

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**

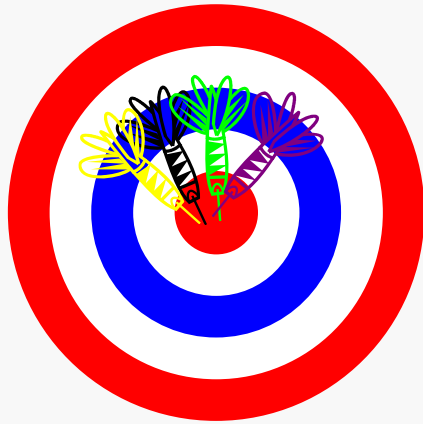
- **statistika:**

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

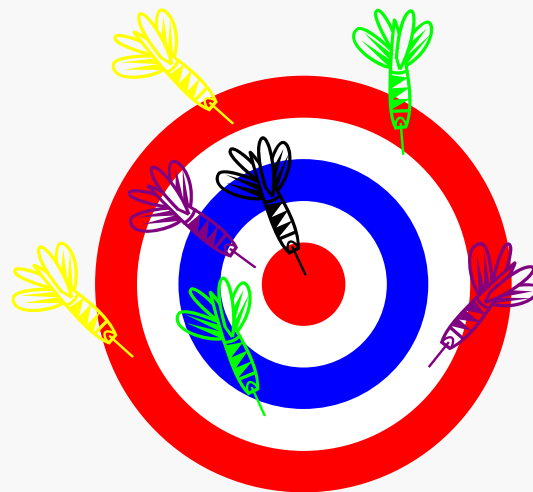
Uobičajene oznake:

	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*)

- karakteristične vrijednosti oko kojih se grupiraju mjereni podaci

ARITMETIČKA SREDINA

MEDIJAN (središnja vrijednost)

MOD (dominantna ili tipična vrijednost)

GEOMETRIJSKA SREDINA

HARMONIJSKA SREDINA

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?

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

Imamo dva niza podataka:

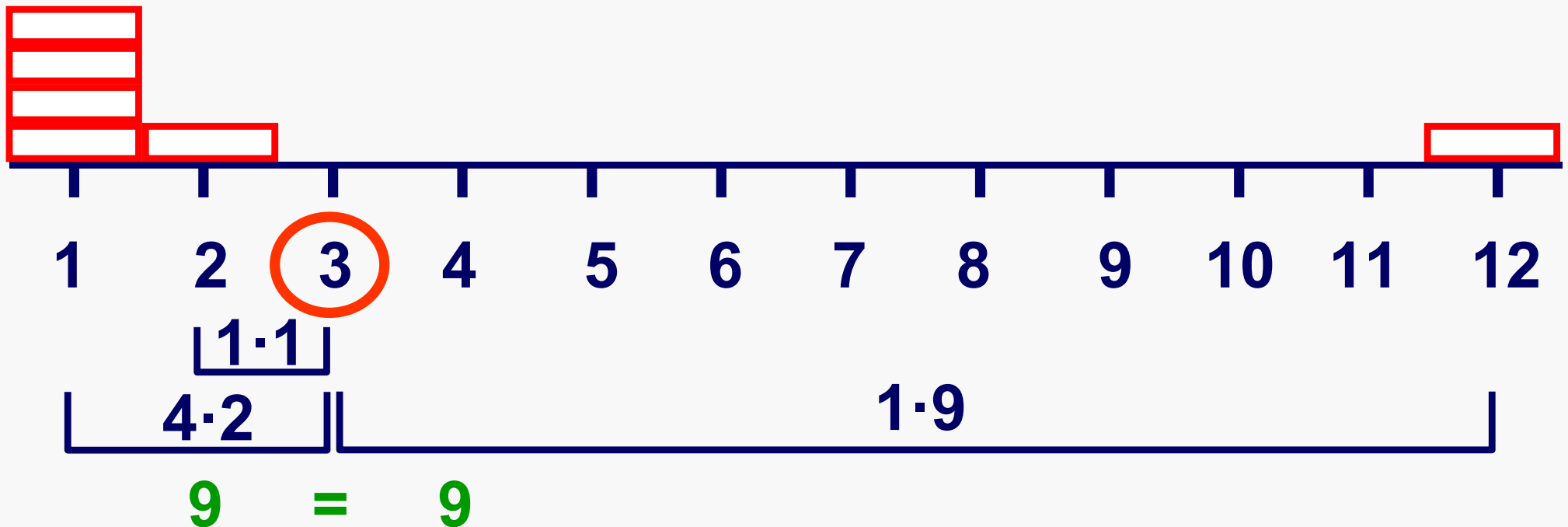
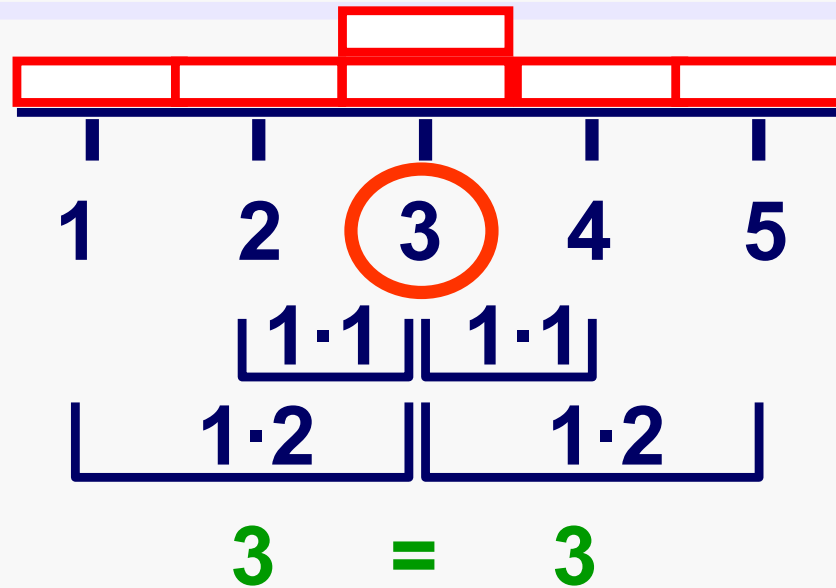
A:	1	2	3	3	4	5	
B:	1	1	1	1	1	2	12

aritmetička sredina niza A = 3

aritmetička sredina niza B = 3

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

- aritmetička sredina je težište distribucije



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

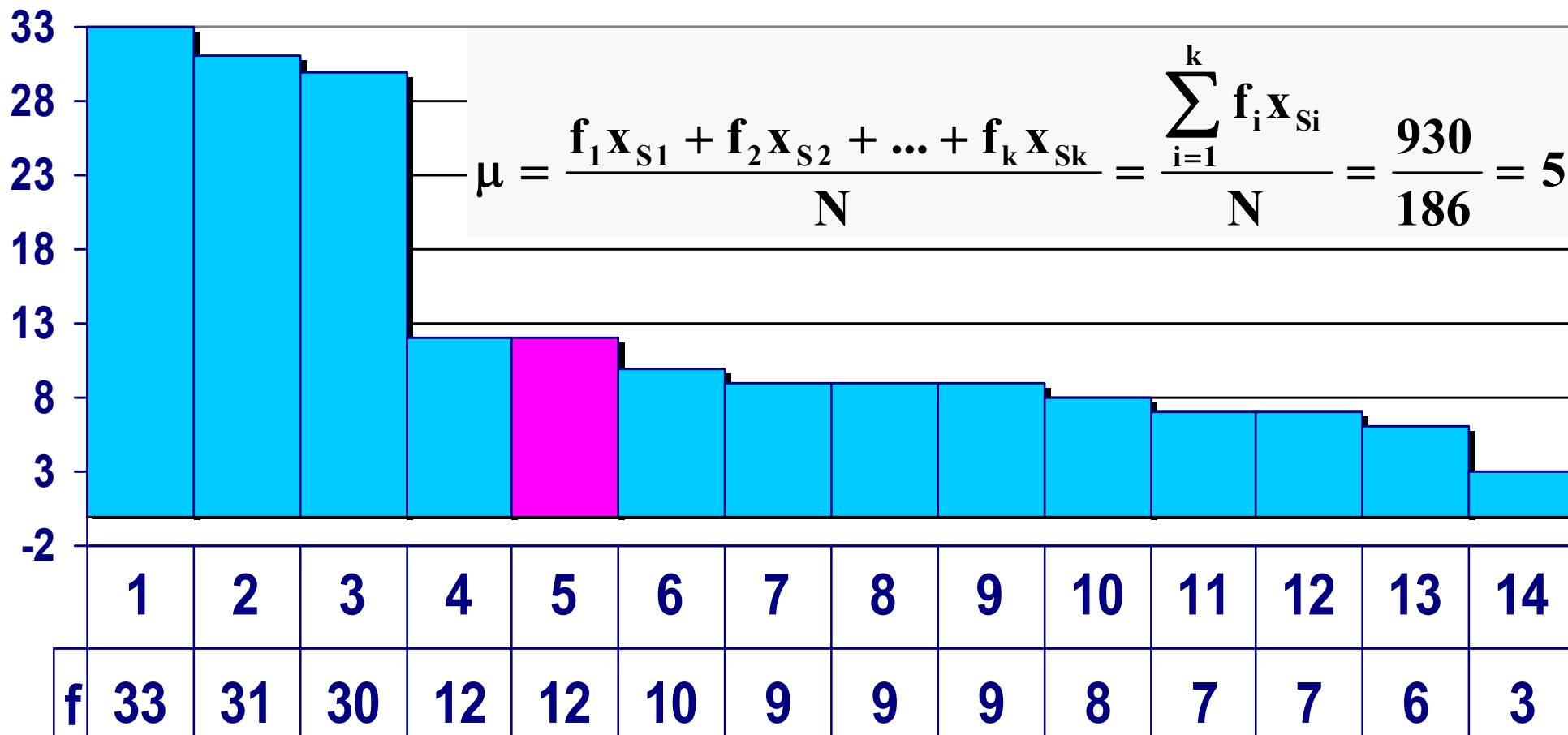
Kolika je aritmetička sredina podataka u sljedećoj tablici:

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 (vrijednost) dobivena iz
 N_i ("ponder", težina) mjerenja
 n ... broj skupina mjerenja (broj mjerenja)

Dva studenta postigla su sljedeći uspjeh u zimskom semestru studija:

PREDMET	OCJENA		ECTS
	STUDENT 1	STUDENT 2	
Medicinska kemija i biokemija 2	5	2	9
Temelji neuroznanosti	5	3	8
Uvod u medicinsku informatiku	2	5	1
Uvod u medicinsku statistiku	3	5	2

Koji je student postigao bolji uspjeh u zimskom semestru?

PREDMET	OCJENA		ECTS
	STUDENT 1	STUDENT 2	
Medicinska kemija i biokemija 2	5	2	9
Temelji neuroznanosti	5	3	8
Uvod u medicinsku informatiku	2	5	1
Uvod u medicinsku statistiku	3	5	2

$$\text{uspjeh}_{s1} = \frac{5 \cdot 9 + 5 \cdot 8 + 2 \cdot 1 + 3 \cdot 2}{20} = \frac{93}{20} = 4,65$$

$$\text{uspjeh}_{s2} = \frac{2 \cdot 9 + 3 \cdot 8 + 5 \cdot 1 + 5 \cdot 2}{20} = \frac{57}{20} = 2,85$$

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

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

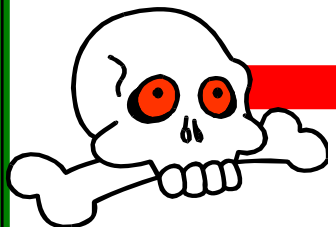
$$\bar{x}_2 = 23 \text{ cm} \quad ; \quad N_2 = 60$$

a) Kolika je zajednička aritmetička sredina?

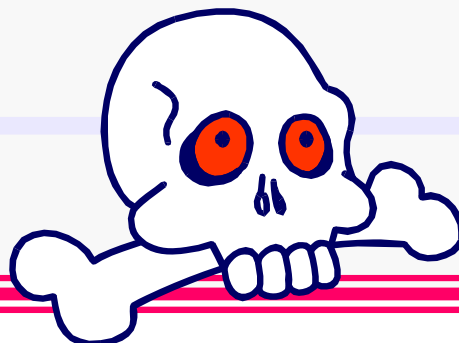
b) Kolika je zajednička aritmetička sredina za $N_1=60; N_2=15$?

$$\text{a) } \bar{x}_{\text{zaj}} = \frac{20 \cdot 15 + 23 \cdot 60}{15 + 60} = \frac{300 + 1380}{75} = 22,4 \text{ cm}$$

$$\text{b) } \bar{x}_{\text{zaj}} = \frac{20 \cdot 60 + 23 \cdot 15}{15 + 60} = \frac{1200 + 354}{75} = 20,6 \text{ cm}$$



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:

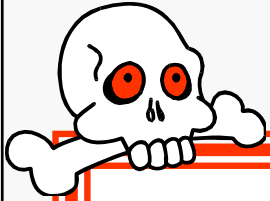
A: 2 2,5 3,5 3 4

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

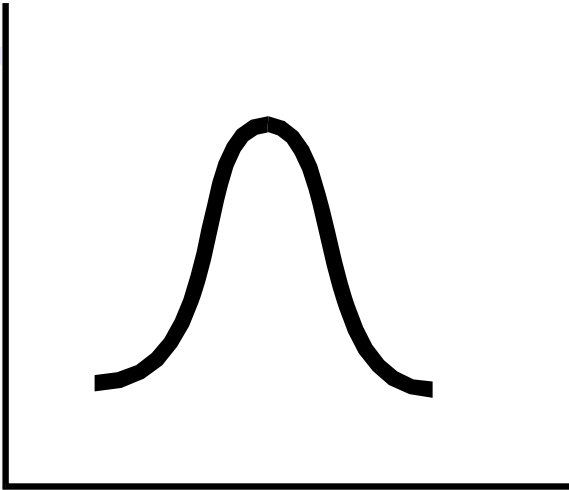
niz A: medijan.... **Me = 3** arit. sred. $\bar{X} = 3$

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

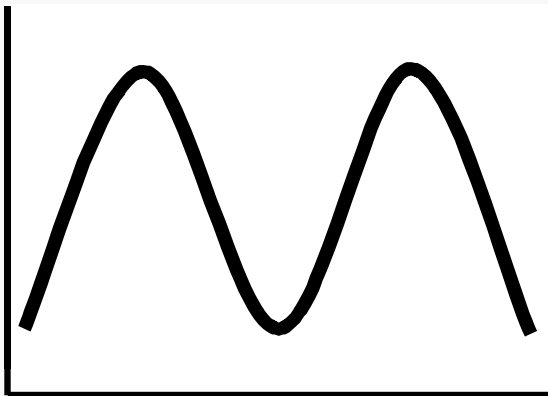
MOD (dominantna vrijednost)

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

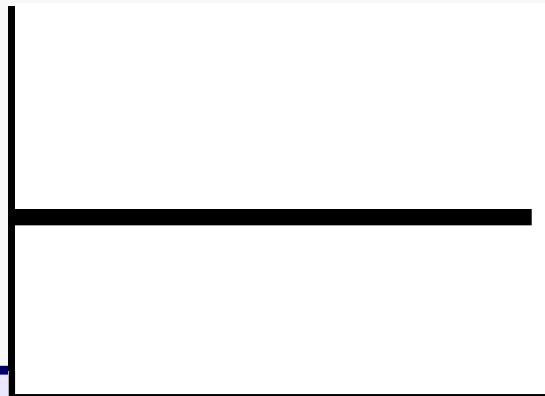
oznaka: Mo



UNIMODALNA



BIMODALNA



UNIFORMNA

MJERE SREDINE PREMA LJESTVICI MJERENJA

LJESTVICA MJERENJA	MJERA SREDINE
NOMINALNA	MOD
ORDINALNA	MOD MEDIJAN
INTERVALNA/OMJERNA (asimetrična raspodjela)	MOD MEDIJAN
INTERVALNA/OMJERNA (simetrična raspodjela)	MOD MEDIJAN ARITMETIČKA SREDINA

MJERE RASPRŠENJA (VARIJABILNOSTI)

(mjere disperzije)

RASPON

KVANTILE

VARIJANCA

STANDARDNA DEVIJACIJA

KOEFICIJENT VARIJABILNOSTI

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

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

ZA UZORAK

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

- 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 različitih svojstava neke grupe ispitanika
 - b) razlike u varijabilnosti istog svojstva u različitim grupama ispitanika

OPIS VARIJABLE

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

MJERE ZA OCJENU OBLIKA RAZDIOBE

MOMENTI RAZDIOBE

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

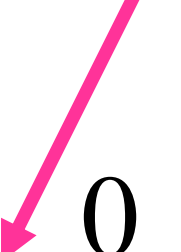
MOMENT n-tog REDA (n-ti moment)

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

NULTI I PRVI MOMENT - konstante

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

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

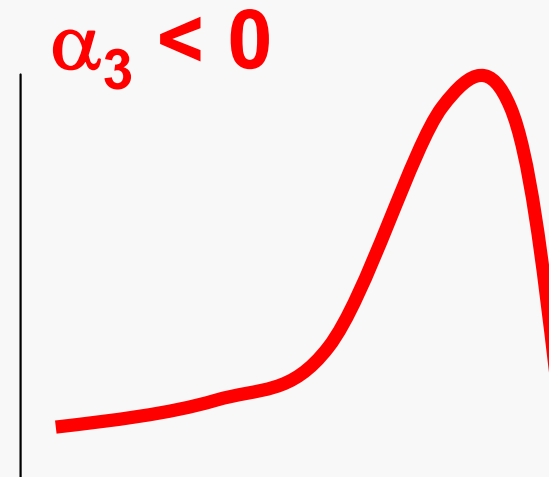
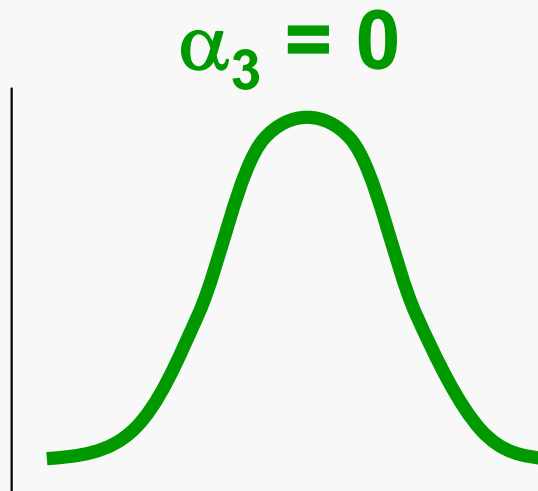
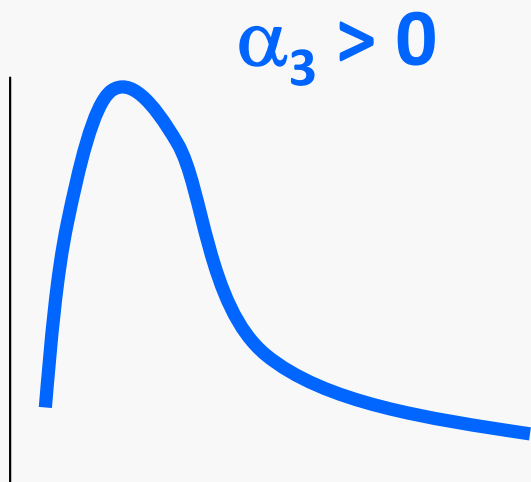
$$\mu_3 = \frac{\sum_{i=1}^N (x_i - \mu)^3}{N}$$

- 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)



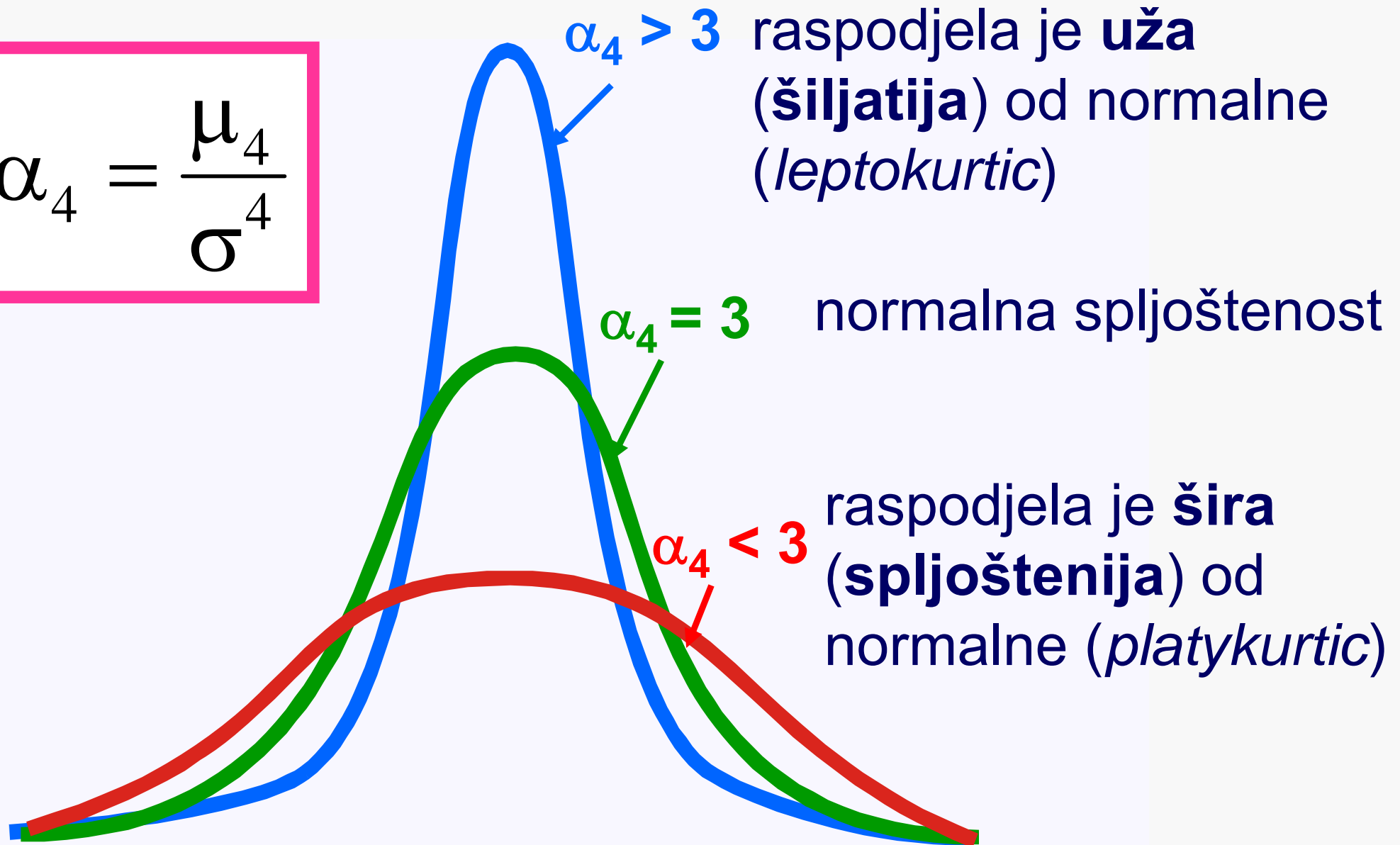
ČETVRTI MOMENT

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

- koristi se za mjeru spljoštenosti

KOEFICIJENT SPLJOŠTENOSTI (*coefficient of kurtosis*)

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



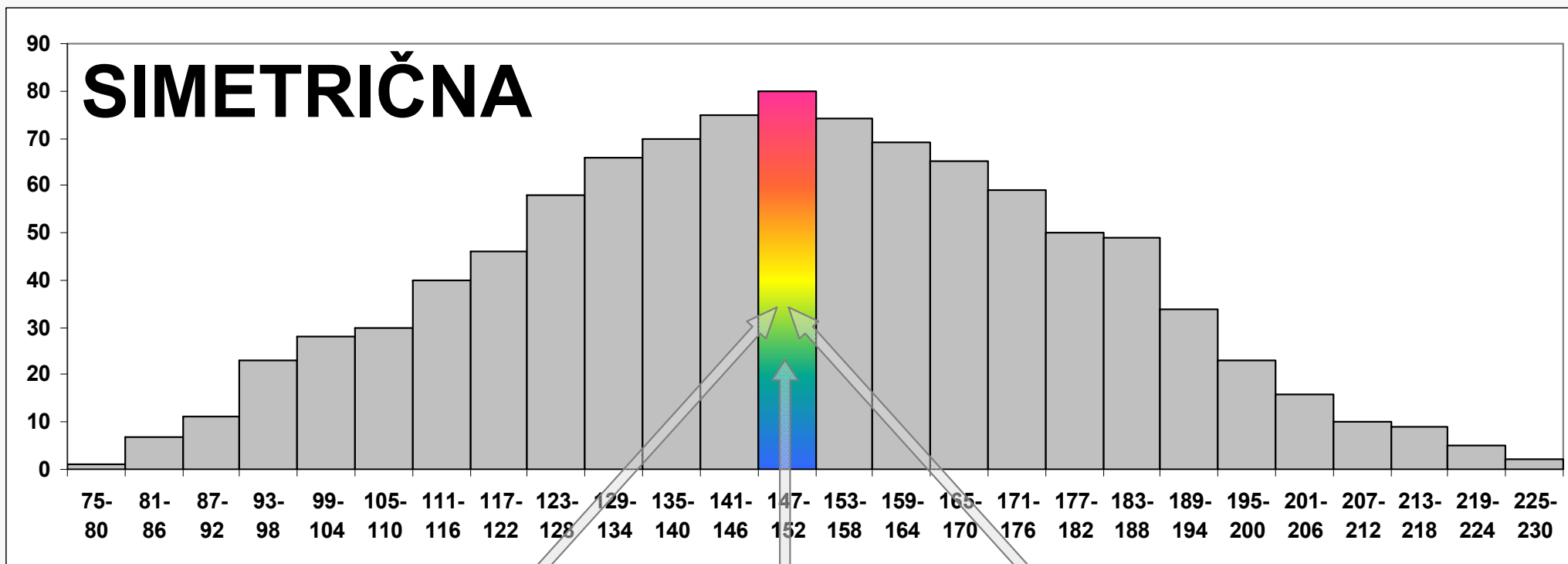
KOEFICIJENT SPLJOŠTENOSTI (*coefficient of kurtosis*)

- 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

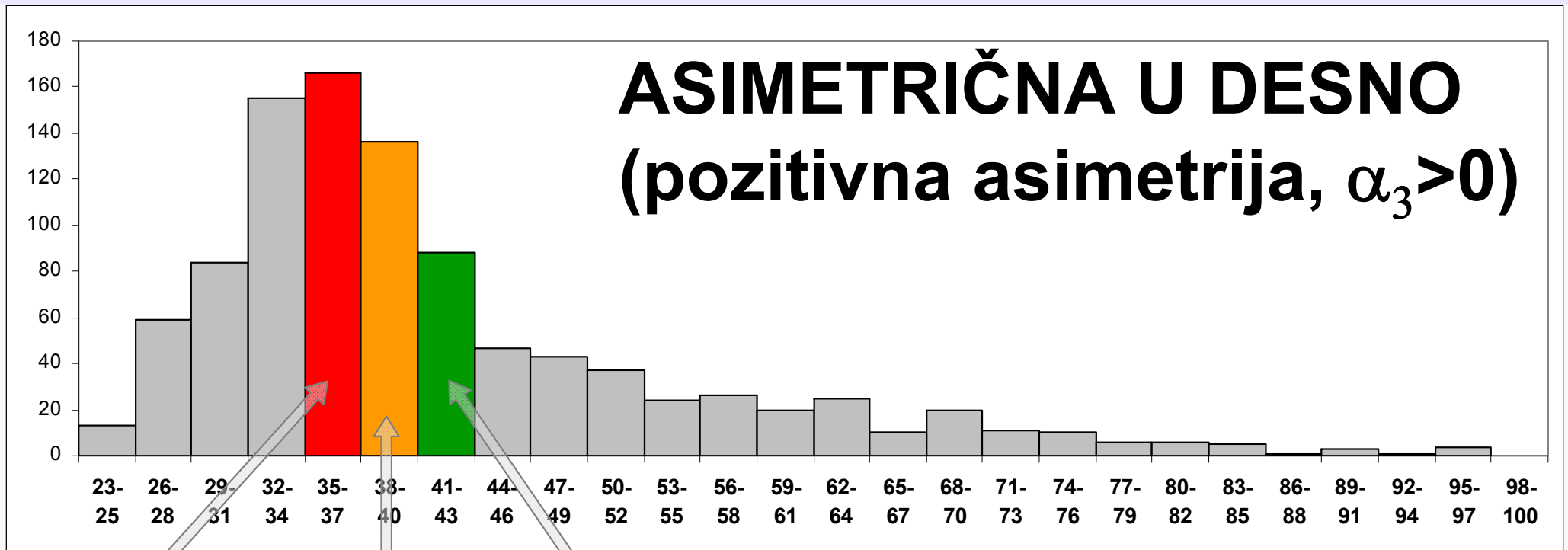
ODNOS MJERA SREDINE I ASIMETRIJE



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

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

ODNOS MJERA SREDINE I ASIMETRIJE



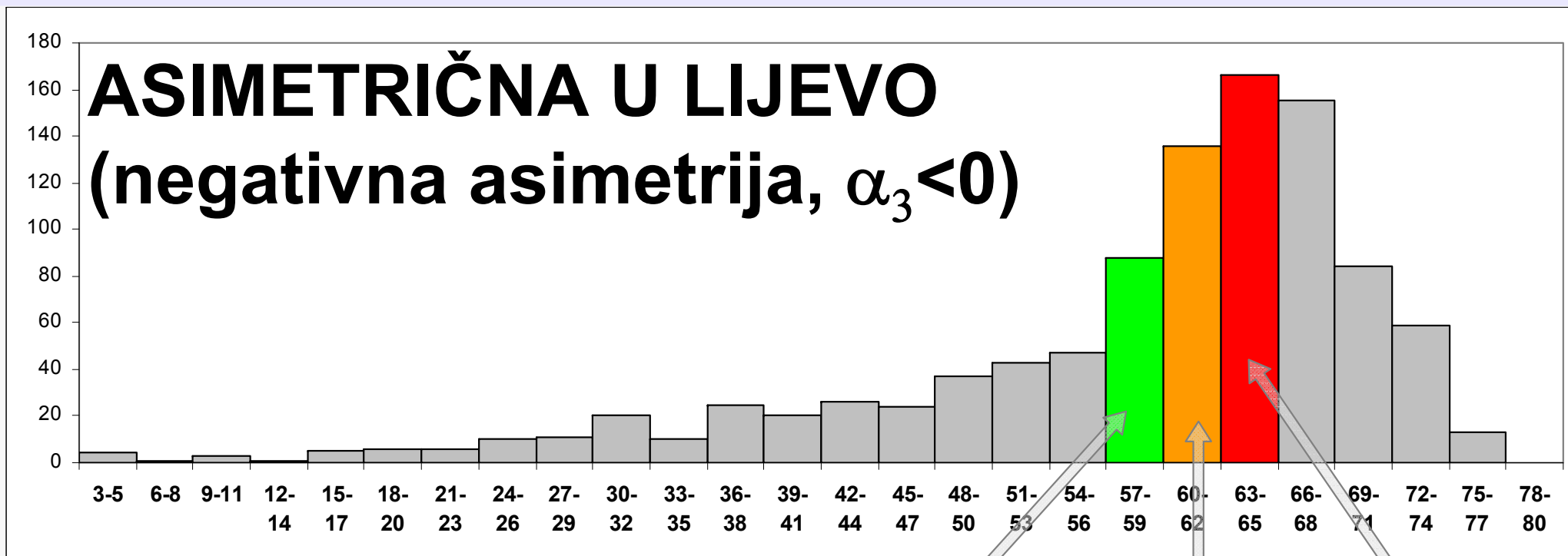
Mo = 37

$\bar{X} = 42$

Me = 38

Mo < Me < \bar{X}

ODNOS MJERA SREDINE I ASIMETRIJE



$$\bar{x} = 58$$

$$Mo = 63$$

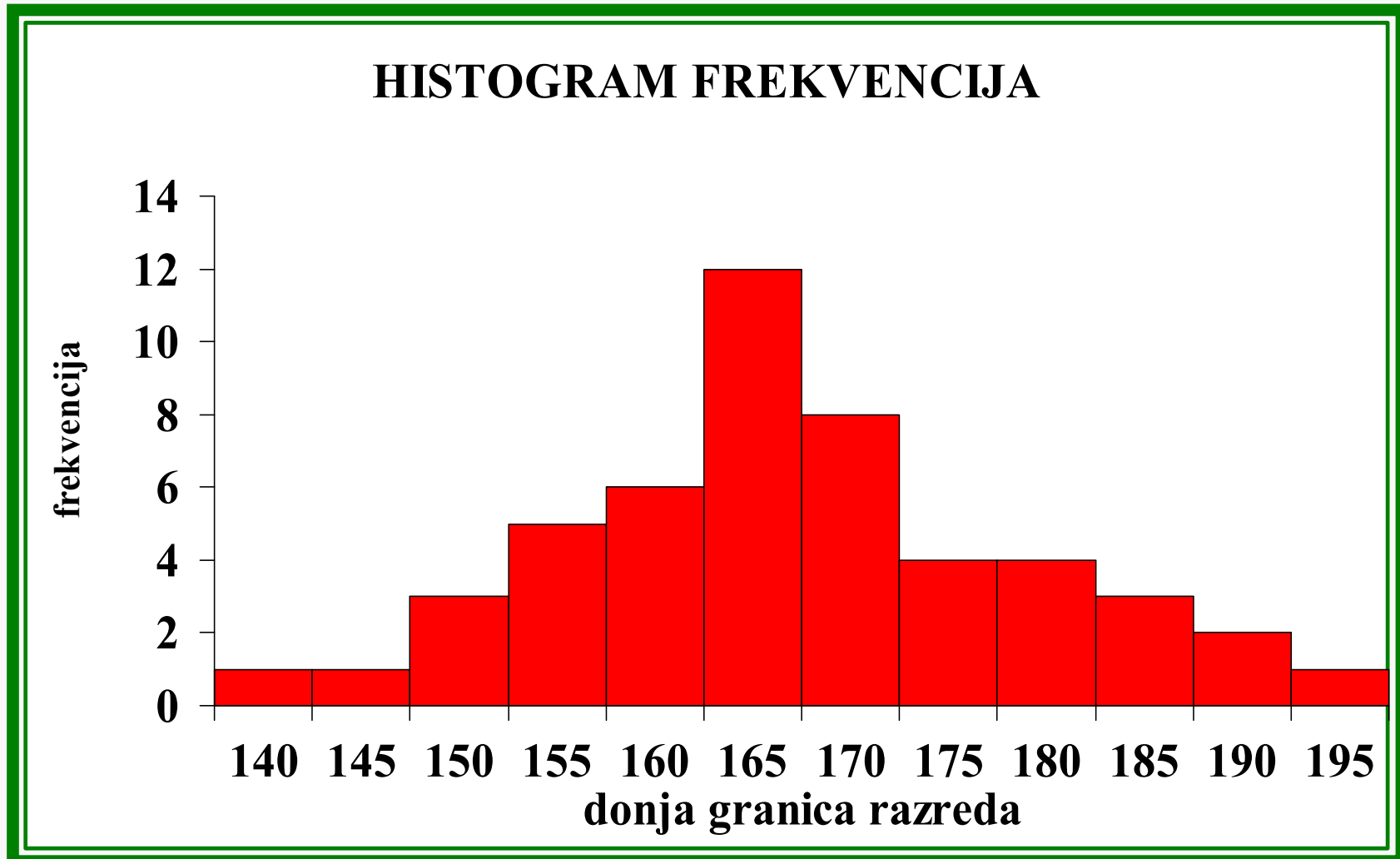
$$Me = 62$$

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

GRAFIČKO PRIKAZIVANJE PODATAKA

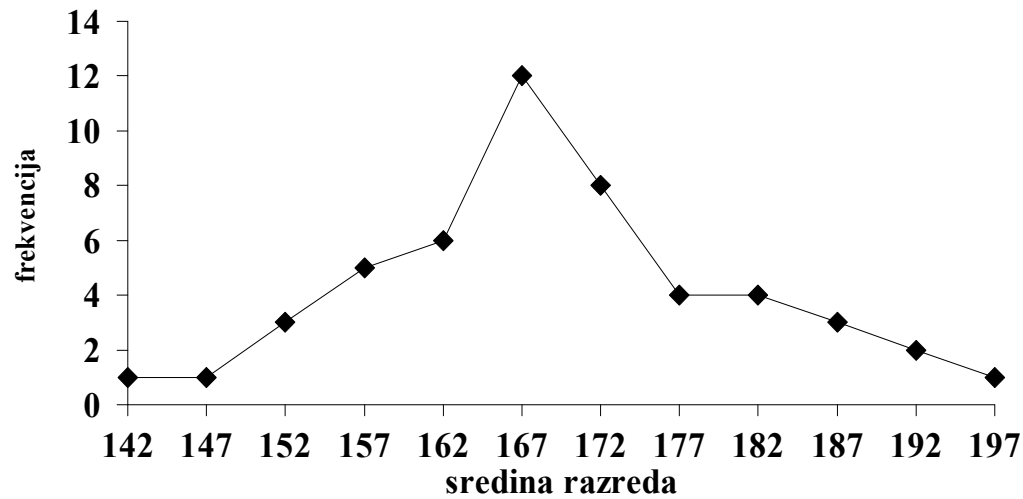
RAZDIOBA OBILJEŽJA

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

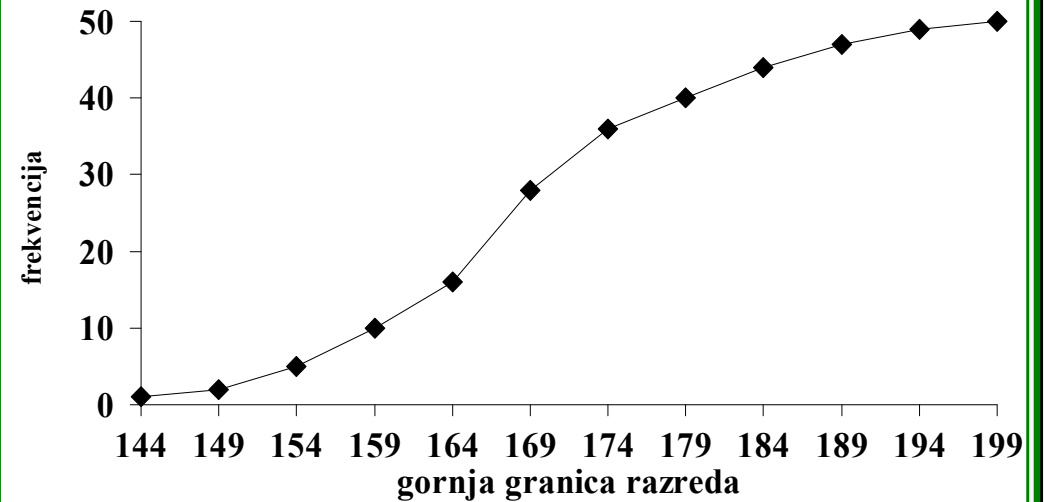


RAZDIoba OBILJEŽJA

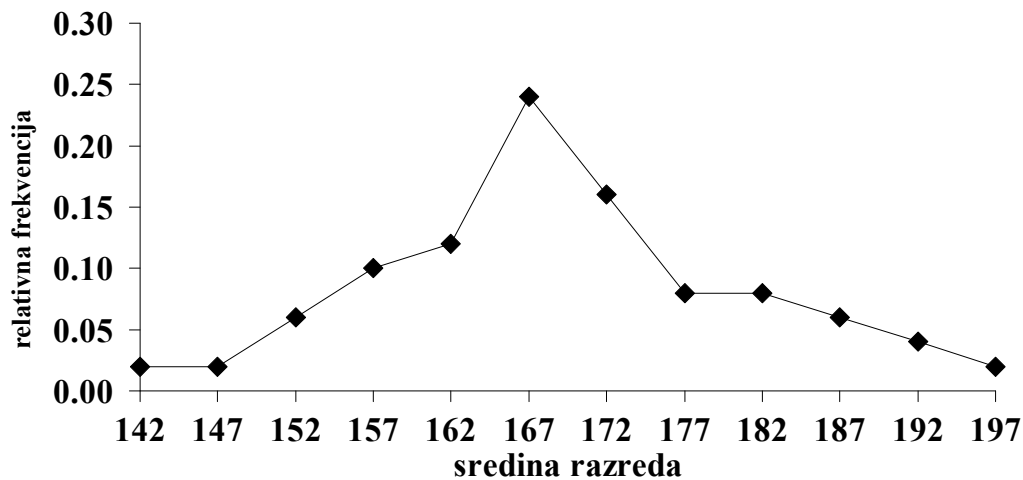
POLIGON FREKVENCIJA



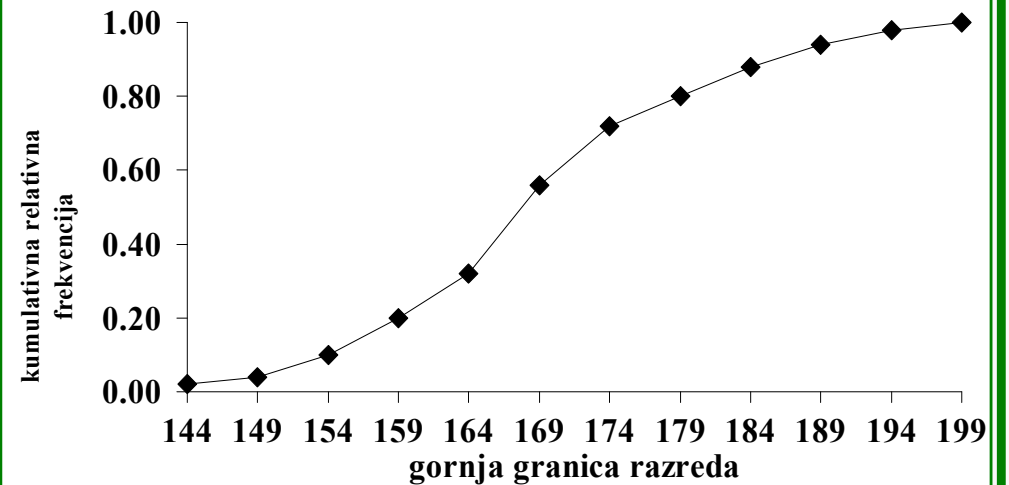
POLIGON KUMULATIVNIH FREKVENCIJA



POLIGON RELATIVNIH FREKVENCIJA

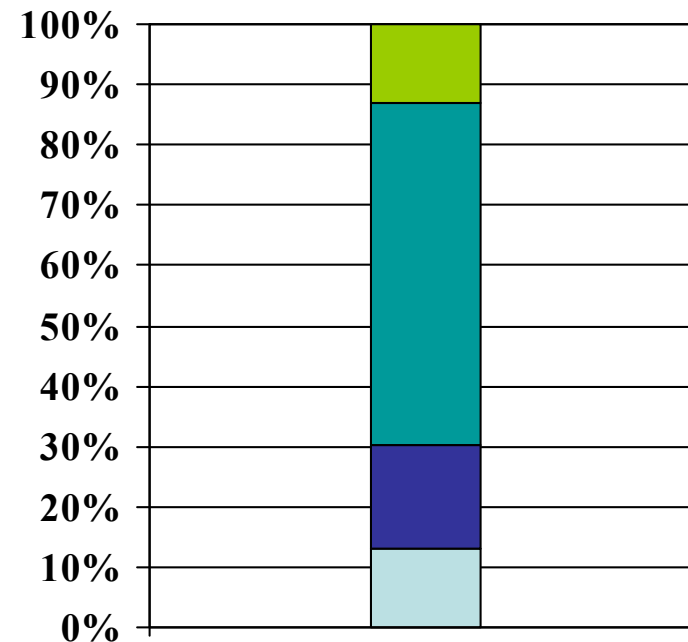


POLIGON KUMULATIVNIH RELATIVNIH FREKVENCIJA



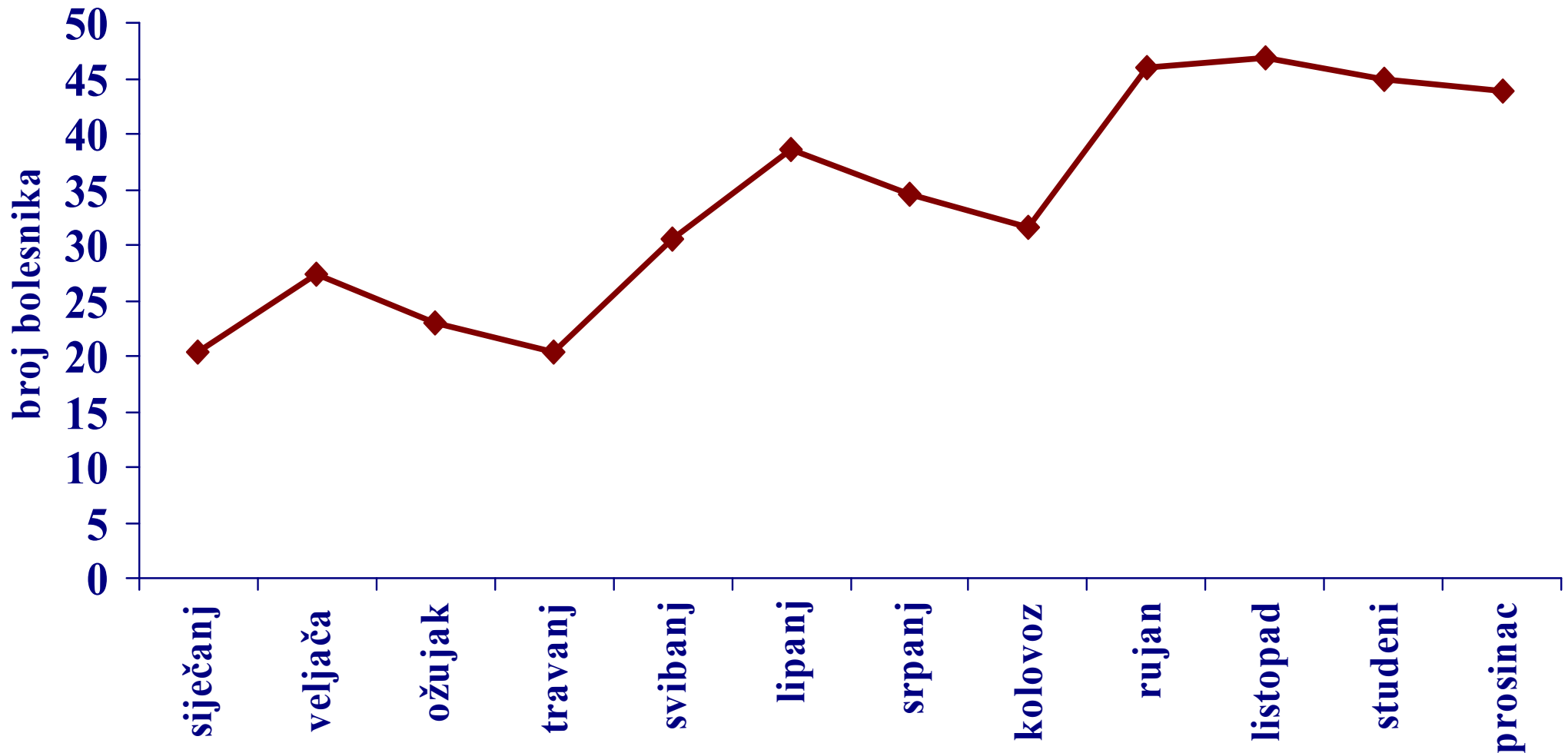
STRUKTURA OBILJEŽJA

- 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

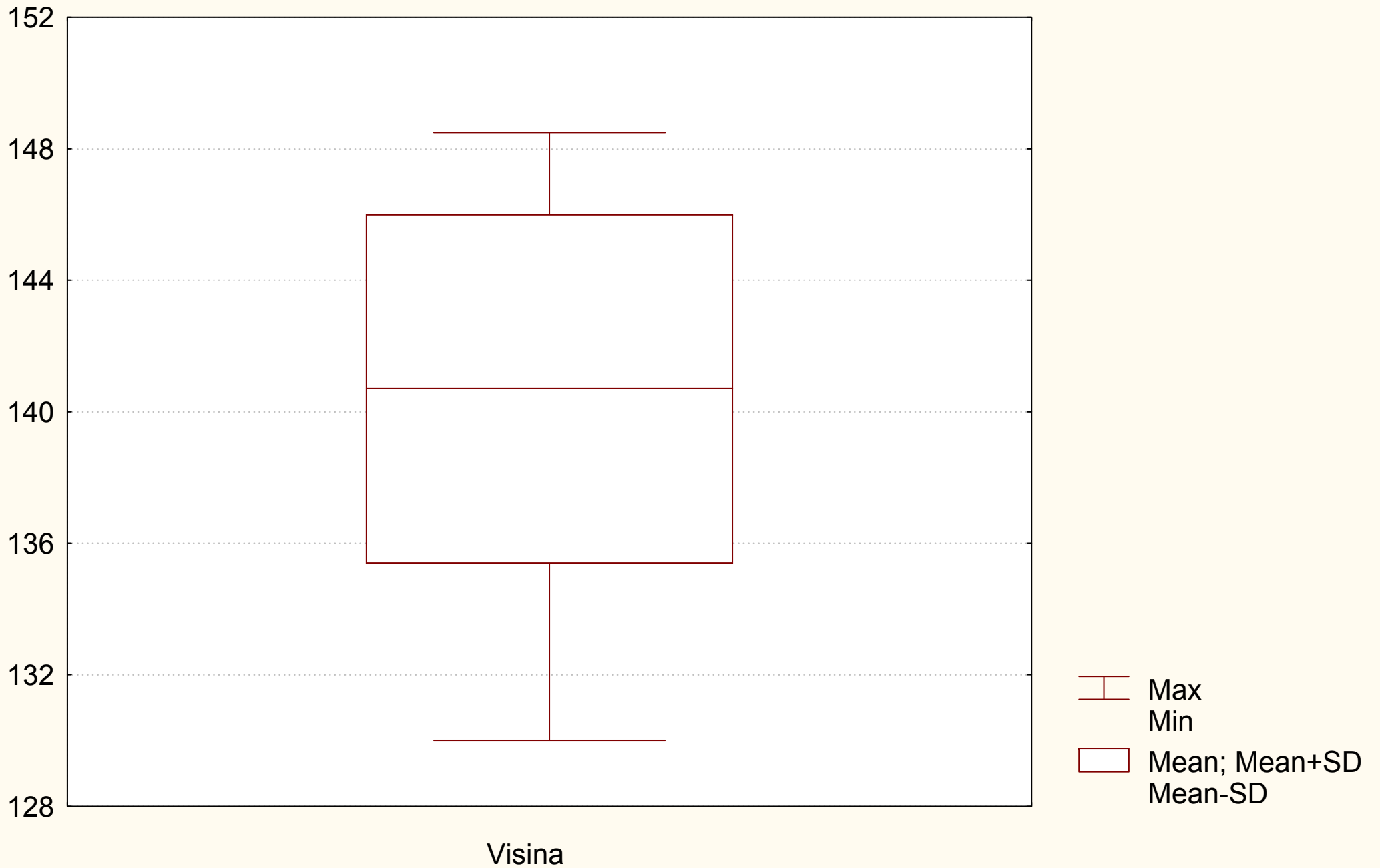
- prikazuju se linijskim grafikonom (*“line chart”*)
- na apscisu se nanose vremenski intervali, a na ordinatu vrijednost promatrane varijable



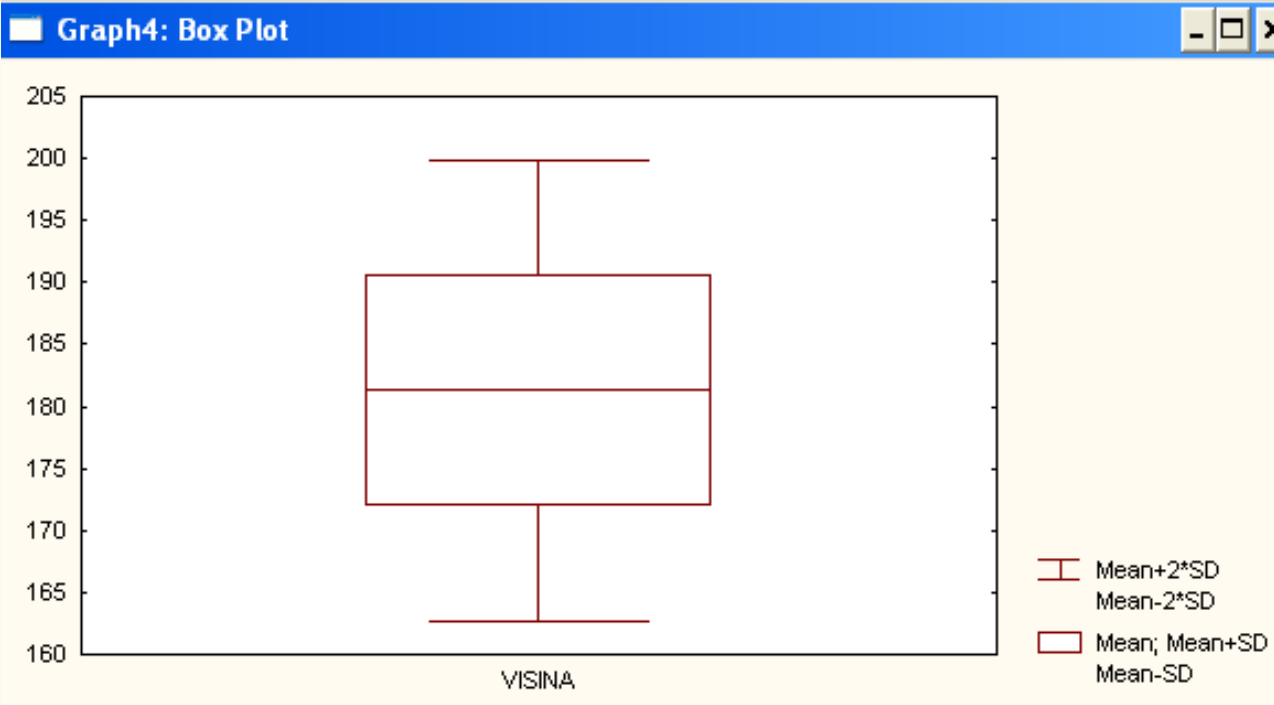
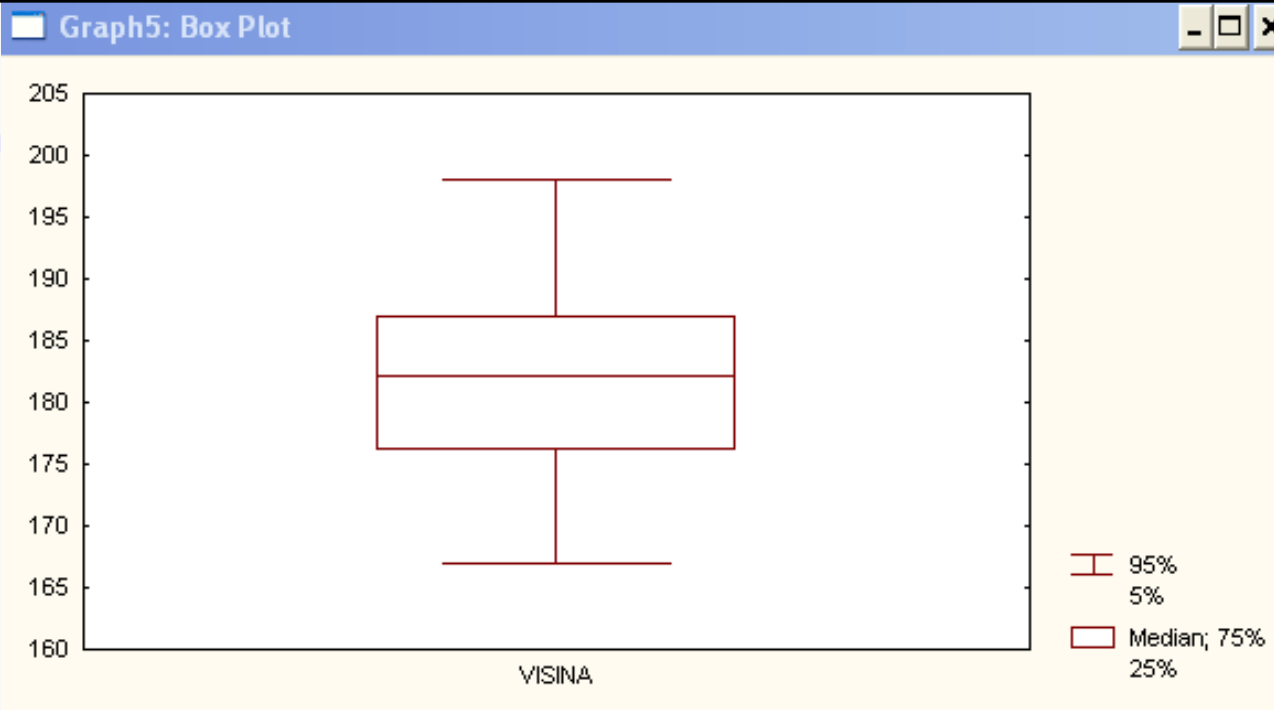
OSNOVNE MJERE SREDINE I RASPRŠENJA VARIJABLE

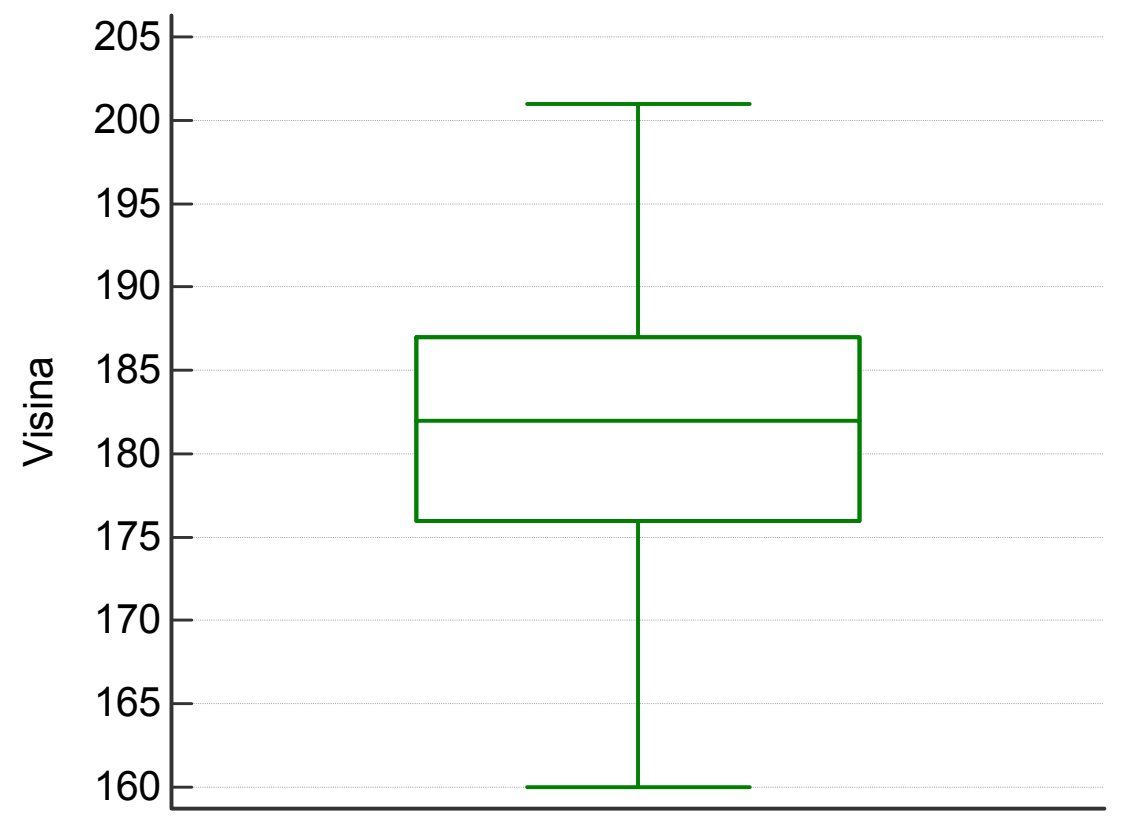
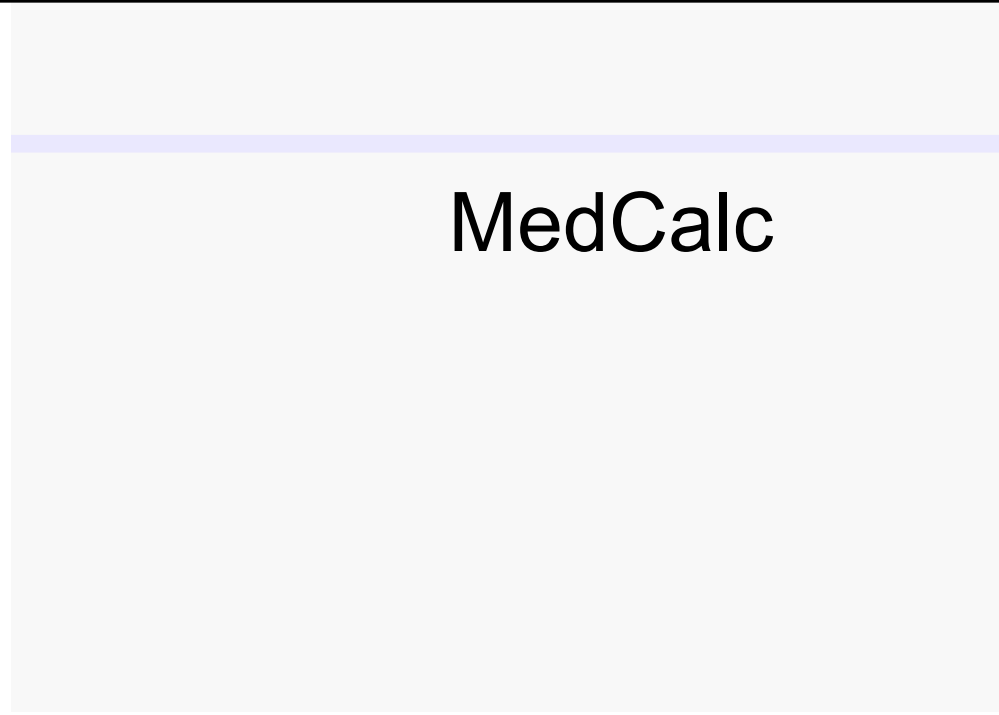
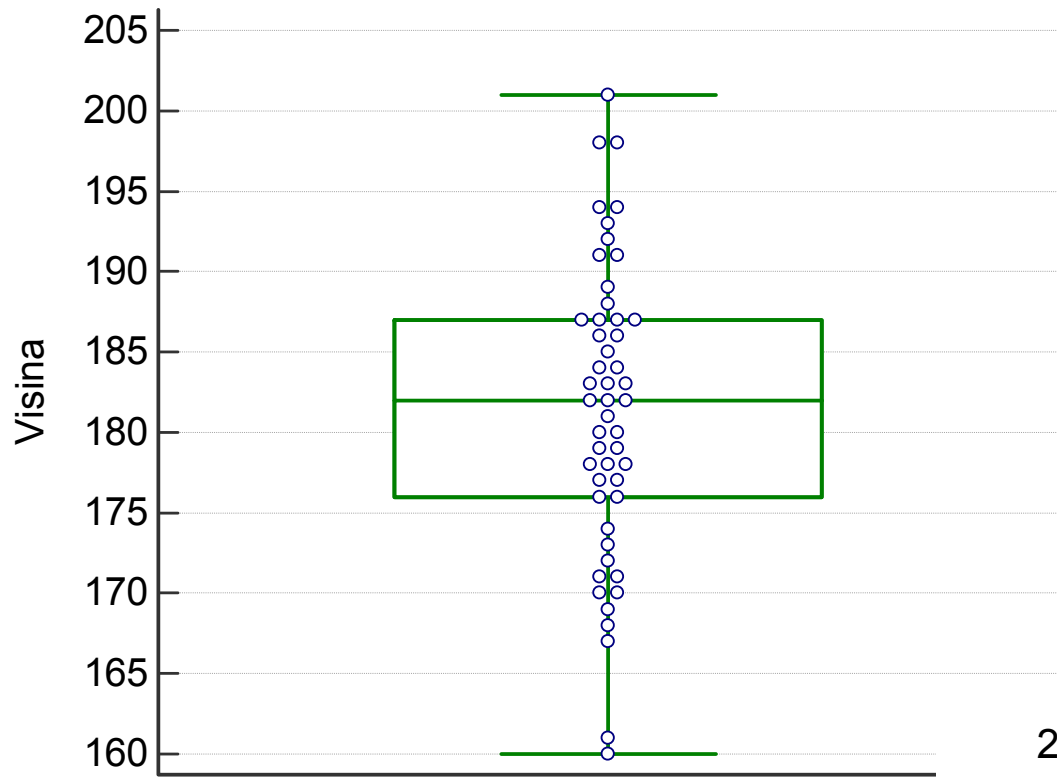
- “kutija i brkovi” grafikon (“Box-and-Whisker” plot)
- najčešće prikazuje kombinacije:

sredina	kutija	brkovi
aritmetička sredina	standardna devijacija	1,5 ili 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

