

Deskriptivna statistika: frekvencije

Prirodoslovno-matematički Fakultet, Sveučilište u Zagrebu

9. listopada 2026.

Učitavanje podataka

Učitajte podatke `diabetes.csv` iz datoteke Vježbe(26./27.) i spremite ih pod nazivom `diabetes` (sve malim slovima).

Učitavanje podataka

Učitajte podatke `diabetes.csv` iz datoteke Vježbe(26./27.) i spremite ih pod nazivom `diabetes` (sve malim slovima). Pogledajte kako podaci izgledaju

- `View(diabetes)` – otvara tablicu sa podacima u zasebnom prozoru
- `names(diabetes)` – ispisuje u konzoli naziv stupaca
- `str(diabetes)` – ispisuje tip svakog stupca i prvih par podataka

Učitavanje podataka

Učitajte podatke `dijabetes.csv` iz datoteke Vježbe(26./27.) i spremite ih pod nazivom `dijabetes` (sve malim slovima). Pogledajte kako podaci izgledaju

- `View(dijabetes)` – otvara tablicu sa podacima u zasebnom prozoru
- `names(dijabetes)` – ispisuje u konzoli naziv stupaca
- `str(dijabetes)` – ispisuje tip svakog stupca i prvih par podataka

Diskretni podaci

Zanima nas koliko je ženskih osoba, a koliko muških sudjelovalo u ispitivanju. Želimo izračunati frekvencije - koristimo naredbu `table()`. Ta naredba nam daje apsolutne frekvencije.

```
dij_f <- table(dijabetes$gender)
dij_f
```

Diskretni podaci

Zanima nas koliko je ženskih osoba, a koliko muških sudjelovalo u ispitivanju. Želimo izračunati frekvencije - koristimo naredbu `table()`. Ta naredba nam daje apsolutne frekvencije.

```
dij_f <- table(dijabetes$gender)
dij_f
```

Kako biste izračunali relativne frekvencije?

Diskretni podaci

Zanima nas koliko je ženskih osoba, a koliko muških sudjelovalo u ispitivanju. Želimo izračunati frekvencije - koristimo naredbu `table()`. Ta naredba nam daje apsolutne frekvencije.

```
dij_f <- table(dijabetes$gender)
dij_f
```

Kako biste izračunali relativne frekvencije?

```
uk_broj <- sum(dij_f)
uk_broj <- length(dijabetes$gender)
dij_rf <- dij_f/uk_broj
dij_rf
```

Diskretni podaci

Zanima nas koliko je ženskih osoba, a koliko muških sudjelovalo u ispitivanju. Želimo izračunati frekvencije - koristimo naredbu `table()`. Ta naredba nam daje apsolutne frekvencije.

```
di_j_f <- table(di_jabetes$gender)
di_j_f
```

Kako biste izračunali relativne frekvencije?

```
uk_broj <- sum(di_j_f)
uk_broj <- length(di_jabetes$gender)
di_j_rf <- di_j_f/uk_broj
di_j_rf
```

Kumulativne frekvencije računamo pozivom funkcije `cumsum()` nad tablicom.

Diskretni podaci - grafički prikaz

- `pie()` – *pie chart*
- `barplot()` – stupčasti dijagram
- `plot(..., type="l")` – linijski dijagram

Kao argument moramo dati **frekvencije**, a ne originalne podatke – ove funkcije očekuju dobiti visine stupića!

Diskretni podaci - grafički prikaz

- `pie()` – *pie chart*
- `barplot()` – stupčasti dijagram
- `plot(..., type="l")` – linijski dijagram

Kao argument moramo dati **frekvencije**, a ne originalne podatke – ove funkcije očekuju dobiti visine stupića!

Zadatak

Grafički prikažite apsolutne frekvencije, te potom relativne frekvencije podataka o

Diskretni podaci - grafički prikaz

- `pie()` – *pie chart*
- `barplot()` – stupčasti dijagram
- `plot(..., type="l")` – linijski dijagram

Kao argument moramo dati **frekvencije**, a ne originalne podatke – ove funkcije očekuju dobiti visine stupića!

Zadatak

Grafički prikažite apsolutne frekvencije, te potom relativne frekvencije podataka o

Česte naredbe za "uljepšavanje" grafova su:

- `xlab="ime x-osi", ylab="ime y-osi".`
- `main = "naziv grafa"`
- `col = "blue"`

Zadaci

Rješavanje zadataka.

Neprekidni podaci

Ako nas zanima kako se kreću podaci za neprekidne podatke, koristimo histogram.

Zadatak

Pozovite naredbu `hist(dijabetes$glyhb)`. Kako su distribuirani podaci?

Neprekidni podaci

Ako nas zanima kako se kreću podaci za neprekidne podatke, koristimo histogram.

Zadatak

Pozovite naredbu `hist(dijabetes$glyhb)`. Kako su distribuirani podaci?

Zadatak

Pozovite naredbu `hist(dijabetes$height)`. Komentirajte što vidite.

Neprekidni podaci

Ako nas zanima kako se kreću podaci za neprekidne podatke, koristimo histogram.

Zadatak

Pozovite naredbu `hist(dijabetes$glyhb)`. Kako su distribuirani podaci?

Zadatak

Pozovite naredbu `hist(dijabetes$height)`. Komentirajte što vidite.

U histogramu možemo birati broj razreda naredbom `breaks`. Na primjer `hist(dijabetes$glyhb, breaks=20)` daje više stupića. Prebrojite koliko ih dobijete.

Neprekidni podaci - graf

Zadatak

Pozovite sljedeću naredbu

```
h <- hist(dijabetes$height)
```

Istražite što je sve spremljeno u h.

Neprekidni podaci - graf

Zadatak

Pozovite sljedeću naredbu

```
h <- hist(dijabetes$height)
```

Istražite što je sve spremljeno u h.

Obično želimo da je površina svih stupića histograma jednaka 1 (razlog će biti jasan nakon što naučimo više vjerojatnosti). Da bismo to dobili crtamo relativne frekvencije i prilagođavamo širine stupića.

Iduće dvije naredbe su ekvivalentne.

```
hist(dijabetes$height, freq=FALSE)
```

```
hist(dijabetes$height, probability=TRUE)
```

Zadaci

Rješavanje zadataka.

Mjere sredine

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 sredine

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 - klasična mjera sredine. Jako osjetljiva na podatke koji odstupaju. Npr. izračunajte aritmetičku sredinu od $a \leftarrow c(1, 1, 2, 1, 1000)$.

Mjere sredine

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 - klasična mjera sredine. Jako osjetljiva na podatke koji odstupaju. Npr. izračunajte aritmetičku sredinu od $a \leftarrow c(1, 1, 2, 1, 1000)$.
- 3 medijan - položajna mjera sredine. Podatke sortiramo od najmanjeg do najvećeg te odaberemo središnji.

Koristimo funkcije `mean()` i `median()`.

Mjere sredine

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 - klasična mjera sredine. Jako osjetljiva na podatke koji odstupaju. Npr. izračunajte aritmetičku sredinu od $a \leftarrow c(1, 1, 2, 1, 1000)$.
- 3 medijan - položajna mjera sredine. Podatke sortiramo od najmanjeg do najvećeg te odaberemo središnji.

Koristimo funkcije `mean()` i `median()`.

Kako ih računamo ako su nam samo dostupne tablice frekvencija?