

# Analiza podataka o vinu

---

Podaci o vinima dostupni su u datoteci "vino.csv". Istraživanje prati različite karakteristike vina i njihov utjecaj na cijenu.

1. Koliko opservacija i varijabli sadrži skup podataka? Koji su nazivi varijabli i kojeg su tipa s obzirom na mjernu skalu?
2. Ima li nedostajućih vrijednosti u podacima?
3. Izračunajte Pearsonov koeficijent korelacije za varijable **KisaZima** (količina kiše tijekom zime) i **FrPop** (broj stanovnika Francuske).
4. Kakvog je smjera i jačine veza između **KisaZima** i **FrPop**?
5. Izračunajte korelaciju između svih varijabli.
6. Između kojih dviju različitih varijabli postoji najjača (pozitivna ili negativna) korelacija i koliko iznosi?
7. Prikažite **korelacijsku matricu**.
8. Između kojih varijabli postoji najjača pozitivna, a između kojih najjača negativna korelacija? (prema bojama na korelogramu)
9. Kreirajte model linearne regresije koji objašnjava **Cijena** ovisno o **FrPop**. Spremite ga u `model1`.
10. Koliko iznosi procijenjena vrijednost koeficijenta linearne regresije?
11. Što predstavlja taj koeficijent?
12. Je li **FrPop** značajna varijabla za model?
13. Koliki je udio varijacije **Cijena** objašnjen modelom?
14. Prikažite podatke i **regresijski pravac** na grafu
  - Postavite:
    - **x-os**: "Broj stanovnika"
    - **y-os**: "Cijena"
    - **Naslov**: "Ovisnost cijene vina o broju stanovnika"
    - **Regresijski pravac**: crvene boje
15. Kreirajte model linearne regresije **Cijena** ovisno o varijabli s najjačom povezanošću. Spremite ga u `model2`.
16. Je li odabrana varijabla značajna za model?
17. Koji model bolje objašnjava varijaciju **Cijena**?
18. Kreirajte model linearne regresije sa svim varijablama. Spremite ga u `model3`.

19. Je li varijabla **Starost** značajna? Postoji li problem multikolinearnosti?
20. Kreirajte model bez varijable **Godina**. Spremite ga u `model4`.
21. Koliki je udio varijacije **Cijena** objašnjen modelom?
22. Jesu li sve varijable značajne (razina značajnosti 0.05)?
23. Kreirajte model samo sa značajnim varijablama (2). Spremite ga u `model5`.
24. Koliki je udio varijacije **Cijena** objašnjen modelom?
25. Učitajte datoteku "vinoTest.csv" u `vinoTest`.
26. Pomoću `model5` predvidite cijene za nove podatke.
27. Koliko se stvarne vrijednosti razlikuju od predviđenih?
28. Jesu li vrijednosti precijenjene ili podcijenjene?

*Dodatno:*

- Izvršite automatski odabir varijabli za `model6` (`regsubsets()`, `plot(subset, scale = "adjr2")`)
- Koliki je udio varijacije **Cijena** objašnjen modelom?