

Statističko opisivanje podataka

Nedelja 3 - Vežbe

Dr Nikola N. Grubor

Statističko opisivanje podataka

3 Nedelja - Vežbe

Relativni brojevi

Mere Centralne Tendencije

Mere Varijabiliteta

EZR

Relativni brojevi, $R_b = \frac{V_r}{V_b}$

Ovog meseca je umrlo manje ljudi nego prethodnog.

$$\frac{80}{100} \times 100 = 80.0\%$$

Mortalitet je porastao za 10% u odnosu na prethodni mesec.

$$\frac{70}{78} \times 100 = 89.7\%$$

Struktura postelja na odeljenjima?

Jedinica	Bolnica A	Bolnica B	Ukupno
Interno	70	80	150
Ginekološko	40	140	180
Hirurško	80	100	180
Pedijatrijsko	50	160	210
Ukupno	240	480	720

Procenti po redovima

$$\frac{70}{150} \times 100 = 47\%$$

Struktura odeljenja u bolnicama?

Jedinica	Bolnica A	Bolnica B	Ukupno
Iнтерно	70	80	150
Гинеколошко	40	140	180
Хируршко	80	100	180
Педијатријско	30	160	210
Укупно	240	480	720

Procenti po kolonama

$$\frac{70}{240} \times 100 = 29\%$$

Struktura postelja po odeljenjima i bolnicama?

Jedinica	Bolnica A	Bolnica B	Ukupno
Iнтерно	70	80	150
Гинеколошко	40	140	180
Хируршко	80	100	180
Педијатријско	50	160	210
Укупно	240	480	720

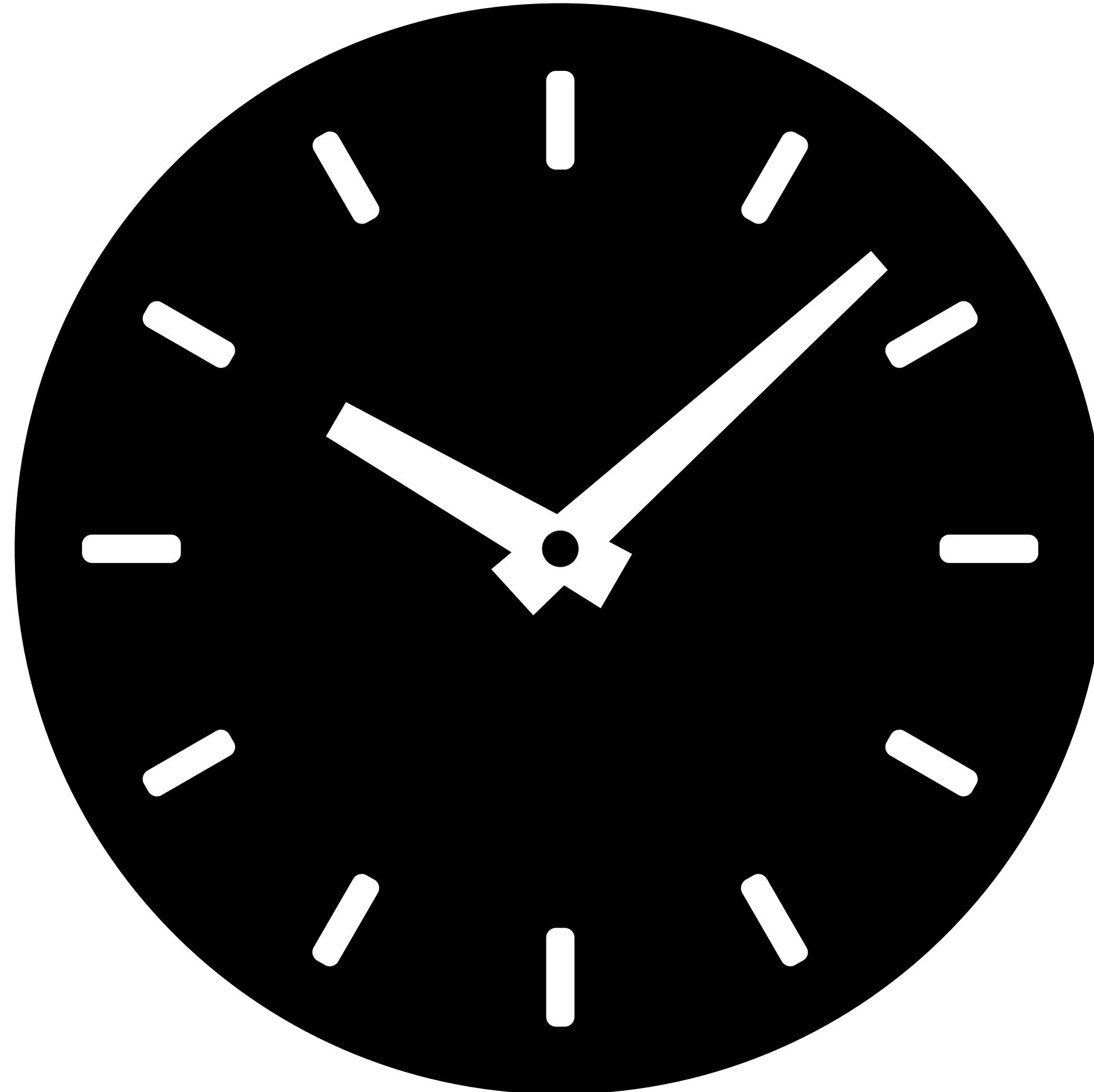
Totalni procenti

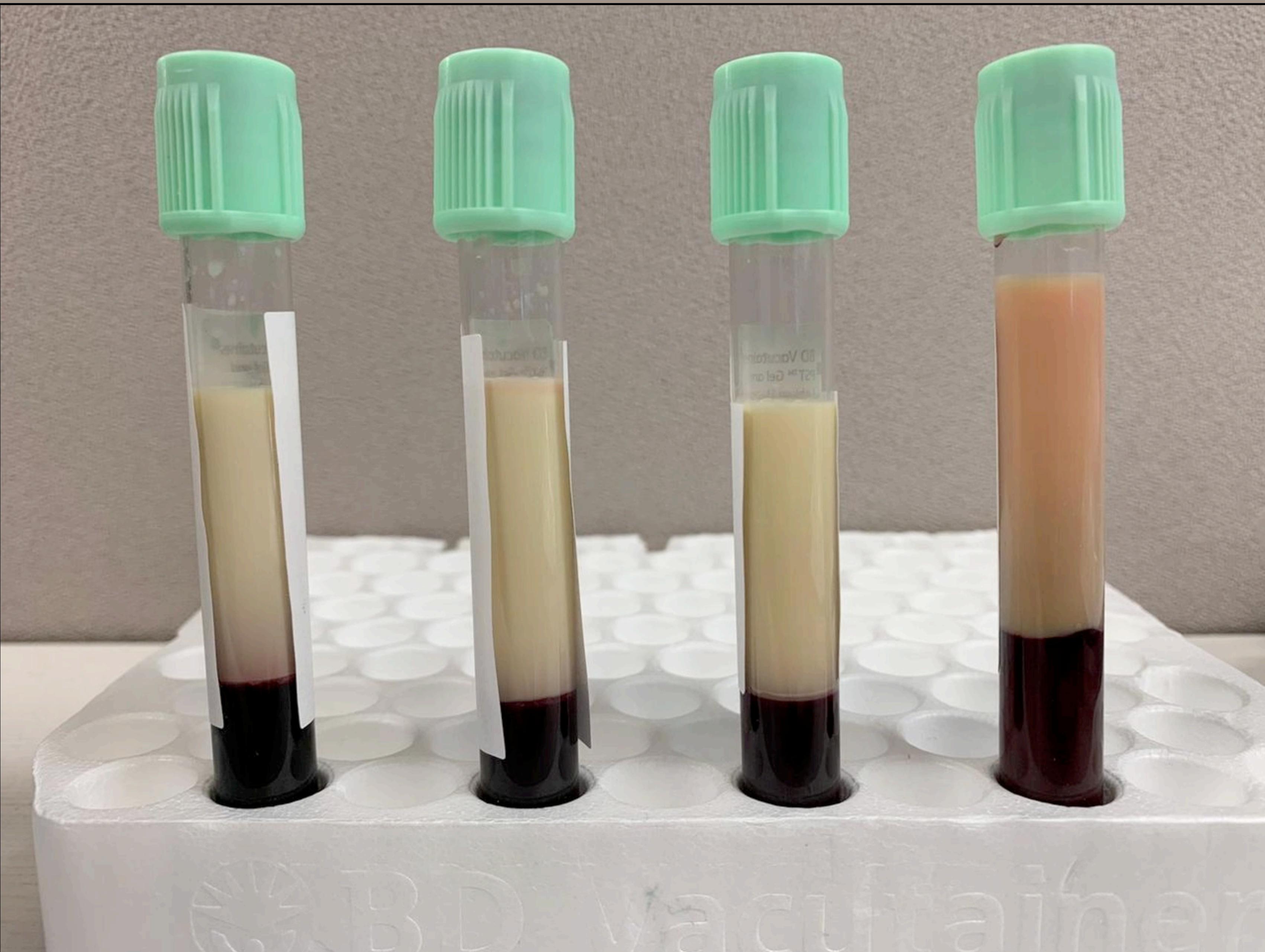
$$\frac{70}{720} \times 100 = 10\%$$

Relativni brojevi za vremenske serije

Indeksi dinamike

- Bazni
- Lančani



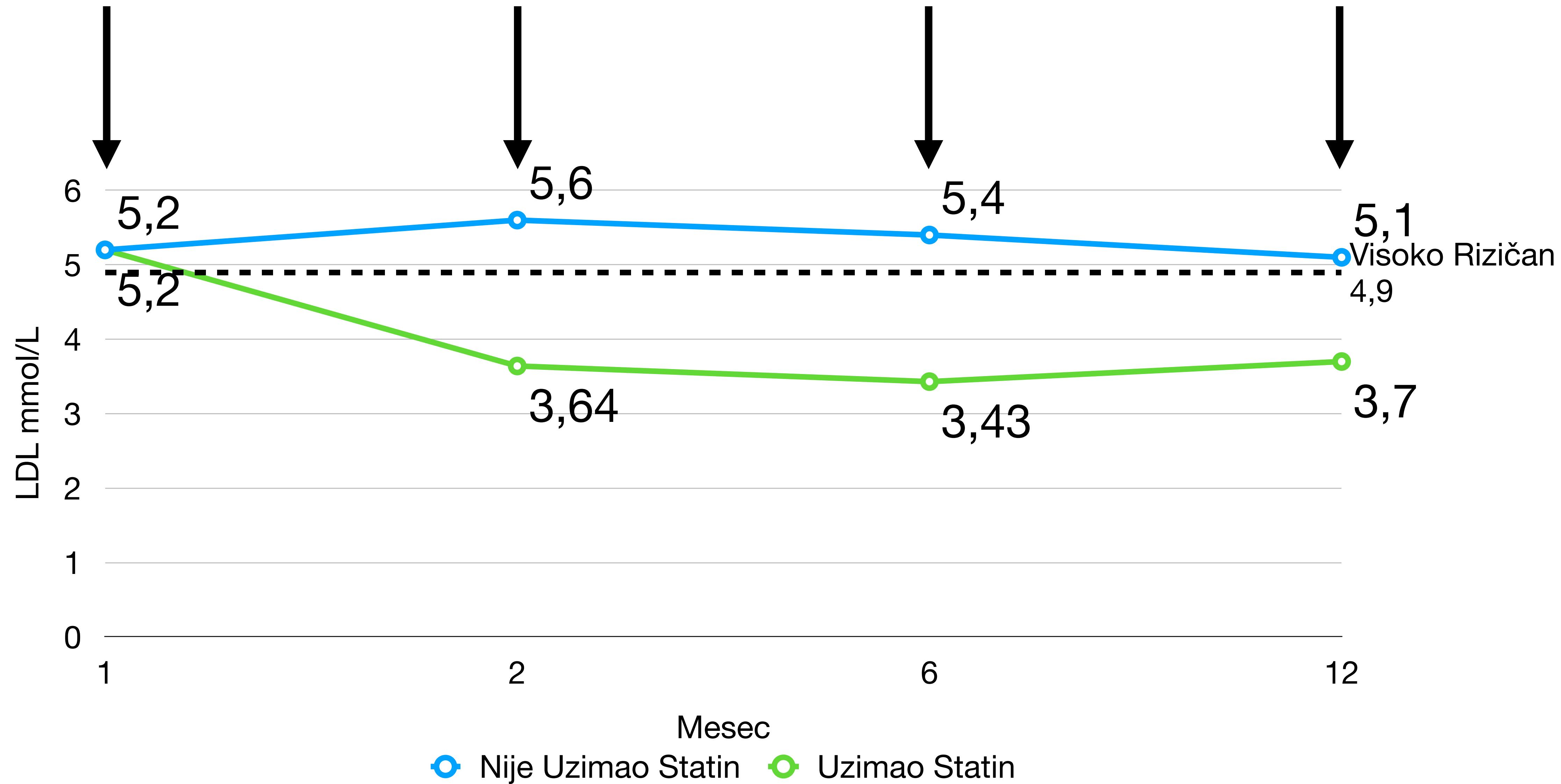


Prvi pregled

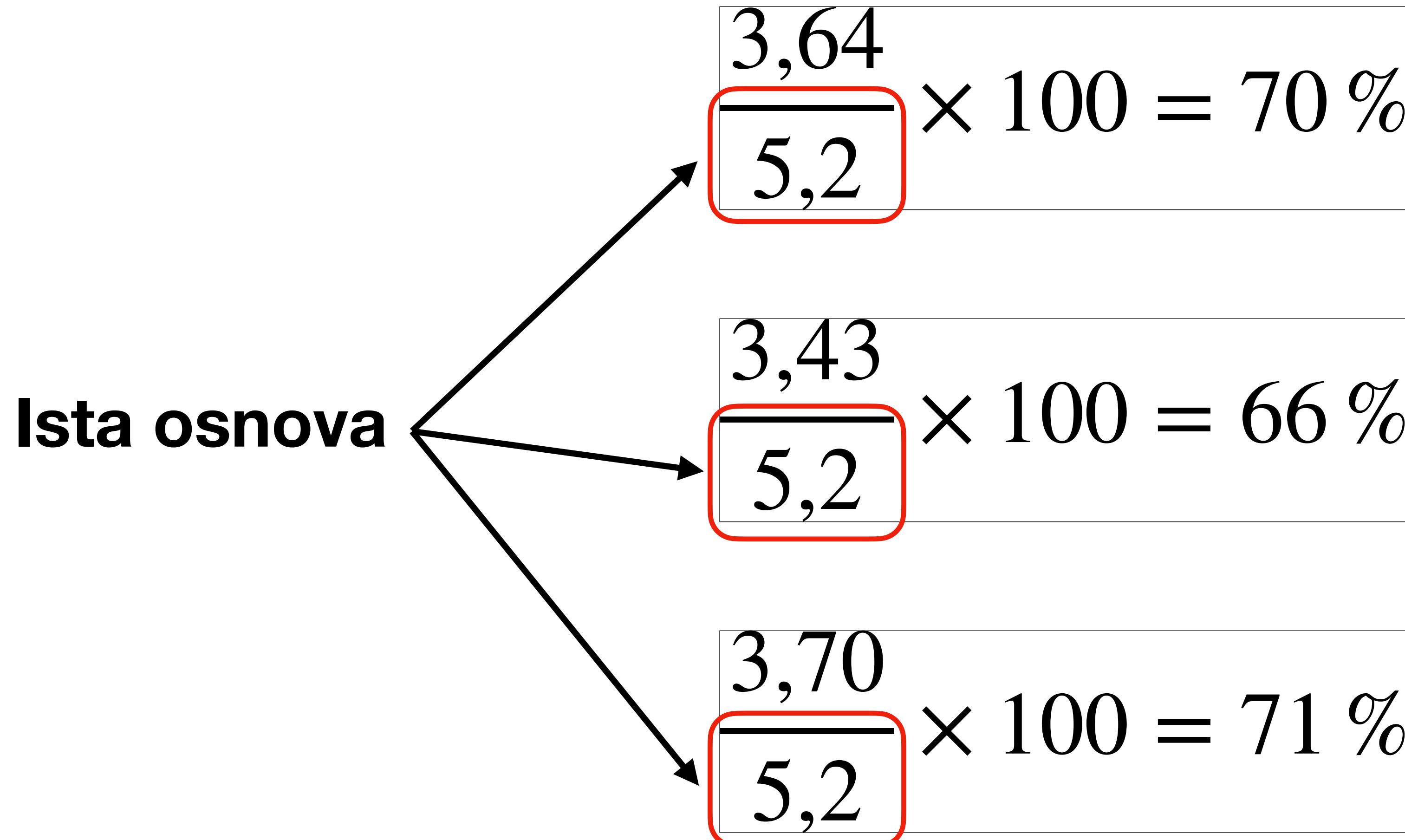
$$\frac{3,64}{5,2} \times 100 = 70\%$$

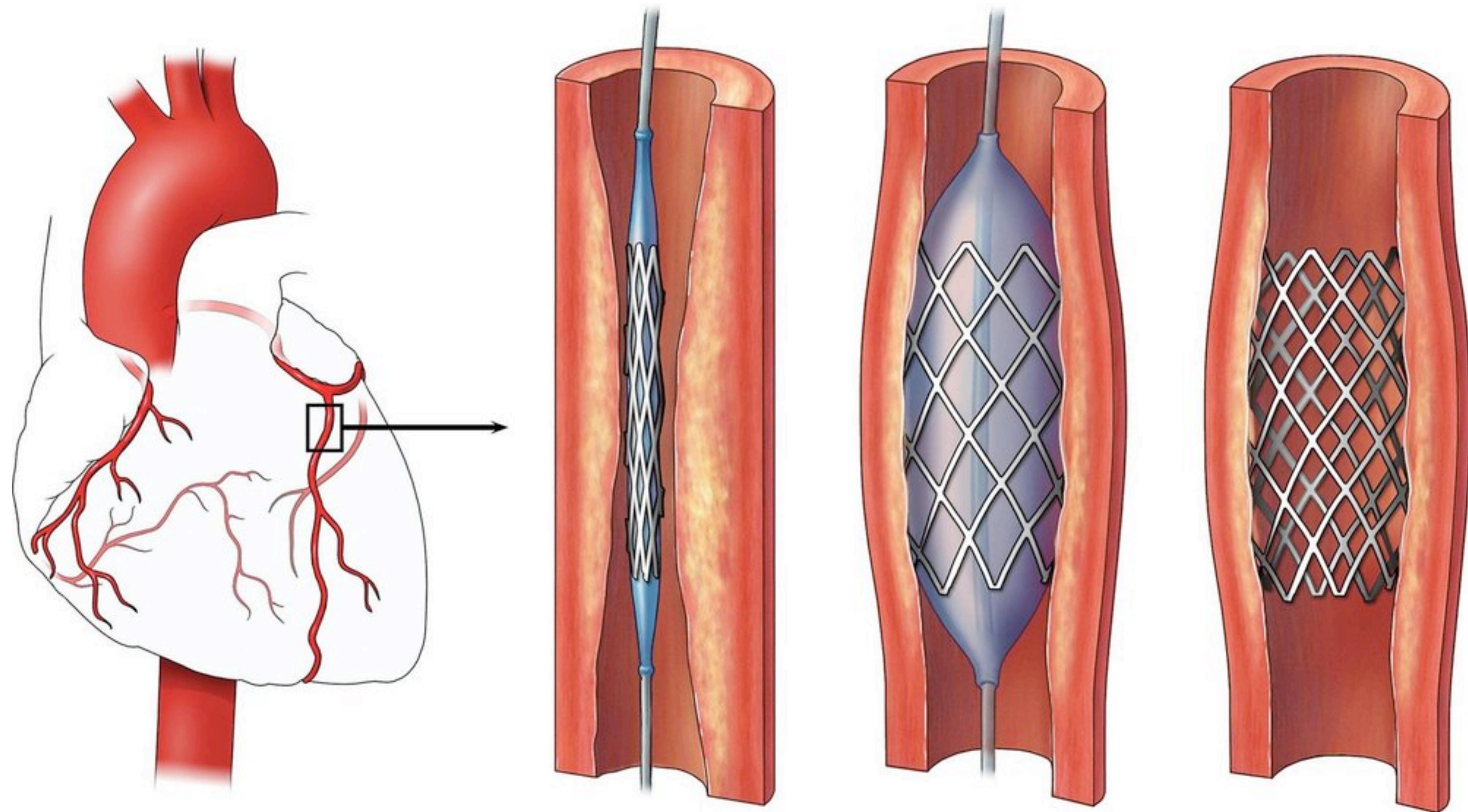
$$\frac{3,43}{5,2} \times 100 = 66\%$$

$$\frac{3,70}{5,2} \times 100 = 71\%$$



Bazni Indeks Dinamike

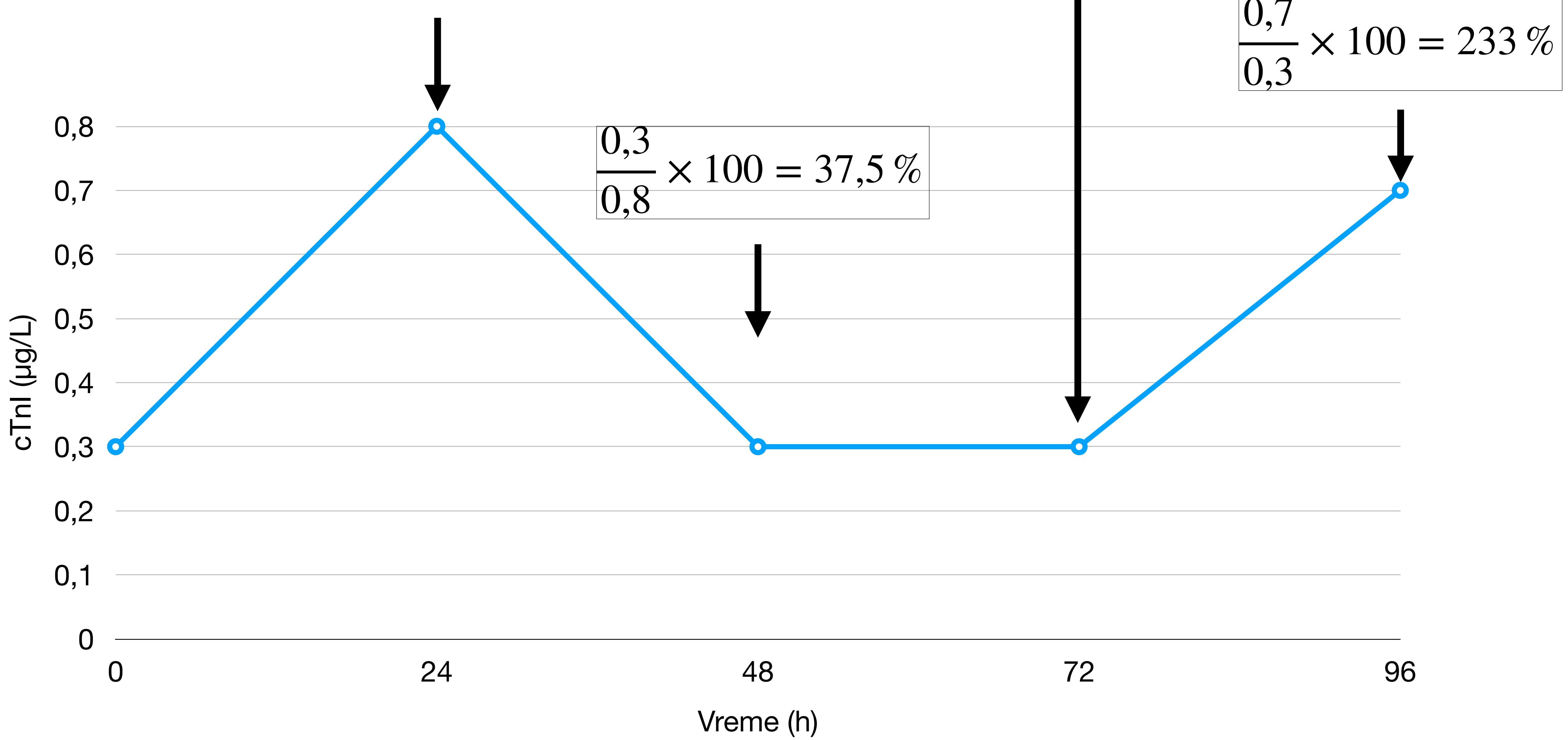




$$\frac{0,8}{0,3} \times 100 = 266\%$$

$$\frac{0,3}{0,3} \times 100 = 100\%$$

$$\frac{0,7}{0,3} \times 100 = 233\%$$



Lančani indeks dinamike

Prethodna osnova

$$\frac{0,8}{0,3} \times 100 = 266\%$$

$$\frac{0,3}{0,8} \times 100 = 37,5\% \quad \leftarrow \text{U padu}$$

$$\frac{0,3}{0,3} \times 100 = 100\% \quad \leftarrow \text{Stabilno}$$

$$\frac{0,7}{0,3} \times 100 = 233\% \quad \leftarrow \text{U porastu}$$

Relativni brojevi

Statistički koeficijenti

- Natalitet
- Mortalitet
- Incidencija
- Prevalencija

$$N = \frac{R}{St} \times 1000$$

$$M = \frac{U}{St} \times 1000$$

$$I = \frac{O_n}{St} \times 100000$$

$$P = \frac{O_u}{St} \times 100000$$



Mere Centralne Tendencije

Mere Centralne Tendencije

Matematičke

- Aritmetička sredina
- Geometrijska sredina
- Harmonijska sredina

Pozicione

- Medijana
- Mod (modus, tipična vrednost)

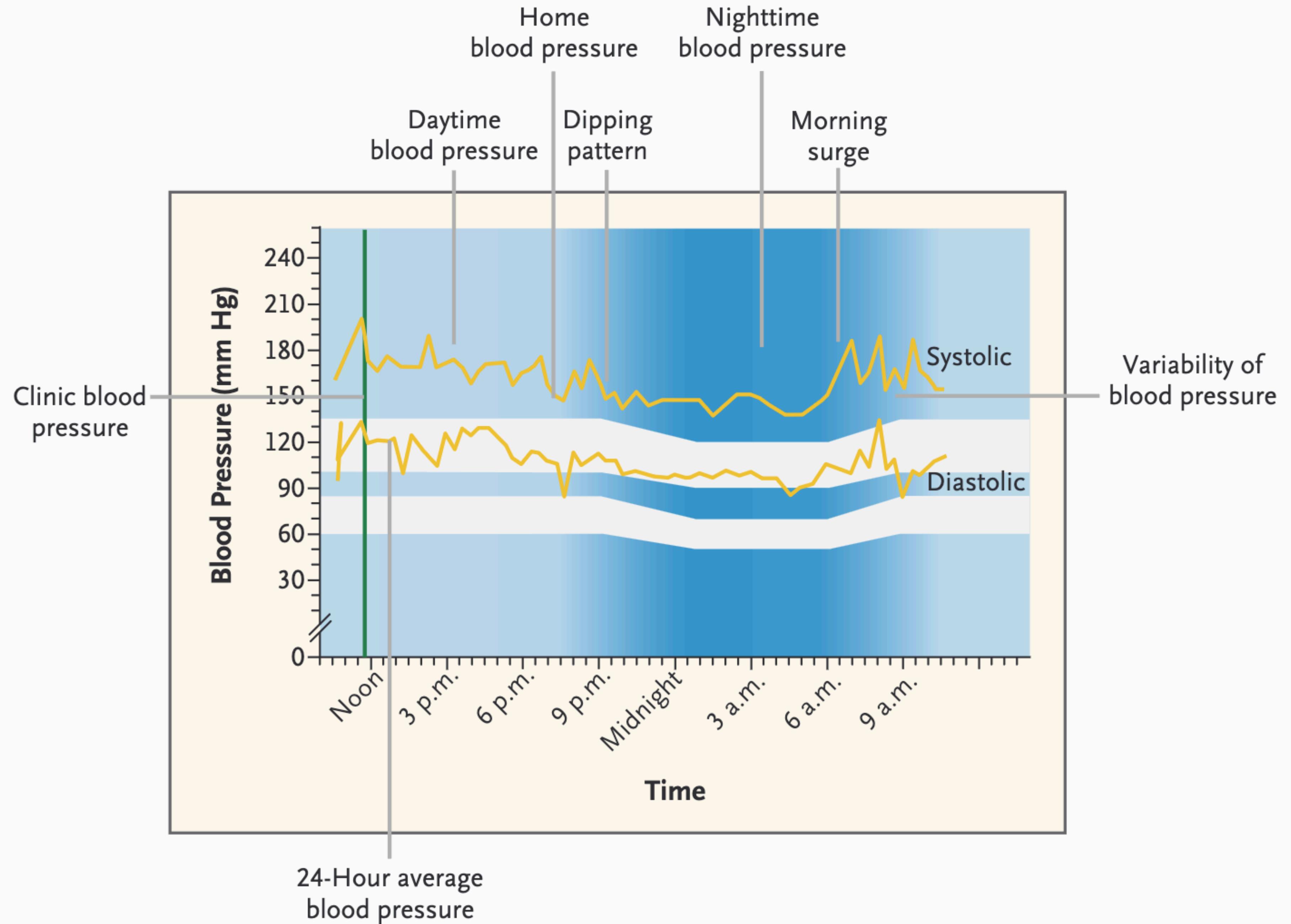
Mere Centralne Tendencije

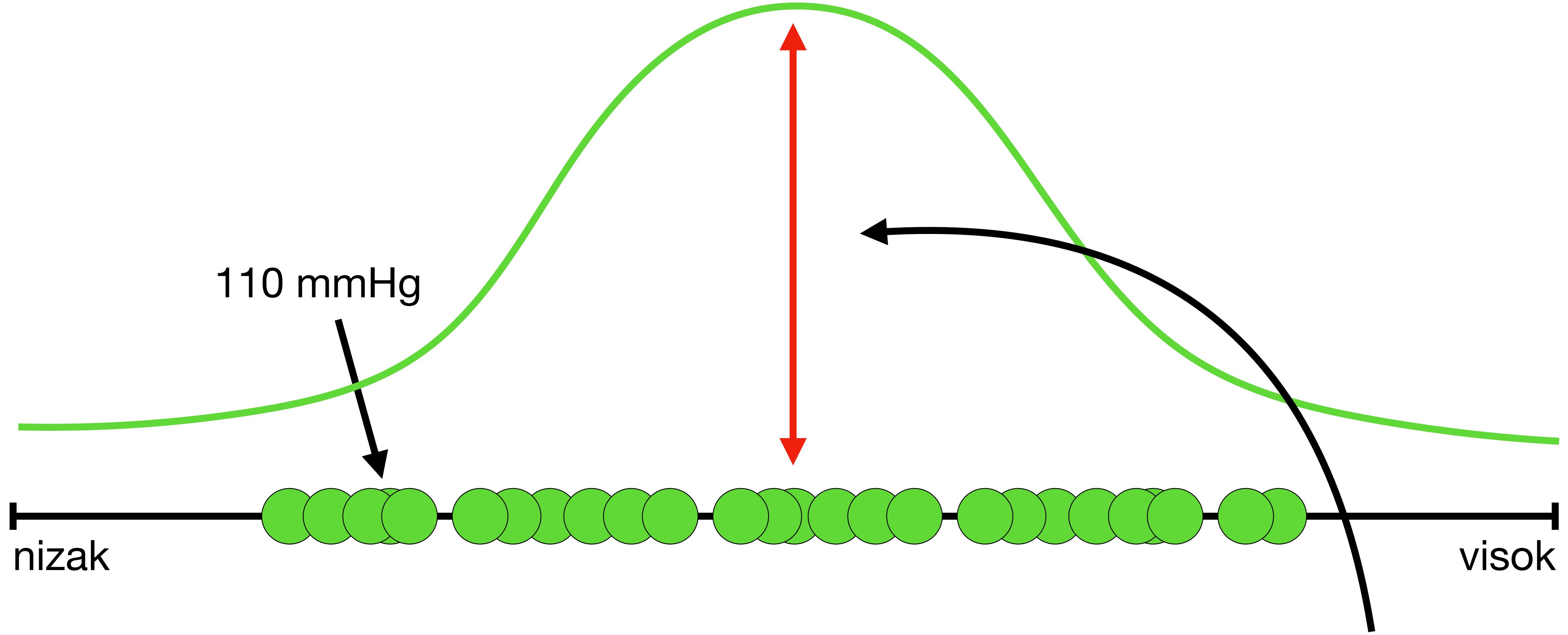
Matematičke

- Aritmetička sredina
- Geometrijska sredina
- Harmonijska sredina

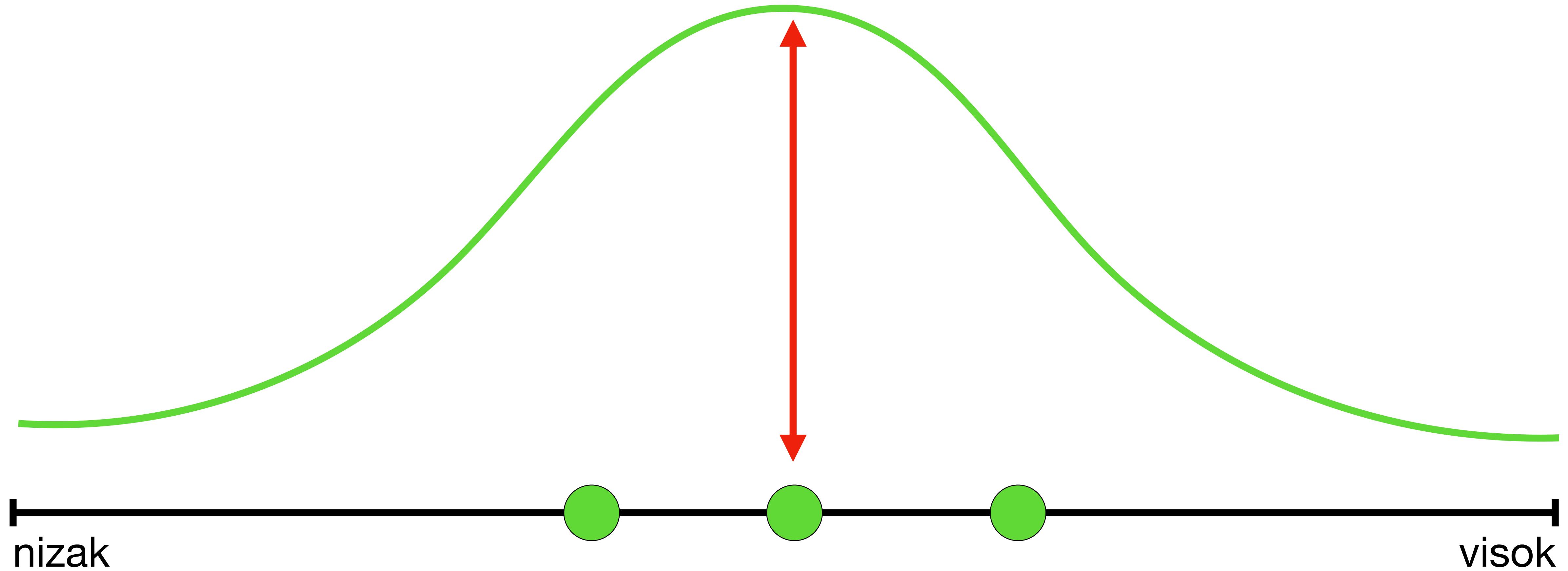
Pozicione

- Medijana
- Mod (modus, tipična vrednost)





$$\text{Stvarni prosečni pritisak} = \frac{110 + 115 + 130 + 123 + \dots + 145}{604800} = 120 \text{ mmHg}$$



$$\text{Procenjen prosečan pritisak} = \frac{110 + 115 + 130}{3} = 118.3 \text{ mmHg}$$

Procenjena aritmetička sredina

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

do poslednjeg podatka

neki podatak

od 1. podatka

broj podataka

saberi sve podatke

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Aritmetička sredina

Za Grupisane Podatke

Hemoglobin (g/L)	f
106-110	3
111-115	7
116-120	11
121-125	16
126-130	9
131-135	4

$$\bar{x}_{interval} = \frac{106 + 110}{2}$$

Aritmetička sredina Za Grupisane Podatke

$$\bar{x}_{interval} = \frac{106 + 110}{2}$$

Hemoglobin (g/L)	f	x	fx
106-110	3	109	324
111-115	7	113	791
116-120	11	118	1298
121-125	16	123	1968
126-130	9	128	1152
131-135	4	133	532
Σ	50		6065

$$f \times \bar{x}_{interval} = 7 \times 113$$

Aritmetička sredina Za Grupisane Podatke

$$\bar{x}_{interval} = \frac{106 + 110}{2}$$

Hemoglobin (g/L)	f	x	fx
106-110	3	108	324
111-115	7	113	791
116-120	11	118	1298
121-125	16	123	1968
126-130	9	128	1152
131-135	4	133	532
Σ	50		6065

$$f \times \bar{x}_{interval} = 7 \times 113$$

$$\bar{x} = \frac{\sum f \times x}{\sum f} = \frac{6065}{50} = 121.3 \text{ g/L}$$

Aritmetička sredina

Aritmetičnih sredina

Škola	Broj učenika u školi	BMI (aritmetička sredina)
A	255	22.7
B	307	19.4
C	511	20.9

$$\bar{x} = \frac{\sum n\bar{x}}{\sum n} = \frac{255 \times 22.7 + 307 \times 19.4 + 511 \times 20.9}{255 + 307 + 511} = 21.1$$

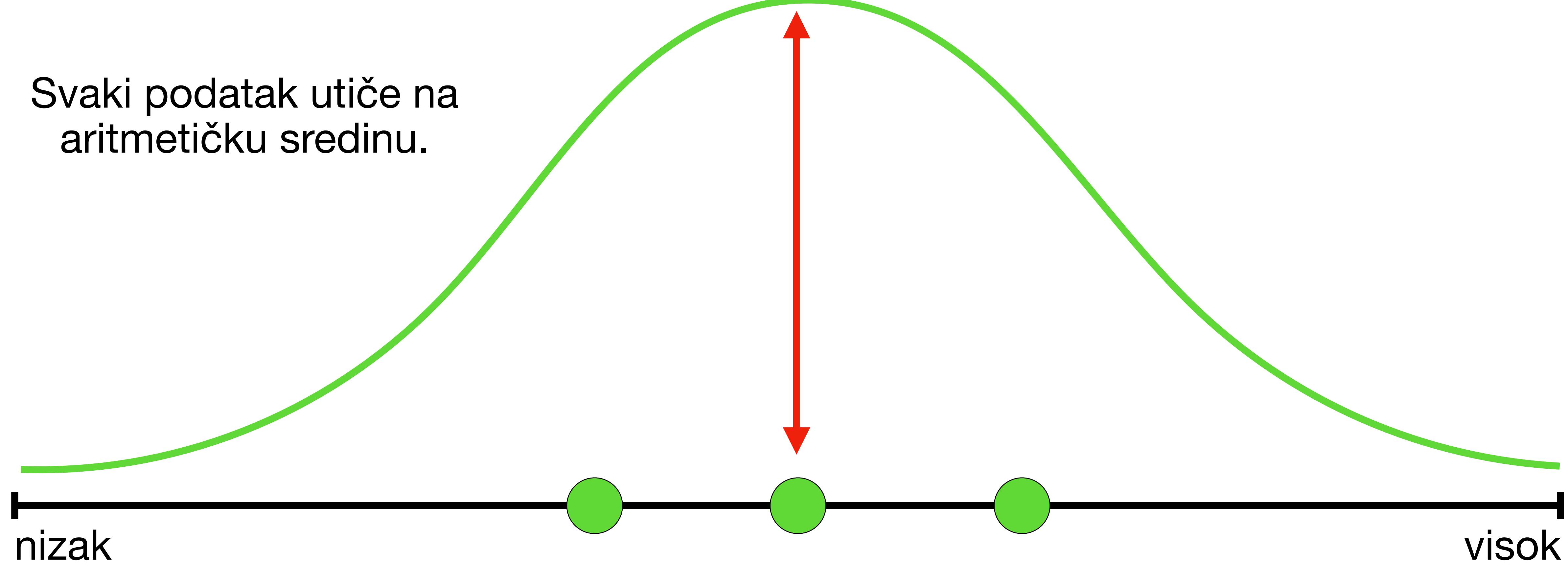
The diagram illustrates the calculation of the weighted arithmetic mean. It shows three blue arrows pointing from the school names (A, B, C) to their corresponding student counts (255, 307, 511). It also shows three red arrows pointing from the schools back to their respective BMI values (22.7, 19.4, 20.9), which are then multiplied by their respective student counts.

Aritmetička sredina

Prednosti i Mane

- Predstavlja sve vrednosti
- Jednostavna za izračunavanje
- Ne može da se koristi sa kategorijalnim podacima
- Ekstremne vrednosti značajno utiču na nju

Svaki podatak utiče na aritmetičku sredinu.



$$\text{Procenjen prosečan pritisak} = \frac{110 + 115 + 130}{3} = 118.3 \text{ mmHg}$$

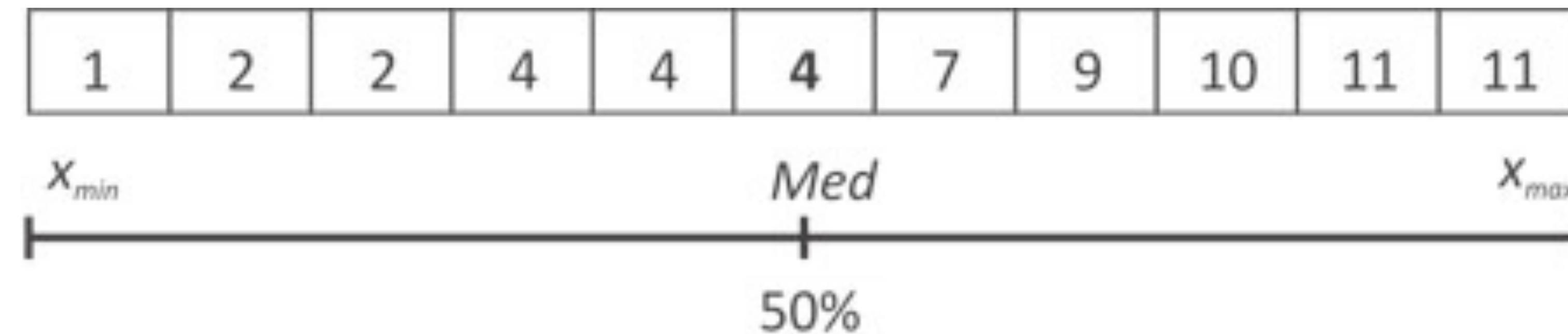
Ekstremne vrednosti imaju veći uticaj.



$$\text{Procenjen prosečan pritisak} = \frac{110 + 115 + 220}{3} = 148.3 \text{ mmHg}$$

Medijana

**Medijana ili centralna vrednost predstavlja srednju pozicionu vrednost.
Deli niz podataka poređanih po veličini na dva jednaka dela.**



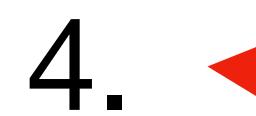
Medijana

5, 8, 11, 10, 14, 20, 1

Medijana

1. 2. 3. 4.

1, 5 , 8, 10, 11, 14, 20



$$me_p = \frac{n+1}{2} = \frac{7+1}{2} = 4$$


Medijana

110, 112, 115, 120, 130, 155

$$\frac{115 + 120}{2} = 117.5 \text{ mmHg}$$

Medijana

neparno

1 ,4, 8

parno

$$\begin{array}{r} 1,4,8,10 \\ \underbrace{4+8}_{2} = 6 \end{array}$$

Medijana

Prednosti i Mane

- Ne zavisi od ekstremnih vrednosti
- Otkriva srednju poziciju
- Manje pogodna za analizu
- Ignoriše uticaj vrednosti

Mod

Mod (tipična vrednost) je vrednost podatka sa najvećom učestalošću.

1	2	2	4	4	4	7	9	10	11	11
---	---	---	---	---	---	---	---	----	----	----

x_{min} Mo x_{max}

Mod

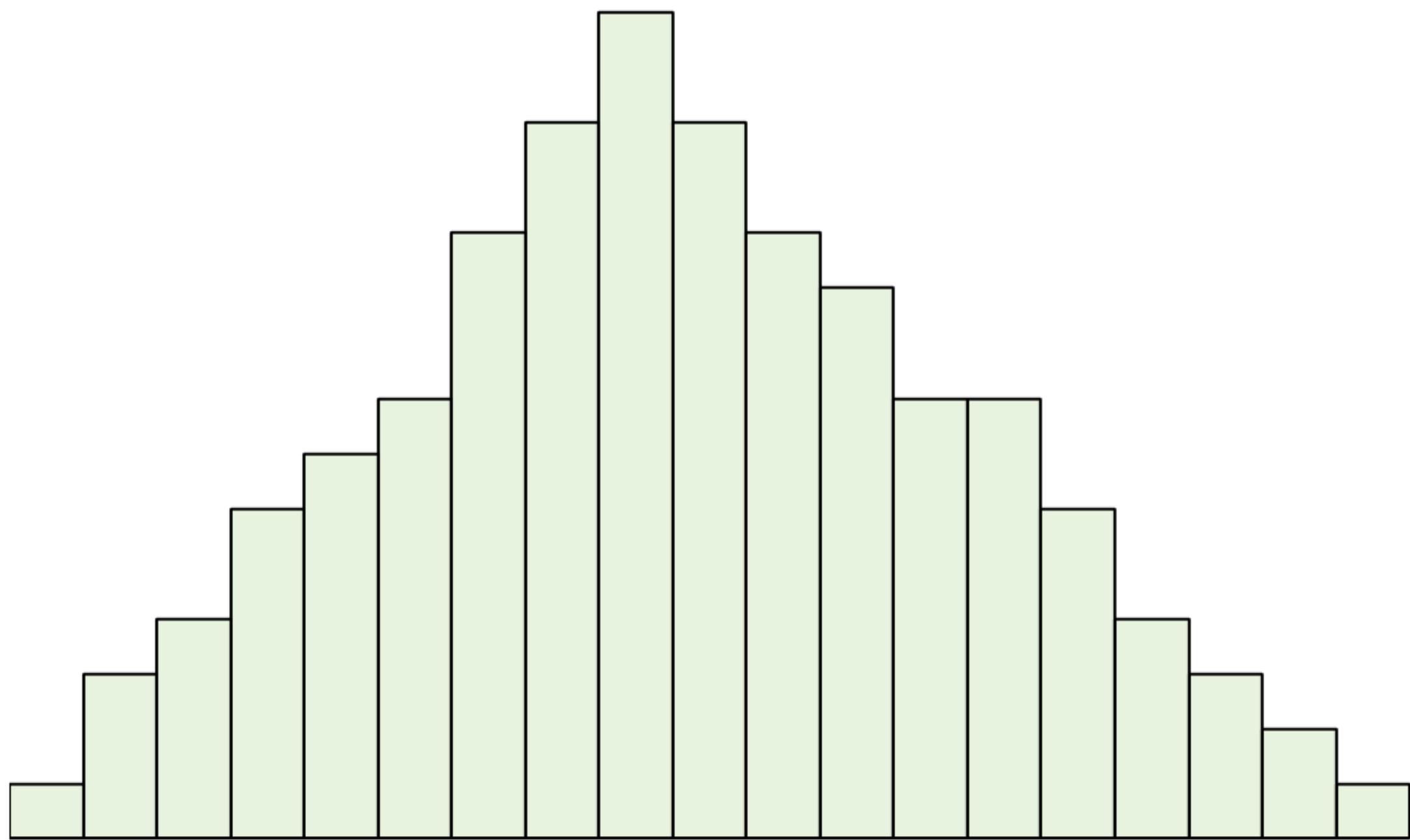
2, 0, 5, 4, 4, 3, 1, 1, 3, 1, 2, 0, 2, 1, 3, 2, 2, 3, 2, 1

Mod

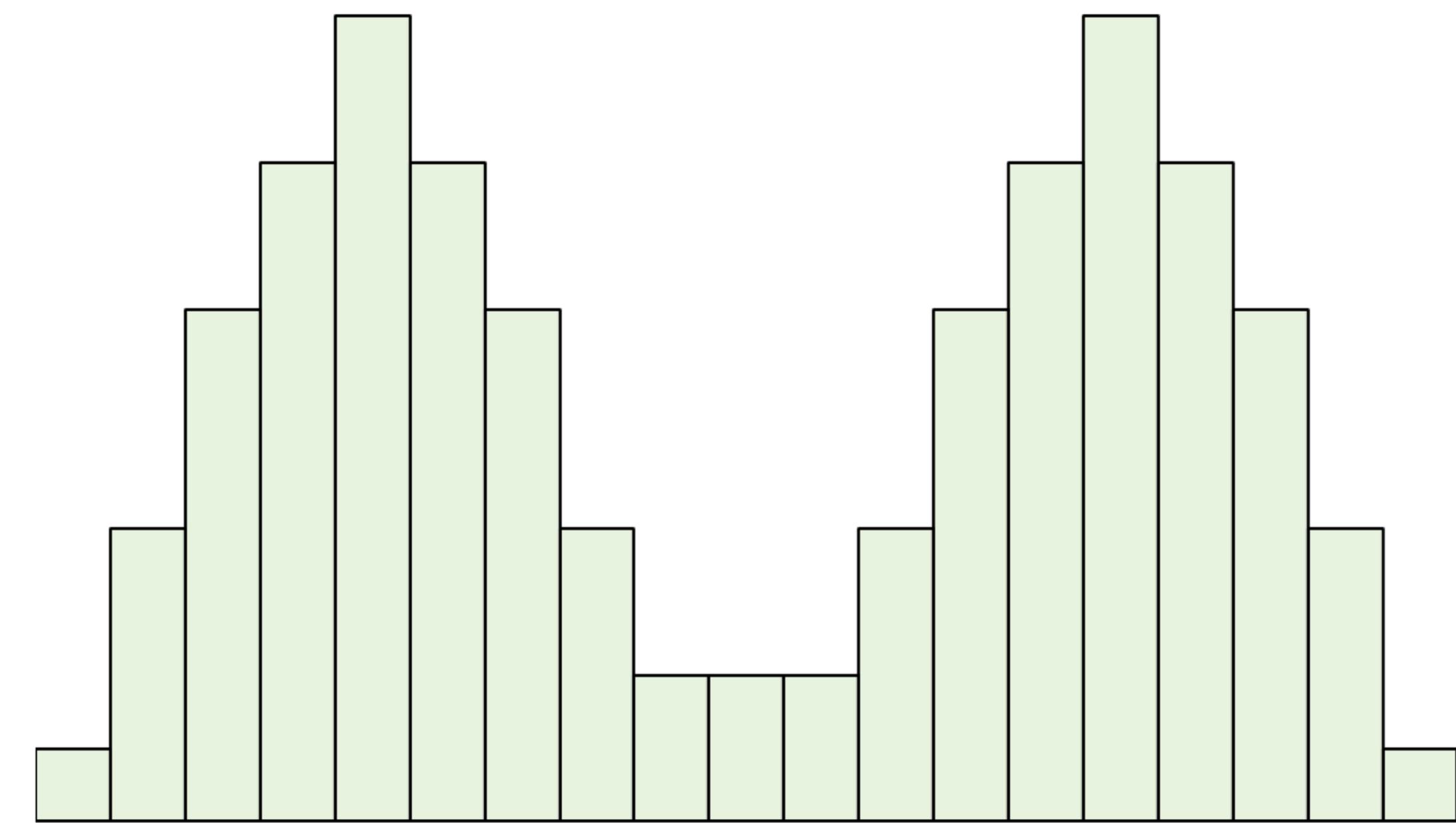
0, 0, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 5

$$mod = 2$$

Unimodalna raspodela



Bimodalna raspodela

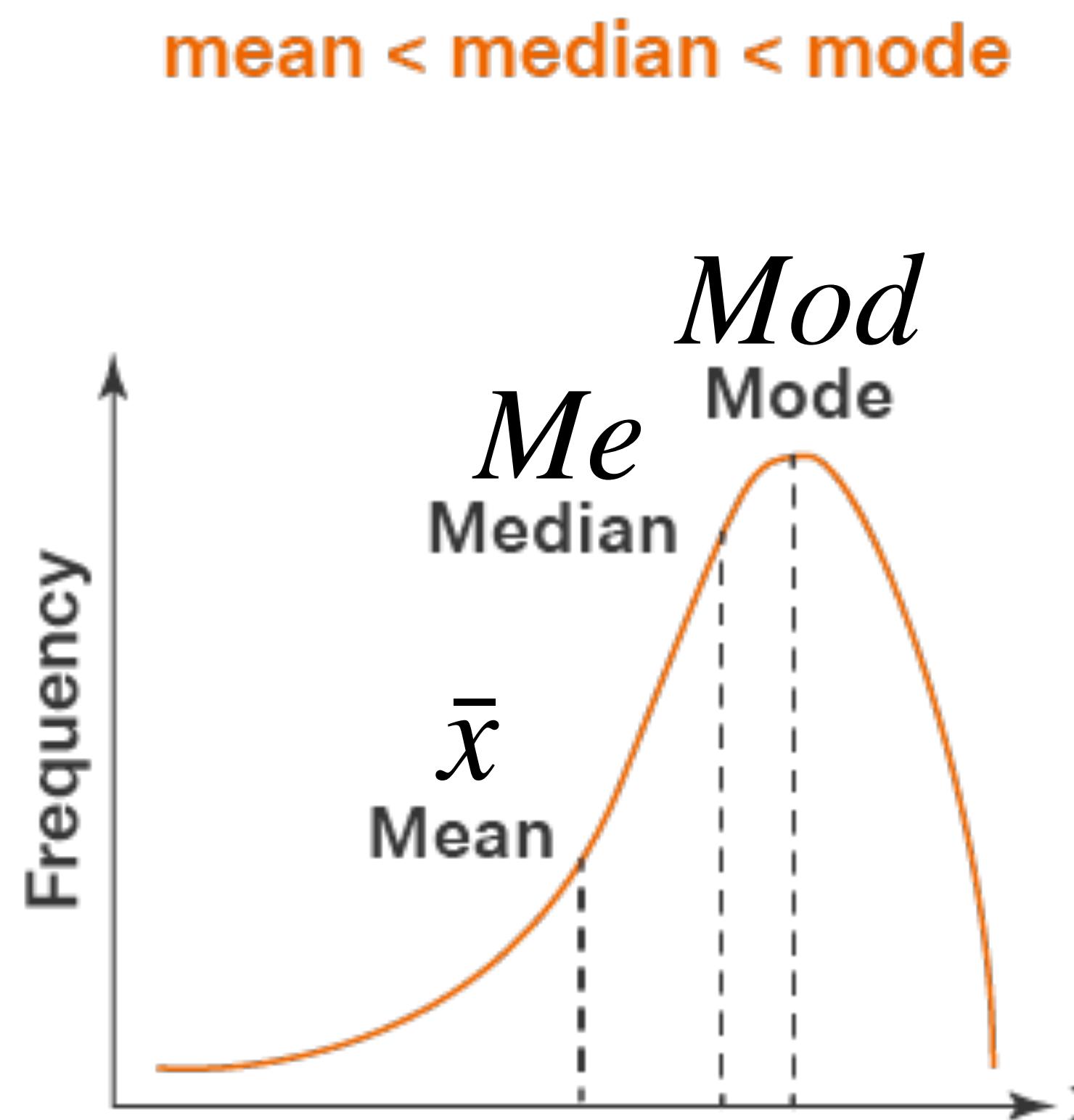


Mod

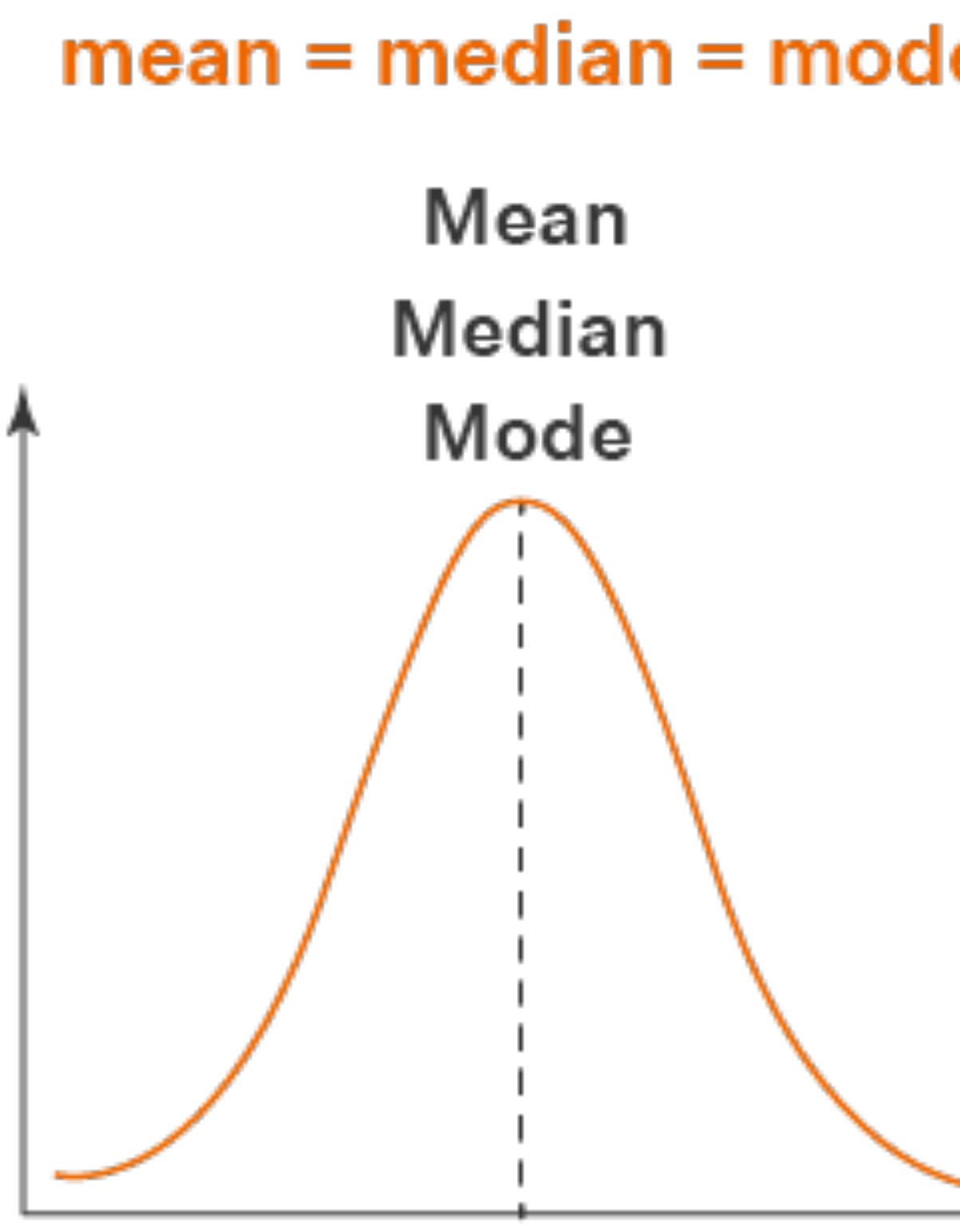
Prednosti i Mane

- Nije pod uticajem ekstremnih vrednosti
- Jedina mera za nominalne podatke
- Može postojati više od jednog moda
- Ne može se odrediti kada postoji bar dva podatka sa istim vrednostima
- Nije pogodan za analizu
- Ignoriše podatke

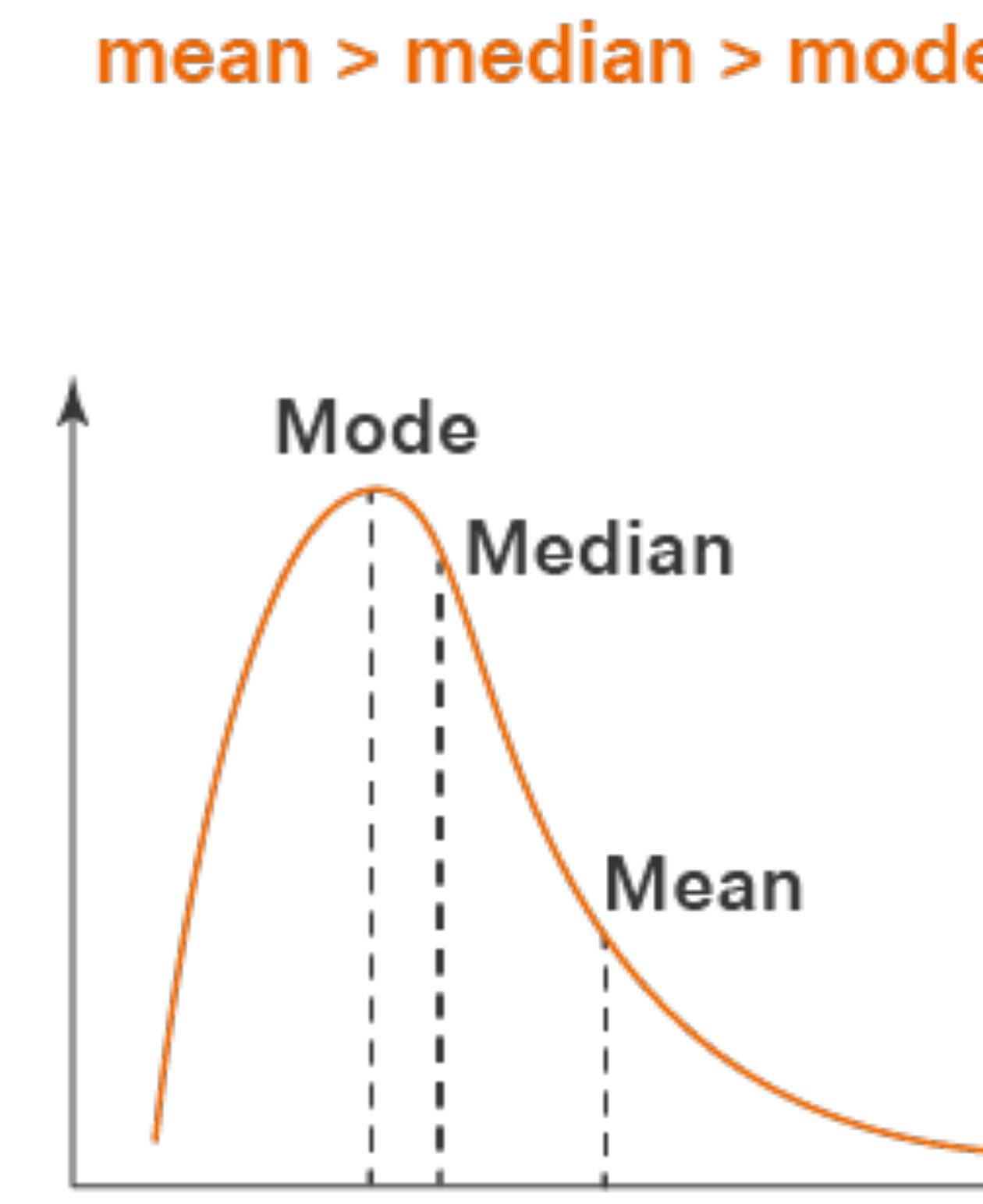
Odnos Mera Centralne Tendencije



Negatively Skewed
levo iskošena



Symmetrical Distribution



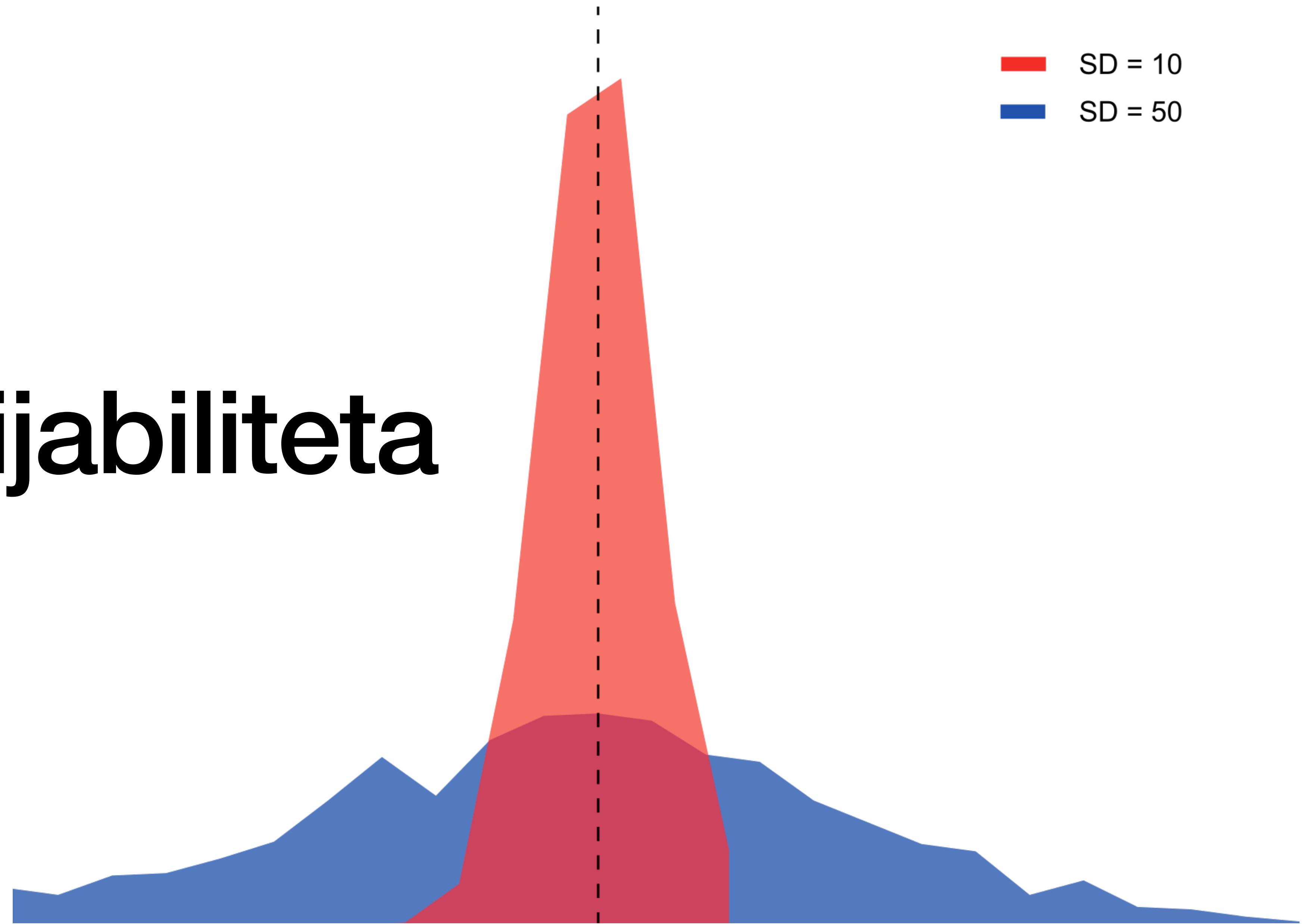
Positively Skewed
desno iskošena

Mere Centralne Tendencije

Upotreba

- Numerički podaci **bez ekstremnih vrednosti** = aritmetička sredina
- Numerički + **ekstremne vrednosti** = medijana
- Ordinalni = medijana
- Ordinalni, **<3 kategorije** = mod
- Nominalni = mod

Mere Variabiliteta



Mere varijabiliteta opisuju variranje vrednosti skupa podataka oko srednje vrednosti.

Mere Varijabiliteta

Apsolutne

- Interval (raspon) varijacije
- Kvantili, percentili, decili, kvartili
- Srednje absolutno odstupanje
- Varijansa
- Standardna devijacija

Relativne

- Koeficijent varijacije
- Standardizovane (normalizovane) vrednosti

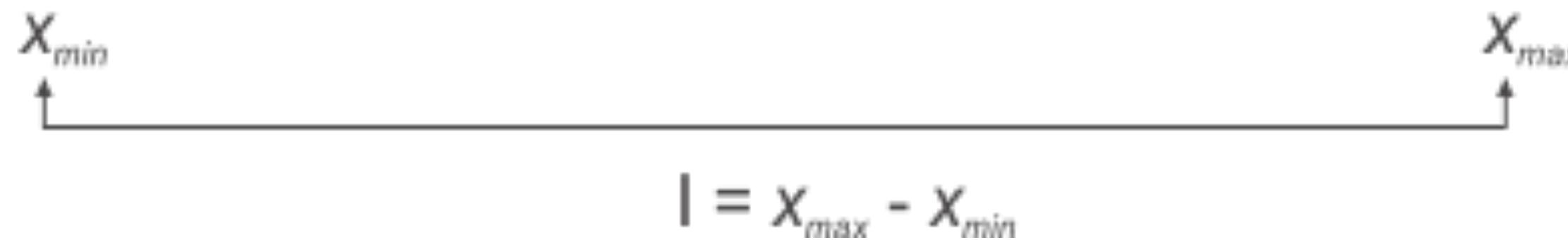
Mere Varijabiliteta

Interval Varijacije

Redni broj

1	2	3	4	5	6	7	8	9	10
165	169	171	172	174	175	177	181	183	190

Visina (cm)



$$I = x_{max} - x_{min} = 190 - 165 = 25 \text{ cm}$$

Mere Varijabiliteta

Interval Varijacije - Prednosti i Mane

- Jednostavan
- Zavisi od ekstremnih vrednosti
- Ne daje informaciju o grupisanju podataka
- Vrednost raste sa povećanjem uzorka

Mere Varijabiliteta

Kvantili

$$T_{99\%} \geq 37.7^\circ C$$

Redni broj	1	2	3	4	5	6	7	8	9	10	11
Telesna visina	165	169	171	172	174	175	177	181	183	190	192

$k = 100, k-1 = 99$

percentili



$k = 10, k-1 = 9$

decili



$k = 4, k-1 = 3$

kvartili



50% Opservacija

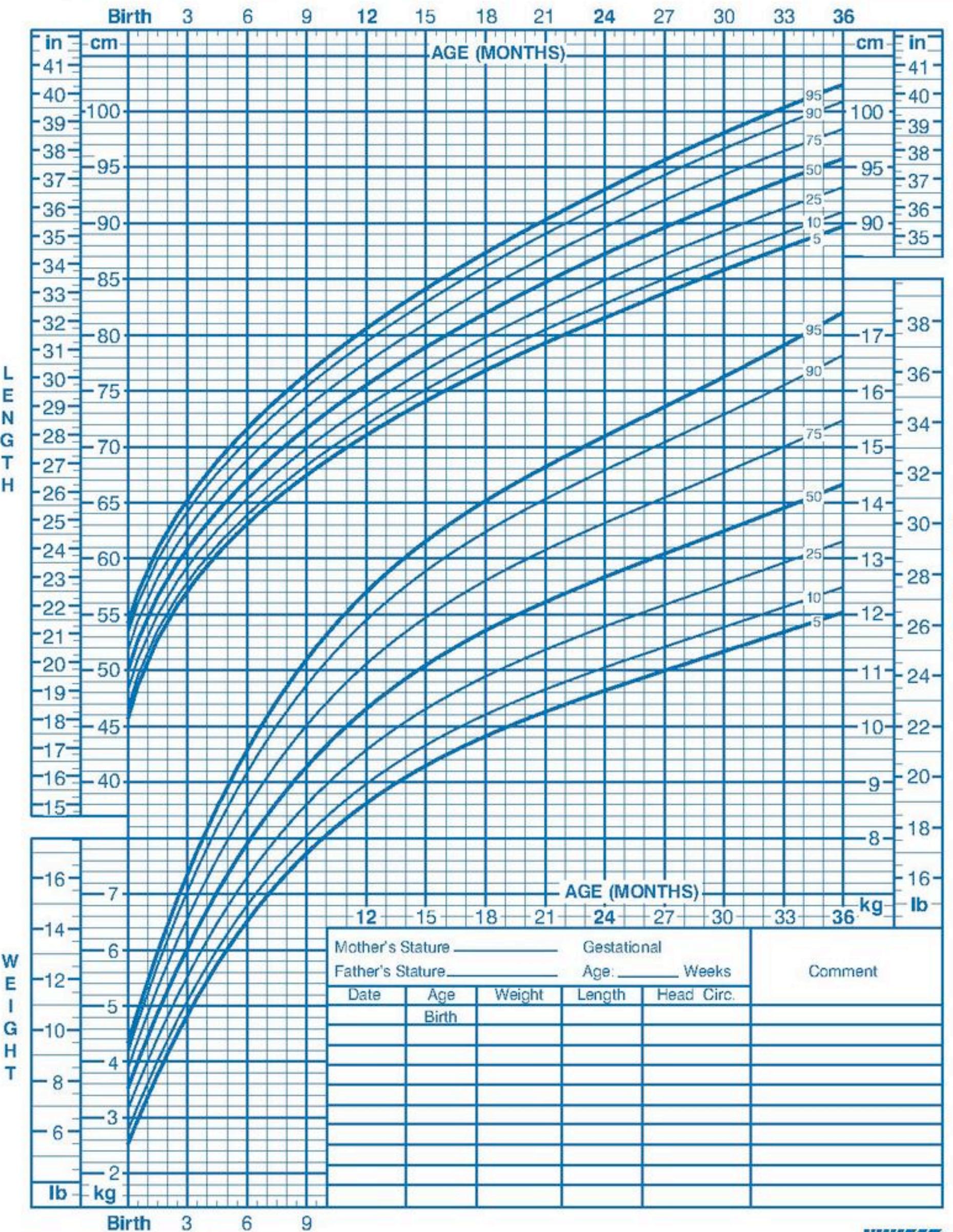
Interkvartilni opseg
 $I_Q = Q_3 - Q_1$

Birth to 36 months: Boys

Length-for-age and Weight-for-age percentiles

NAME _____

RECORD # _____



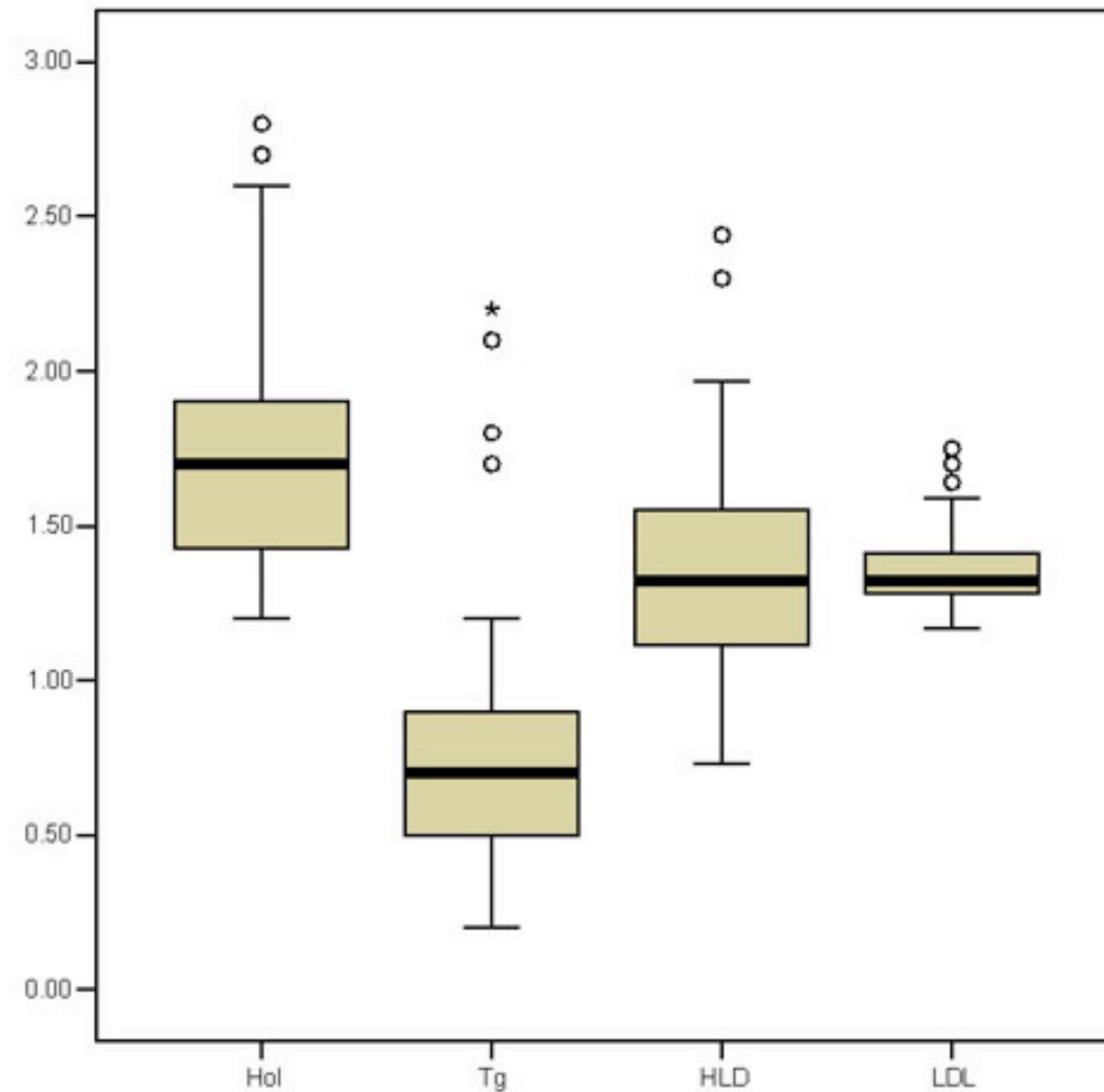
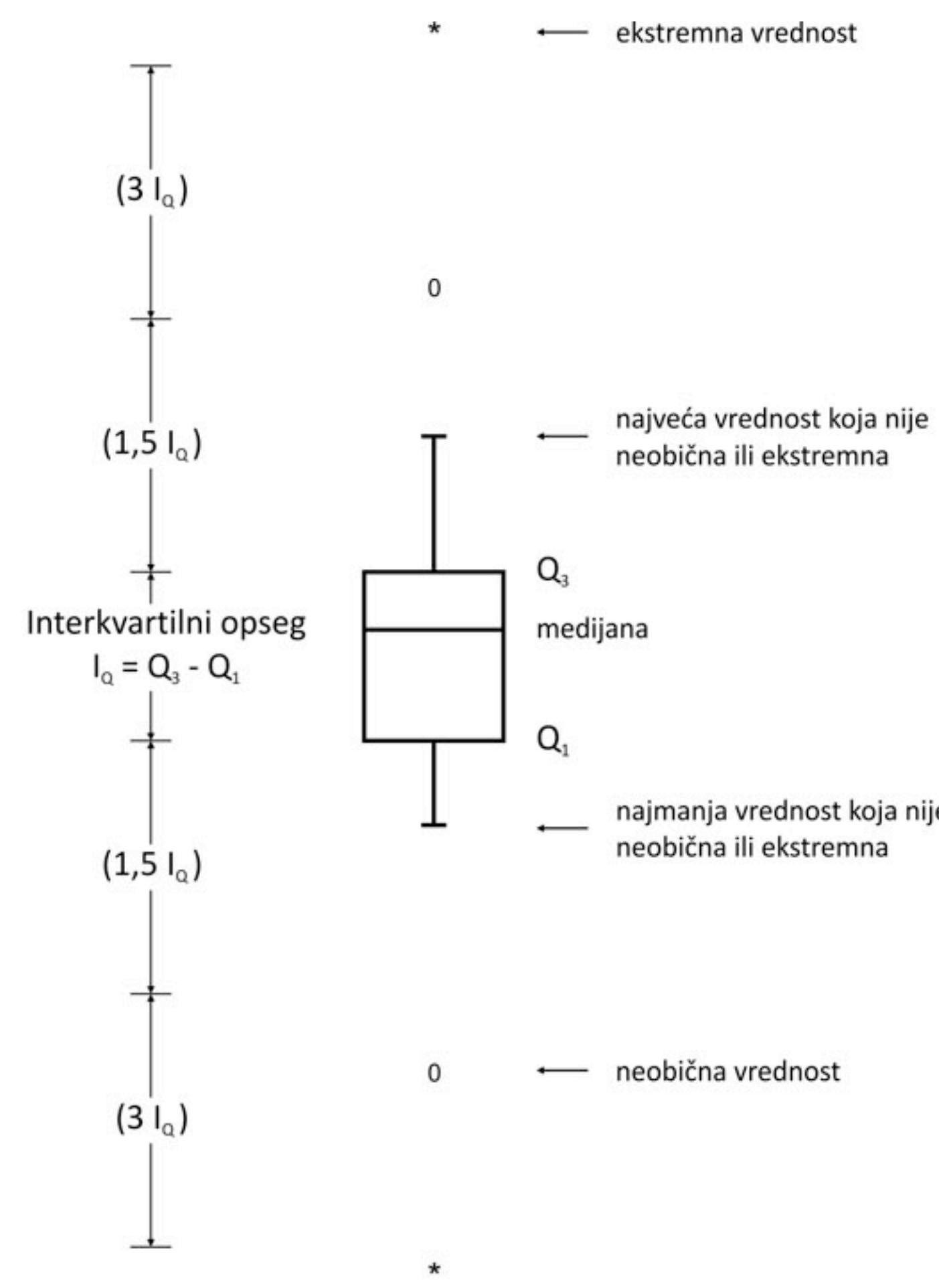
Published May 30, 2000 (modified 4/20/01).

SOURCE: Developed by the National Center for Health Statistics in collaboration with
the National Center for Chronic Disease Prevention and Health Promotion (2000).
<http://www.cdc.gov/growthcharts>

SAFER • HEALTHIER • PEOPLE™

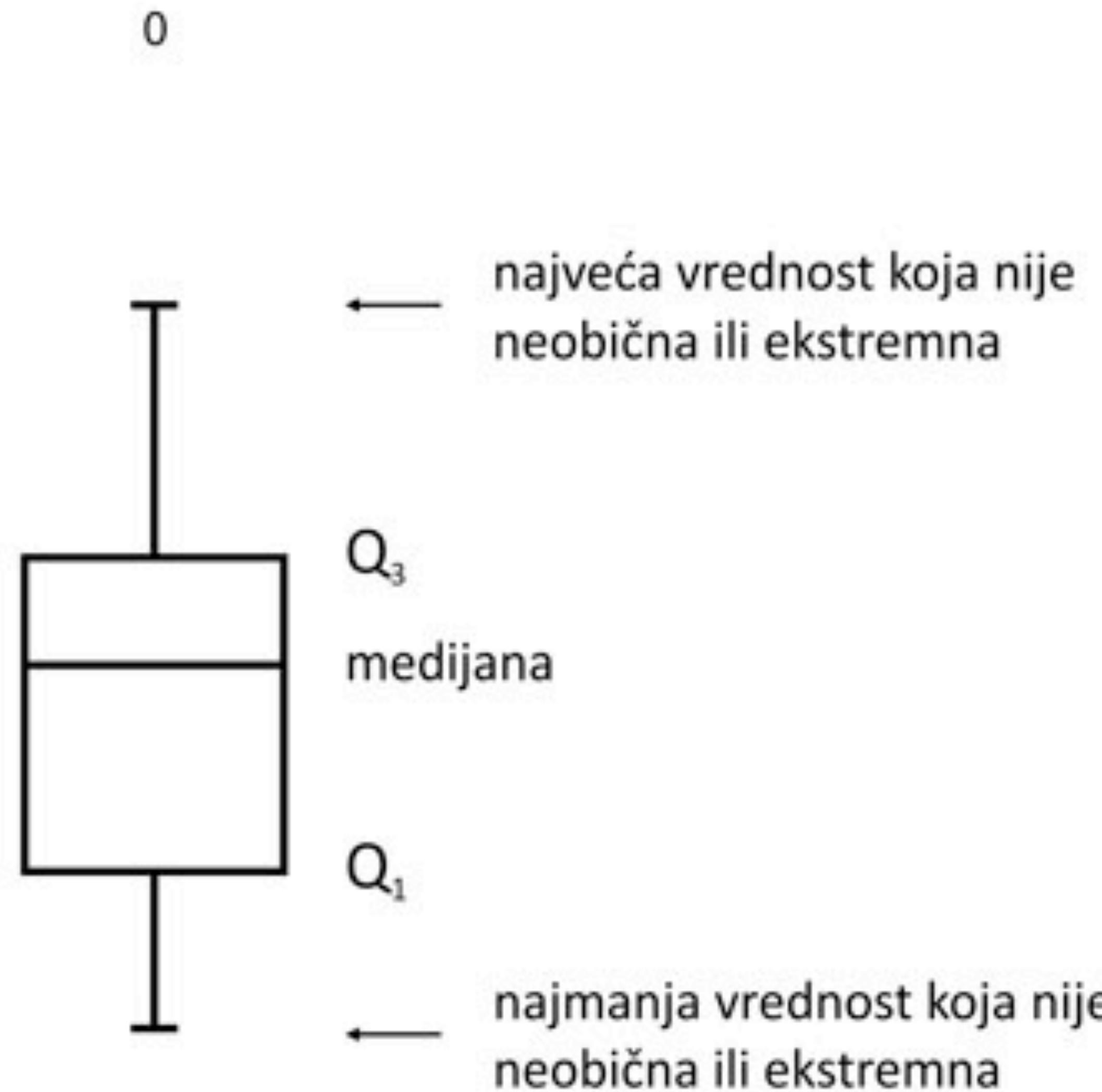
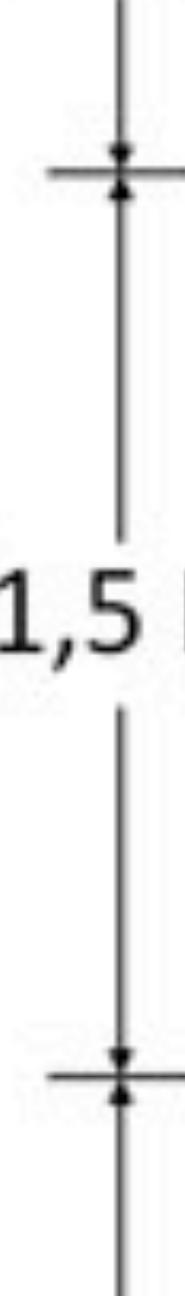
Mere Varijabiliteta

Grafikon kutije (Box-and-Whisker-Plot, Box-plot, B-P)



$(1,5 I_Q)$

Interkvartilni opseg

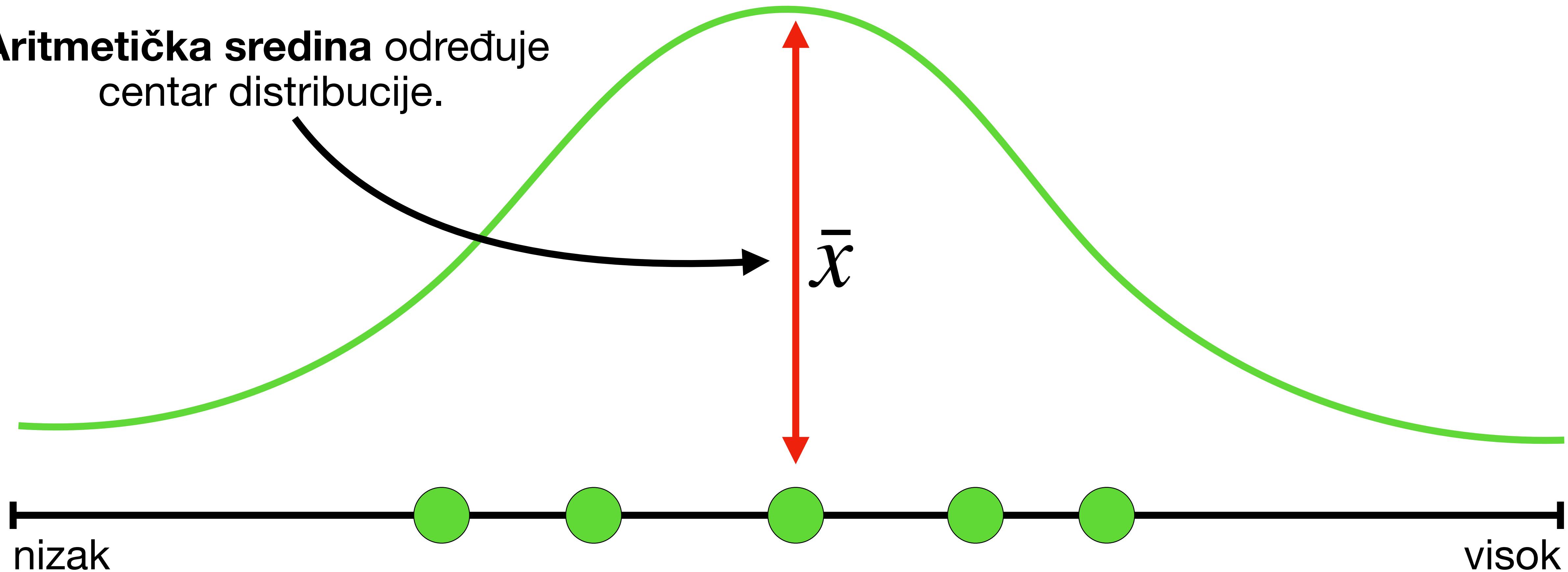
$$I_Q = Q_3 - Q_1$$


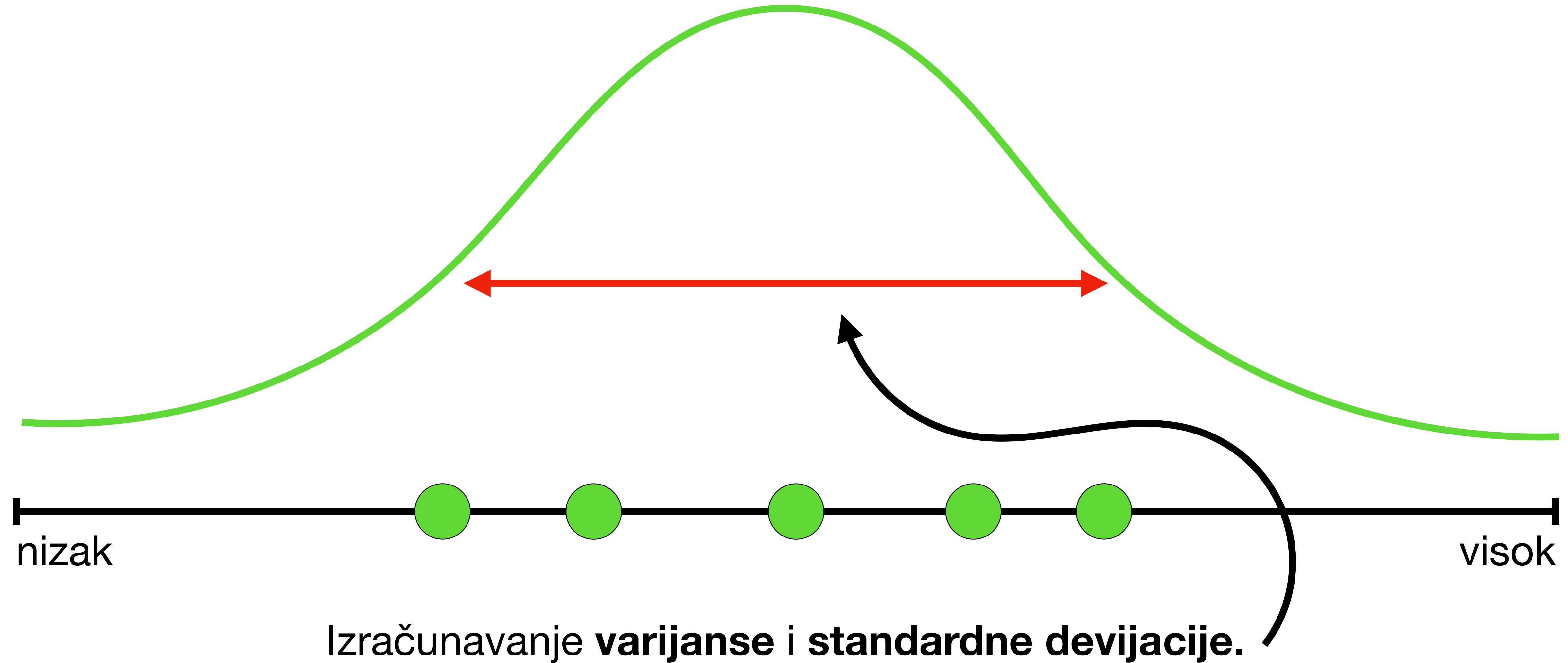
Mere Variabiliteta

Apsolutne

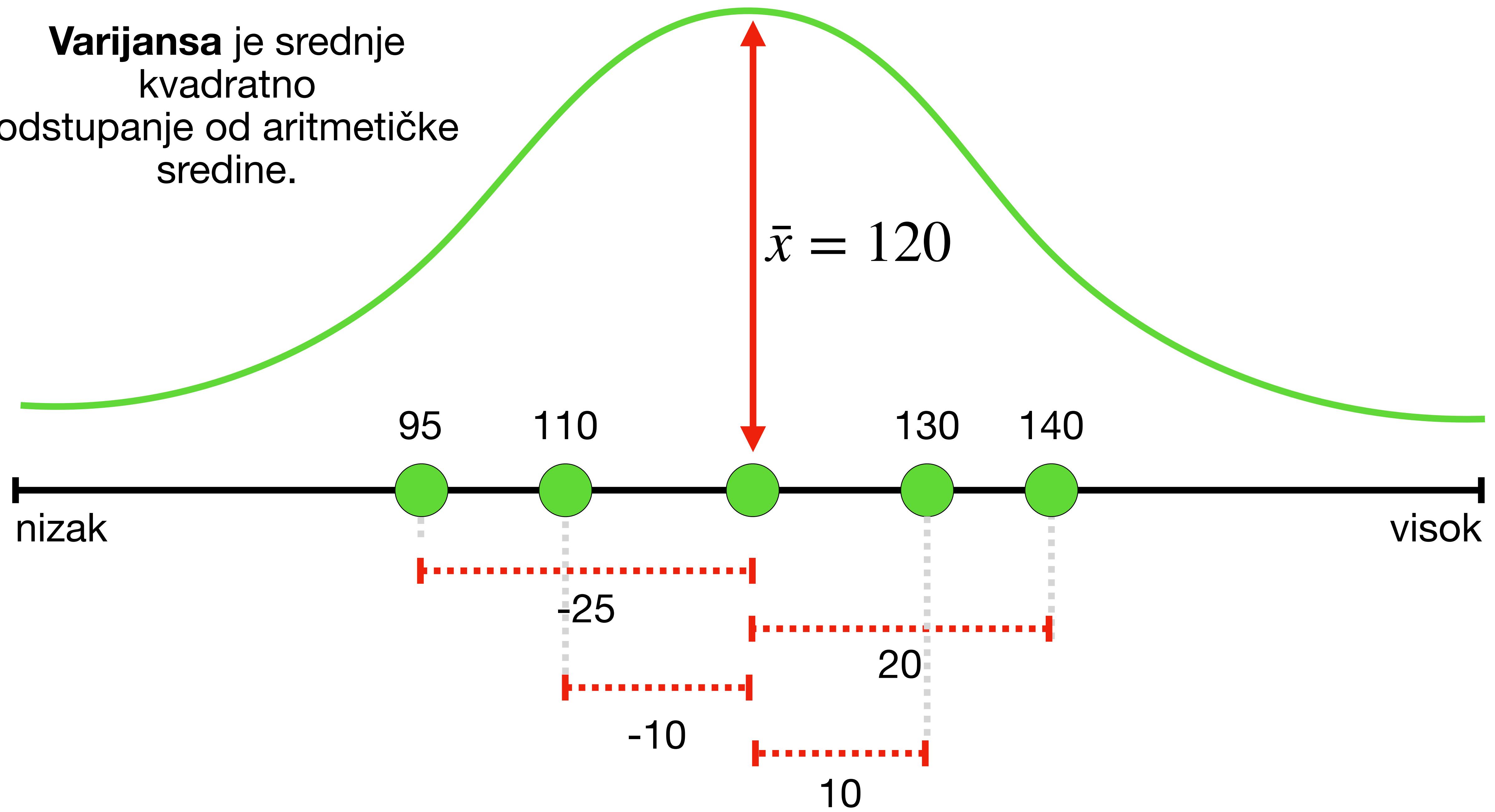
- Varijansa
- Standardna devijacija

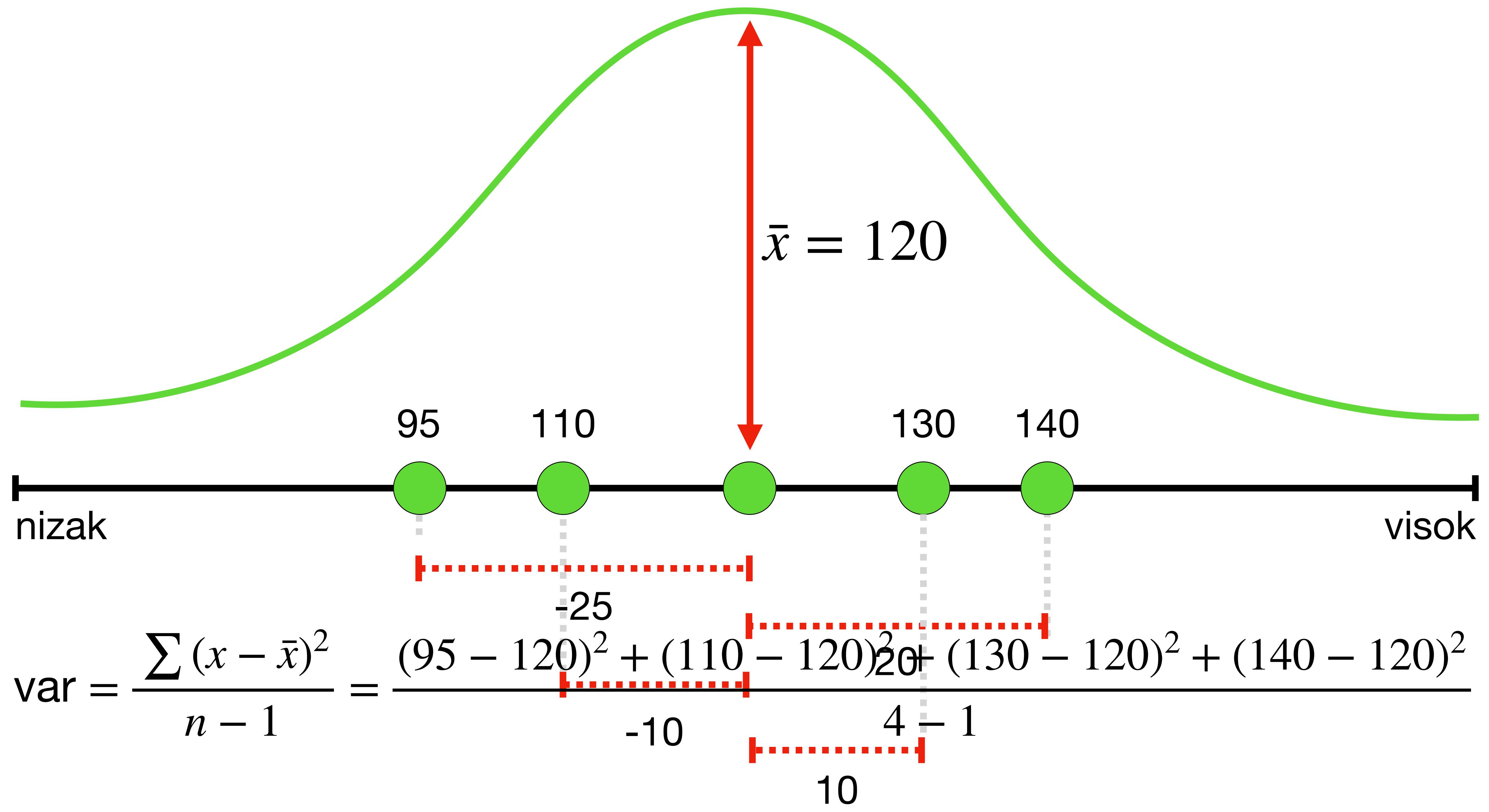
Aritmetička sredina određuje centar distribucije.





Varijansa je srednje kvadratno odstupanje od aritmetičke sredine.





$$\text{var} = \frac{\sum (x - \mu)^2}{n - 1} = \frac{(95 - 120)^2 + (110 - 120)^2 + (130 - 120)^2 + (140 - 120)^2}{4 - 1}$$

$$\text{varijansa} = sd^2 = \frac{\sum (x - \bar{x})^2}{n - 1} \text{ mmHg}^2$$

$$\text{varijansa} = sd^2 = \frac{\sum (x - \bar{x})^2}{n - 1} \text{ mmHg}^2$$

$$sd = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} \text{ mmHg}$$

Mere Varijabiliteta

Koeficijent Varijacije, Relativna Mera

$$CV = \frac{sd}{\bar{x}} \times 100 \%$$

$\leq 30\%$ Homogeni podaci

$> 30\%$ Heterogeni podaci

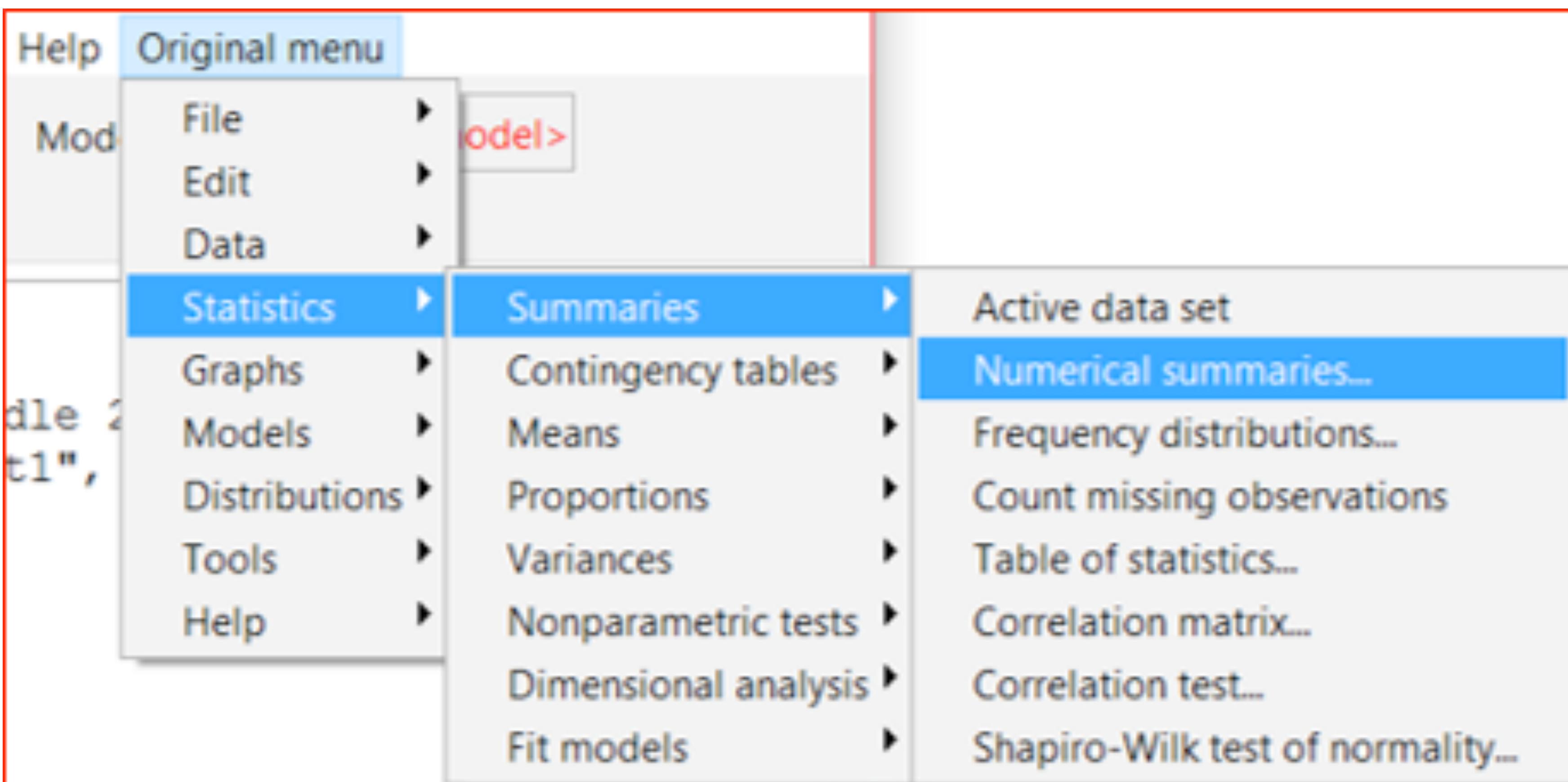
Mere Varijabiliteta

Z-Vrednost, Relativna Mera

$$z = \frac{x_i - \bar{x}}{sd}$$

Koliko je *sd* podatak udaljen od aritmetičke sredine?





Deskripcija Podataka

1. Učitati bazu podataka: **zadovoljstvo zdravstvenom zaštitom.xlsx**
2. Za varijablu **starost** izračunati **mere centralne tendencije**
 1. Aritmetičku sredinu, standardnu devijaciju, koef. varijacije
 2. Mod
 3. Interval varijacije
3. Izračunati **mere centralne tendencije** za **starost** grupisane **po polu**