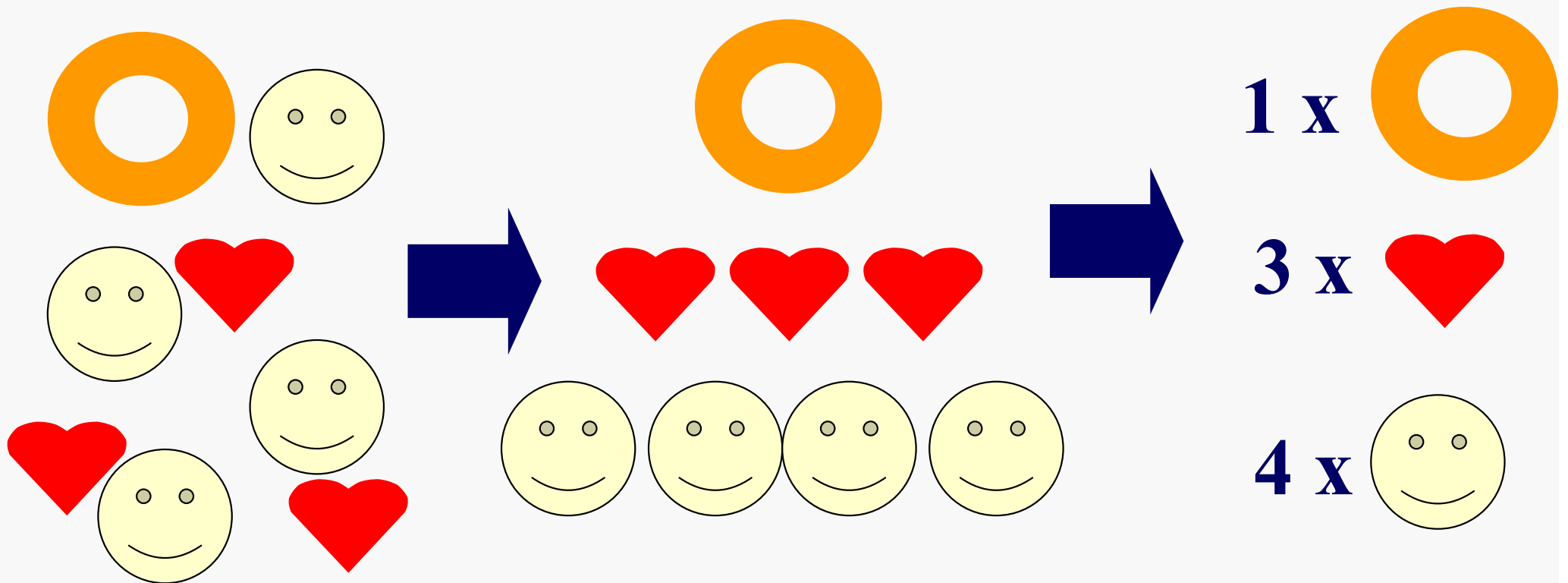
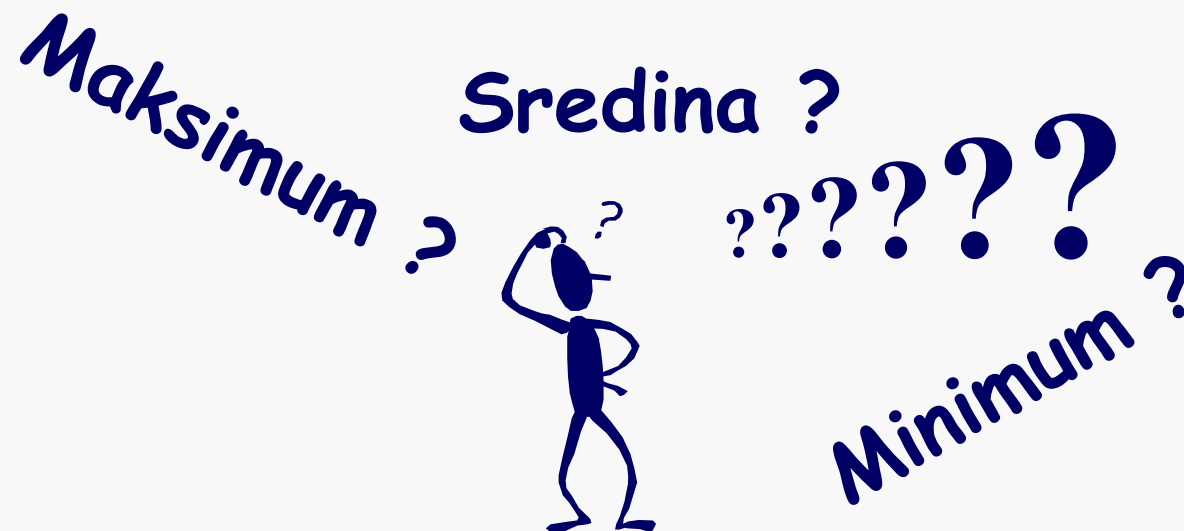


RAZDILOBA OBILJEŽJA



Na jednom čovjeku izvršeno je 50 mjerenja vremena reakcije. Dobiveni su sljedeći podaci (u tisućinkama sekunde):

196	173	186	189	173	165	167	160	140	174
180	151	157	164	154	169	190	180	163	157
169	167	165	160	177	165	157	177	159	175
166	173	185	177	184	183	162	192	174	162
165	172	158	169	146	170	171	169	168	153



TABLICA FREKVENCIJA

- tablica u kojoj su originalni podaci sažeti u određeni broj kategorija (*razreda*) koje su opisane numerički izraženim granicama

raspon (interval) razreda

– razlika granica razreda

sredina razreda

- broj koji najbolje reprezentira dani razred

računanje sredine razreda:

- *diskretne varijable:*

suma granica razreda / 2

- *kontinuirane varijable:*

suma donjih granica razreda / 2

apsolutna frekvencija razreda(f)

- broj podataka koji pripadaju intervalu tog razreda

kumulativna frekvencija razreda(cf)

- broj podataka čija je vrijednost manja ili jednaka gornjoj granici razreda

relativna frekvencija razreda (rf)

- apsolutna frekvencija razreda podijeljena s ukupnim brojem podataka

kumulativna relativna frekvencija razreda (crf)

- kumulativna frekvencija razreda podijeljena s ukupnim brojem podataka

za dani razred:

- **apsolutna frekvencija:**
 - koliko mjerenja ima vrijednosti iz intervala tog razreda
- **apsolutna kumulativna frekvencija:**
 - koliko mjerenja ima vrijednost manju ili jednaku gornjoj granici tog razreda
- **relativna frekvencija:**
 - koliki udio (postotak) mjerenja od ukupnog broja mjerenja ima vrijednost iz intervala tog razreda
- **kumulativna relativna frekvencija:**
 - koliki udio (postotak) mjerenja od ukupnog broja mjerenja ima vrijednost manju ili jednaku gornjoj granici tog razreda

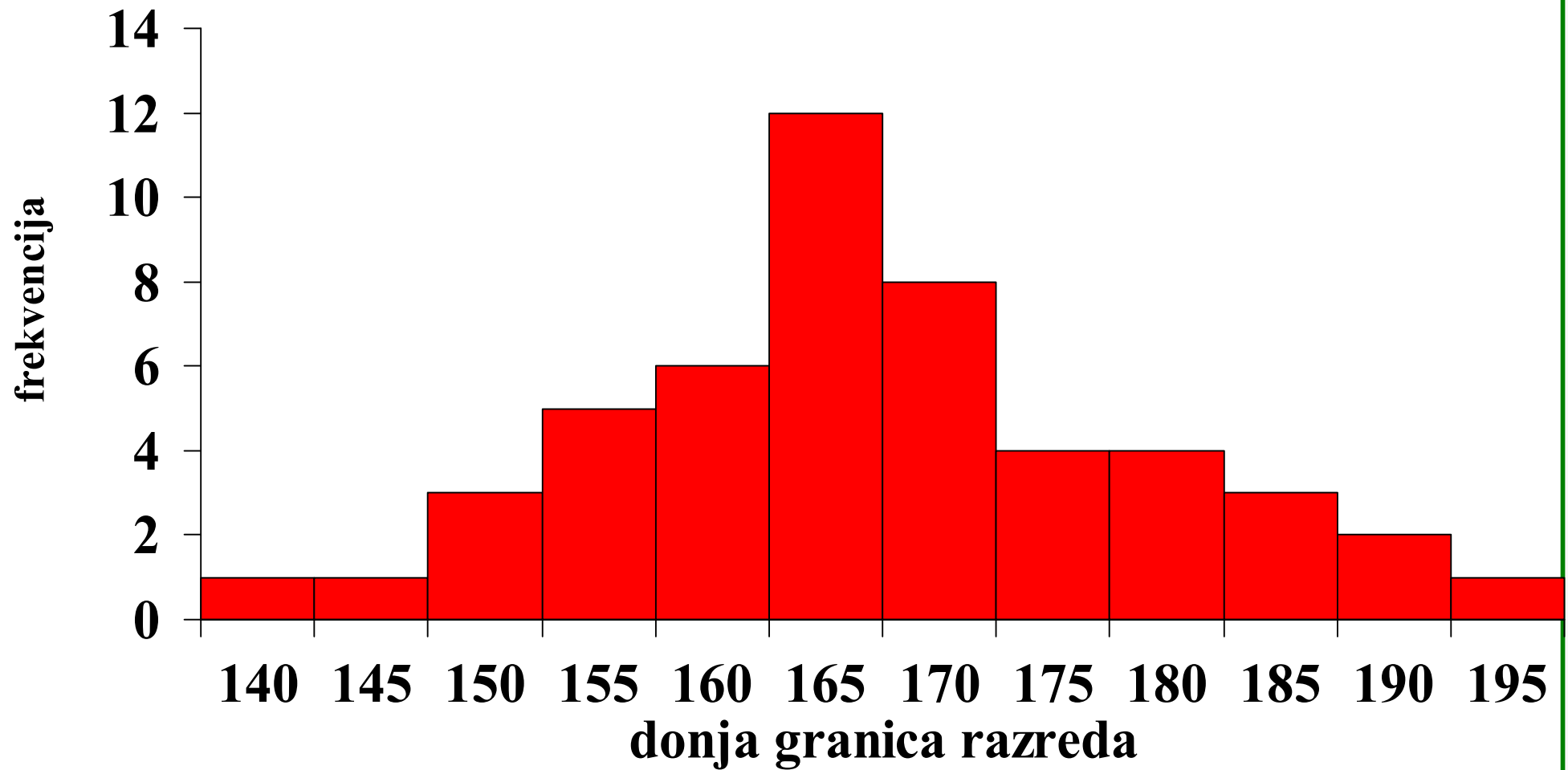
Tablica frekvencija podataka iz primjera

min = 140**max = 196**

Broj r.	Granice r.	Sredina r.	f	cf	rf	crf
1	140 – 144	142	1	1	0,02	0,02
2	145 – 149	147	1	2	0,02	0,04
3	150 – 154	152	3	5	0,06	0,10
4	155 – 159	157	5	10	0,10	0,20
5	160 – 164	162	6	16	0,12	0,32
6	165 – 169	167	12	28	0,24	0,56
7	170 – 174	172	8	36	0,16	0,72
8	175 – 179	177	4	40	0,08	0,80
9	180 – 184	182	4	44	0,08	0,88
10	185 – 189	187	3	47	0,06	0,94
11	190 – 194	192	2	49	0,04	0,98
12	195 – 199	197	1	50	0,02	1,00
UKUPNO :			50		1,00	

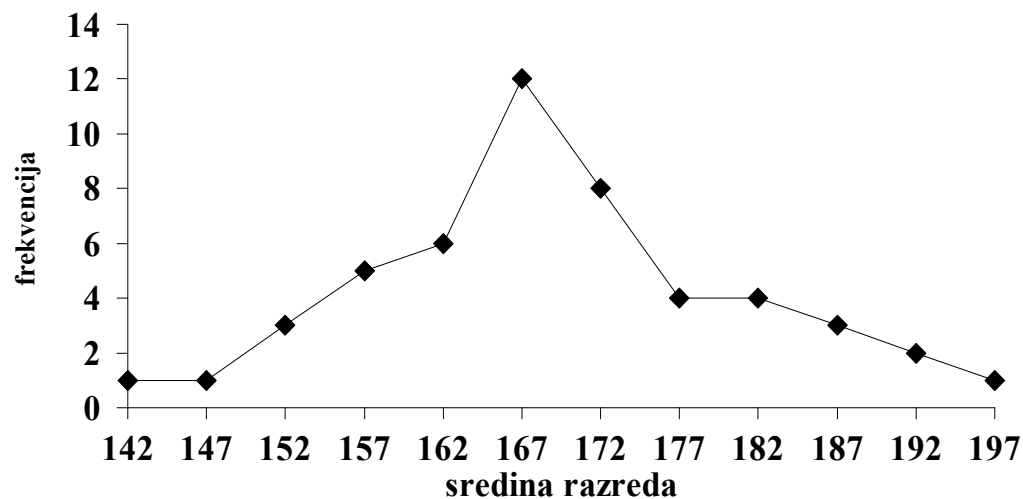
RAZDIoba FREKVENCIJA

HISTOGRAM FREKVENCIJA

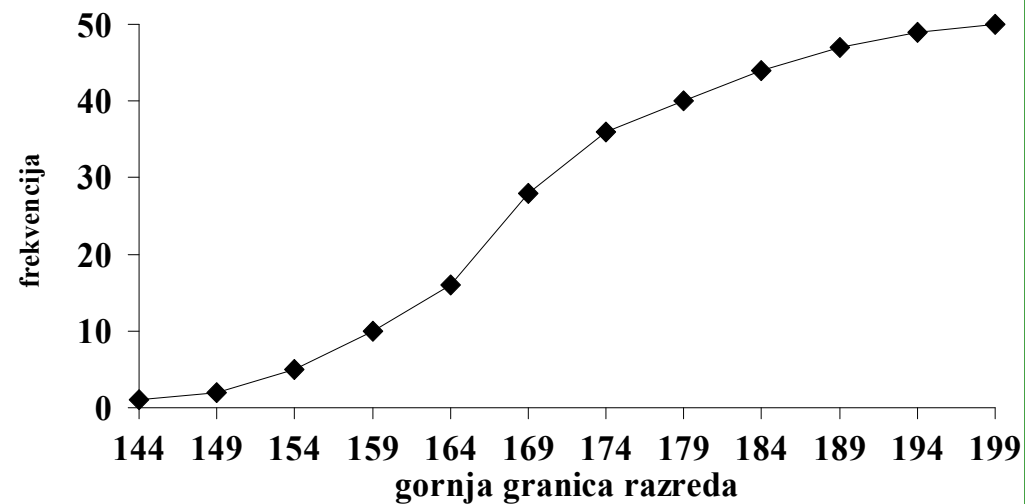


GRAFIČKI PRIKAZ RAZDIOBE

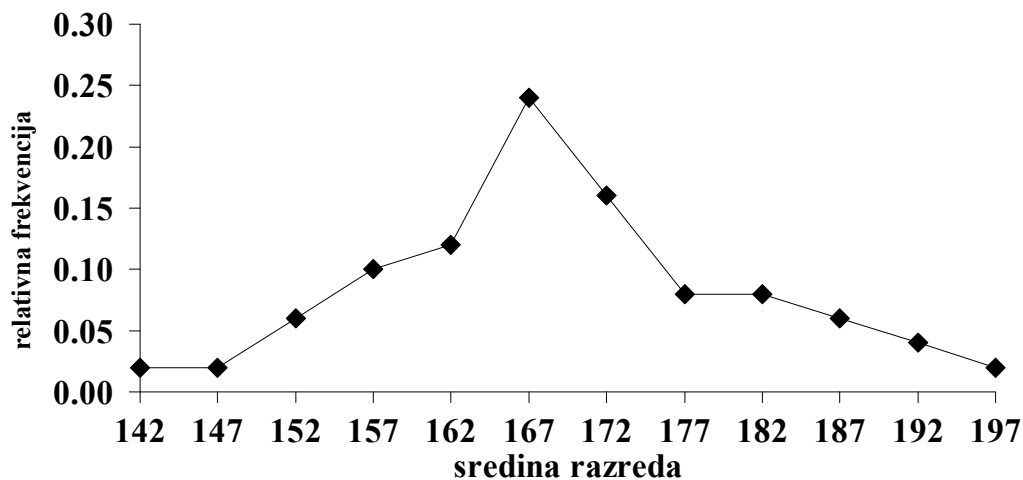
POLIGON FREKVENCIJA



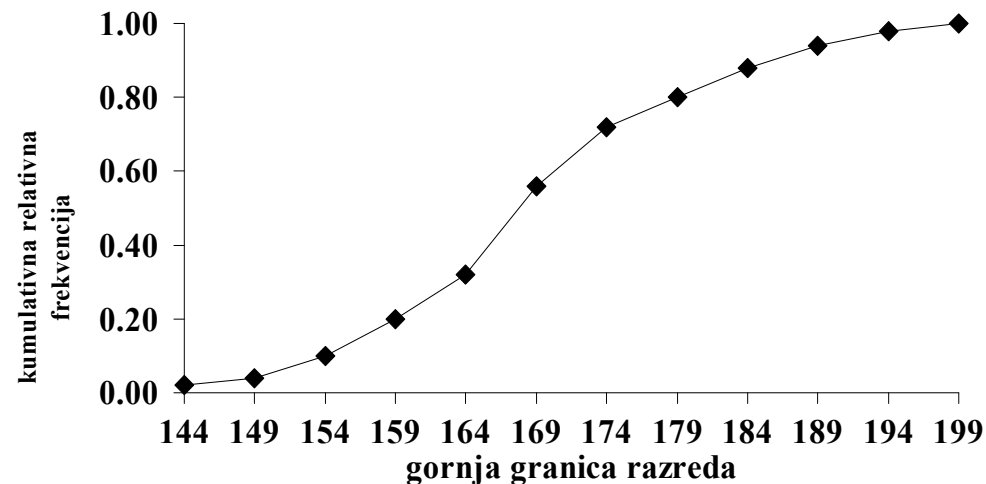
POLIGON KUMULATIVNIH FREKVENCIJA



POLIGON RELATIVNIH FREKVENCIJA



POLIGON KUMULATIVNIH RELATIVNIH FREKVENCIJA



Stablo i list (stem-and-leaf)

f	stablo	list
1.00	14 .	0
1.00	14 .	6
3.00	15 .	134
5.00	15 .	77789
6.00	16 .	002234
12.00	16 .	555567789999
8.00	17 .	01233344
4.00	17 .	5777
4.00	18 .	0034
3.00	18 .	569
2.00	19 .	02
1.00	19 .	6

PAŽNJA !

– broj razreda

- **preveliki broj razreda => male frekvencije ili prazni razredi**
- **premali broj razreda => razredi jako sažeti => izgubljeno puno informacija**
- **uobičajeno: 10-20 razreda (ovisno o broju i prirodi podataka)**
- **kod nominalnih varijabli:**

broj kategorija=broj razreda

PAŽNJA !

– granice razreda

- najmanje na onoj točnosti na kojoj je izvršeno mjerenje
- određene tako da **SVAKI PODATAK PADNE U SAMO JEDAN OD RAZREDA!**

OPIŠIVANJE RAZDIOBE PODATAKA

PARAMETAR I STATISTIKA

POPULACIJA



1.UZORAK



2.UZORAK

▪
▪
▪



n-ti UZORAK

PARAMETAR I STATISTIKA



**aritmetička sredina
visine populacije
= 175.4**



**aritmetička sredina
visine 1. uzorka
= 172.2**



**aritmetička sredina
visine 2. uzorka
= 178.1**



**aritmetička sredina
visine n-tog uzorka
= 173.7**

PARAMETAR I STATISTIKA

- **parametar:**

- vrijednost (obično nepoznata) koja predstavlja neku karakteristiku populacije
- unutar populacije, parametar je nepromjenljiva vrijednost koja **NE VARIRA**

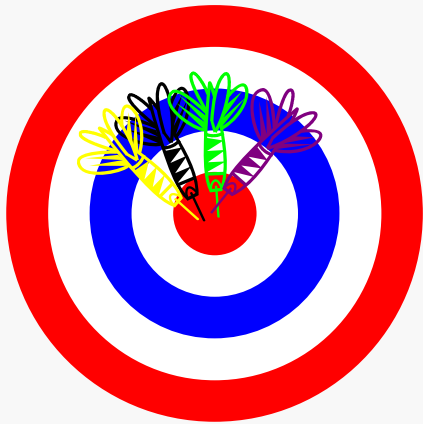
- **statistika:**

- veličina izračunata iz podataka izmjerenih na uzorku
- vrijednost statistike **MIJENJA SE** od uzorka do uzorka

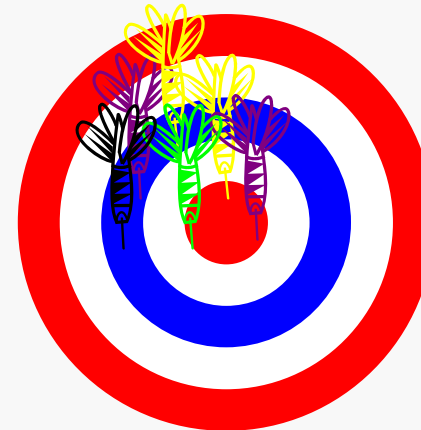
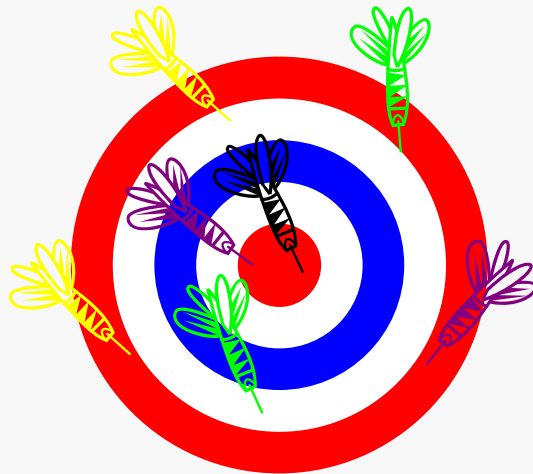
Uobičajene oznake:

	OCJENA PARAMETRA (STATISTIKA)	PARAMETAR POPULACIJE
ARITMETIČKA SREDINA	\bar{X}	μ
STANDARDNA DEVIJACIJA	s	σ
PROPORCIJA	p	π

OPISIVANJE RAZDIOBE PODATAKA



- sredina
- varijabilnost
- oblik



MJERE SREDINE (centralne tendencije)

(srednje vrijednosti, prosjeci, mjere lokacije)

ARITMETIČKA SREDINA

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

MEDIJAN (središnja vrijednost)

- vrijednost koja se u nizu podataka poredanih po veličini nalazi na srednjem mjestu
- dijeli podatke "na pola"

MOD (dominantna ili tipična vrijednost)

GEOMETRIJSKA SREDINA

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

HARMONIJSKA SREDINA

$$H = \frac{N}{\sum_{i=1}^N \frac{1}{x_i}}$$

ARITMETIČKA SREDINA

- oznake: \bar{X} uzorak μ populacija

ARITMETIČKA SREDINA INDIVIDUALNIH PODATAKA

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

x_i ... vrijednosti mjenenog obilježja

N ... ukupan broj podataka

Primjer

Kolika je aritmetička sredina niza podataka:

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

$$\begin{aligned}\bar{x} &= \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{6} = \\ &= \frac{1 + 2 + 3 + 3 + 4 + 5}{6} = \frac{18}{6} = 3\end{aligned}$$

Kolika je aritmetička sredina niza podataka:

1, 1, 1, 1, 2, 12?

$$\bar{x} = \frac{1 + 1 + 1 + 1 + 2 + 12}{6} = \frac{18}{6} = 3$$

A:	1	2	3	3	4	5
B:	1	1	1	1	2	12

aritmetička sredina niza A = 3

aritmetička sredina niza B = 3

- loše opisuje niz B**
- veliki utjecaj ekstremne vrijednosti (12)**

ARITMETIČKA SREDINA GRUPIRANIH PODATAKA

$$\mu = \frac{f_1 x_{s1} + f_2 x_{s2} + \dots + f_k x_{sk}}{N} = \frac{\sum_{i=1}^k f_i x_{si}}{N}$$

f_i ... frekvencija i-tog razreda

x_{si} ... sredina i-tog razreda

k ... broj razreda

N ... ukupan broj podataka

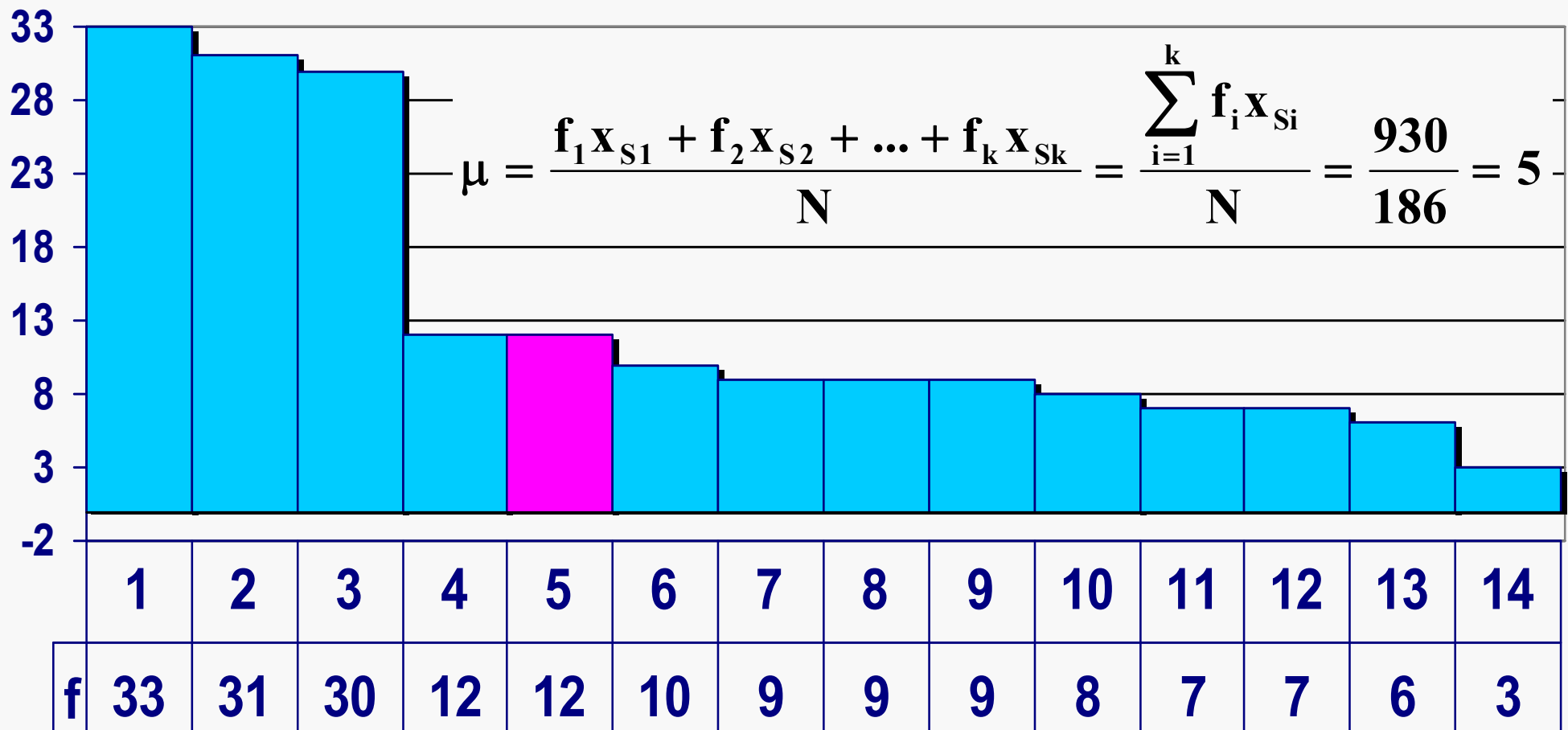
Kolika je aritmetička sredina podataka u sljedećoj tablici:

x	1	2	3	4	5	6	7	8	9	10	11	12	13	14
f	33	31	30	12	12	10	9	9	9	8	7	7	6	3
f·x	33	62	90	48	60	60	63	72	81	80	77	84	78	42

$$\sum_{i=1}^{14} f_i = 186$$

$$\sum_{i=1}^{14} f_i \cdot x_i = 930$$

$$\bar{x} = \frac{930}{186} = 5$$



➤ **razdioba frekvencija nije simetrična**
 => primjer lošeg opisivanja rezultata
 aritmetičkom sredinom

ZAJEDNIČKA ARITMETIČKA SREDINA (aritmetička sredina aritmetičkih sredina, ponderirana aritmetička sredina)

$$\bar{x}_{\text{zaj}} = \frac{\bar{x}_1 N_1 + \bar{x}_2 N_2 + \dots + \bar{x}_n N_n}{\sum_{i=1}^n N_i} = \frac{\sum_{i=1}^n \bar{x}_i N_i}{\sum_{i=1}^n N_i}$$

x_i ... aritmetička sredina dobivena iz N_i mjerenja
 n ... broj skupina mjerenja

Dva studenta studija sestrinstva postigla su sljedeći uspjeh u V semestru studija:

PREDMET	OCJENA		ECTS
	STUDENT 1	STUDENT 2	
Osnove istraživačkog rada	2	5	4
Zdravstvena njega odraslih II	5	2	9
Zdravstvena njega psihijatrijskih bolesnika	5	3	8
Klinička medicina III	3	5	4

Koji je student postigao bolji uspjeh u V semestru?

PREDMET	OCJENA		ECTS
	STUDENT 1	STUDENT 2	
Osnove istraživačkog rada	2	5	4
Zdravstvena njega odraslih II	5	2	9
Zdravstvena njega psihijatrijskih bolesnika	5	3	8
Klinička medicina III	3	5	4
Σ	15	15	25

$$\text{uspjeh}_{s_1} = \frac{2 \cdot 4 + 5 \cdot 9 + 5 \cdot 8 + 3 \cdot 4}{25} = \frac{105}{25} = 4.20$$

$$\text{uspjeh}_{s_2} = \frac{5 \cdot 4 + 2 \cdot 9 + 3 \cdot 8 + 5 \cdot 4}{25} = \frac{82}{25} = 3.28$$

U dva navrata vršeno je mjerenje neke dužine i dobiveni su slijedeći rezultati:

$$\bar{x}_1 = 20\text{cm} \quad ; \quad N_1 = 15$$

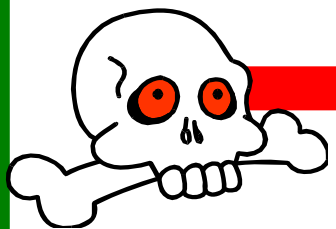
$$\bar{x}_2 = 23\text{cm} \quad ; \quad N_2 = 60$$

a) Kolika je zajednička aritmetička sredina?

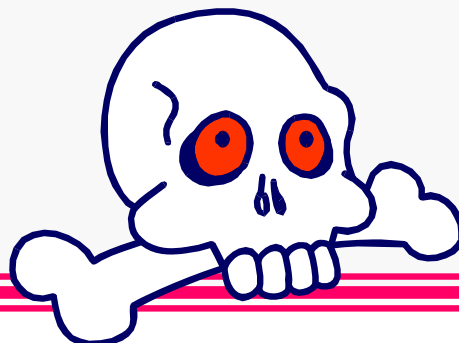
b) Kolika je zajednička aritmetička sredina za $N_1=60; N_2=15$?

$$\text{a) } \bar{x}_{\text{zaj}} = \frac{20 \cdot 15 + 23 \cdot 60}{15 + 60} = \frac{300 + 1380}{75} = 22.4 \text{ cm}$$

$$\text{b) } \bar{x}_{\text{zaj}} = \frac{20 \cdot 60 + 23 \cdot 15}{15 + 60} = \frac{1200 + 354}{75} = 20.6 \text{ cm}$$



aritmetička sredina osjetljiva je ne samo na vrijednost nego i na broj podataka



Aritmetička sredina nema smisla, tj. nije dobar reprezentant podataka ako je:

- **razdioba asimetrična**
- **broj podataka mali, a varijabilnost velika (velike razlike u vrijednostima podataka)**

Imamo niz podataka:

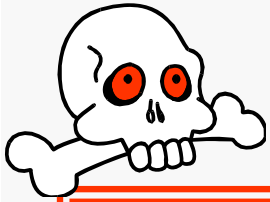
A: 2 2.5 3.5 3 4

Kolika je suma odstupanja pojedinačnih vrijednosti od aritmetičke sredine? Kolika je suma kvadrata odstupanja pojedinačnih vrijednosti od aritmetičke sredine te od vrijednosti 2; 4; 5?

$$(2-3)+(2.5-3)+(3.5-3)+(3-3)+(4-3) = -1-0.5+0.5+0+1 = 0$$

x_i	$(x_i-3)^2$	$(x_i-2)^2$	$(x_i-4)^2$	$(x_i-5)^2$
2	1	0	4	9
2.5	0.25	0.25	2.25	6.25
3	0	1	1	4
3.5	0.25	2.25	0.25	2.25
4	1	4	0	1
Σ	2.5	7.5	7.5	22.5

SVOJSTVA ARITMETIČKE SREDINE



$$\sum_{i=1}^N (x_i - \mu) = 0$$

$$\sum_{i=1}^N (x_i - \mu)^2 < \sum_{i=1}^N (x_i - a)^2, \forall a \neq \mu$$

MEDIJAN (središnja vrijednost)

- vrijednost koja se u nizu podataka poredanih po veličini nalazi točno u sredini – **središnja vrijednost po položaju**
- **vrijednost medijana:**
 - **za neparan N :** vrijednost koja se nalazi na $(N+1)/2$ mjestu
 - **za paran N :** sredina vrijednosti podataka koji se nalaze na $N/2$ i $(N+2)/2$ mjestu

MEDIJAN (središnja vrijednost)

- prednosti:
 - na vrijednost medijana ***ne utječu ekstremne vrijednosti***
 - ⇒ **pogodan** kao mjera centralne tendencije kod **asimetričnih raspodjela**

oznaka: M_e (**C**, M_d)

PRIMJER

Za nizove podataka iz primjera:

A: 1 2 3 3 4 5

B: 1 1 1 1 2 12

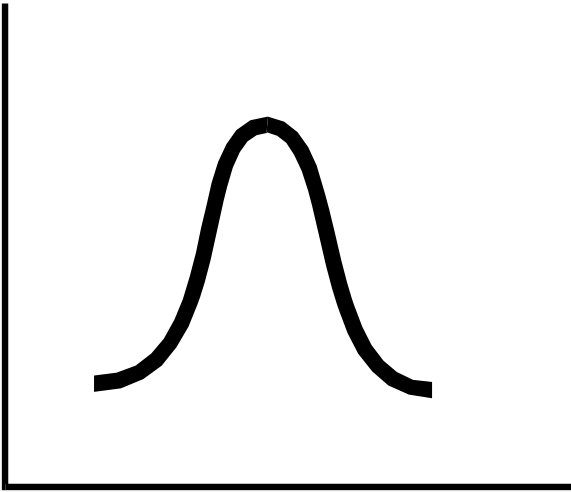
niz A: medijan.... **Me=3** arit. sred. **$\bar{X} = 3$**

niz B: medijan.... **Me=1** arit. sred. **$\bar{X} = 3$**

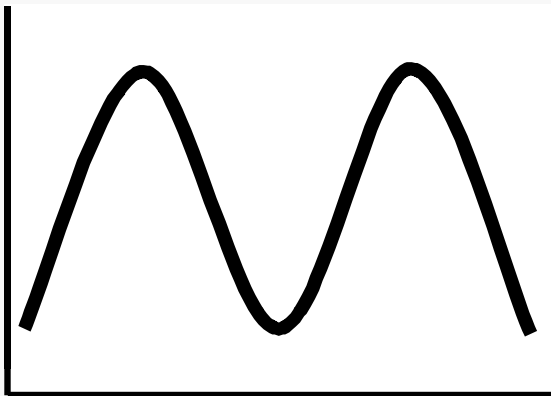
MOD (dominantna vrijednost)

- vrijednost koja se u nizu mjerenja **najčešće** javlja (dominira svojom frekvencijom)
- na mod ne utječu **ni broj ni veličina** podataka, već **samo frekvencija**

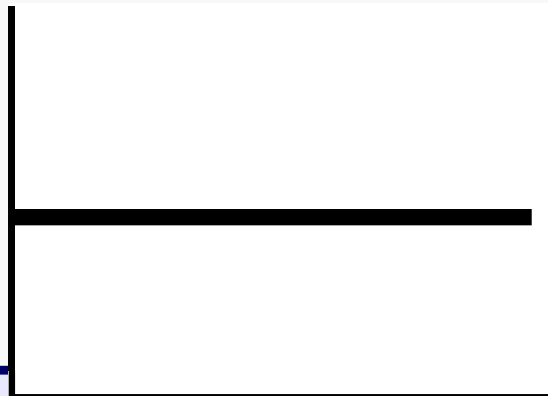
oznaka: Mo



UNIMODALNA



BIMODALNA



UNIFORMNA

PAŽNJA !

PARAMETRIJSKA MJERA SREDINE

NEPARAMETRIJSKE MJERE SREDINE

	Aritmetička sredina	Medijan	Mod
Ljestvica	Najmanje intervalna	Najmanje ordinalna	Sve ljestvice
Prisustvo ekstremnih vrijednosti	NIJE POGODNA!	Ne smetaju	Ne smetaju
Razdioba	Normalna ili najmanje simetrična	Nije važno	Unimodalna

MJERE VARIJABILNOSTI

(mjere disperzije, raspršenja)

RASPON - max-min

KVANTILE - mjere varijabilnosti po položaju

- najčešće: kvartile (Q_1 - 25 %, Q_3 - 75 %)

VARIJANCA

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

STANDARDNA DEVIJACIJA

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

KOEFICIJENT VARIJABILNOSTI

$$K.V. = \frac{s}{\bar{x}} \cdot 100$$

RASPON

$$R = \max - \min$$

- nedostaci:
 - uzima u obzir samo dvije ekstremne vrijednosti koje uopće ne moraju biti karakteristične za promatranu varijablu
 - ovisi o broju opažanja (veći broj opažanja => veći raspon)

KVANTILE

- **mjere varijabilnosti po položaju**
- **kvartile, decile, centile**
- donja kvartila (Q_1 ili 25%)
 - vrijednost podatka koji stoji na centralnom mjestu polovice podataka nižih od medijana
- gornja kvartila (Q_3 ili 75%)
 - vrijednost podatka koji stoji na centralnom mjestu polovice podataka viših od medijana
- **Q_2 - medijan**

centila	obuhvat jedinica promatranja
prva	1%
druga	2%
treća	3%
....	

decila	obuhvat jedinica promatranja
prva	10%
druga	20%
treća	30%
....	

VARIJANCA

- prosječno kvadratno odstupanje od aritmetičke sredine

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

**VARIJANCA
POPULACIJE**

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

**VARIJANCA
UZORKA**

STANDARDNA DEVIJACIJA

ZA POPULACIJU

ZA UZORAK

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

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

- služi za ocjenu pojedinih rezultata oko aritmetičke sredine
- izražava se uz aritmetičku sredinu
- obično je

$$2s < \text{raspon} < 6s$$

KOEFICIJENT VARIJABILNOSTI

ZA POPULACIJU

$$K.V. = \frac{\sigma}{\mu} \cdot 100$$

ZA UZORAK

$$K.V. = \frac{s}{\bar{x}} \cdot 100$$

- relativna standardna devijacija
- govori o HOMOGENOSTI promatranog obilježja
- koristan je ako želimo znati:
 - a) razlike u varijabilnosti svojstava neke grupe ispitanika
 - b) razlike u varijabilnosti istog svojstva u različitim grupama ispitanika

MJERE ZA OCJENU OBLIKA RAZDIOBE

MOMENTI RAZDIOBE

- uzastopne mjere prosječnih odstupanja od aritmetičke sredine nultog, prvog, drugog, trećeg i višeg reda

MOMENT n-tog REDA (n-ti moment)

$$\mu_n = \frac{\sum_{i=1}^N (x_i - \mu)^n}{N}$$

NULTI I PRVI MOMENT

$$\mu_0 = \frac{\sum_{i=1}^N (x_i - \mu)^0}{N} = \frac{\sum_{i=1}^N 1}{N} = \frac{N}{N} = 1$$

prvo svojstvo aritmetičke sredine

$$\mu_1 = \frac{\sum_{i=1}^N (x_i - \mu)^1}{N} = \frac{0}{N} = 0$$

DRUGI MOMENT

$$\mu_2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} = \sigma^2$$

VARIJANCA

TREĆI MOMENT

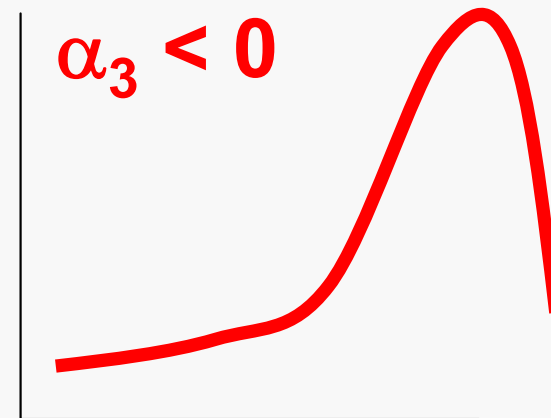
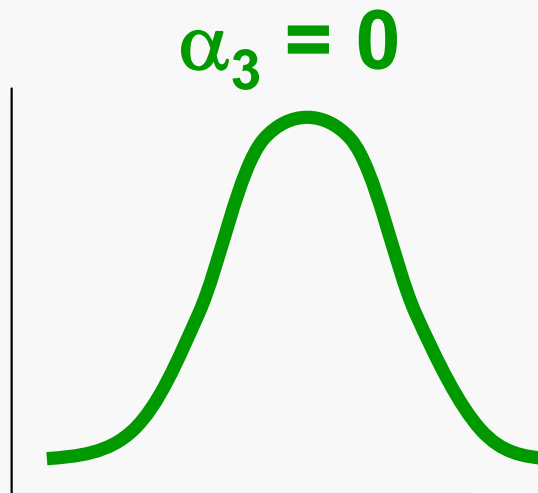
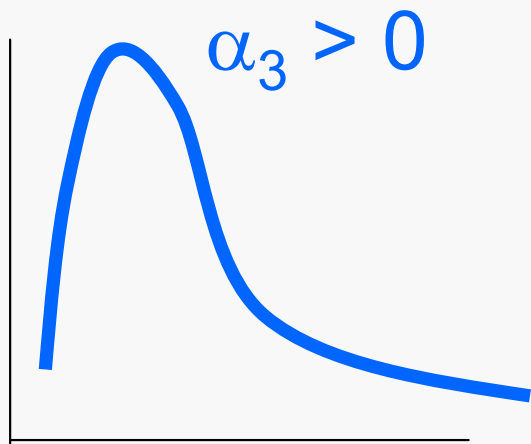
$$\mu_3 = \frac{\sum_{i=1}^N (x_i - \mu)^3}{N}$$

- za simetrične raspodjele $\mu_3 = 0$

KOEFICIJENT ASIMETRIJE (*coefficient of skewness*)

$$\alpha_3 = \frac{\mu_3}{\sigma^3}$$

- $\alpha_3 > 0$ asimetrija udesno (pozitivna asimetrija)
- $\alpha_3 < 0$ asimetrija ulijevo (negativna asimetrija)



ČETVRTI MOMENT

$$\mu_4 = \frac{\sum_{i=1}^N (x_i - \mu)^4}{N}$$

- koristi se za mjeru spljoštenosti

KOEFICIJENT SPLJOŠTENOSTI (*coefficient of kurtosis*)

$$\alpha_4 = \frac{\mu_4}{\sigma^4}$$

$\alpha_4 > 3$ raspodjela je **uža**
(**šiljatija**) od normalne
(*leptokurtic*)

$\alpha_4 = 3$ normalna spljoštenost

$\alpha_4 < 3$ raspodjela je **šira**
(**spljoštenija**) od
normalne (*platykurtic*)

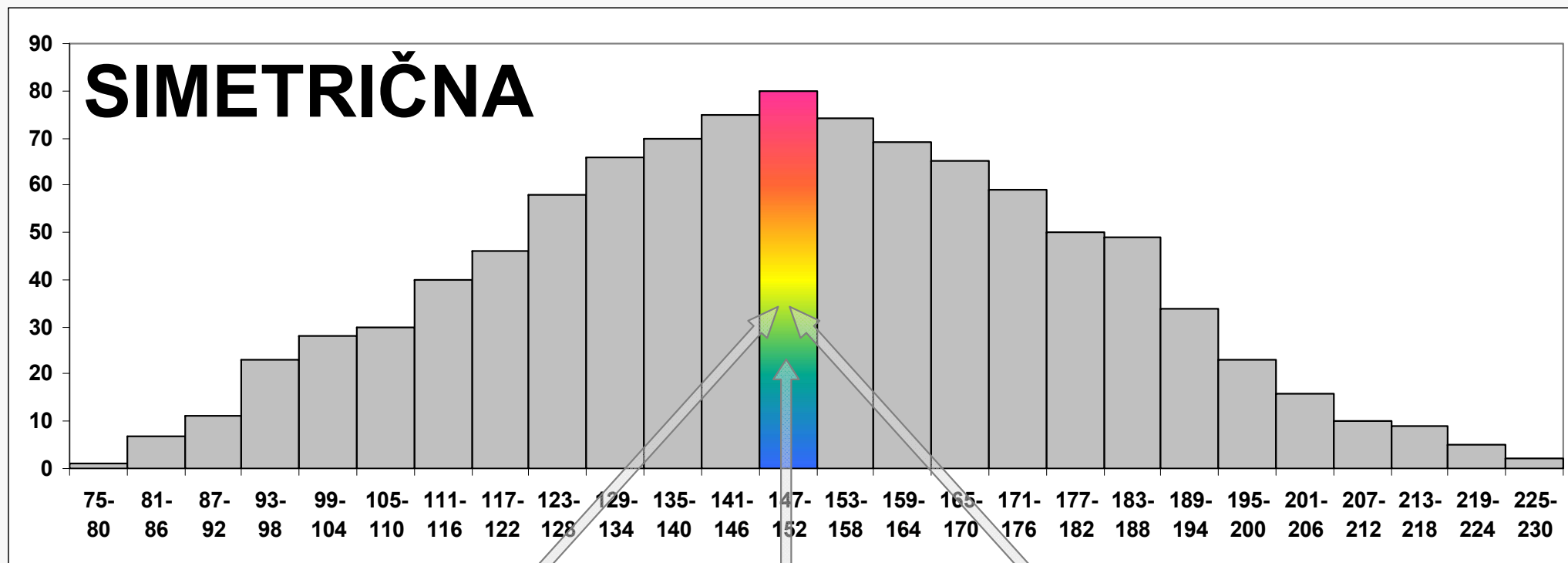
KOEFICIJENT SPLJOŠTENOSTI (*coefficient of kurtosis*)

- statistički programi koeficijent spljoštenosti prikazuju kao **eksces spljoštenosti** (kurtosis excess)

$$\text{eksces spljoštenosti} = \frac{\mu_4}{\sigma^4} - 3$$

- za normalnu raspodjelu
eksces spljoštenosti = 0

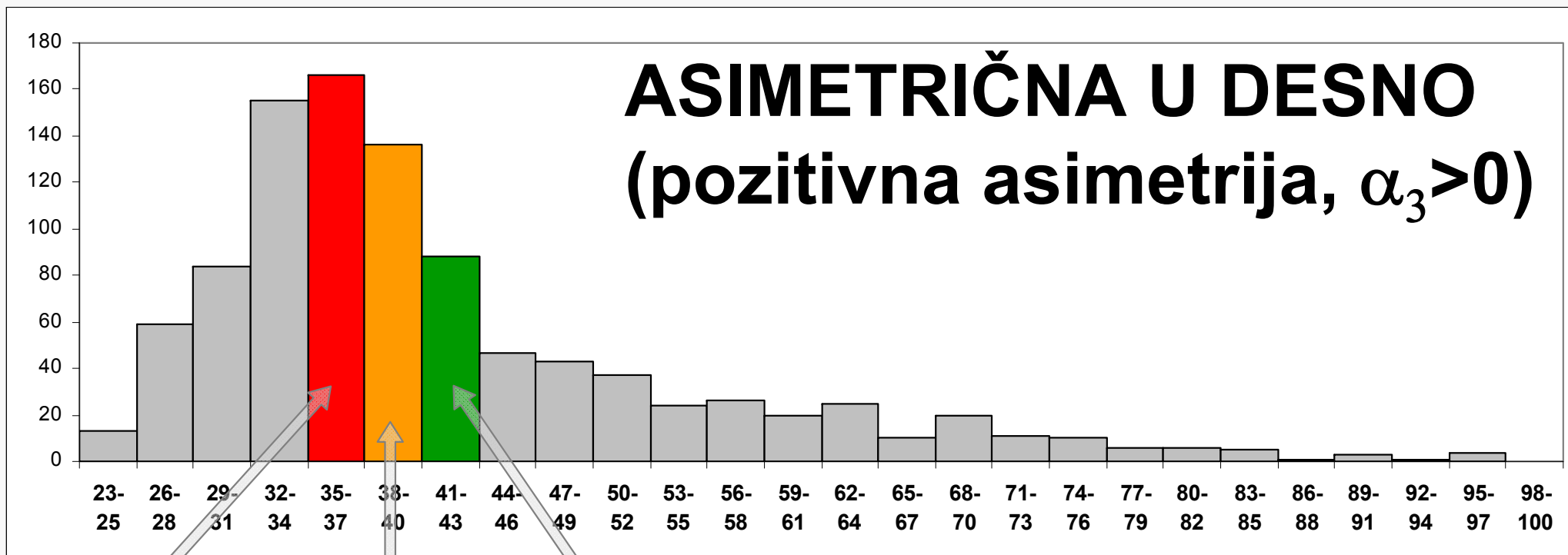
ODNOS MJERA SREDINE I ASIMETRIJE



$$\bar{X} = 150 \quad Me = 150 \quad Mo = 150$$

$$\bar{X} = Me = Mo$$

ODNOS MJERA SREDINE I ASIMETRIJE



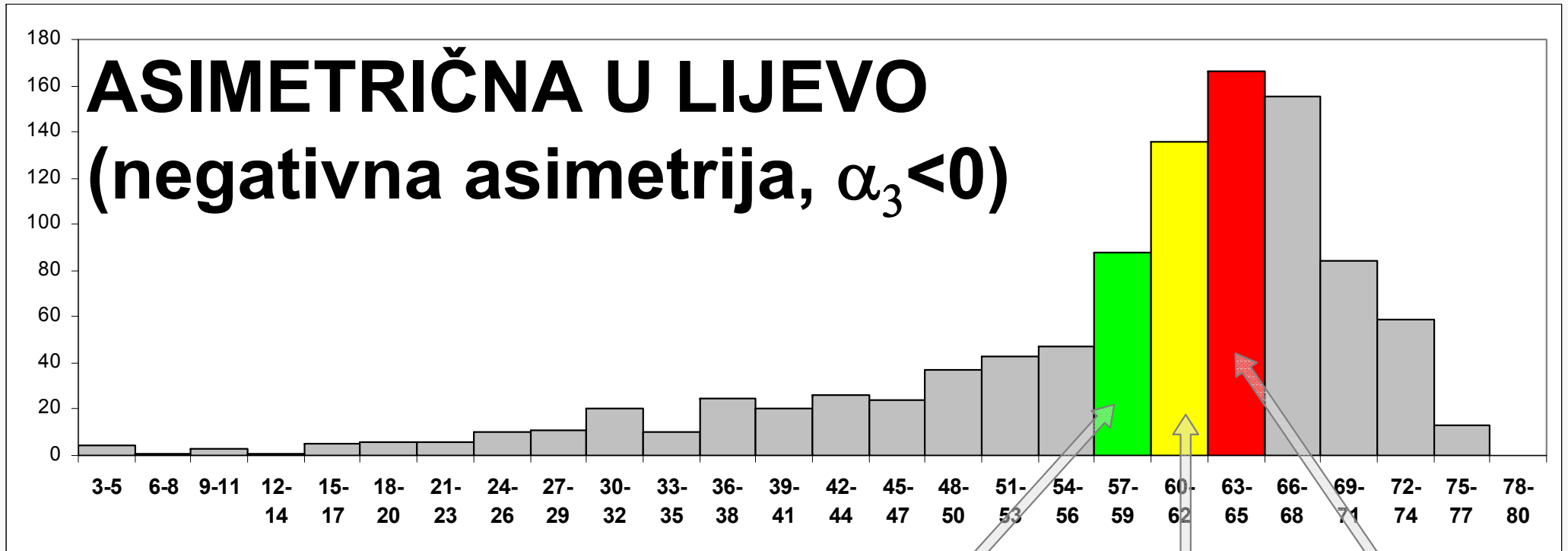
Mo = 37

$\bar{X} = 42$

Me = 38

Mo < Me < \bar{X}

ODNOS MJERA SREDINE I ASIMETRIJE



$$\bar{X} = 58$$

$$Mo = 63$$













$$Me = 62$$

$$\bar{x} < Me < Mo$$

MJERE SREDINE PREMA LJESTVICI MJERENJA

Ljestvica mjerenja	MOD	MEDIJAN	ARITMETIČKA SREDINA
NOMINALNA	✓	✗	✗
ORDINALNA	✓	✓	✗
INTERVALNA/OMJERNA (asimetrična raspodjela)	✓	✓	✗
INTERVALNA/OMJERNA (simetrična raspodjela)	✓	✓	✓

MJERE VARIJABILNOSTI PREMA LJESTVICI MJERENJA

Ljestvica mjerenja	RASPON	KVARTILE	VARIJANCA
			STANDARDNA DEVIJACIJA KOEFICIJENT VARIJABILNOSTI
NOMINALNA			
ORDINALNA			
INTERVALNA/OMJERNA (asimetrična raspodjela)			
INTERVALNA/OMJERNA (simetrična raspodjela)			

MJERE VARIJABILNOSTI

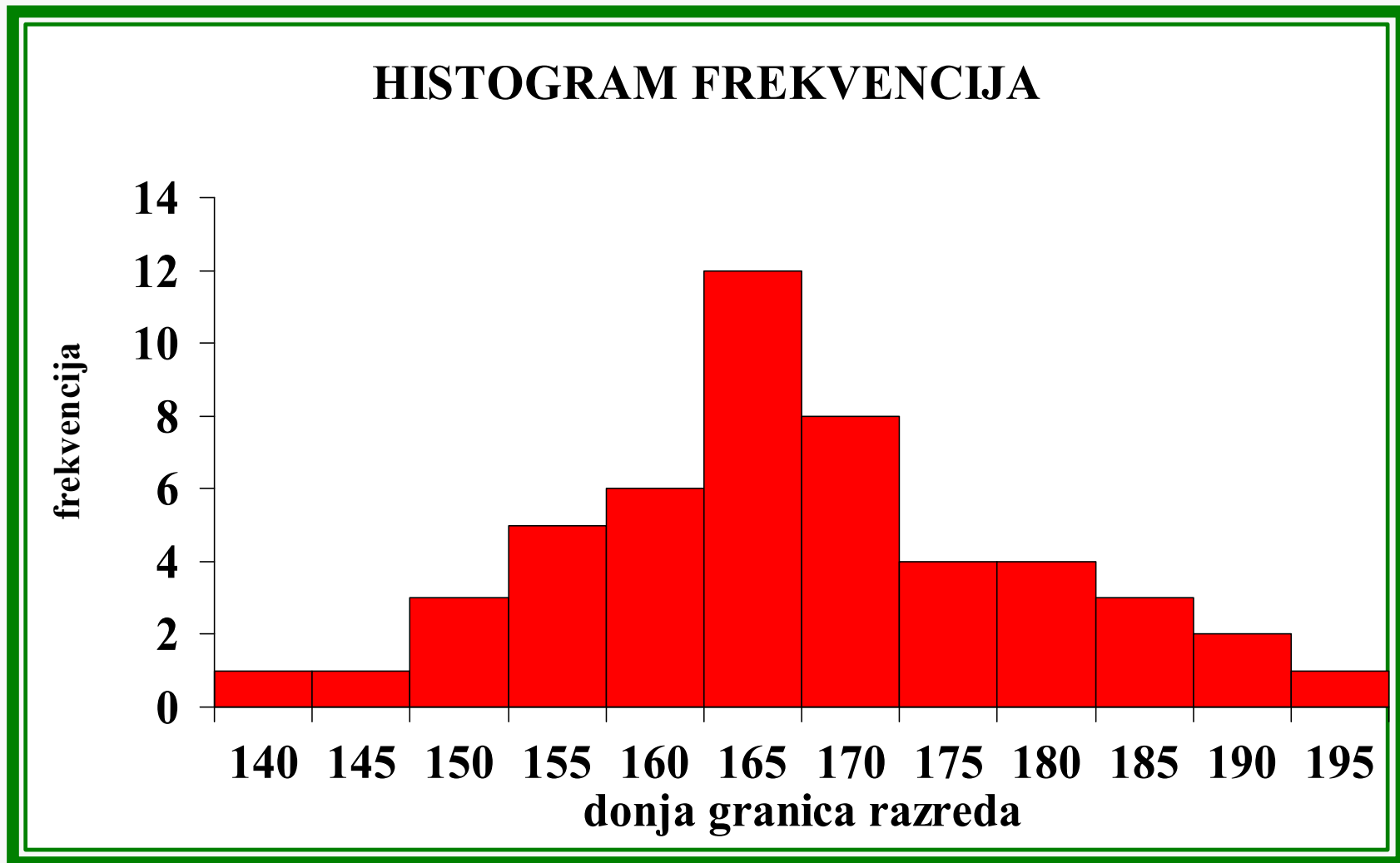
- iskazuju se uz odgovarajuću mjeru sredine

MJERA SREDINE	MJERA VARIJABILNOSTI
ARITMETIČKA SREDINA	STANDARDNA DEVIJACIJA
MEDIJAN	GRANICE INTERKVARTILNOG RASPONA (25%-75%)

GRAFIČKI PRIKAZ

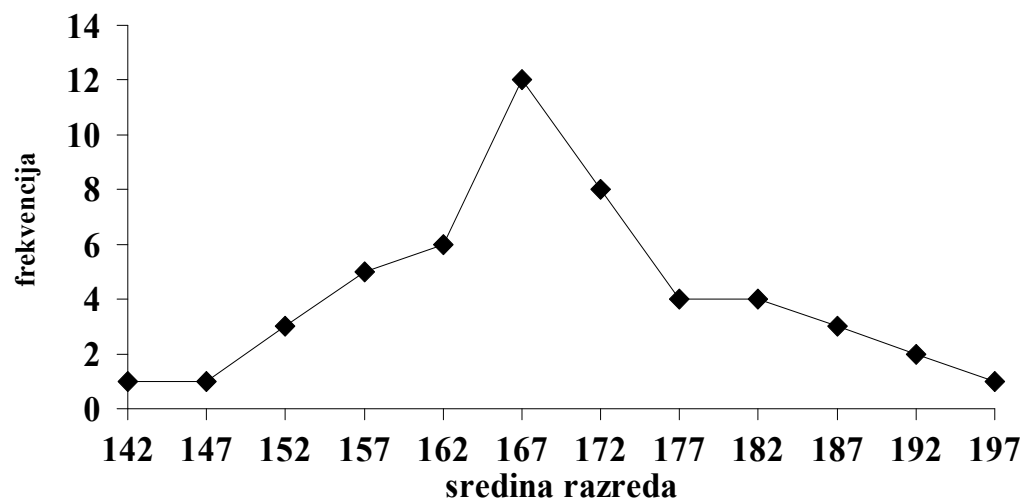
RAZDIoba OBILJEŽJA

- prikazuje se histogramom (stupičasti grafikon, “*column chart*”) ili poligonom frekvencija (linijski grafikon, “*line chart*”)

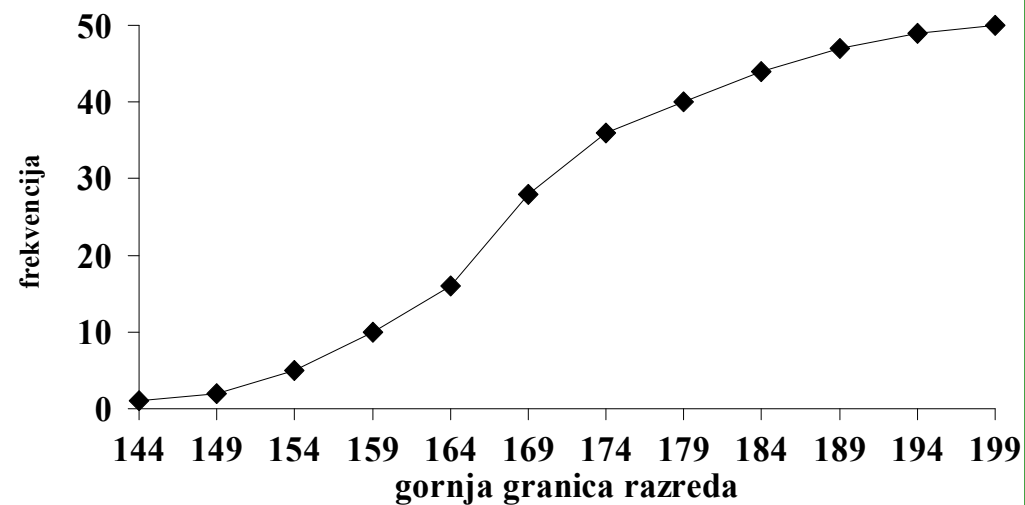


RAZDIoba OBILJEŽJA

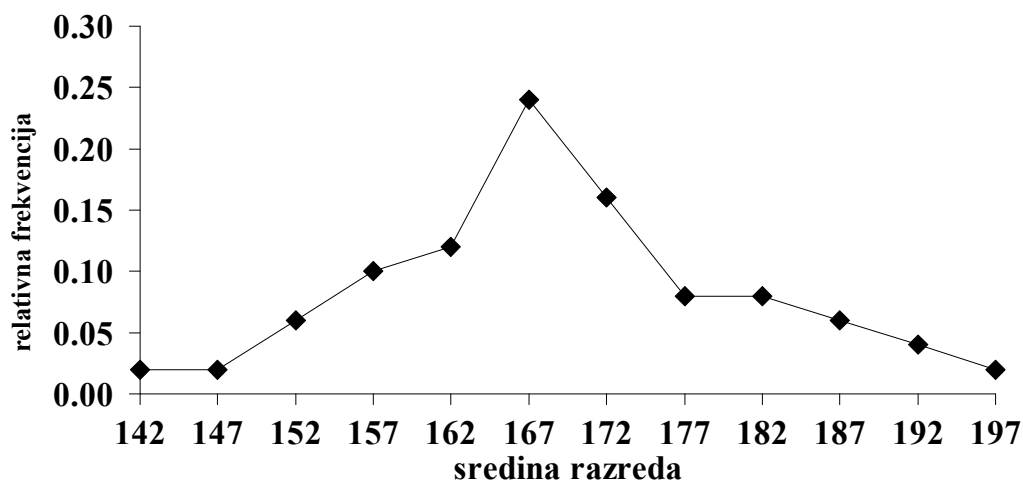
POLIGON FREKVENCIJA



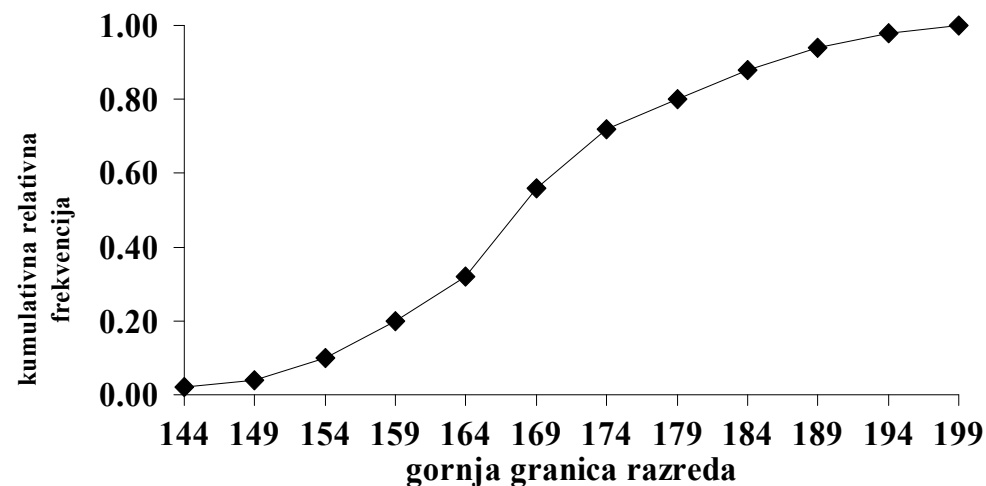
POLIGON KUMULATIVNIH FREKVENCIJA



POLIGON RELATIVNIH FREKVENCIJA

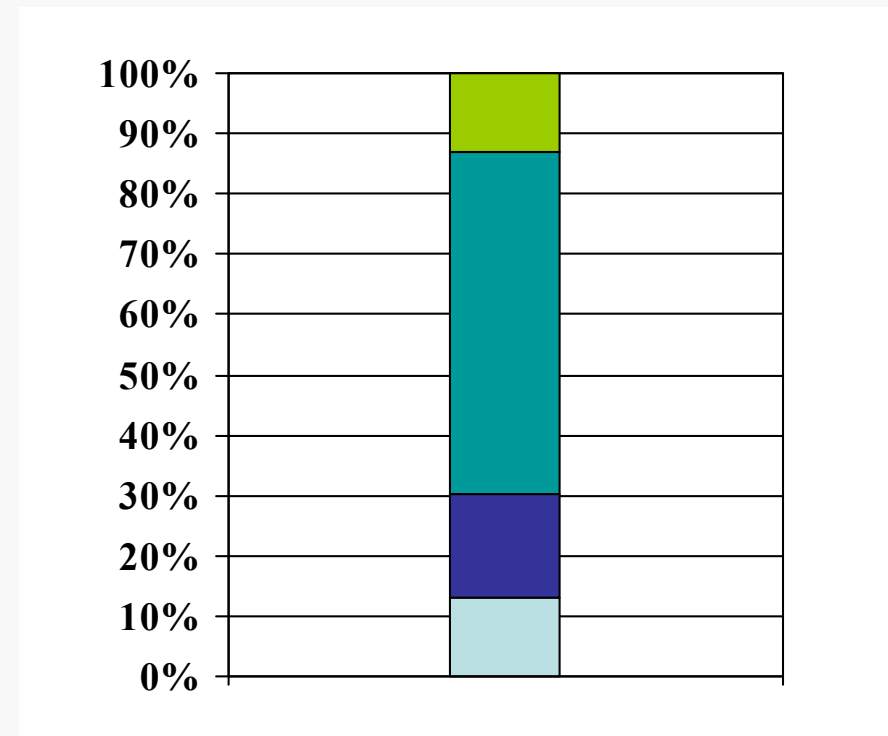
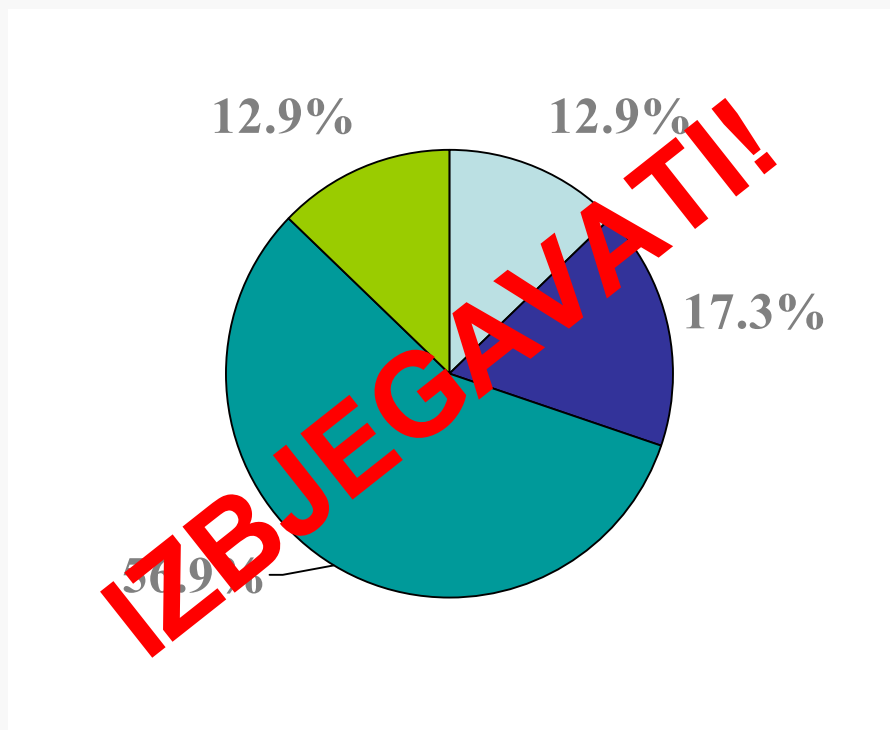


POLIGON KUMULATIVNIH RELATIVNIH FREKVENCIJA



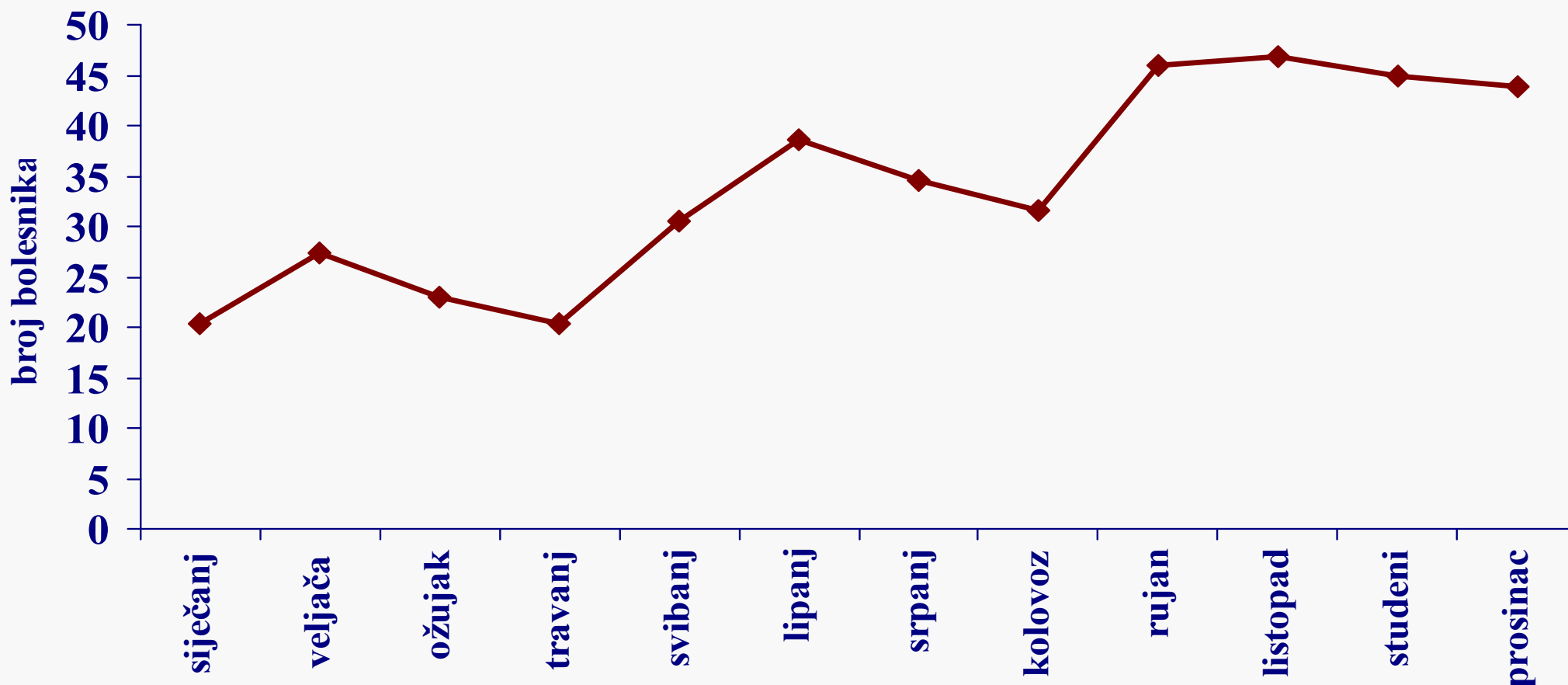
STRUKTURA OBILJEŽJA

- pokazuje udio pojedinih kategorija u ukupnom broju promatranja/mjerenja
- prikazuje se kružnim grafikonom (“torta”, “*pie chart*”) ili složenim stupičastim grafikonom (“*100% stacked column chart*”)



PROMJENE U VREMENU

- prikazuju se linijskim grafikonom (*“line chart”*)
- na apscisu se nanose vremenski intervali, a na ordinatu vrijednost promatrane varijable

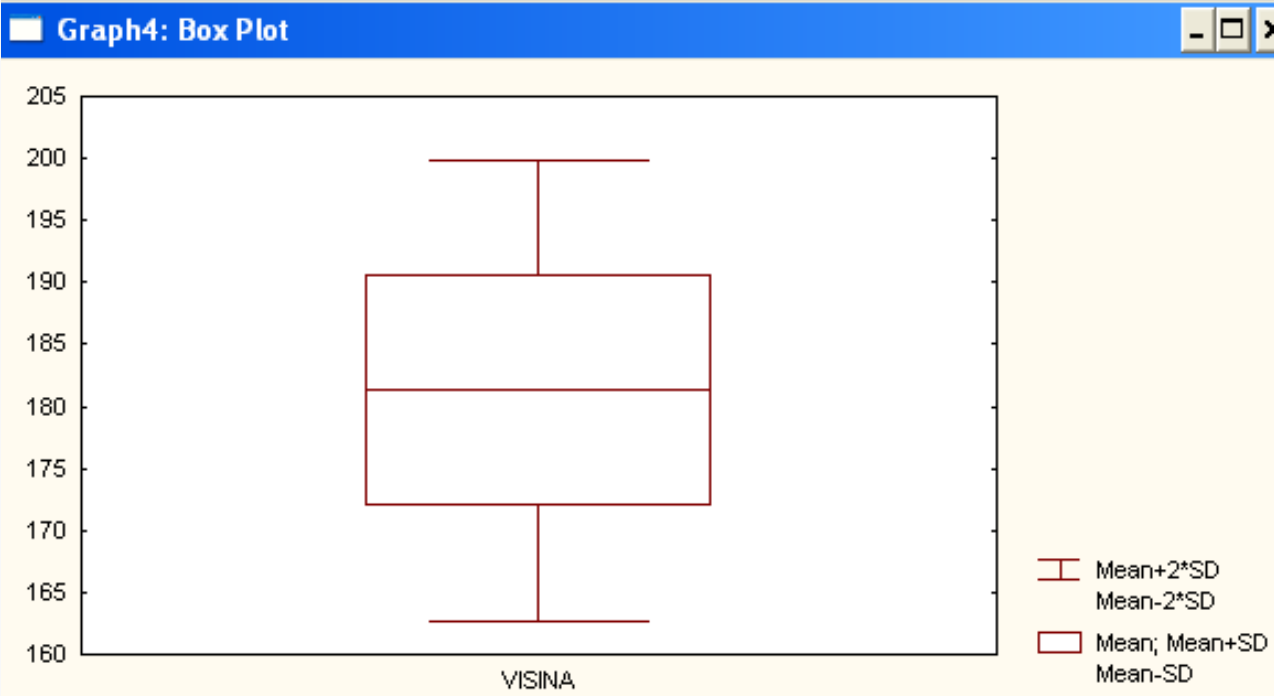
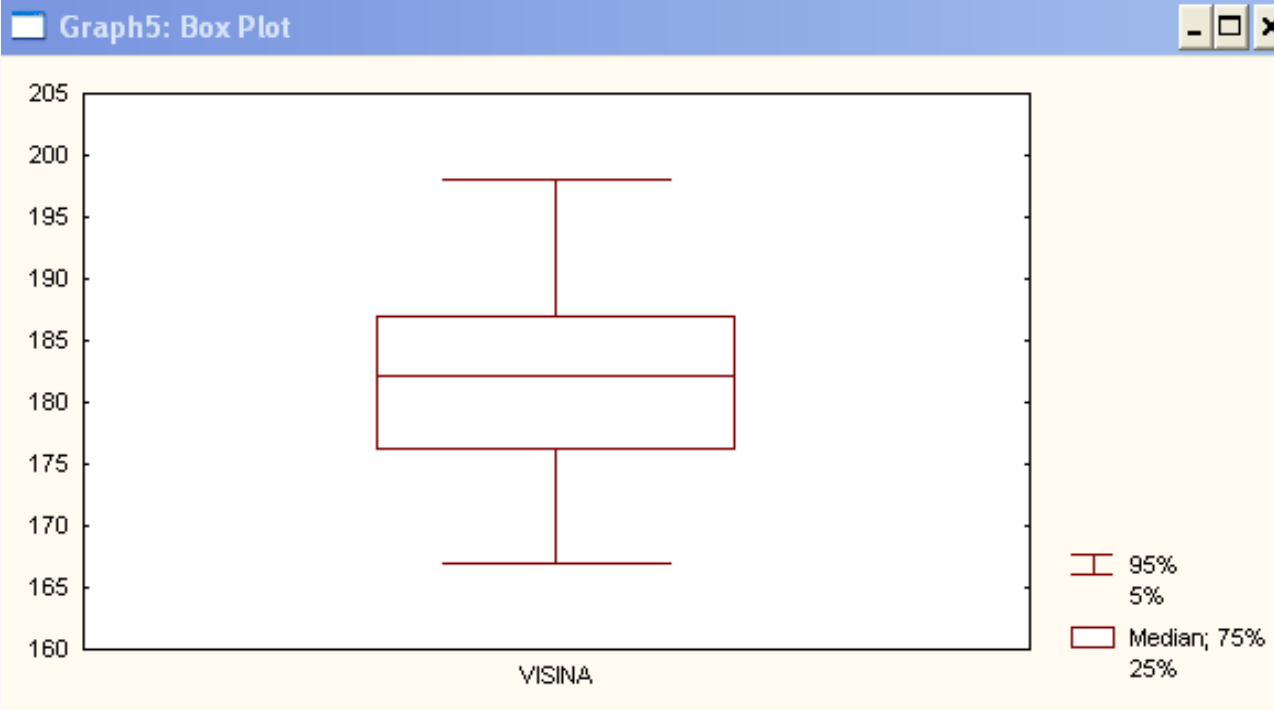


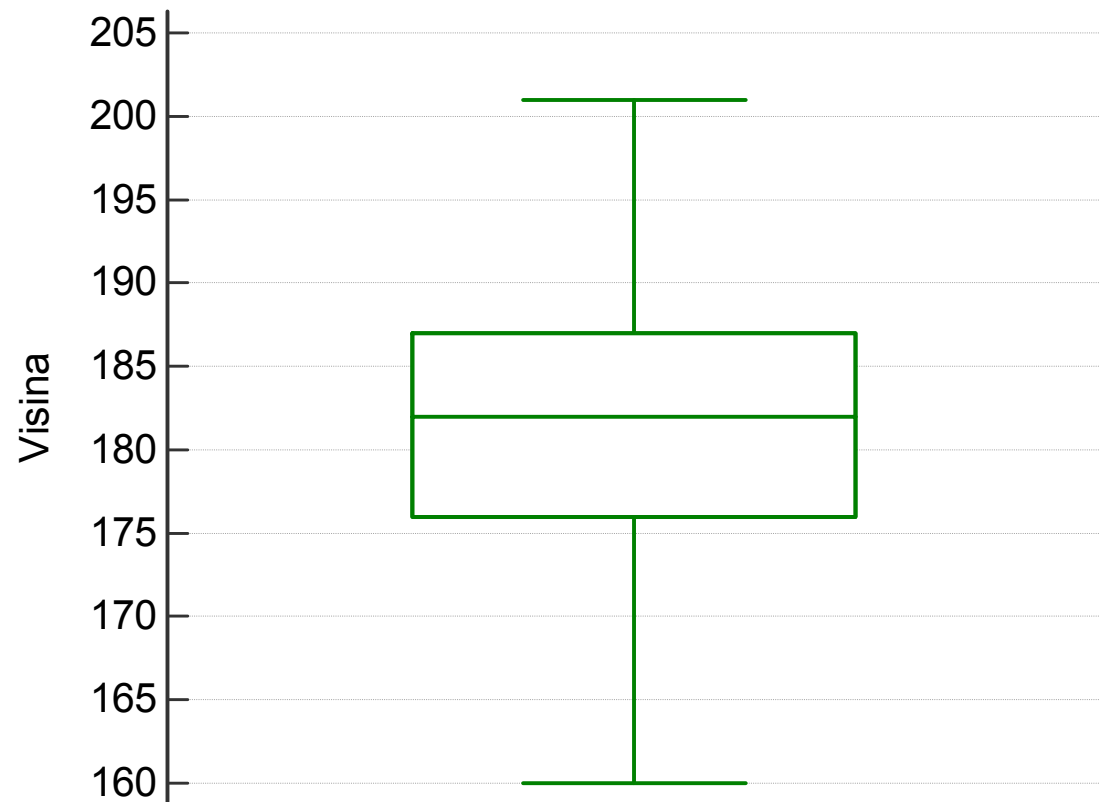
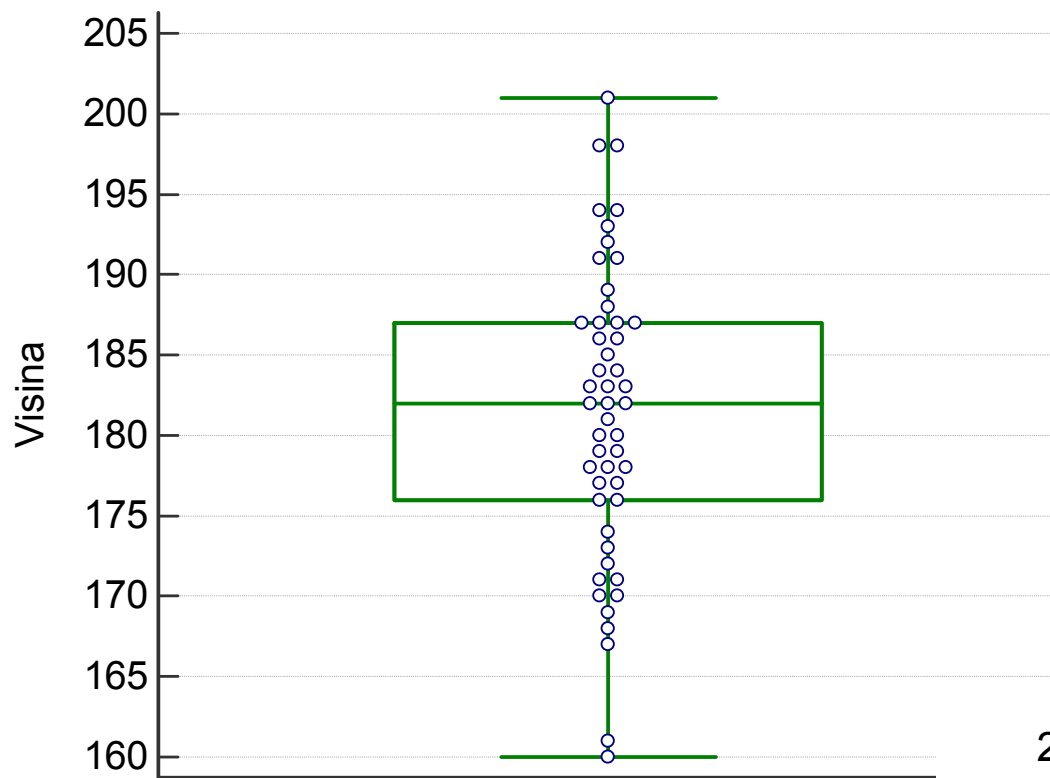
OSNOVNE MJERE SREDINE I RASPRŠENJA

- “kutija i brkovi” grafikon (“Box-and-Whisker” plot)
- najčešće prikazuje kombinacije:

sredina	kutija	brkovi
aritmetička sredina	standardna devijacija	2 SD ili raspon
medijan	25% - 75%	5% - 95% ili raspon

Statistica





POVEZANOST DVIJU VARIJABLI

- raspršni grafikon (korelacijski, “scatter graph”)
- svaka točka predstavlja par vrijednosti promatranih varijabli

