataka:

15 2 21 13 8 17.

**Rješenje.** Za određivanje rangova podatke moramo prvo posložiti od najmanjeg do najvećeg:

2 8 13 15 17 21,

a zatim im pridružimo rang:

Vrijednost	2	8	13	15	17	21
Rang	1	2	3	4	5	6

Npr. Rang za vrijednost 15 je 4.



**Primjer.** Odredite rangove za zadani skup podataka:

15 2 21 13 8 17 15 8 15.

**Rješenje.** Promatramo niz podataka:

2 8 8 13 15 15 15 17 21,

a zatim im pridružimo rang:

Vrijednost	2	8	8	13	15	15	15	17	21
	1	2	3	4	5	6	7	8	9
Rang	1	2.5	2.5	4	6	6	6	8	9

Npr. rang za vrijednost 8 je  $(2+3)/2 = 2.5$  a za 15:  $(5+6+7)/3=6$ .

## Centili (percentili)

Percentili (centili) dijele niz podataka na 100 jednakih dijelova.

**$p$ -ti percentil** je broj koji niz podataka dijeli tako da je  $p\%$  podataka manje od  $p$ -tog percentila a  $(100 - p)\%$  veće od njega.

Medijan = 50-ti percentil

Donji kvartil = 25-ti percentil

Gornji kvartil = 75-ti percentil

**Decili** - dijele niz podataka na 10 jednakih dijelova.

Zajednički naziv za percentile, decile, kvartile i medijan je **kvantili** ili **fraktili**.

# Standardizirana varijabla

Za varijablu  $X$  čija je srednja vrijednost  $\mu$  i standardna devijacija  $\sigma$  standardizirana varijabla je varijabla

$$Z = \frac{X - \mu}{\sigma}.$$

Vrijednost standardizirane varijable: z-vrijednost, standardizirana vrijednost

Standardizirana varijabla pokazuje koliko je standardnih devijacija vrijednost podatka udaljena od srednje vrijednosti.

Standardizirana varijabla  $Z$  zadovoljava:

$$\mu_Z = 0 \quad \text{i} \quad \sigma_Z = 1.$$

(Srednja vrijednost je 0 a standardna devijacija 1).

Jer je

$$Z = \frac{1}{\sigma}X - \frac{\mu}{\sigma},$$

vrijedi

$$\mu_Z = \frac{1}{\sigma}\mu_X - \frac{\mu}{\sigma} = \frac{1}{\sigma}\mu - \frac{\mu}{\sigma} = 0$$

i

$$\sigma_Z = \frac{1}{\sigma}\sigma_X = \frac{1}{\sigma}\sigma = 1.$$

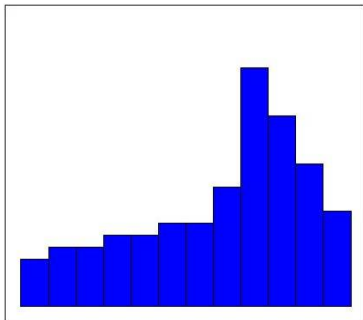
# Čebiševljeva nejednakost

Između  $\mu - k\sigma$  i  $\mu + k\sigma$  ( $k > 1$ ) nalazi se najmanje

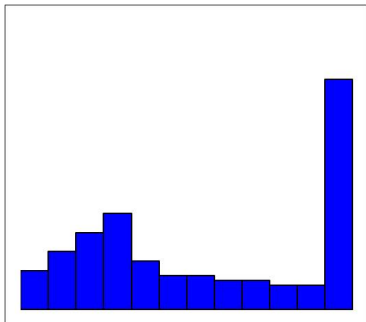
$$\left(1 - \frac{1}{k^2}\right) \cdot 100\%$$

članova populacije.

$k$	Min % unutar $k\sigma$ od $\mu$	Max % izvan $k\sigma$ od $\mu$
1	0.00	100.00
$\sqrt{2}$	50.00	50.00
1.5	55.56	44.44
2	75.00	25.00
3	88.89	11.11
4	93.75	6.25
5	96.00	4.00
6	97.22	2.78
7	97.96	2.04
8	98.44	1.56
9	98.77	1.23
10	99.00	1.00



Unimodalna  
distribucija



Bimodalna  
distribucija

**Unimodalna distribucija** - samo jedan vrh

**Biimodalna distribucija** - dva vrha

## Vysochanskij-Petuninova nejednakost

Za unimodalne distribucije postoje bolje ocjene od Čebiševljeve nejednakosti.

Ukoliko je distribucija varijable unimodalna, tada između  $\mu - k\sigma$  i  $\mu + k\sigma$  ( $k > 1$ ) nalazi se najmanje

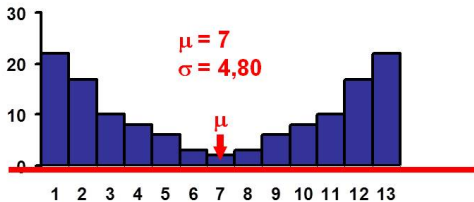
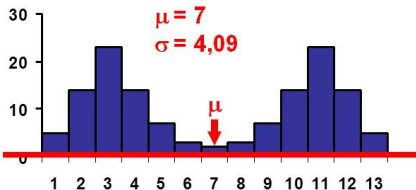
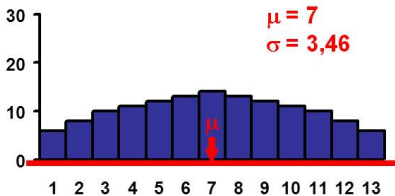
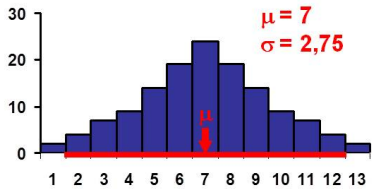
$$\left(1 - \frac{4}{9k^2}\right) \cdot 100\%$$

članova populacije.



Udio populacije(%) između  $\mu - k\sigma$  i  $\mu + k\sigma$  (unimodalna razdioba)

$k$	Čebiševljeva nejednakost	Vysochanskij- Petuninova nejednakost
1	0.00	55.56
$\sqrt{2}$	50.00	77.78
1.5	55.56	80.25
2	75.00	88.89
3	88.89	95.06
4	93.75	97.22
5	96.00	98.22
6	97.22	98.77
7	97.96	99.09
8	98.44	99.31
9	98.77	99.45
10	99.00	99.56



$\mu - 2\sigma$

$\mu + 2\sigma$

**Primjer.** U utrci slaloma na Svjetskom kupu u Kranjskoj Gori 15.3.2015. Marcel Hirscher je u prvoj vožnji postigao vrijeme od 52.22 s a u drugoj vožnji 50.17 s.

Kakvo je vrijeme postigao u odnosu na druge natjecatelje?  
Kakav je ukupni rezultat?

	1. vožnja	2. vožnja
Vrijeme	52.22	50.17
Rang	19	1
Relativni rang	63.3 %	3.3 %
z-vrijednost	0.4308	-1.516

Ukupni rezultat:

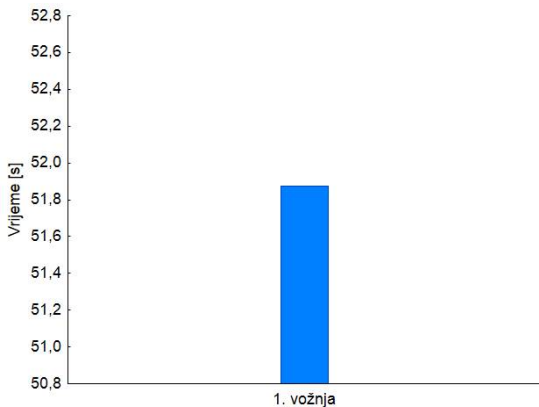
Vrijeme: 102.39 s

Rang: 6

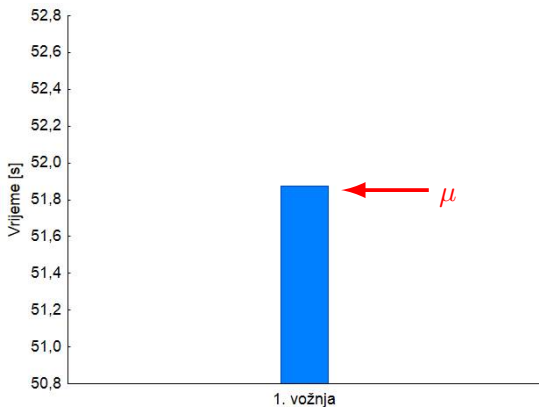
# Grafički prikaz mjera centralne tendencije i disperzije

- Stupčasti graf
- Linijski graf
- "Box-whiskers" graf

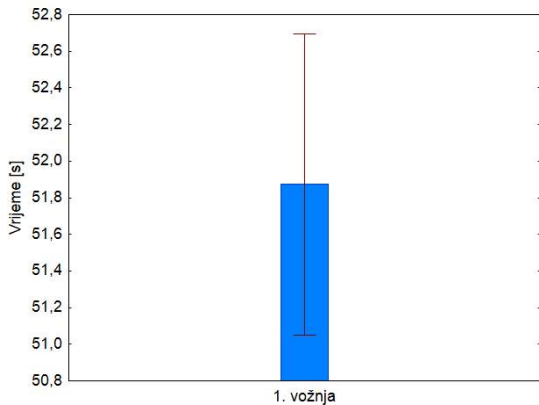
# Stupčasti graf (engl. *bars*)



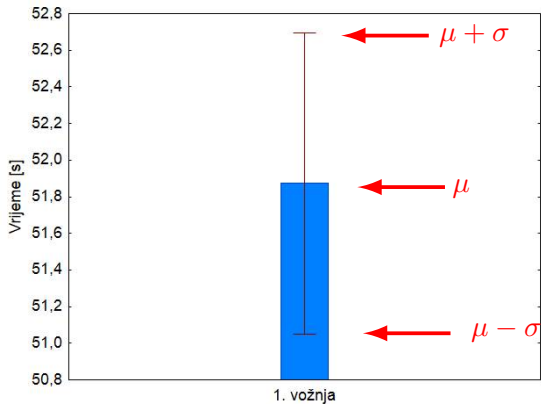
# Stupčasti graf (engl. *bars*)



# Stupčasti graf (engl. *bars*)

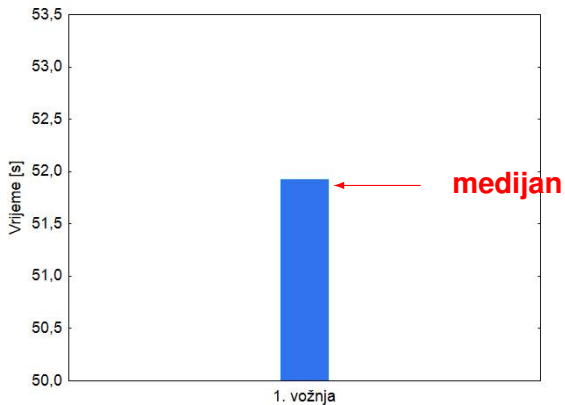


# Stupčasti graf (engl. *bars*)

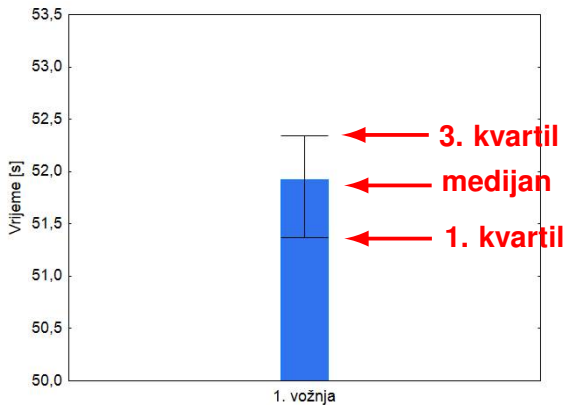




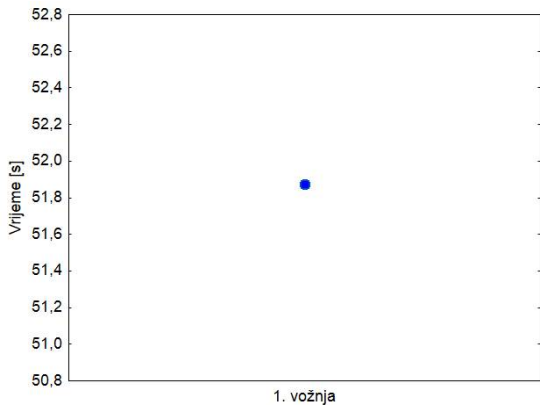
# Stupčasti graf (engl. *bars*)



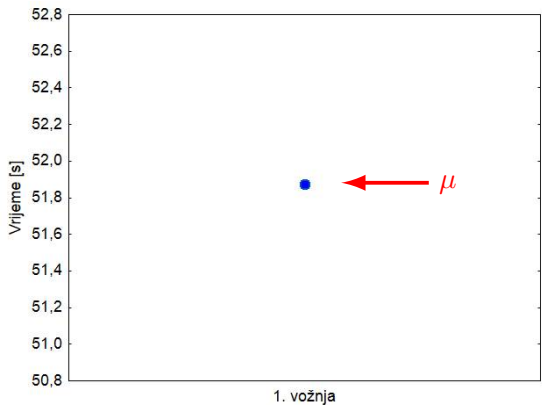
# Stupčasti graf (engl. *bars*)



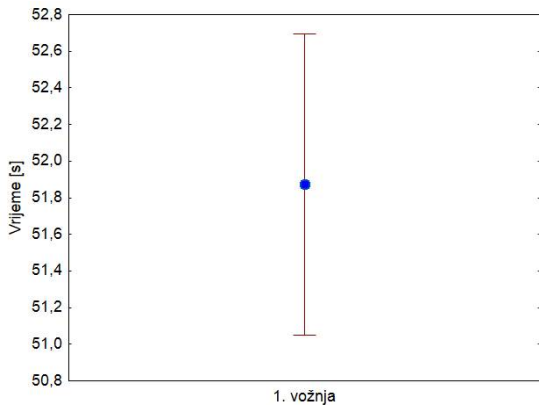
# Linijski graf



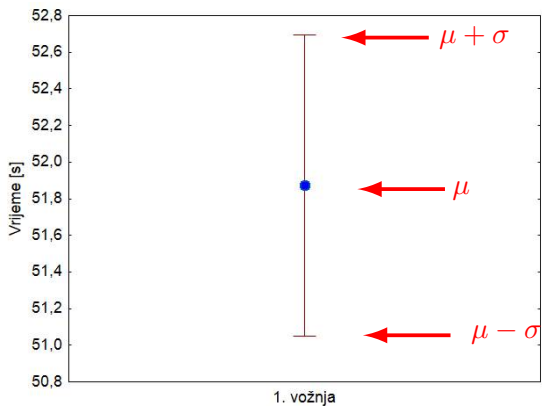
# Linijski graf



# Linijski graf



# Linijski graf

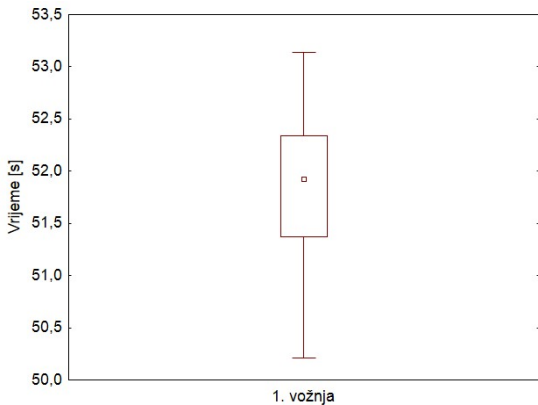


# Linijski graf

Na isti način se mogu prikazati i

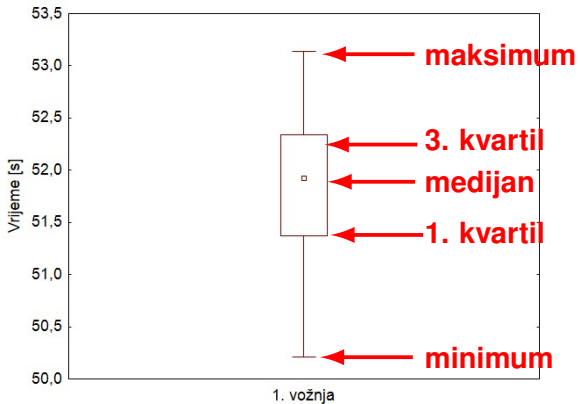
- medijan
- medijan, 1. kvartil, 3. kvartil
- srednja vrijednost, minimum, maksimum
- medijan, minimum, maksimum

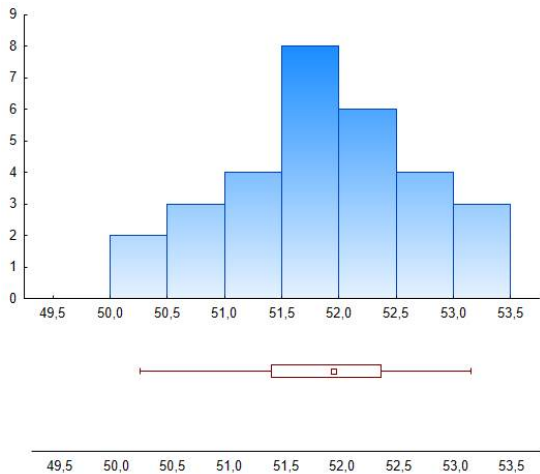
# "Box-whiskers" graf





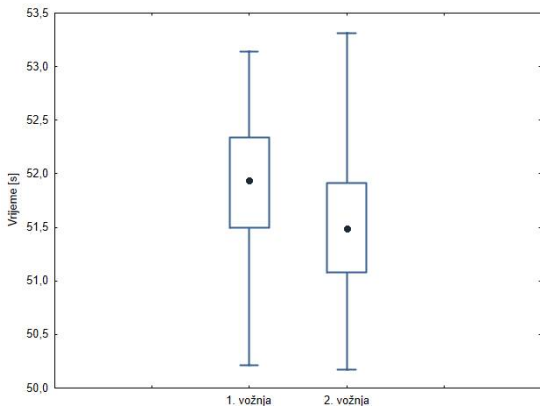
# "Box-whiskers" graf





## Primjer. Kranjska Gora 15.3.2015.

**Slika 1.** Rezultati utrke u slalomu (medijan, interkvartilni raspon, minimum, maksimum)



# Statistica

STATISTICA 64 - Spreadsheet1

File Home Edit View Format Statistics Data Mining Graphs Tools Data

Feature Finder

New Open Save Project Add to Workbook Add to Report Add to Word Add to Workspace Analysis Bar Macro Options Check Out Check In Discard SharePoint Tile Horizontally Tile Vertically Cascade Close All Arrange Icons Switch Windows

File Output Tools SharePoint Windows

Data: Spreadsheet1 (10v by 10c)

	1	2	3	4	5	6	7	8	9	10
	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										

## Podaci:

- Unos podataka
- Otvaranje postojeće datoteke s podacima (.sta)
- Kopiranje iz drugih aplikacije (Excel)

STATISTICA 64 - [Data: vjezba-1.sta (10v by 30c)]

File Home Edit View Format Statistics Data Mining Graphs Tools Data

Arial 10 Cells

Width Uncloak Height Uncloak New from Selection

AutoFit Uncloak All AutoFit Uncloak All Format Manager

Cloak Cloak Cloak Formats

Format Variables Cases Block Spreadsheets

	1 Var1	2 Var2	3 Var3	4 Var4	5 Var5	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10
1	50,51	50,75								
2	50,4	51,1								
3	50,21	51,91								
4	51,2	51,08								
5	50,81	51,52								
6	52,22	50,17								
7	51,57	51,16								
8	51,37	51,41								
9	51,5	51,31								
10	51,55	51,31								
11	51,91	51,08								
12	52,27	50,74								
13	51,56	51,49								
14	52,34	50,89								
15	51,9	51,61								
16	52,83	50,74								
17	51,94	51,86								
18	51,87	52,03								
19	52,19	51,8								
20	52,23	51,8								
21	51,95	52,1								
22	52,78	51,32								
23	53,13	50,98								
24	52,22	52,19								
25	51,12	53,31								
26	52,54	52,02								
27	53,08	51,62								
28	52,97	51,93								
29	53,14	52,54								
30	50,86	55,74								

## Uređivanje podataka:

- Brisanje nepotrebnih varijabli:
  - Označiti varijable koje želimo obrisati
  - → izbornik Edit/Delete/Variables
- Promjena imena varijabli (dvostruki 'klik' na ime varijable)

STATISTICA 64 - [Data: Spreadsheet1\* (10v by 30c)]

File Home **Edit** View Format Statistics Data Mining Graphs Tools Data

Clipboard: Cut, Copy, Paste, Format, Clear, Delete

Find/Replace: Find, Replace, Repeat, Go To, Copy Down, Copy Right

Standardize: Variables, Cases

Insert: Variables, Cases, OLE Object

OLE: DDE, OLE, Object Links

Feature Finder Options

	1	2	3	4	5	6	7	8	9	10
	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10
1	50,51	50,75								
2	50,4	51,1								
3	50,21	51,91								
4	51,2	51,08								
5	50,81	51,52								
6	52,22	50,17								
7	51,57	51,16								
8	51,37	51,41								
9	51,5	51,31								
10	51,55	51,31								
11	51,91	51,08								
12	52,27	50,74								
13	51,56	51,49								
14	52,34	50,89								
15	51,9	51,61								
16	52,83	50,74								
17	51,94	51,86								
18	51,87	52,03								
19	52,19	51,8								
20	52,23	51,8								
21	51,95	52,1								
22	52,78	51,32								
23	53,13	50,98								
24	52,22	52,19								
25	51,12	53,31								
26	52,54	52,02								
27	53,08	51,62								
28	52,97	51,93								
29	53,14	52,54								
30	50,86	55,74								

Delete Variables

From variable:

To variable:

Double-click on variable field or press F2 to select from list.

For Help, press F1

Spreadsheet1 C1.V3

Set OFF Weight OFF CAP N

Windows taskbar: Start, Chrome, Firefox, Statistica, PowerPoint

System tray: Volume, Network, Date/Time



STATISTICA 64 - [Data: vjezba-1.sta (2v by 30c)]

File Home Edit View Format Statistics Data Mining Graphs Tools Data

Clipboard/Data Find/Replace Fill Standardize Insert Links

Feature Finder Options

Cut Copy Paste Select All Clear Format Delete Move Screen Catcher Find Repeat Go To Replace Copy Down Copy Right Random Values Copy OLE Object DDE OLE Object

	1	2
	1. vožnja	Var 2
1	50.51	50.75
2	50.4	51.1
3	50.21	51.91
4	51.2	51.08
5	50.81	51.52
6	52.22	50.17
7	51.57	51.16
8	51.37	51.41
9	51.6	51.31
10	51.55	51.31
11	51.91	51.08
12	52.27	50.74
13	51.56	51.49
14	52.34	50.89
15	51.9	51.61
16	52.83	50.74
17	51.94	51.86
18	51.87	52.03
19	52.19	51.8
20	52.23	51.8
21	51.95	52.1
22	52.78	51.32
23	53.13	50.98
24	52.22	52.19
25	51.12	53.31
26	52.54	52.02
27	53.08	51.62
28	52.97	51.93
29	53.14	52.54
30	50.86	55.74

Variable 2

Arial 10 B I U x² x³

Name: Var 2 Type: Double OK

Measurement Type: Auto Length: 8 Cancel

Excluded  Label  Case State MD code: -99999998 << >>

Display format

- General
- Number
- Date
- Time
- Scientific
- Currency
- Percentage
- Fraction
- Custom

All Specs... Text Labels... Values/Stats... Properties... [Bundles]...

Long name (label or formula) Functions ):  Function guide

Labels: Use any text. Formulas: Must begin with an = sign.  
 Use variable names or V1, V2, ... V30 is the case number.  
 Examples: (a) = mean(v1+v2, exp(7), AGE) (b) = v1+v2, comment (after).  
 In case of conflict, variable names take precedence over variable text values.  
 Specify text values by appending \$, as in "value\$".

## Deskriptivna statistika

- Statistics
- Basic Statistics
- Descriptive statistics

STATISTICA 64 - [D]

Home Edit View Format **Statistics** Data Mining Graphs Tools Data

Basic Statistics Regression ANOVA Nonparametrics Distribution Fitting More Distributions

Advanced Models Mult/Exploratory Power Analysis Neural Nets PLS, PCA, ... Variance QC Charts Multivariate Predictive

Base Advanced/Multivariate Indu:

	1	2
	1. vožnja	2. vožnja
1	50,51	50,75
2	50,4	51,1
3	50,21	51,91
4	51,2	51,08
5	50,81	51,52
6	52,22	50,17
7	51,57	51,16
8	51,37	51,41
9	51,5	51,31
10	51,55	51,31
11	51,91	51,08
12	52,27	50,74
13	51,56	51,49
14	52,34	50,89
15	51,9	51,61
16	52,83	50,74
17	51,94	51,86
18	51,87	52,03
19	52,19	51,8
20	52,23	51,8
21	51,95	52,1
22	52,78	51,32
23	53,13	50,98
24	52,22	52,19

Basic Statistics and Tables: vjezb... ?

Quick

- Descriptive statistics
- Correlation matrices
- t-test, independent, by groups
- t-test, independent, by variables
- t-test, dependent samples
- t-test, single sample
- Breakdown & one-way ANOVA
- Breakdown; non-factorial tables
- Frequency tables
- Tables and banners
- Multiple response tables
- Difference tests: r, %, means
- Probability calculator

OK

Cancel

Options

Open Data

SELECT CASES

Descriptive Statistics: vjezba-1.sta

Variables: none

Quick | Advanced | Robust | Normality | Prob. & Scatterplots | Categ. plots | Options

Summary: Statistics | Graphs 1 | Graphs 2

Frequency tables | Histograms

Box & whisker plot for all variables

Graphical comparative summary display

Summary

Cancel

Options

By Group...

SELECT CASES   W

Wgtdd momnts

DF =

W-1  N-1

MD deletion

Casewise

Pairwise

Descriptive Statistics: vjezba-1.sta

Variables: none

Quick | **Advanced** | Robust | Normality | Prob. & Scatterplots | Categ. plots | Options

Summary: Statistics | Graphs 1 | Graphs 2

Frequency tables | Histograms

Box & whisker plot for all variables

Graphical comparative summary display

Summary

Cancel

Options

By Group...

SELECT CASES   W

Wghtd momnts

DF =  W-1  N-1

MD deletion  Casewise  Pairwise

Odabir varijabli za analizu.

Descriptive Statistics: vjezba-1.sta

Select the variables for the analysis

1 - 1. vožnja  
2 - 2. vožnja

2 - 2. vožnja

Select All Spread Zoom

Select variables:  
1-2

Show appropriate variables only

OK  
Cancel  
[ Bundles ]...

Use the "Show appropriate variables only" option to pre-screen variable lists and show categorical and continuous variables. Press F1 for more information.

Summary  
Cancel  
Options  
By Group...

Summary: Statistics  
Frequency tables  
Box & whisker plot  
Graphical comparative

Wghtd momnts  
F =  
W-1 N-1  
ID deletion  
Casewise  
Pairwise

Descriptive Statistics: Spreadsheet1

Variables: ALL

Quick | Advanced | Robust | Normality | Prob. & Scatterplots | Categ. plots | Options

Summary: Statistics | Graphs 1 | Graphs 2

Frequency tables | Histograms

Box & whisker plot for all variables

Graphical comparative summary display

Summary

Cancel

Options

By Group...

SELECT CASES  W  W

Wghtd momnts

DF =  
 W-1  N-1

MD deletion  
 Casewise  
 Pairwise

File Home Edit View Format Statistics Data Mining Graphs Tools Data Workbook

Paste Cut Copy Clear Format Delete Move Screen Catcher Find Find/Replace Go To Random Values Copy Down Copy Right Variables Cases Variables Cas Standardize Inse

Clipboard/Data Find/Replace Fill Standardize Inse

Workbook1\*  
Basic Statistics/Tab  
Descriptive sta  
Descriptiv

Descriptive Statistics (vjezba-1.sta)						
Variable	Valid N	Mean	Minimum	Maximum	Std.Dev.	
1. vožnja	30	51,87233	50,21000	53,14000	0,820877	
2. vožnja	30	51,65033	50,17000	55,74000	0,993158	

Nema medijana, kvartila,...



Descriptive Statistics: Spreadsheet1

Variables: ALL

Quick **Advanced** Robust Normality Prob. & Scatterplots Categ. plots Options

Summary: Statistics Graphs 1 Graphs 2

Frequency tables Histograms

Box & whisker plot for all variables

Graphical comparative summary display

Summary

Cancel

Options

By Group...

SELECT CASES  W  W

Wghtd momnts

DF =

W-1  N-1

MD deletion

Casewise

Pairwise

Descriptive Statistics: Spreadsheet1

Variables: ALL

Quick | Advanced | Robust | Normality | Prob. & Scatterplots | Categ. plots | Options

Summary: Statistics G1 G2

Compute

Location, valid N

- Valid N
- % valid obsvn.
- Mean
- Sum
- Median
- Mode
- Geom. mean
- Harm. mean

Variation, moments

- Standard Deviation
- CI for Sample SD Interval: 95,00 %
- Coefficient of variation
- Variance
- Std. err. of mean
- Conf. limits for means Interval: 95,00 %
- Skewness
- Std. err., Skewness
- Kurtosis
- Std. err., Kurtosis

Percentiles, ranges

- Minimum & maximum
- Lower & upper quartiles
- Percentile boundaries
- First: 10,00 % Second: 90,00 %
- Boundaries 10 20 30 40 50 60 70 80 90
- Intervals: 10
- Range  Quartile range

Select all stats | Reset

Save settings as default

Summary

Cancel

Options

By Group...

SELECT CASES   W

Wghtd momnts

DF =

- W-1
- N-1

MD deletion

- Casewise
- Pairwise

Descriptive Statistics: vjezba-1.sta

Variables: ALL

Quick | Advanced | Robust | Normality | Prob. & Scatterplots | Categ. plots | Options

Summary: Statistics | G1 | G2

Compute

Location, valid N

- Valid N
- % valid obsvn.
- Mean
- Sum
- Median
- Mode
- Geom. mean
- Harm. mean

Variation, moments

- Standard Deviation
- CI for Sample SD  
Interval: 95,00 %
- Coefficient of variation
- Variance
- Std. err. of mean
- Conf. limits for means  
Interval: 95,00 %
- Skewness
- Std. err., Skewness
- Kurtosis
- Std. err., Kurtosis

Percentiles, ranges

- Minimum & maximum
- Lower & upper quartiles
- Percentile boundaries
- First: 10,00 % Second: 90,00 %
- Boundaries: 10 20 30 40 50 60 70 80 90
- Intervals: 10
- Range
- Quartile range

Select all stats | Reset

Save settings as default

Summary

Cancel

Options

By Group...

SELECT CASES   W

Wgtdt momnts

DF =

- W-1
- N-1

MD deletion

- Casewise
- Pairwise

STATISTICA - [Workbook1\* - Descriptive Statistics (vjezba-1.sta)]

File Home Edit View Format Statistics Data Mining Graphs Tools Data Workbook

Clipboard/Data: Paste, Cut, Copy, Clear, Format, Delete, Move, Screen Catcher

Find/Replace: Find, Go To, Replace, Repeat

Fill: Random Values, Copy Down, Copy Right

Standardize: Variables, Cases

Insert: Variables, Cases, OLE Object

Links: DDE, OLE, Object

Workbook1\*
 

- Basic Statistics/Te
- Descriptive stz
- Descriptiv

Descriptive Statistics (vjezba-1.sta)											
Variable	Valid N	Mean	Median	Minimum	Maximum	Lower Quartile	Upper Quartile	Range	Quartile Range	Std.Dev.	
1. vožnja	30	51,87233	51,92500	50,21000	53,14000	51,37000	52,34000	2,930000	0,970000	0,820877	
2. vožnja	30	51,65033	51,50500	50,17000	55,74000	51,08000	51,93000	5,570000	0,850000	0,993158	

# Crtanje grafova

## Stupčasti graf (histogram)

STATISTICA 64 - [Data: vježba-1.sta\* (2v by 30c)]

File Edit View Format Statistics Data Mining **Graphs** Tools Data

Histogram Scatterplot Means Box Variability Line 2D 3D Seq. 3D XYZ Matrix Icons Categorized Batch (By Group) Stock Data Graphs Input Data Graphs User-defined Multi-Graph Layouts

	1	2
	1. vožnja	2. vožnja
1	50.51	50.75
2	50.4	51.1
3	50.21	51.91
4	51.2	51.08
5	50.81	51.52
6	52.22	50.17
7	51.57	51.16
8	51.37	51.41
9	51.5	51.31
10	51.55	51.31
11	51.91	51.08
12	52.27	50.74
13	51.56	51.49
14	52.34	50.89
15	51.9	51.61
16	52.83	50.74
17	51.94	51.86
18	51.87	52.03
19	52.19	51.8
20	52.23	51.8
21	51.95	52.1
22	52.78	51.32
23	53.13	50.98
24	52.22	52.19
25	51.12	53.31
26	52.54	52.02
27	53.08	51.62
28	52.97	51.93
29	53.14	52.54
30	50.86	55.74

2D Histograms

Quick Advanced Appearance Categorized Options 1 Options 2

Variables: Variable: none

Graph type: Regular Multiple

Intervals:
 

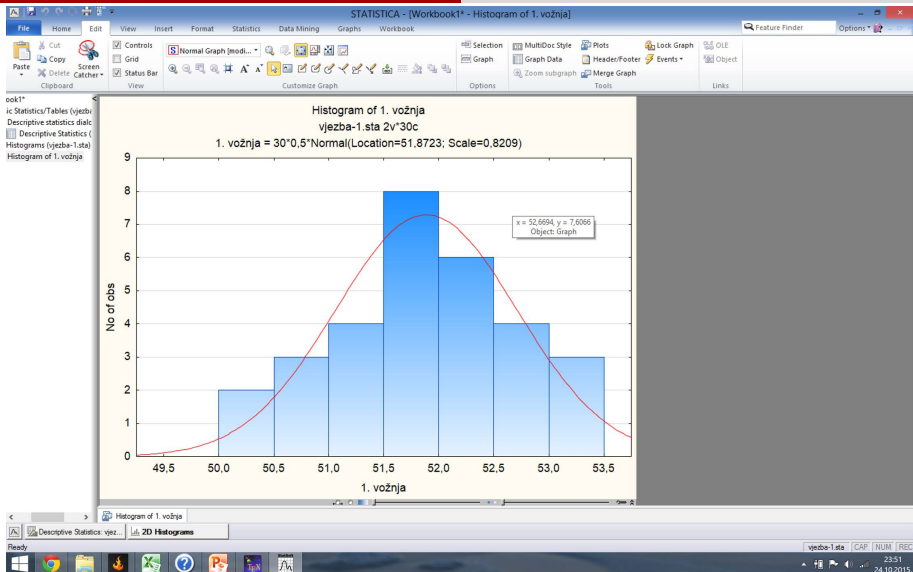
- Integer mode  Auto
- Unique values
- Unsorted  Asc  Desc
- Categories: 10
- Codes: none

Fit type:  Normal

OK Cancel Options By Group Sel Cond Case Weights Graphs Gallery Updating: Auto

Descriptive Statistics: vjez... 2D Histograms

vježba-1.sta C3.V2 51.91 Sel OFF Weight OFF CAF NUM REC 23:50 24.10.2015



Crvenu krivulju treba ukloniti.

Desnim klikom miša otvara se izbornik i izaberemo 'Graph Options'.

STATISTICA - [Workbook1\* - Histogram of 1. vožnja]

File Home Edit View Insert Format Statistics Data Mining Graphs Workbook

Selection Graph MultiDoc Style Plots Lock Graph OLE  
Graph Data Header/Footer Events Object  
Zoom subgraph Merge Graph Link

Graph Options

Plot: 1: 1. vožnja Fit: 1 Add new fit Delete

Fit type: Normal

Fitting data

Scale: X

Range: Full Range Min: 0 Max: 0

Line

Pattern... Resolution: Normal

Options

Polynomial Quadratic Stiffness

Logarithm base: 10 Min: 0,25 Max

No. of periods: 5

Exp smooth const: 0,5

Use default Gaussian kernel bandwidth Bandwidth: 0,5

Style: S Plot 1 Fit attributes

OK Cancel

Graph

- Window
- Layout
- Titles/Text
- Plot
  - General
  - Bars
  - Histogram
  - Pies
  - Fitting
  - Legend Bands
  - Custom Function
- Axis
  - General
  - Title
  - Scaling
  - Major Units
  - Minor Units
  - Scale Values
  - Custom Units
  - Reference Lines

STATISTICA

Delete current fit?

Yes No

Histogram of 1. vožnja

vjezba-1.sta 2v\*30c

1. vožnja = 30\*0,5\*Normal(Location=51,8723)

No of obs

49,5 50,0 50,5 51,0 51,5 52,0

1. vožnja

Histogram of 1. vožnja

Descriptive Statistics - vjez... 2D Histograms

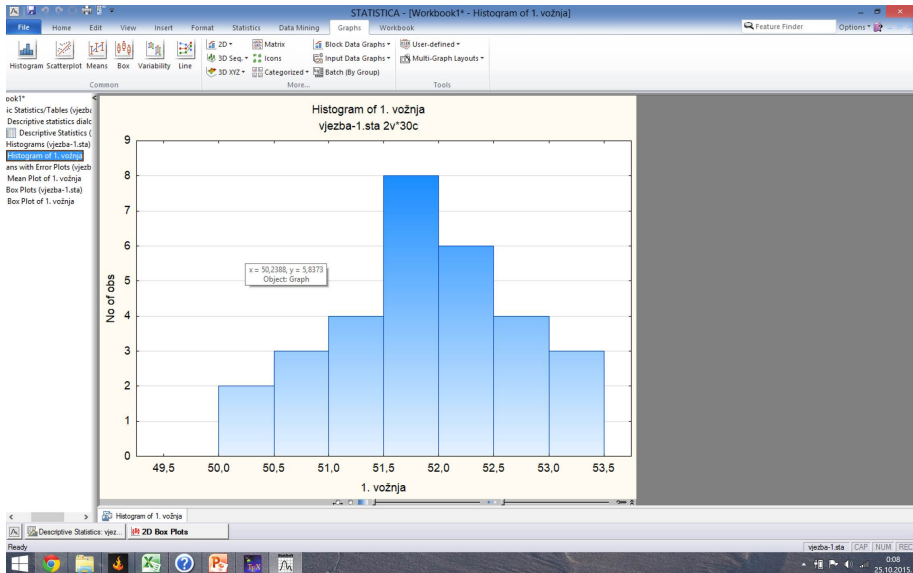
Ready

vjezba-1.sta CAP NUM REC

23:52

24.10.2015





# Linijski (točkasti) graf

STATISTICA 64 - [Data: vježba-1.sta\* (2v by 30c)]

File Home Edit View Format Statistics Data Mining Graphs Tools Data Feature Finder Options

Histogram Scatterplot Means Box Variability Line 2D 3D Seq. Matrix Block Data Graphs User-defined 3D XYZ Icons Input Data Graphs Multi-Graph Layouts Categorized Batch (By Group)

Common More... Tools

	1	2
	1. vožnja	2. vožnja
1	50,51	50,75
2	50,4	51,1
3	50,21	51,91
4	51,2	54,66
5	50,81	
6	52,22	
7	51,57	
8	51,37	
9	51,5	
10	51,56	
11	51,91	
12	52,27	
13	51,56	
14	52,34	
15	51,9	
16	52,83	
17	51,94	
18	51,87	
19	52,19	
20	52,23	
21	51,95	
22	52,78	
23	53,13	
24	52,22	52,19
25	51,12	53,31
26	52,54	52,02
27	53,08	51,62
28	52,97	51,93
29	53,14	52,54
30	50,86	55,74

Select Variables for Means with Error Plots

1- 1. vožnja  
2- 2. vožnja

1- 1. vožnja  
2- 2. vožnja

OK Cancel [Bundles]...

Use the "Show appropriate variables only" option to pre-screen variable lists and show categorical and continuous variables. Press F1 for more information.

Select All Spread Zoom Select All Spread Zoom

Dependent variables: 1

Grouping variable:

Show appropriate variables only

Means with Error Plots

Quick Advanced Appearance Categorized Options 1 Options 2

Graph type: Whiskers Regular Variables High-Low Close Multiple Dependent variables: none Grouping variable: none

Grouping intervals: Integer mode Auto Unique values Unsorted Asc Desc Codes: 10 Categories: none

Layout: Separate Overlaid

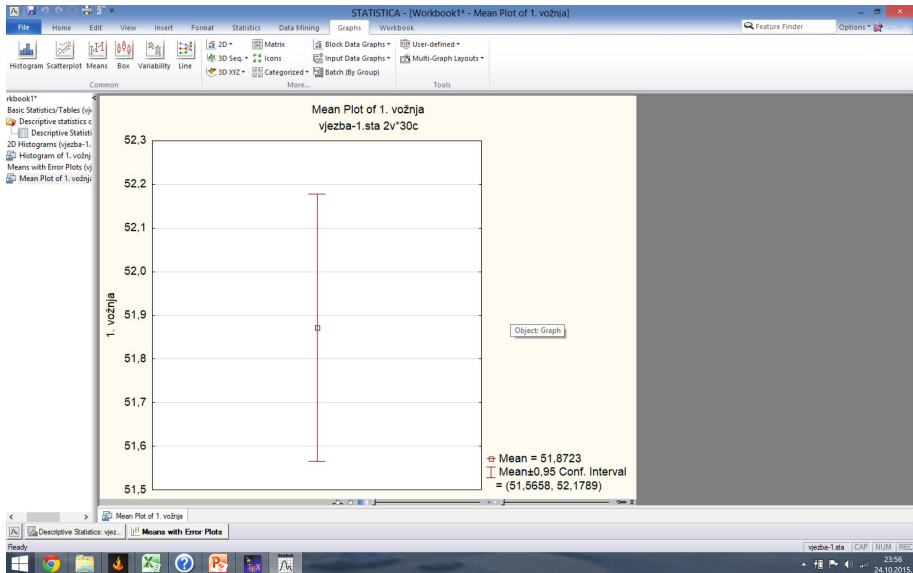
Mean+-3S Conf. Interval Mean Mean-3S Conf. Interval Whisker Value: Conf. Interval Probability: .95

Frequency/severity plot

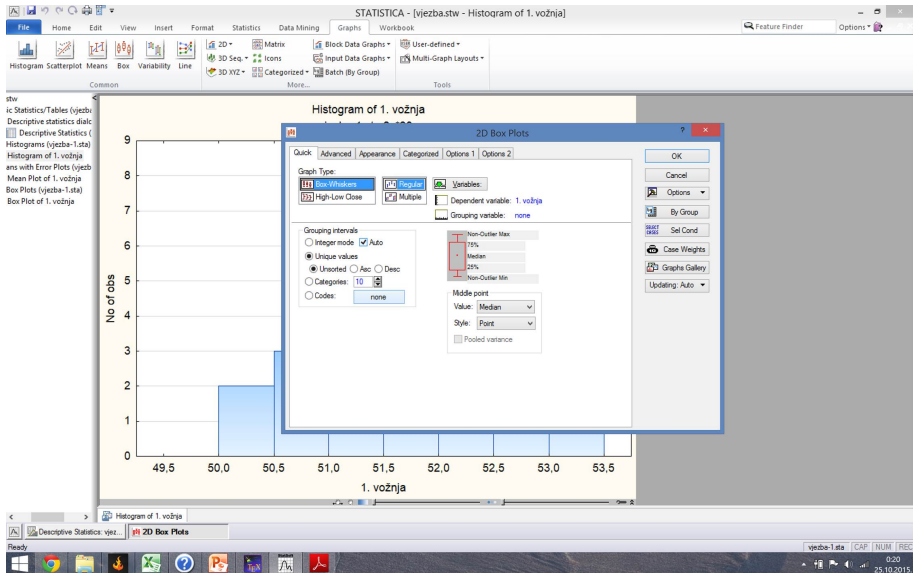
OK Cancel Options By Group Case Weights Graphs Gallery Updating: Auto

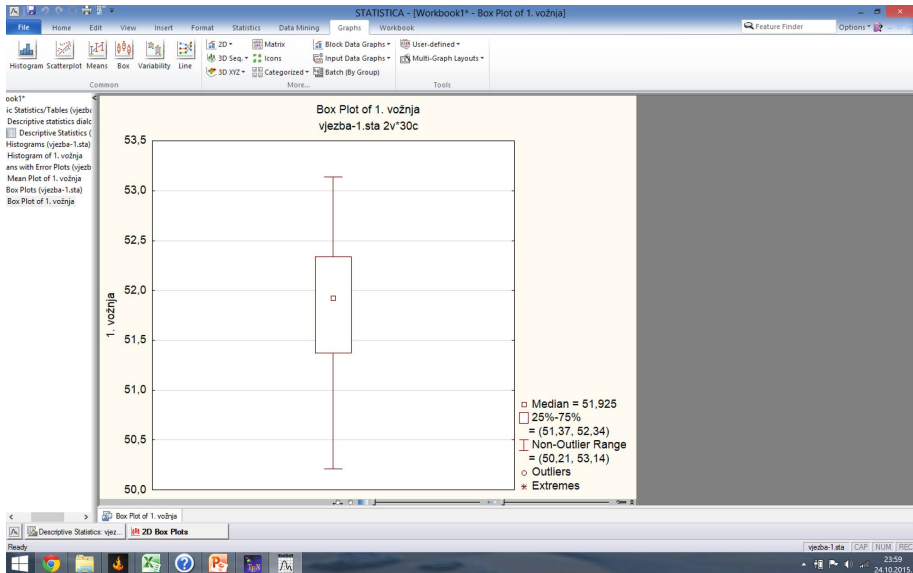
Descriptive Statistics: vjež... Means with Error Plots

For Help, press F1 vježba-1.sta C3.V2 51.91 Sel OFF | Weight: OFF CAP NUM REC 23:55 24.10.2015



# "Box-whiskers" graf





**Primjer.** Usporedite 1. i 2. vožnju slaloma na Svjetskom kupu u Kranjskoj Gori 15.3.2015.

	1. vožnja	2. vožnja
Srednja vrijednost	51.87	51.65
Standardna devijacija	0.82	0.99
Medijan	51.92	51.51
1. kvartil	51.37	51.08
3. kvartil	52.34	51.93
IQR	0.97	0.85
Minimum	50.21	50.17
Maksimum	53.14	55.74
Raspon	2.93	5.57

2. vožnja je bila kraća (srednja vrijednost i medijan).  
Ali ima veći raspon.



## VAŽNA NAPOMENA.

Varijanca je definirana s:

$$\frac{1}{n} \sum_k (x_k - \mu)^2.$$

U Statistici se varijanca računa kao

$$\frac{1}{n-1} \sum (x_k - \mu)^2.$$

Prava vrijednost standardne devijacije za 1. vožnju je:

$$0.82 \cdot \sqrt{\frac{29}{30}} = 0.81$$

Razlika je minimalna i nećemo je korigirati.

Razlog za ovakvo računanje bit će objašnjeno u sljedećim poglavljima.

Natjecatelj	Vrijeme 1. vožnje
HERBST Reinfried	51.86
HARGIN Mattias	51.91
CHODOUNSKY David	51.93
AERNI Luca	52.02
GRANGE Jean-Baptiste	52.03
NORDBOTTEN Jonathan	52.10
RAICH Benjamin	52.19
KUERNER Miha	52.54
GROSS Stefano	53.31
THALER Patrick	55.74

Patrick Thaler ima znatno veće vrijeme od ostalih.

**Ekstremna vrijednost (engl. *outlier*)** - vrijednost znatno veća ili manja od ostalih podataka

Detekcija ekstremnih vrijednosti:

Podaci izvan intervala [medijan-1.5 IQR, medijan+1.5 IQR]

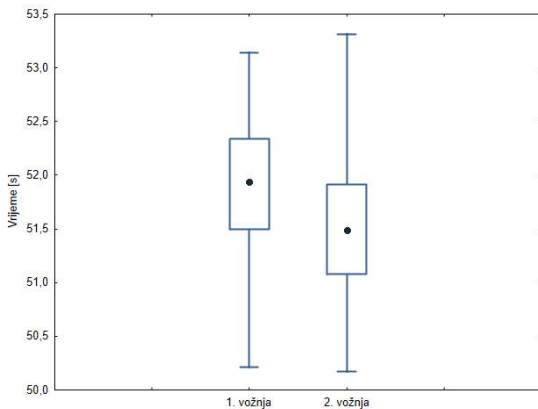
Patrick Thaler : 55.74

$$\text{medijan}+1.5 \text{ IQR} = 51.51 + 1.5 \cdot 0.85 = 52.785$$

## Ponovljena analiza bez ekstremnih vrijednosti (Patrick Thaler)

	1. vožnja	2. vožnja
Srednja vrijednost	51.91	51.51
Standardna devijacija	0.81	0.64
Medijan	51.94	51.49
1. kvartil	51.50	51.08
3. kvartil	52.34	51.91
IQR	0.84	0.83
Minimum	50.21	50.17
Maksimum	53.14	53.31
Raspon	2.93	3.14

**Slika.** Usporedba 1. i 2. vožnje (medijan, IQR, min, max)



**Primjer.** U slalomu se ukupan rezultat dobije na osnovu zbroja vremena dvije vožnje. Kada je jedna vožnja duža, odnosno ima veću standardnu devijaciju onda rezultat u toj vožnji značajnije utječe na ukupan rezultat.

Jedan od načina za ujednačavanje utjecaja dva rezultata je promatrati standardizirane vrijednosti i ukupan rezultat bazirati na zbroju standardiziranih vrijednosti.

Ovo ćemo demonstrirati na primjeru slalomske utrke u Kranjskoj Gori 15.3.2015.

Natjecatelj	1. vožnja	2. vožnja	1. vožnja (z)	2. vožnja (z)	Ukupno (z)	Rang
KRISTOFFERSEN Henrik	50,51	50,75	-1,720	-1,195	-2,915	1
RAZZOLI Giuliano	50,40	51,10	-1,855	-0,644	-2,499	2
HARGIN Mattias	50,21	51,91	-2,089	0,631	-1,458	5
LARSSON Markus	51,20	51,08	-0,871	-0,676	-1,546	4
KHOROSHILOV Alexander	50,81	51,52	-1,351	0,017	-1,334	6
HIRSCHER Marcel	52,22	50,17	0,385	-2,108	-1,723	3
PINTURAUULT Alexis	51,57	51,16	-0,415	-0,550	-0,965	7
FOSS-SOLEVAAG Sebastian	51,37	51,41	-0,661	-0,156	-0,818	8
NEUREUTHER Felix	51,50	51,31	-0,501	-0,314	-0,815	9
OPFER Fritz	51,55	51,31	-0,440	-0,314	-0,753	11
YULE Daniel	51,91	51,08	0,003	-0,676	-0,672	12
LIZEROUX Julien	52,27	50,74	0,447	-1,211	-0,764	10
MOELGG Manfred	51,56	51,49	-0,427	-0,030	-0,458	13
ZAMPA Adam	52,34	50,89	0,533	-0,975	-0,442	14
MUFFAT-JEANDET Victor	51,90	51,61	-0,009	0,158	0,150	16
MYHRER Andre	52,83	50,74	1,136	-1,211	-0,075	15
HERBST Reinfried	51,94	51,86	0,040	0,552	0,592	17
GRANGE Jean-Baptiste	51,87	52,03	-0,046	0,820	0,774	19
LINDH Calle	52,19	51,80	0,348	0,458	0,806	21
BAECK Axel	52,23	51,80	0,397	0,458	0,855	22
NORDBOTTEN Jonathan	51,95	52,10	0,053	0,930	0,982	23
STRASSER Linus	52,78	51,32	1,074	-0,298	0,776	20
HAUGEN Leif Kristian	53,13	50,98	1,505	-0,833	0,672	18
RAICH Benjamin	52,22	52,19	0,385	1,071	1,456	24

Natjecatelj	1. vožnja	2. vožnja	1. vožnja	2. vožnja	Ukupno	Rang
			(z)	(z)		
GROSS Stefano	51,12	53,31	-0,969	2,834	1,865	<b>27</b>
AERNI Luca	52,54	52,02	0,779	0,804	1,583	<b>25</b>
SKUBE Matic	53,08	51,62	1,444	0,174	1,618	<b>26</b>
CHODOUNSKY David	52,97	51,93	1,308	0,662	1,970	28
KUERNER Miha	53,14	52,54	1,517	1,622	3,140	29

- standardizacija utječe na ukupan poredak
- standardizacija ovisi o svim rezultatima. Iznimno veliki pojedinačni rezultat pove-q'cava varijancu te smanjuje važnost utrke u ukupnom rezultatu.
- mogu se koristiti drugi oblici standardizacije (npr. IQR/medijan, standardna devijacija najboljih 15 rezultata,...)

Podaci:

<http://www.fis-ski.com/alpine-skiing/events-and-places/results/>



# Vježba.

Ponovite ovu analizu na podacima neke druge utrke svjetskog kupa.

## Rješavanje zadatka u Statistici

Za kreiranje standardiziranih skorova trebamo kopirati rezultate 1. i 2. vožnje u nove varijable (Z1 i Z2):

Edit / Variables+

- add - dodaje nove varijable
- copy - kopira varijable u nove varijable

Kreiramo novu varijablu Z ukupno (tu će se nalaziti Z1+Z2)

STATISTICA 64 - vjezba-1b.sta

File Home Edit View Format Statistics Data Mining Graphs Tools Data

Clipboard/Data: Paste, Copy, Format, Delete, Select All, Clear, Move, Screen Catcher

Find/Replace: Find, Replace, Repeat, Go To

Fill: Random Values, Copy Down, Copy Right

Standardize: Variables, Cases

Add: Variables, Cases

Copy: Variables, Cases

OLE Object: OLE, Object, Links

Feature Finder

Data: vjezba-1b.sta (2v by 29c)

	1	2
	1 vožnja	2 vožnja
1	50,51	50,75
2	50,4	51,1
3	50,21	51,91
4	51,2	51,08
5	50,81	51,52
6	52,22	50,17
7	51,57	51,16
8	51,37	51,41
9	51,5	51,31
10	51,56	51,31
11	51,91	51,08
12	52,27	50,74
13	51,56	51,49
14	52,34	50,89
15	51,9	51,61
16	52,83	50,74
17	51,94	51,86
18	51,87	52,03
19	52,19	51,8
20	52,23	51,8
21	51,95	52,1

For Help, press F1

vjezba-1b.sta

Sel OFF Weight OFF

Windows Taskbar: Start, Chrome, Word, PowerPoint, Firefox, VLC, File Explorer, Taskbar

System Tray: Volume, Network, System Clock

Data: vjezba-1b.sta (2v by 29c)

	1 1. vožnja	2 2. vožnja
1	50,51	50,75
2	50,4	51,1
3	50,21	51,91
4	51,2	51,08
5	50,81	51,52
6	52,22	50,17
7	51,57	51,16
8	51,37	51,41
9	51,5	51,31
10	51,55	51,31
11	51,91	51,08
12	52,27	50,74
13	51,56	51,49
14	52,34	50,89
15	51,6	51,2

Copy Variables ? x

From variable: 1. vožnja

To variable: 2. vožnja

Insert after: 2. vožnja

Double-click on variable field or press F2 to select from list.

Standardizacija varijable:

Standardiziramo varijable Z1 i Z2.

Sada se u njima nalaze vremena 1. i 2. vožnje.

→ Data / Standardize

STATISTICA 64 - vjezba-1b.sta

File Home Edit View Format Statistics Data Mining Graphs Tools Data Feature Finder

Recalculate  Standardize  I/I Stack  Box-Cox  ETL+  Sort  Specs  Shift  External Data+  Sampling  Direct  Input  
 Transforms  Transpose+  Filter/Recode+  Reporting Tables  Names  All Specs  Bundles  Merge  Seed  
 Rules Builder  Rank  Date/Time  Auto Filter+  Cases+  Text Labels  Variables+  Subset  Verify+  
 Transformations Cases Variables Manage Mode

Data: vjezba-1b.sta\* (4v by 29c)

	1	2	3	4
	1. vožnja	2. vožnja	Z1	Z2
10	51.55	51.31	51.55	51.31
11	51.91	51.08	51.91	51.08
12	52.27	50.74	52.27	50.74
13	51.56	51.49	51.56	51.49
14	52.34	50.89	52.34	50.89
15	51.9	51.61	51.9	51.61
16	52.83	50.74	52.83	50.74
17	51.94	51.86	51.94	51.86
18	51.87	52.03	51.87	52.03
19	52.19	51.8	52.19	51.8
20	52.23	51.8	52.23	51.8
21	51.95	52.1	51.95	52.1
22	52.78	51.32	52.78	51.32
23	53.13	50.98	53.13	50.98
24	52.22	52.19	52.22	52.19
25	51.12	53.31	51.12	53.31
26	52.54	52.02	52.54	52.02
27	53.08	51.62	53.08	51.62
28	52.97	51.93	52.97	51.93
29	53.14	52.54	53.14	52.54

Standardization of Values ?

Variables: NONE

Cases: ALL

Weight: OFF

Values of the selected variables will be standardized (within variables). Non-selected cases will be excluded from all calculations.

OK Cancel

Select variables ?

1 - 1. vožnja  
2 - 2. vožnja  
3 - Z1  
4 - Z2

OK  
Cancel  
[Bundles]...

Select All Spread Zoom

Select variables for standardization:  
3-4

Show appropriate variables only

For Help, press F1

vjezba-1b.sta Sel OFF Weight OFF

Windows taskbar icons: Start, Chrome, Word, PowerPoint, Firefox, VLC, Steam, Taskbar Search, System Tray (Volume, Network, Power, Time)

Računanje zbroja standardiziranih skorova:

Kreiramo novu varijablu Z ukupno.

Računamo  $Z1+Z2 \rightarrow$  Data / Transforms

STATISTICA 64 - vježba-1b.sta

File Home Edit View Format Statistics Data Mining Graphs Tools Data

Recalculate Standardize Stack Box-Cox ETL Sort Specs Shift External Data Sampling Direct Input  
 Transforms Recode Transpose Filter/Recode Reporting Tables Names All Specs Bundles Merge Seed  
 Rules Builder Rank Date/Time Auto Filter Transformations Cases Text Labels Variables Subset Verify Mode

Data: vježba-1b.sta\* (5v by 29c)

	1 1. vožnja	2 2. vožnja	3 Z1	4 Z2	5 Z ukupno
1	50.51	50.75	-1.720	-1.195	
2	50.4	51.1	-1.855	-0.644	
3	50.21	51.91	-2.089	0.631	
4	51.2	51.08	-0.871	-0.676	
5	50.81	51.52	-1.351	0.017	
6	52.22	50.17	0.385	-2.108	
7	51.57	51.16	-0.415	-0.550	
8	51.37	51.41	-0.661	-0.156	
9	51.5	51.31	-0.501	-0.314	
10	51.55	51.31	-0.440	-0.314	
11	51.91	51.08	0.003	-0.676	
12	52.27	50.74	0.447	-1.211	
13	51.56	51.49	-0.427	-0.030	
14	52.34	50.89	0.533	-0.975	
15	51.9	51.61	-0.009	0.158	
16	52.83	50.74	1.136	-1.211	
17	51.94	51.86	0.040	0.552	
18	51.87	52.03	-0.046	0.820	
19	52.19	51.8	0.348	0.458	
20	52.23	51.8	0.397	0.458	
21	51.95	52.1	0.053	0.930	

Batch Transformation Formulas

Formulas:  
 $Z\ ukupno = z1 + z2$

OK  
 Cancel  
 Open...  
 Save As...  
 Functions...  
 Function guide

Refer to variables by name (Gender, AGE...) or number (v1, v2, v3...). v0 is case number. Some functions support long lists of variables and/or ranges of variables. Use logical operators to create conditional expressions.

Examples:  $v1 = v2 + v3$ ; my comments (everything typed after a semicolon is interpreted as comment)  
 $Measure1 = mean(v2, v3, sort(v3), v10:v20)$   
 $NewVar1 = Hypot(v1, v2)$ ; If no variable named NewVar1 exists, create and compute a new variable  
 $TextVar$(20) = v1 + v2$ ; Create a new text variable of length 20 by concatenating text variables v1 and v2

How to handle Spreadsheet formulas:  
 Ignore (do not recalculate) Spreadsheet formulas  Recalculate all variables

Place these Batch formulas in the Spreadsheet after recalculation (replace Spreadsheet formulas if they already exist)  
 Append new variable(s) for assignments to unrecognized names on left side of equations



Računanje ranga za Z ukupno:

Varijablu Z ukupno kopiramo u novu varijablu Rang.

Računamo rangove za Z ukupno: → Data / Rank

STATISTICA 64 - vježba-1b.sta

File Home Edit View Format Statistics Data Mining Graphs Tools Data

Recalculate Standardize Stack Box-Cox ETL Sort Specs Shift External Data Sampling Direct Input  
 Transforms Recode Transpose Filter/Recode Reporting Tables Names All Specs Bundles Merge Seed  
 Rules Builder Rank Date/Time Auto Filter Text Labels Variables Subset Verify

Transformations Cases Variables Manage Mode

Data: vježba-1b.sta\* (6v by 29c)

	1 1. vožnja	2 2. vožnja	3 Z1	4 Z2	5 Z ukupno	6 Rang
1	50,51	50,75	-1,720	-1,195	-2,915	-2,915
2	50,4	51,1	-1,855	-0,644	-2,499	-2,499
3	50,21	51,91	-2,089	0,631	-1,458	-1,458
4	51,2	51,08	-0,871	-0,676	-1,546	-1,546
5	50,81	51,52	-1,351	0,017	-1,334	-1,334
6	52,22	50,17	0,385	-2,108	-1,723	-1,723
7	51,57	51,16	-0,415	-0,550	-0,965	-0,965
8	51,37	51,41	-0,661	-0,156	-0,818	-0,818
9	51,5	51,31	-0,501	-0,314	-0,815	-0,815
10	51,55	51,31	-0,440	-0,314	-0,753	-0,753
11	51,91	51,08	0,003	-0,676	-0,672	-0,672
12	52,27	50,74	0,447	-1,211	-0,764	-0,764
13	51,56	51,49	-0,427	-0,030	-0,458	-0,458
14	52,34	50,89	0,533	-0,975	-0,442	-0,442
15	51,9	51,61	-0,009	0,158	0,150	0,150
16	52,83	50,74	1,136	-1,211	-0,075	-0,075
17	51,94	51,86	0,040	0,552	0,592	0,592
18	51,87	52,03	-0,046	0,820	0,774	0,774
19	52,19	51,8	0,348	0,458	0,806	0,806
20	52,23	51,8	0,397	0,458	0,855	0,855
21	51,95	52,1	0,053	0,930	0,982	0,982

Rank Order Values

Variables: Rang

Cases: ALL

Weight: OFF

Assign rank 1 to  
 smallest value  
 largest value

Ranks for ties  
 mean  sequential  
 low  high

Type of ranks  
 regular  fractional  fractional as %

OK Cancel

STATISTICA 64 - vjezba-1b.sta

File Home Edit View Format Statistics Data Mining Graphs Tools Data Feature Finder

Arial 10 Cells

Width Uncloak Height Uncloak New from Selection  
 AutoFit Uncloak All AutoFit Uncloak All Format Manager  
 Cloak Cloak Cloak Formats\* Layouts

Format Variables Cases Block Spreadsheets

Data: vjezba-1b.sta\* (6v by 29c)

	1	2	3	4	5	6
	1. vožnja	2. vožnja	Z1	Z2	Z ukupno	Rang
1	50.51	50.75	-1.720	-1.195	-2.915	1
2	50.4	51.1	-1.855	-0.644	-2.499	2
3	50.21	51.91	-2.089	0.631	-1.458	5
4	51.2	51.08	-0.871	-0.676	-1.546	4
5	50.81	51.52	-1.351	0.017	-1.334	6
6	52.22	50.17	0.385	-2.108	-1.723	3
7	51.57	51.16	-0.415	-0.550	-0.965	7
8	51.37	51.41	-0.661	-0.156	-0.818	8
9	51.5	51.31	-0.501	-0.314	-0.815	9
10	51.55	51.31	-0.440	-0.314	-0.753	11
11	51.91	51.08	0.003	-0.676	-0.672	12
12	52.27	50.74	0.447	-1.211	-0.764	10
13	51.56	51.49	-0.427	-0.030	-0.458	13
14	52.34	50.89	0.533	-0.975	-0.442	14
15	51.9	51.61	-0.009	0.158	0.150	16
16	52.83	50.74	1.136	-1.211	-0.075	15
17	51.94	51.86	0.040	0.552	0.592	17
18	51.87	52.03	-0.046	0.820	0.774	19
19	52.19	51.8	0.348	0.458	0.806	21
20	52.23	51.8	0.397	0.458	0.855	22
21	51.95	52.1	0.053	0.930	0.982	23
22	52.78	51.32	1.074	-0.298	0.776	20
23	53.13	50.98	1.505	-0.833	0.672	18
24	52.22	52.19	0.385	1.071	1.456	24
25	51.12	53.31	-0.969	2.834	1.865	27
26	52.54	52.02	0.779	0.804	1.583	25
27	53.08	51.62	1.444	0.174	1.618	26
28	52.97	51.93	1.308	0.662	1.970	28
29	53.14	52.54	1.517	1.622	3.140	29



For Help, press F1

vjezba-1b.sta C2,V6

Z Sel OFF Weight OFF



**SVAKA ANALIZA ZAPOČINJE  
UNIVARIJATNOM ANALIZOM!**