

Vježbe 3

Zadatak 1. Zadana su tri uzorka:

```
visina = c(158, 162, 163, 165, 166, 167, 168, 168, 169,
           170, 171, 172, 174, 175, 178, 182)
krpelji = c(0, 0, 1, 1, 1, 2, 2, 2, 3, 3, 4, 5, 6, 8,
            12, 25)
trudnoca = c(28, 33, 35, 37, 38, 38, 39, 39, 39, 40, 40,
             40, 40, 41, 41, 42)
```

- (a) Za svaki uzorak nacrtajte histogram i dijagram pravokutnika. Opišite oblik razdiobe i repove.
- (b) Za svaki uzorak izračunajte aritmetičku sredinu i medijan i prikazite ih na odgovarajućim histogramima. Povežite njihov odnos s oblikom razdiobe.
- (c) Izračunajte koeficijent asimetrije α_3 za sva tri uzorka. Ugrađena funkcija za α_3 ne postoji.

Zadatak 2. Tijekom dva tjedna u listopadu mjerena je srednja dnevna temperatura zraka u Zagrebu ($^{\circ}\text{C}$):

12.4, 14.1, 9.8, 11.5, 15.2, 13.0, 8.7, 10.9, 16.3, 14.8, 12.1, 11.0, 7.5, 13.6.

- (a) Unesite podatke u vektor `celzij` te izračunajte aritmetičku sredinu, medijan, varijancu i standardnu devijaciju.
- (b) Pretvorite podatke u stupnjeve Fahrenheita, $^{\circ}F = 1.8 \cdot ^{\circ}C + 32$. Prije računanja predvidite iste četiri mjere u $^{\circ}F$, a zatim ih provjerite u R-u.
- (c) Nacrtajte dijagram raspršenja temperature u $^{\circ}F$ u ovisnosti o temperaturi u $^{\circ}C$ (koristeći naredbu `plot(fahrenheit ~ celzij)`) te izračunajte koeficijent korelacije. Objasnite rezultat.
- (d) Bez računanja odredite korelaciju između temperature u $^{\circ}C$ i u kelvinima. Provjerite u R-u.
- (e) Standardizirajte podatke u obje mjerne jedinice. Što uočavate?

Zadatak 3. U jedanaest jezera na području zahvaćenom kiselim kišama izmjeren je pH vode:

```
pH <- c(4.3, 4.9, 5.4, 5.8, 6.2, 6.5, 6.8, 7.0, 7.1,
        7.4, 7.9)
```

pH je određen koncentracijom vodikovih iona $[\text{H}^+]$ (u mol/L): $\text{pH} = -\log_{10} [\text{H}^+]$, odnosno $[\text{H}^+] = 10^{-\text{pH}}$.

- (a) Izračunajte aritmetičku sredinu i medijan pH vrijednosti.
- (b) Izračunajte koncentracije vodikovih iona te njihovu aritmetičku sredinu i medijan:

```
H <- 10^(-pH)
```

Dobivene vrijednosti pretvorite natrag u pH (naredba `-log10()`) i usporedite s rezultatima iz (a). Što se dogodilo s aritmetičkom sredinom, a što s medijanom? Usporedite sa Zadatkom 2(b).

- (c) Nacrtajte histograme pH vrijednosti i koncentracija $[H^+]$ te za oba uzorka izračunajte koeficijent asimetrije α_3 . Kako je transformacija promijenila oblik razdiobe?
- (d) Nacrtajte dijagram raspršenja koncentracije $[H^+]$ u ovisnosti o pH i izračunajte koeficijent korelacije. Koncentracija je potpuno određena pH vrijednošću. Zašto korelacija ipak nije jednaka 1 ili -1 ? Usporedite sa Zadatkom 2(c).

Zadatak 4. Mjerena je razina proteina toplinskog stresa y kod kukaca uzgajanih na različitim temperaturama, pri čemu je x odstupanje temperature od optimalne za tu vrstu ($^{\circ}C$):

```
x <- c(-3, -2.5, -2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2, 2.5, 3)
y <- c(10.4, 5.5, 3.6, 2.0, 0.4, 0.1, 0.4, 0.2, 1.1, 3.6, 4.2, 7.9, 10.4)
```

- (a) Nacrtajte dijagram raspršenja i opišite vezu između x i y .
- (b) Izračunajte koeficijent korelacije. Objasnite dobivenu vrijednost.
- (c) Izračunajte korelaciju posebno za $x > 0$ i za $x < 0$. Odgovarajuće vrijednosti možete filtrirati pomoću

```
pozitivnix = x[x > 0]
pozitivniy = y[x > 0]

negativnix = x[x < 0]
negativniy = y[x < 0]
```

- (d) Izračunajte korelaciju između x^2 i y .
- (e) Smijemo li iz korelacije blizu nule zaključiti da varijable nisu povezane?

Zadatak 5. Za jedan primorski grad zabilježeni su po mjesecima prosječna temperatura zraka ($^{\circ}C$), prodaja sladoleda (u tisućama kornata) i broj intervencija spasilaca zbog utapanja:

```
temperatura <- c(8, 9, 12, 16, 20, 25, 28, 28, 23, 18, 13, 9)
sladoled <- c(1.2, 1.5, 2.8, 4.1, 6.5, 9.8, 12.4, 12.1, 7.9, 4.6, 2.2, 1.4)
utapanja <- c(0, 0, 1, 1, 3, 6, 9, 8, 4, 1, 0, 0)
```

- (a) Nacrtajte dijagram raspršenja broja utapanja u ovisnosti o prodaji sladoleda i izračunajte koeficijent korelacije.

- (b) Bi li zabrana prodaje sladoleda smanjila broj utapanja? Obrazložite.
- (c) Izračunajte korelaciju temperature s preostalim dvjema varijablama i protumačite rezultate.
- (d) Za svaki od sljedećih primjera objasnite zašto korelacija ne znači nužno uzročnu vezu:
 - i. kod djece u osnovnoj školi veći broj cipela povezan je s boljim rezultatom na testu čitanja;
 - ii. u općinama s više gnijezda roda rađa se više djece.
- (e) Kako biste pokusom provjerili uzrokuje li neko gnojivo veći prinos biljke?