

1. domaća zadaća

Statistički praktikum 2, ožujak 2026.

U datoteci `putovanja.xlsx` nalaze se podaci aranžmana jedne turističke agencije. Varijable su sljedeće:

- `putovanje_id` - oznaka aranžmana,
- `cijena` - cijena aranžmana po osobi,
- `trajanje` - trajanje putovanja (u danima),
- `recenzije` - prosječna ocjena koju su tom aranžmanu dali korisnici, na skali od 0 do 100,
- `zvjezdice` - broj zvjezdica smještaja na putovanju,
- `udaljenost` - približna udaljenost (u kilometrima) odredišta od Hrvatske,
- `temperatura` - prosječna dnevna temperatura na odredištu za vrijeme putovanja, u °C,
- `popunjenost` - postotak (0 – 100) popunjenosti kapaciteta putovanja,
- `sezona` - godišnje doba realizacije putovanja,
- `polazak` - grad polaska na putovanje.

Zadaci (bodova)

1. Učitajte i proučite podatke. Grafički i inferencijalno testirajte normalnost distribucije `cijena`.
2. Promotrite varijable `cijena` i `trajanje`. Vizualno, čine li vam se povezane? Testirajte to. Možemo li pomoću jedne varijable uspješno predviđati vrijednost druge? A koristiti linearni regresijski model?
3. Promotrite zavisnost recenzije putnika o prosječnoj temperaturi na putovanju. Pokušajte modelirati tu zavisnost najjednostavnijim linearnim modelom. Što primjećujete, kako se ponašaju greške, a kako procijenjene vrijednosti? Predložite neka poboljšanja tog modela. Postoje li točke u podacima koje odstupaju od ostalih ili značajno utječu na modele?
4. Vlasniku agencije se čini da recenzije za pojedino putovanje ovise o broju zvjezdica korištenog smještaja te udaljenost odredišta od Hrvatske. Provjerite to. Jeste li `zvjezdice` promatrali kao kvalitativnu ili kvantitativnu varijablu? Zašto? Što se događa ako modeliramo i interakcije? Promotrite 95% pouzdane intervale za koeficijente modela. Što zaključujete? Postoji li multikolinearnost među prediktorima u ovom modelu? Predložite barem dva načina na koja biste to provjerili. Što primjećujete?
5. Imalo bi smisla da i drugi faktori utječu na zadovoljstvo putnika i recenzije koje putnici ostavljaju. Konstruirajte "puni" linearni model za varijablu odziva `recenzije` (ne morate modelirati interakcije). Bi li vlasnici agencije mogli koristiti taj model za predviđanje zadovoljstva putnika nekim novim predloženim aranžmanom. 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.
6. Agencija želi dodati novi aranžman putovanja u Maroko na 10 dana. Polazak bi bio iz Zagreba, a cijena 1190 eura. Imaju ugovor s hotelom od 4 zvjezdice i žele postići što bolji uspjeh, odnosno što bolje recenzije za novi paket. Po vašem najboljem modelu iz točke 5, ako bi aranžman bio potpuno rasprodan, je li bolje to putovanje planirati u travnju ili listopadu 2026. godine?