

Vježbe na računalu: deskriptivna statistika u programu R

Zadatak 1. Mjere sredine za diskretne podatke.

Na uzorku od 20 studenata zabilježen je broj položenih ispita u jednom ispitnom roku:

4, 5, 3, 6, 4, 5, 4, 7, 3, 4, 5, 4, 6, 2, 5, 4, 3, 5, 4, 6.

- (a) Unesite podatke u vektor `ispiti`.
- (b) Izračunajte aritmetičku sredinu i medijan.
- (c) Odredite mod, odnosno vrijednost ili vrijednosti s najvećom frekvencijom. Nemojte pretpostaviti da ugrađena funkcija za mod postoji; izradite tablicu frekvencija pomoću `table()`.
- (d) Nacrtajte stupčasti dijagram frekvencija `barplot()`. Objasnite koja od triju mjera sredine najbolje opisuje “tipičan” broj položenih ispita i zašto.

Zadatak 2. Učinak ekstremne vrijednosti.

Sljedeći podaci predstavljaju vrijeme, u minutama, potrebno 15 osoba da riješe kraći računalni zadatak:

18, 20, 19, 21, 22, 18, 20, 19, 23, 21, 20, 19, 18, 22, 58.

- (a) Unesite podatke u vektor `vrijeme` te izračunajte aritmetičku sredinu, medijan i mod. Izračunajte i varijancu te gornji i donji kvartil.
- (b) Izračunajte iste tri mjere nakon uklanjanja opažanja 58. Usporedite rezultate.
- (c) Nacrtajte dijagram pravokutnika za sve podatke `boxplot()`. Na grafu označite opažanje koje R prepoznaje kao potencijalnu izdvojenu vrijednost.
- (d) Objasnite, na temelju izračuna i dijagrama, zašto je medijan često prikladnija mjera središta od aritmetičke sredine kada uzorak sadržava ekstremne vrijednosti.

Zadatak 3. Histogram neprekidnih podataka: vrijeme putovanja.

Pretpostavite da je vrijeme putovanja na posao, mjereno u minutama, zabilježeno za 30 osoba:

22.4, 24.1, 25.7, 26.0, 26.8, 27.2, 27.9, 28.3, 28.7, 29.1,
29.4, 29.9, 30.2, 30.6, 31.0, 31.5, 31.8, 32.1, 32.6, 33.0,
33.5, 34.2, 34.8, 35.1, 35.7, 36.4, 37.2, 38.0, 39.6, 44.8.

- Unesite podatke u vektor `putovanje` i izračunajte aritmetičku sredinu te medijan. Imaju li podaci mod? Obrazložite odgovor.
- Nacrtajte histogram s razredima širine 5 minuta, s granicama razreda 20, 25, 30, 35, 40 i 45. Usporedite ga s histogramom koji R stvara pri automatskom izboru razreda.
- Na histogram dodajte okomitu crtu u položaju aritmetičke sredine i drugu u položaju medijana.
(Možete upotrijebiti `abline(v = ..., col = ..., lwd = 2)`.)
- Odredite varijancu uzorka, te gornji i donji kvartil.
- Nacrtajte i dijagram pravokutnika za iste podatke. Opišite oblik distribucije i komentirajte utjecaj najveće vrijednosti na aritmetičku sredinu i medijan.

Zadatak 4. Diskretni nasuprot neprekidnim podacima.

Promatraju se dva obilježja za 40 kućanstava:

- `broj_djece`: broj djece u kućanstvu;
- `potrosnja`: mjesečna potrošnja električne energije, u kWh.

Podaci su:

```
broj_djece <- c(0,1,2,1,3,2,0,1,2,2,1,0,3,1,2,1,4,0,2,1,
               1,2,0,3,2,1,0,2,1,2,3,1,0,2,1,1,2,0,3,1)
potrosnja <- c(182,195,201,208,214,218,221,225,229,233,
               236,240,244,247,251,255,259,263,267,271,
               274,279,283,287,292,297,301,306,311,316,
               322,329,337,346,358,371,389,412,445,486)
```

- Za varijablu `broj_djece` odredite tablicu frekvencija, aritmetičku sredinu, medijan i mod. Prikažite podatke stupčastim dijagramom; objasnite zašto je on prikladniji od histograma za ovo obilježje.
- Za varijablu `potrosnja` izračunajte aritmetičku sredinu i medijan te nacrtajte histogram i dijagram pravokutnika. Koristite najmanje dva različita izbora broja razreda histograma, primjerice `breaks = 5` i `breaks = 10`.
- Usporedite grafičke prikaze za varijable `broj_djece` i `potrosnja`. Objasnite razliku između diskretnog i neprekidnog obilježja te kako se ta razlika odražava na izbor grafa.

- (d) Napišite dvije do tri rečenice o tome u kojim okolnostima biste, za svaku od promatranih varijabli, prednost dali aritmetičkoj sredini, a u kojima medijanu.

Zadatak 5. Dijagram pravokutnika i usporedba dviju skupina.

U tablici su dane vrijednosti završnog rezultata (od najviše 100 bodova) za dvije skupine studenata.

Skupina A	62, 71, 68, 74, 75, 69, 73, 77, 70, 72, 66, 76
Skupina B	51, 58, 60, 62, 64, 65, 67, 69, 70, 72, 74, 88

- (a) Unesite rezultate u dva vektora, **A** i **B**. Izračunajte za svaku skupinu aritmetičku sredinu, medijan i mod. Ako mod nije jedinstven ili ne postoji u uobičajenom smislu, navedite to jasno.
- (b) Izradite jedan zajednički dijagram pravokutnika za skupine **A** i **B**, primjerice naredbom `boxplot(A, B, names = c("A", "B"))`.
- (c) Odredite prvi i treći kvartil te interkvartilni raspon za obje skupine.
- (d) Usporedite skupine s obzirom na središte i raspršenost rezultata. Objasnite zašto sama aritmetička sredina nije dovoljna za potpunu usporedbu skupina.