

Deskriptivna statistika: frekvencije

Prirodoslovno-matematički Fakultet, Sveučilište u Zagrebu

16. listopada 2026.

Mjere centralne tendencije

N = duljina uzorka (broj podataka)!

Ako nas zanima središnja vrijednost, možemo ju odrediti na više načina.

- ① **mod** - vrijednost koja se najčešće pojavljuje u podacima. Možemo ju pronaći uz pomoć tablice frekvencija.

Mjere centralne tendencije

N = duljina uzorka (broj podataka)!

Ako nas zanima središnja vrijednost, možemo ju odrediti na više načina.

- 1 **mod** - vrijednost koja se najčešće pojavljuje u podacima. Možemo ju pronaći uz pomoć tablice frekvencija.
- 2 **aritmetička sredina** (srednja vrijednost) - klasična mjera sredine. Jako osjetljiva na podatke koji odstupaju. Npr. izračunajte aritmetičku sredinu od `a <- c(1,1,2,1,1000)`.

Mjere centralne tendencije

N = duljina uzorka (broj podataka)!

Ako nas zanima središnja vrijednost, možemo ju odrediti na više načina.

- 1 **mod** - vrijednost koja se najčešće pojavljuje u podacima. Možemo ju pronaći uz pomoć tablice frekvencija.
- 2 **aritmetička sredina** (srednja vrijednost) - klasična mjera sredine. Jako osjetljiva na podatke koji odstupaju. Npr. izračunajte aritmetičku sredinu od `a <- c(1,1,2,1,1000)`.
- 3 **medijan** - položajna mjera sredine. Podatke sortiramo od najmanjeg do najvećeg te odaberemo središnji.

Koristimo funkcije `mean()` i `median()`. Za određivanje moda kreiramo tablicu frekvencija `table()`.

Mjere centralne tendencije

N = duljina uzorka (broj podataka)!

Ako nas zanima središnja vrijednost, možemo ju odrediti na više načina.

- 1 **mod** - vrijednost koja se najčešće pojavljuje u podacima. Možemo ju pronaći uz pomoć tablice frekvencija.
- 2 **aritmetička sredina** (srednja vrijednost) - klasična mjera sredine. Jako osjetljiva na podatke koji odstupaju. Npr. izračunajte aritmetičku sredinu od `a <- c(1,1,2,1,1000)`.
- 3 **medijan** - položajna mjera sredine. Podatke sortiramo od najmanjeg do najvećeg te odaberemo središnji.

Koristimo funkcije `mean()` i `median()`. Za određivanje moda kreiramo tablicu frekvencija `table()`.

Zadatak

Izračunajte aritmetičku sredinu, medijan i mod podataka
2, 5, 7, 8, 5, 3, 3, 1, 0, 2, 4, 3, 5, 7.

Mjere centralne tendencije

N = duljina uzorka (broj podataka)!

Ako nas zanima središnja vrijednost, možemo ju odrediti na više načina.

- 1 **mod** - vrijednost koja se najčešće pojavljuje u podacima. Možemo ju pronaći uz pomoć tablice frekvencija.
- 2 **aritmetička sredina** (srednja vrijednost) - klasična mjera sredine. Jako osjetljiva na podatke koji odstupaju. Npr. izračunajte aritmetičku sredinu od `a <- c(1,1,2,1,1000)`.
- 3 **medijan** - položajna mjera sredine. Podatke sortiramo od najmanjeg do najvećeg te odaberemo središnji.

Koristimo funkcije `mean()` i `median()`. Za određivanje moda kreiramo tablicu frekvencija `table()`.

Zadatak

Izračunajte aritmetičku sredinu, medijan i mod podataka
2, 5, 7, 8, 5, 3, 3, 1, 0, 2, 4, 3, 5, 7.

Kako ih računamo ako su nam samo dostupne tablice frekvencija?

Mjere sredine iz tablice frekvencija

Vrijednost	0	1	2	3	4	5	7	8
Frekvencija	1	1	2	3	1	3	2	1

Mjere sredine iz tablice frekvencija

Vrijednost	0	1	2	3	4	5	7	8
Frekvencija	1	1	2	3	1	3	2	1

- aritmetička sredina (f_i apsolutne frekvencije, r_i relativne)

$$\frac{\sum f_i x_i}{\sum f_i} = \frac{\sum f_i x_i}{N}, \quad \text{odnosno} \quad \sum r_i x_i$$

Mjere sredine iz tablice frekvencija

Vrijednost	0	1	2	3	4	5	7	8
Frekvencija	1	1	2	3	1	3	2	1

- aritmetička sredina (f_i apsolutne frekvencije, r_i relativne)

$$\frac{\sum f_i x_i}{\sum f_i} = \frac{\sum f_i x_i}{N}, \quad \text{odnosno} \quad \sum r_i x_i$$

- medijan - iz kumulativnih (relativnih) frekvencija

Zadatak

Izračunajte aritmetičku sredinu, medijan i mod za podatke dane tablicom

Vrijednost	3	-1	5	2	1
Frekvencija	4	2	2	3	3

Tablica frekvencija s intervalnim rasponima

Interval	Frekvencija (f)	Kumulativna frekvencija
10 – 20	3	3
20 – 30	5	8
30 – 40	8	16
40 – 50	4	20
Ukupno	N = 20	

Aritmetička sredina? Medijan? Mod?

Tablica frekvencija s intervalnim rasponima

Interval	Frekvencija (f)	Kumulativna frekvencija
10 – 20	3	3
20 – 30	5	8
30 – 40	8	16
40 – 50	4	20
Ukupno	N = 20	

Aritmetička sredina? Medijan? Mod?

- aritmetička sredina - računamo kao da svi podaci unutar razreda imaju istu srednju vrijednost (sredina intervala)
- medijan = $L_{med} + \frac{(N+1)/2 - (kf_{preth} + 1)}{f_{med}} \times \text{širina}$
- mod ne možemo → modalni razred!

Tablica frekvencija s intervalnim rasponima

Interval	Frekvencija (f)
0.5 – 1.5	8
1.5 – 2.5	5
2.5 – 3.5	2
3.5 – 4.5	10

Zadatak

Za gornju tablicu odredite aritmetičku sredinu, medijan i modalni razred.

Druge mjere centralne tendencije

- Težinska (ponderirana) srednja vrijednost. Za svaki podatak $x_1, \dots, x_N$ imamo težine $w_1, \dots, w_N$ (npr. označavaju koliko bitan/točan pojedini podatak). Koristimo

$$\mu_w = \frac{\sum w_i x_i}{\sum w_i}$$

Druge mjere centralne tendencije

- Težinska (ponderirana) srednja vrijednost. Za svaki podatak $x_1, \dots, x_N$ imamo težine $w_1, \dots, w_N$ (npr. označavaju koliko bitan/točan pojedini podatak). Koristimo

$$\mu_w = \frac{\sum w_i x_i}{\sum w_i}$$

- Geometrijska sredina

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

Druge mjere centralne tendencije

- Težinska (ponderirana) srednja vrijednost. Za svaki podatak $x_1, \dots, x_N$ imamo težine $w_1, \dots, w_N$ (npr. označavaju koliko bitan/točan pojedini podatak). Koristimo

$$\mu_w = \frac{\sum w_i x_i}{\sum w_i}$$

- Geometrijska sredina

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

- Harmonijska sredina

$$h = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_N}}$$

Druge mjere centralne tendencije

- Težinska (ponderirana) srednja vrijednost. Za svaki podatak $x_1, \dots, x_N$ imamo težine $w_1, \dots, w_N$ (npr. označavaju koliko bitan/točan pojedini podatak). Koristimo

$$\mu_w = \frac{\sum w_i x_i}{\sum w_i}$$

- Geometrijska sredina

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

- Harmonijska sredina

$$h = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_N}}$$

Zadatak

Za podatke 3, 4, 5, 1, 1, 2, 3, 1, 5 izračunajte geometrijsku i harmonijsku sredinu. Ako su pripadne težine podataka (redom) 0.5, 0.5, 2.1, 0.6, 0.6, 0.5, 0.3, 0.5, 3.4, odredite tež. srednju vrijednost.

Mjere raspršenosti

Odstupanje i -tog podatke od aritmetičke sredine je $x_i - \mu$. Kako mjeriti ukupno odstupanje?

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

Mjere raspršenosti

Odstupanje i -tog podatke od aritmetičke sredine je $x_i - \mu$. Kako mjeriti ukupno odstupanje?

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

Ideja - gledamo $(x_i - \mu)^2$ ili $|x_i - \mu|$.

- varijanca ($\text{var}()$) je srednjekvadratno odstupanje od srednje vrijednosti.

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

Mjere raspršenosti

Odstupanje i -tog podatke od aritmetičke sredine je $x_i - \mu$. Kako mjeriti ukupno odstupanje?

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

Ideja - gledamo $(x_i - \mu)^2$ ili $|x_i - \mu|$.

- varijanca ($\text{var}()$) je srednjekvadratno odstupanje od srednje vrijednosti.

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

- standardna devijacija ($\text{sd}()$) = korijen iz varijance

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

Mjere raspršenosti

Funkcije `var()` i `sd()` u R-u dijele sa $N - 1$ - moramo korigirati!

$$\sigma^2 = \text{var}(x) * (N - 1)/N, \quad \sigma = \text{sd}(x) * \text{sqrt}((N - 1)/N)$$

Mjere raspršenosti

Funkcije `var()` i `sd()` u R-u dijele sa $N - 1$ - moramo korigirati!

$$\sigma^2 = \text{var}(x) * (N - 1)/N, \quad \sigma = \text{sd}(x) * \text{sqrt}((N - 1)/N)$$

- Koeficijent varijacije - govori nam koliko je srednje raspršenje veliko u odnosu na aritmetičku sredinu

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

Zadatak

Za podatke 2, 3, 7, 7, -5, 3, -1, -2 odredite varijancu, standardnu devijaciju i koeficijent varijacije.

Mjere raspršenosti za tablične podatke

Vrijednost	0	1	2	3	4	5	7	8
Frekvencija	1	1	2	3	1	3	2	1

Mjere raspršenosti za tablične podatke

Vrijednost	0	1	2	3	4	5	7	8
Frekvencija	1	1	2	3	1	3	2	1

- varijanca

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

Mjere raspršenosti za tablične podatke

Vrijednost	0	1	2	3	4	5	7	8
Frekvencija	1	1	2	3	1	3	2	1

- varijanca

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

- standardna devijacija

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

Zadatak

Izračunajte varijancu i standardnu devijaciju za podatke dane tablicom

Vrijednost	3	-1	5	2	1
Frekvencija	4	2	2	3	3

Kvartili

Promotrimo podatke 3, 5, 1, 4, 1, 6, 10, 10, 8. Onda je

$$x_1 = 3, x_2 = 5, \dots, x_8 = 10, x_9 = 8.$$

Kada ih sortiramo dobit ćemo 1, 1, 3, 4, 5, 6, 8, 10, 10. Ako se referiramo na drugi podatak u sortiranom uzorku (drugi podatak po redu), pišemo $x_{(2)}$.

- Koliko iznosi $x_{(5)}$?

Kvartili

Promotrimo podatke 3, 5, 1, 4, 1, 6, 10, 10, 8. Onda je

$x_1 = 3, x_2 = 5, \dots, x_8 = 10, x_9 = 8$.

Kada ih sortiramo dobit ćemo 1, 1, 3, 4, 5, 6, 8, 10, 10. Ako se referiramo na drugi podatak u sortiranom uzorku (drugi podatak po redu), pišemo $x_{(2)}$.

- Koliko iznosi $x_{(5)}$?
- Koliko iznosi medijan?

Kvartili

Promotrimo podatke 3, 5, 1, 4, 1, 6, 10, 10, 8. Onda je $x_1 = 3, x_2 = 5, \dots, x_8 = 10, x_9 = 8$.

Kada ih sortiramo dobit ćemo 1, 1, 3, 4, 5, 6, 8, 10, 10. Ako se referiramo na drugi podatak u sortiranom uzorku (drugi podatak po redu), pišemo $x_{(2)}$.

- Koliko iznosi $x_{(5)}$?
- Koliko iznosi medijan?

Osim medijana, možemo gledati i tzv. **kvartile**. Donji kvartil je podatak od kojeg je 25% podataka manje, dok je gornji kvartil podatak od kojeg je 75% podataka manje.

Kvartili

Promotrimo podatke 3, 5, 1, 4, 1, 6, 10, 10, 8. Onda je

$x_1 = 3, x_2 = 5, \dots, x_8 = 10, x_9 = 8$.

Kada ih sortiramo dobit ćemo 1, 1, 3, 4, 5, 6, 8, 10, 10. Ako se referiramo na drugi podatak u sortiranom uzorku (drugi podatak po redu), pišemo $x_{(2)}$.

- Koliko iznosi $x_{(5)}$?
- Koliko iznosi medijan?

Osim medijana, možemo gledati i tzv. **kvartile**. Donji kvartil je podatak od kojeg je 25% podataka manje, dok je gornji kvartil podatak od kojeg je 75% podataka manje.

Ako nas zanima podatak proizvoljnog, realnog redoslijeda koristimo sljedeću formulu. Neka je $s \in \mathbb{R}$, i zapišemo ga kao $s = k + r$, gdje je k najveći mogući cijeli broj, a r "ostatak". Na primjer, $s = 3.4 = 3 + 0.4$. Vrijedi

$$x_{(s)} = x_{(k+r)} = (1 - r)x_{(k)} + rx_{(k+1)} = x_{(k)} + r(x_{(k+1)} - x_{(k)})$$

Kvartili

Promotrimo podatke 3, 5, 1, 4, 1, 6, 10, 10, 8. Onda je

$x_1 = 3, x_2 = 5, \dots, x_8 = 10, x_9 = 8$.

Kada ih sortiramo dobit ćemo 1, 1, 3, 4, 5, 6, 8, 10, 10. Ako se referiramo na drugi podatak u sortiranom uzorku (drugi podatak po redu), pišemo $x_{(2)}$.

- Koliko iznosi $x_{(5)}$?
- Koliko iznosi medijan?

Osim medijana, možemo gledati i tzv. **kvartile**. Donji kvartil je podatak od kojeg je 25% podataka manje, dok je gornji kvartil podatak od kojeg je 75% podataka manje.

Ako nas zanima podatak proizvoljnog, realnog redoslijeda koristimo sljedeću formulu. Neka je $s \in \mathbb{R}$, i zapišemo ga kao $s = k + r$, gdje je k najveći mogući cijeli broj, a r "ostatak". Na primjer, $s = 3.4 = 3 + 0.4$. Vrijedi

$$x_{(s)} = x_{(k+r)} = (1 - r)x_{(k)} + rx_{(k+1)} = x_{(k)} + r(x_{(k+1)} - x_{(k)})$$

Kvartili i kvantili

- medijan $= x_{(\frac{N+1}{2})}$
- donji (prvi) kvartil $Q_1 = x_{(\frac{1}{4}(N+1))}$
- gornji (treći) kvartil $Q_3 = x_{(\frac{3}{4}(N+1))}$
- **interkvartil** daje nam raspršenost središnjih 50% podataka

$$IQ = Q_3 - Q_1$$

- **Koeficijent kvartilne devijacije** je relativna mjera raspršenosti, uvijek će biti između 0 i 1

$$VQ = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Kvartili i kvantili

- medijan $= x_{(\frac{N+1}{2})}$
- donji (prvi) kvartil $Q_1 = x_{(\frac{1}{4}(N+1))}$
- gornji (treći) kvartil $Q_3 = x_{(\frac{3}{4}(N+1))}$
- **interkvartil** daje nam raspršenost središnjih 50% podataka

$$IQ = Q_3 - Q_1$$

- **Koeficijent kvartilne devijacije** je relativna mjera raspršenosti, uvijek će biti između 0 i 1

$$VQ = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Zadatak

Zadani su podaci 3, 5, -1, 2, 2, 3, 4, 9. Odredite medijan, prvi kvartil, treći kvartil, interkvartil i koeficijent kvartilne devijacije. *Napomena: za sortiranje podataka možete koristiti funkciju `sort()`.*

Ostale mjere raspršenosti

- Srednje apsolutno odstupanje - umjesto $(x_i - \mu)^2$ imamo $|x_i - \mu|$,

$$MAD = \frac{1}{N} \sum_i |x_i - \mu|$$

Ostale mjere raspršenosti

- Srednje apsolutno odstupanje - umjesto $(x_i - \mu)^2$ imamo $|x_i - \mu|$,

$$MAD = \frac{1}{N} \sum_i |x_i - \mu|$$

- Raspon varijacije $\text{range}()$

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

Ostale mjere raspršenosti

- Srednje apsolutno odstupanje - umjesto $(x_i - \mu)^2$ imamo $|x_i - \mu|$,

$$MAD = \frac{1}{N} \sum_i |x_i - \mu|$$

- Raspon varijacije $\text{range}()$

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

Zadatak

Za podatke s prethodne stranice izračunajte raspon varijacije i srednje apsolutno odstupanje.