

Vježbe 13. Korelacija

Zadatak 1: Udaljenost i gustoća naseljenosti

U datoteci "nasljenost.csv" nalaze se podaci o gustoći naseljenosti (broj stanovnika po m^2) uz pripadne udaljenost od centra grada (u km). Očekujemo kako gustoća naseljenosti opada s većom udaljenošću od grada.

1. Procijenite parametre linearne regresije i nacrtajte regresijski pravac $y = ax + b$ s podacima.
2. Provedite test značajnosti linearne regresije.
3. Koliko je jaka korelacija? Koliki udio ukupne varijacije je objašnjen našim modelom?
4. Procijenite gustoću naseljenosti za kvart koji je udaljen 2.4 km od centra.

Zadatak 2: Veličina i cijena stana (DZ)

U datoteci "cijenastana.csv" nalaze se podaci o veličinama stanova i njihovima cijenama po kvadratu. Ispitajte ovisnost cijene o veličini.

1. Procijenite parametre linearne regresije i nacrtajte regresijski pravac s podacima.
2. Provedite test značajnosti linearne regresije.
3. Koliko je jaka korelacija? Koliki udio ukupne varijacije je objašnjeno našim modelom?
4. Procijenite cijena stana po kvadratu za stan od $60 m^2$.

Zadatak 3: Automobili

Skup podataka "mtcars" je ugrađena R-ova baza podataka koja sadrži informacije o različitim karakteristikama automobila. Skup podataka ima 32 automobila i 11 varijabli koje opisuju različite tehničke karakteristike automobila. Želimo predvidjeti varijablu mpg (milja po galonu).

- Izračunajte koeficijent determinacije (R^2) za model u kojemu su svih ostalih 10 varijabli prediktori.
- Provedite eliminaciju unazad. U svakom koraku navedite uklonjeni prediktor i njegovu p -vrijednost prije izbacivanja.
- Koristeći funkciju `step()` provedite eliminaciju unazad. Koliko iznose parametri novog modela?

Zadatak 4: Boston (DZ)

Skup podataka "Boston" sadrži informacije o stambenim nekretninama u različitim četvrtima Bostona. Želimo predvidjeti varijablu medv (medijan vrijednosti kuća u \$1000). Baza se nalazi u paketu "MASS".

- Izračunajte koeficijent determinacije (R^2) za model u kojemu su sve varijable ostale prediktori.
- Provedite eliminaciju unazad. U svakom koraku navedite uklonjeni prediktor i njegovu p -vrijednost prije izbacivanja.
- Koristeći funkciju `step()` provedite eliminaciju unazad. Koliko iznose parametri novog modela?