

# Zadaci za vježbu: deskriptivna statistika

---

## 1 Zadaci za vježbu u programu R

### Zadatak 1. Broj potomaka po leglu

Kod 25 ženki laboratorijskog miša zabilježen je broj mladunčadi po leglu:

6, 7, 5, 8, 7, 6, 9, 7, 6, 5, 8, 7, 6, 7, 10, 6, 7, 8, 5, 7, 6, 9, 7, 6, 8.

- (a) Unesite podatke u vektor `leglo` i izradite tablicu frekvencija naredbom `table()`.
- (b) Izračunajte aritmetičku sredinu, medijan i mod broja mladunčadi.
- (c) Izračunajte varijancu (naredba `var()` daje uzoračku varijancu s dijeljenjem s  $n - 1$ ; prokomentirajte tu razliku u odnosu na formulu  $s^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$ ) te standardnu devijaciju.
- (d) Nacrtajte stupčasti dijagram frekvencija broja mladunčadi po leglu i obrazložite zašto histogram nije prikladan izbor za ovo obilježje.

### Zadatak 2. Koncentracija glukoze u krvi

Izmjerene su koncentracije glukoze u krvi (mmol/L) kod 24 ispitanika natshe:

4.6, 4.8, 5.0, 5.1, 5.2, 5.3, 5.3, 5.4, 5.5, 5.5, 5.6, 5.6,  
5.7, 5.7, 5.8, 5.9, 6.0, 6.1, 6.2, 6.4, 6.6, 6.9, 7.3, 8.9.

- (a) Unesite podatke u vektor `glukoza` i izračunajte aritmetičku sredinu, medijan, varijancu i standardnu devijaciju.
- (b) Izračunajte prvi i treći kvartil naredbom `quantile()` te interkvartilni raspon naredbom `IQR()`.
- (c) Nacrtajte histogram s granicama razreda 4, 5, 6, 7, 8 i 9 mmol/L. Na graf dodajte okomite crte u položaju aritmetičke sredine i medijana pomoću `abline()`.
- (d) Nacrtajte dijagram pravokutnika za iste podatke i identificirajte, na temelju pravila 1.5 IQR, opažanje koje R označava kao potencijalnu izdvojenu vrijednost. Komentirajte kako ta vrijednost utječe na aritmetičku sredinu u usporedbi s medijanom.

### Zadatak 3. Vrijeme poluraspada reakcije – usporedba dva katalizatora.

U kemijskom pokusu mjereno je vrijeme (u sekundama) potrebno da koncentracija reaktanta padne na polovicu početne vrijednosti, uz dva različita katalizatora:

|               |                                        |
|---------------|----------------------------------------|
| Katalizator A | 42, 45, 44, 47, 46, 43, 45, 48, 44, 46 |
| Katalizator B | 38, 52, 40, 41, 39, 55, 42, 40, 39, 41 |

- (a) Unesite podatke u vektore `katA` i `katB`. Za svaki katalizator izračunajte aritmetičku sredinu, medijan, varijancu i standardnu devijaciju.
- (b) Izradite zajednički dijagram pravokutnika za oba katalizatora naredbom `boxplot(..., ..., names = ...)`.
- (c) Usporedite relativnu raspršenost rezultata uspoređujući više mjera raspršenosti.
- (d) Na temelju grafa i izračunatih mjera prosudite koji katalizator daje pouzdanije (manje raspršene) rezultate te obrazložite zašto sama aritmetička sredina nije dovoljna za tu prosudbu.

### Zadatak 4. Duljina listova biljne vrste – histogram i oblik distribucije.

Izmjerene su duljine (u milimetrima) 30 listova iste biljne vrste:

28.4, 30.1, 31.5, 32.0, 32.8, 33.2, 33.9, 34.3, 34.7, 35.1,  
35.4, 35.9, 36.2, 36.6, 37.0, 37.5, 37.8, 38.1, 38.6, 39.0,  
39.5, 40.2, 40.8, 41.1, 41.7, 42.4, 43.2, 44.0, 45.6, 50.8.

- (a) Unesite podatke u vektor `listovi` i izračunajte aritmetičku sredinu, medijan, varijancu i standardnu devijaciju.
- (b) Nacrtajte histogram s razredima širine 5 mm (granice 25, 30, 35, 40, 45, 50) i usporedite ga s histogramom dobivenim automatskim izborom razreda u Ru.
- (c) Izračunajte kvartile i interkvartilni raspon te nacrtajte dijagram pravokutnika.
- (d) Opišite oblik distribucije (simetrija, asimetrija) na temelju histograma i usporedbe aritmetičke sredine i medijana.

### Zadatak 5. Koncentracija onečišćivača u uzorcima vode – više razreda histograma.

U 40 uzoraka riječne vode izmjerena je koncentracija nitrata (mg/L):

```
nitrati <- c(1.2, 1.5, 1.8, 2.0, 2.1, 2.3, 2.4, 2.5, 2.6, 2.7,
            2.8, 2.9, 3.0, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7,
            3.8, 3.9, 4.0, 4.1, 4.2, 4.3, 4.5, 4.7, 4.9, 5.1,
            5.3, 5.6, 5.9, 6.2, 6.6, 7.0, 7.5, 8.1, 8.9, 11.4)
```

- (a) Izračunajte aritmetičku sredinu, medijan, mod (ako postoji u smislenom obliku), varijancu i standardnu devijaciju.

- (b) Nacrtajte histograme s parametrima  $\text{breaks} = 5$ ,  $\text{breaks} = 10$  i  $\text{breaks} = 20$  te usporedite kako izbor broja razreda utječe na sliku distribucije.
- (c) Odredite kvartile, interkvartilni raspon te granice za izdvojene vrijednosti prema pravilu 1.5 IQR. Nacrtajte dijagram pravokutnika i označite eventualne izdvojene vrijednosti.
- (d) Rasprava: hidrološki propisi mnogih zemalja definiraju granične vrijednosti nitrata u vodi za piće. Objasnite zašto bi u ovakvom kontekstu medijan i interkvartilni raspon mogli biti informativniji pokazatelji od aritmetičke sredine i standardne devijacije.

## 2 Zadaci za vježbu s olovkom i papirom

### Zadatak 1. Broj otkucaja srca u minuti.

Kod deset laboratorijskih štakora izmjeren je broj otkucaja srca u minuti tijekom mirovanja:

320, 335, 340, 328, 350, 335, 342, 335, 330, 345.

- (a) Poredajte podatke po veličini te odredite aritmetičku sredinu, medijan i mod.
- (b) Odredite gornji, donji kvartil, i interkvartilni raspon, te broj od kojega je 90% podataka manje.
- (c) Usporedite aritmetičku sredinu i medijan te obrazložite koja mjera bolje opisuje tipičnu vrijednost broja otkucaja srca u ovom uzorku.

### Zadatak 2. Masa sjemenki graha – frekvencijska tablica.

Izmjerena je masa (u gramima, zaokružena na cijeli broj) 20 sjemenki graha, a rezultati su prikazani tablicom.

|                 |   |   |   |   |   |
|-----------------|---|---|---|---|---|
| Masa $x$ (g)    | 2 | 3 | 4 | 5 | 6 |
| Frekvencija $f$ | 3 | 6 | 6 | 4 | 1 |

- (a) Provjerite ukupan broj opažanja  $n$  i izračunajte aritmetičku sredinu.
- (b) Odredite medijan i mod na temelju frekvencijske tablice, uz prikaz kumulativnih frekvencija.
- (c) Izračunajte varijancu i standardnu devijaciju.
- (d) Raspisivanjem podataka kao uređenog niza od 20 vrijednosti odredite gornji i donji kvartil i interkvartilni raspon.

### Zadatak 3. pH vrijednosti otopina – kvartili i izdvojene vrijednosti.

U 13 uzoraka otopine izmjerene su sljedeće pH vrijednosti:

6.8, 7.0, 7.1, 7.2, 7.2, 7.3, 7.4, 7.4, 7.5, 7.6, 7.7, 7.9, 9.1.

- (a) Odredite medijan te prvi i treći kvartil primjenom konvencije iz uvodne napomene.
- (b) Izračunajte interkvartilni raspon te donju i gornju granicu za potencijalne izdvojene vrijednosti prema pravilu  $Q_L - 1.5 \text{ IQR}$  i  $Q_U + 1.5 \text{ IQR}$ .
- (c) Provjerite postoji li u uzorku opažanje koje prema tom pravilu predstavlja izdvojenu vrijednost.
- (d) Izračunajte aritmetičku sredinu sa svim opažanjima i bez uočene izdvojene vrijednosti, te usporedite dobivene rezultate s medijanom.

#### **Zadatak 4. Prinos reakcije kod dva postupka sinteze.**

Kod ponovljenih pokusa sinteze istog spoja dvama različitim postupcima zabilježeni su prinosi (u postotcima):

|             |                                |
|-------------|--------------------------------|
| Postupak I  | 78, 81, 79, 83, 80, 82, 79, 81 |
| Postupak II | 70, 90, 75, 85, 72, 88, 74, 86 |

- (a) Za svaki postupak odredite aritmetičku sredinu, medijan i mod (ako postoji).
- (b) Izračunajte varijancu i standardnu devijaciju prinosa za oba postupka.
- (c) Usporedite raspršenost.
- (d) Na temelju izračunatih mjera obrazložite koji je postupak sinteze pouzdaniji (dosljedniji) s obzirom na prinos, iako obje skupine mogu imati sličnu aritmetičku sredinu.

#### **Zadatak 5. Broj bakterijskih kolonija na Petrijevim zdjelicama.**

Na 16 Petrijevih zdjelica nakon inkubacije prebrojane su bakterijske kolonije:

12, 15, 14, 13, 15, 16, 14, 15, 13, 14, 15, 17, 14, 15, 13, 45.

- (a) Poredajte podatke i odredite aritmetičku sredinu, medijan i mod.
- (b) Odredite gornji i donji kvartil, broj od kojega je 10% podataka manje, i interkvartilni raspon.
- (c) Provjerite prema pravilu  $1.5 \text{ IQR}$  predstavlja li najveće opažanje izdvojenu vrijednost.
- (d) Izračunajte varijancu i standardnu devijaciju sa svim podacima, a zatim bez uočene izdvojene vrijednosti. Objasnite zašto su mjere raspršenosti, jednako kao i aritmetička sredina, osjetljive na izdvojene vrijednosti, dok medijan i interkvartilni raspon to nisu (ili su znatno manje osjetljivi).