

2. DOMAĆA ZADAĆA

STATISTIČKI PRAKTIKUM 2

ožujak 2025.

3. zadatak (40 bodova)

U datoteci `university_admission_data.csv` nalaze se podaci za 50 (izmišljenih) studenata iz SAD-a. Varijable su sljedeće

- ▶ `GPA` - Prosječna ocjena učenika na skali od 0 do 4, s dvije decimale.
- ▶ `SAT_score` - Rezultat na SAT ispitima (0-1600), bez decimalnih mjesta. Koristi se kao standardizirani test za prijem na fakultet.
- ▶ `Admission_Test_Score` - Rezultat na prijemnom ispitu (0-340), bez decimalnih mjesta.
- ▶ `Letters_of_Recommendation` - Broj preporuka koje je student dobio, na skali od 1 do 5.
- ▶ `Extracurricular_Activities` - Yes/No varijabla koja označava je li student bio uključen u izvanškolske aktivnosti.
- ▶ `Volunteer_Experience` - Yes/No varijabla koja označava je li student imao iskustva u volontiranju.

- ▶ `Parental_Edu` - Prosječan broj godina obrazovanja roditelja, na skali od 12 (srednja škola) do 18+ (viši stupanj obrazovanja).
- ▶ `Admission_Score` - Konačan rezultat koji se koristi za donošenje odluke o prijemu na fakultet, temeljen na svim prethodnim varijablama.

Želimo analizirati `Admission_Score`, odnosno vjerojatnost da student bude primljen na odabrani studij.

Učitajte i proučite podatke.

Konstruirajte "puni" linearni model za varijablu odziva `Admission_Score` koristeći sve dane varijable prediktore (za kvalitativne varijable ne morate modelirati interakciju). Biste li odabrali taj model za predikciju budućih vjerojatnosti prijema na studij? Zašto?

Proučite statističku povezanost među varijablama. Postoje li ometajuće varijable? Postoje li korelirane nezavisne varijable? Predložite bolje modele te odaberite najbolje varijable prediktore. Obavezno testirajte sve svoje hipoteze vezane uz parametre modela i argumentirajte sve odluke koje ste donijeli.

Ima li vaš optimalan model smisla? Zadovoljava li on sve pretpostave linearnog regresijskog modela?

Za kategorijske varijable koje su vam ostale u optimalnom modelu, modelirajte interakcije s kvantitativnim varijablama poticaja. Što se dogodilo? Interpretirajte rezultate i testirajte svoje zaključke.

Vaš prijatelj nestrpljivo čeka rezultate o prijemu na fakultet pa vas je zamolio da mu svojim najboljim modelom procijenite očekivanu vrijednost njegovog Admission_Score-a. Zna da tijekom srednje škole nije baš puno učio i imao je prosjek 2.9 i zato mu je samo 2 profesora dalo pismo preporuke. Stoga se puno pripremao za završne SAT i prijemne testove i na njima skupio 1258 i 303 boda redom. U slobodno vrijeme je igrao nogomet i pjevao u školskom zboru pa nije imao vremena za volontiranje. Njegovi roditelji su oba završili fakultete (16 godina obrazovanja).