

Zadaci za vježbu: deskriptivna statistika i usporedba podataka

Vježbe 2

1 Zadaci za rješavanje olovkom i papirom

Zadatak 1. Brojčane oznake uzoraka.

Osam uzoraka vode označeno je prema mjestu uzimanja: izvor = 1, potok = 2, jezero = 3. Zabilježene su oznake 1, 2, 2, 3, 1, 2, 3, 2.

- (a) Odredite tip obilježja, tablicu apsolutnih i relativnih frekvencija te mod.
- (b) Nacrtajte stupčasti dijagram. Ima li aritmetička sredina oznaka smisljeno značenje?
- (c) Promijenite oznake u 10, 20 i 90. Koje se informacije o uzorcima mijenjaju, a koje ostaju iste? Obrazložite zašto brojčani zapis ne određuje tip obilježja.

Zadatak 2. Jednak raspon, različita raspršenost.

Brojevi izdanaka na pet biljaka u dvjema skupinama su

$$A : 2, 2, 5, 8, 8, \quad B : 2, 4, 5, 6, 8.$$

- (a) Za obje skupine odredite aritmetičku sredinu, medijan i raspon.
- (b) Izračunajte empirijsku varijancu i standardnu devijaciju.
- (c) Objasnite što varijanca otkriva o ovim podacima, a raspon ne otkriva.

Zadatak 3. Usporedba relativne raspršenosti.

Mase šest plodova vrste A iznose 18, 19, 20, 20, 21, 22 g, a mase šest plodova vrste B iznose 80, 90, 100, 100, 110, 120 g.

- (a) Izračunajte sredinu i standardnu devijaciju u objema skupinama.
- (b) Izračunajte koeficijent varijacije $100s/\bar{x}$ u postotcima.
- (c) Koja skupina ima veću apsolutnu, a koja veću relativnu raspršenost? Objasnite zašto je koeficijent varijacije ovdje smislen.

Zadatak 4. Koliko podataka čuva dijagram pravokutnika?

Duljine sedam korjenčića (mm) su 2, 4, 5, 6, 7, 8, 10.

- (a) Odredite kvartile, medijan, interkvartilni raspon i granice prema pravilu 1.5 IQR. Skicirajte dijagram pravokutnika.
- (b) Sastavite drugi uređeni niz od sedam duljina koji ima isti minimum, kvartile, medijan i maksimum, ali različitu aritmetičku sredinu.
- (c) Izračunajte obje sredine i objasnite zašto ih sam dijagram ne određuje.

Zadatak 5. Spajanje terenskih uzoraka.

Na lokalitetu A izmjerene su duljine 4, 6 cm, a na lokalitetu B 7, 8, 9, 10 cm. Svaki podatak odnosi se na drugu jedinku.

- (a) Izračunajte sredine lokaliteta i sredinu svih šest opažanja.
- (b) Usporedite ukupnu sredinu s običnim prosjekom dviju sredina. Napišite formulu koja uvažava veličine skupina.
- (c) Izračunajte empirijsku varijancu svih opažanja te zasebno za svaki lokalitet. Je li ukupna varijanca jednaka ponderiranom prosjeku varijanci skupina? Objasnite ulogu razlike između njihovih sredina.

2 Zadaci za rad u R-u

Zadatak 1. Sastav dvaju staništa.

U dvama jednakim istraživanim područjima prebrojene su biljke triju skupina:

```
skupina <- c("trave", "grmovi", "drvece")
A <- c(36, 18, 6)
B <- c(30, 30, 40)
```

- (a) Izračunajte ukupne brojeve biljaka i relativne frekvencije po staništu.
- (b) Prikažite apsolutne pa relativne frekvencije usporednim stupčastim dijagramima, koristeći `barplot(rbind(A, B), beside = TRUE)` kao polazište.
- (c) Usporedite udio trave i njihov broj. Zašto se usporedbe razlikuju?

Zadatak 2. Različiti načini računanja kvartila.

Vremena obrade deset uzoraka (min) su 3, 4, 5, 5, 6, 7, 8, 9, 12, 18.

- (a) Izračunajte kvartile s `type = 6` i sa zadanom postavkom `quantile()`. Usporedite položaje s kojih se interpolira.
- (b) Za obje konvencije izračunajte IQR i granice izdvojenih vrijednosti.
- (c) Nacrtajte `boxplot()` i ispišite `boxplot.stats(x)$stats`. Usporedite rubove pravokutnika s izračunatim kvartilima. R-ov dijagram koristi tzv. Tukeyjeve pregibe, koji se mogu razlikovati od obiju konvencija.

Zadatak 3. Histogram s nejednakim širinama razreda.

Koncentracije tvari u vodi (mg/L) zadane su ovako:

```
x <- c(.2, .4, .5, .7, .8, .9, 1.1, 1.2, 1.4, 1.7,
       1.9, 2.1, 2.4, 2.8, 3.2, 3.8, 4.5, 5.7, 6.1, 7.6)
granice <- c(0, 1, 2, 4, 8)
```

- (a) Dobijte frekvencije i gustoće iz objekta:

```
h <- hist(x, breaks = granice, plot = FALSE)
```

- (b) Nacrtajte histogram s `freq = FALSE`. Za svaki stupac izračunajte visinu puta širinu i provjerite da je zbroj površina jednak 1.
- (c) Koji razred ima najveću frekvenciju, a koji najveću gustoću? Objasnite zašto pri nejednakim širinama visina stupca nije relativna frekvencija.

Zadatak 4. Sažimanje ugrađenih podataka.

Ugrađeni skup `iris` sadrži mjerenja cvjetova; `Sepal.Width` je širina lapova u cm, a `Species` vrsta.

- (a) Za svaku vrstu izračunajte broj opažanja, sredinu, medijan, empirijsku standardnu devijaciju i koeficijent varijacije. Možete koristiti `split()` i `sapply()`.
- (b) Nacrtajte zajednički dijagram pravokutnika. Usporedite vrste prema središtu, apsolutnoj i relativnoj raspršenosti širine.
- (c) Iz sredina i veličina skupina rekonstruirajte ukupnu sredinu te je provjerite izravnim izračunom. Zaključke ograničite na opis tih podataka.

Zadatak 5. Gubitak podataka pri grupiranju.

Promatrajte ugrađene podatke `iris$Sepal.Length`, izražene u cm.

- (a) Izračunajte točnu sredinu i empirijsku varijancu te nacrtajte histogram s granicama `seq(4, 8, by = .5)`.
- (b) Svako opažanje zamijenite sredinom njegova razreda. Iz frekvencija i sredina razreda izračunajte približnu sredinu i varijancu.
- (c) Ponovite postupak za granice `seq(4, 8, by = 1)`. U tablici prikazite obje aproksimacije i njihove apsolutne pogreške. Mora li finije grupiranje za svaki skup podataka dati strogo manju pogrešku sredine?