

Statistička teorija telekomunikacija

ZADAĆA 1

Eldar Osmanović
27.05.2025.

1 Uvod

U okviru ove zadaće, analizirani su podaci iz fajla `student-mat.csv`, koji sadrži informacije o studentima koji slušaju predmet Statistička teorija telekomunikacija. Cilj analize je bio istražiti obrasce ponašanja studenata kroz broj izostanaka i uspjeh na predmetu, te izračunati relevantne statističke mjere i vjerovatnoće koje bi mogle ukazati na potencijalne veze između izostanaka i akademskog uspjeha. Pri ovoj analizi je korištena `Pandas` biblioteka u programskom jeziku Python.

Zadatak 1.1

U ovom zadatku cilj je bio izračunati uslovnu vjerovatnoću da je student položio predmet, pod uslovom da je izostao sa 15 ili više sati predavanja.

Prije same analize bilo je potrebno pripremiti podatke. Relevantne kolone su nam kolona `Absences` u kojoj je broj izostanaka za svakog studenta i kolona `G3` u kojoj su ostvareni bodovi. Zbog toga u obzir uzimamo samo podaci = `podaci[["absences", "G3"]]`, gdje su u varijabli `podaci` prethodno učitani podaci tj. `podaci = pd.read_csv("studentmat.csv")`.

Analiza srednje vrijednosti i standardne devijacije je podijeljena u dva dijela: prije i poslije pripreme odnosno filtriranja podataka.

Nakon što su prazna polja za `absences` popunjena najčešće prisutnim brojem izostanaka koji je, uzgred rečeno jednak 0.0, izračunate su srednja vrijednost i standardna devijacija $E[X]_p$ i σ_p gdje indeks p naglašava da su to vrijednosti prije filtriranja. Sa X je označen broj izostanaka. Rezultati koje smo dobili su:

$$\text{srednja} = E[X]_p = 9.944303797468354, \quad \text{std_dev} = \sigma_p = 56.74698441603364.$$

Ukoliko malo pogledamo vrijednosti u koloni `absences`, vidjet ćemo da ova vrijednost standardne devijacije ima smisla jer imamo mnogo vrijednosti koje značajno odstupaju od srednje vrijednosti, što standardna devijacija u suštini i mjeri.

Nakon ovog proračuna, iz analize smo isključili sve redove u kojima je broj izostanaka od srednje vrijednosti udaljen više od jedne standardne devijacije. Konkretno, za ove vrijednosti, filtrirali smo samo one podatke kojima je broj izostanaka u segmentu

$$[-46.80268061856529, 66.69128821350199].$$

Srednja vrijednost $E[X]_f$ i standardna devijacija σ_f (indeks f označava da su sada podaci filtrirani) nakon filtriranja su

$$\text{srednja} = E[X]_f = 5.52051282051282, \quad \text{std_dev} = \sigma_f = 7.227008735523394.$$

Primijetimo da sada standardna devijacija manje odstupa od srednje vrijednosti.

Nakon uspješne pripreme podataka, potrebno je bilo odrediti vjerovatnoću da je student položio predmet ako je izostao sa 15 ili više predavanja. Za studenta se smatra da je položio ako je ostvario minimalno 55 % od maksimalnog broja bodova

```
max_points = podaci["G3"].max()*0.55.
```

Za izračun tražene vjerovatnoće, koristili smo formulu za uslovnu vjerovatnoću

$$p(\text{položio} \mid \text{izostanci} \geq 15) = \frac{p(\text{položio} \cap \text{izostanci} \geq 15)}{p(\text{izostanci} \geq 15)}$$

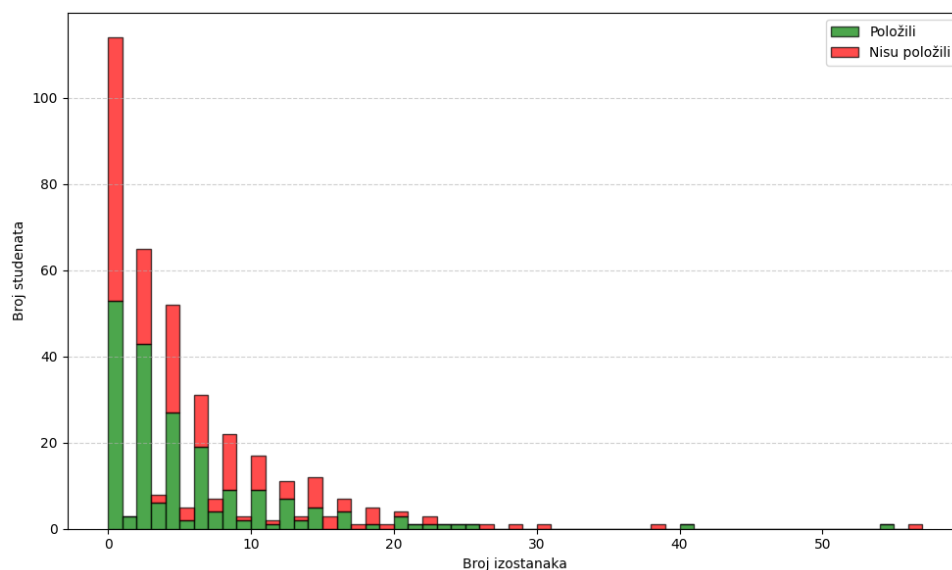
Ključni dio koda je dat u nastavku

```
1 # ...
2 podaci["Polozeno"] = np.where(podaci["G3"] >= max_points, 1 , 0);
3
4 # Broj studenata koji su izostali sa 15 ili više sati
5 ukupno_izostali_15plus = podaci[podaci["absences"] >= 15]
6
7 # Broj koji su i izostali 15 i položili
8 polozili_izostali_15plus =
9     ↪ ukupno_izostali_15plus[ukupno_izostali_15plus["Polozeno"] == 1]
10
11 # Vjerovatnoća
12 vjerovatnoca = len(polozili_izostali_15plus) / len(ukupno_izostali_15plus);
13 # ...
```

Vjerovatnoća do koje smo na kraju došli iznosi

$$vjerovatnoca = p(\text{položio} \mid \text{izostanci} \geq 15) = 0.4286.$$

Ostaje nam još i grafički predstaviti rezultate koristeći histogram. Histogram prikazuje raspodjelu broja izostanaka studenata koji su slušali predmet Statistička teorija telekomunikacija, s posebnim naglaskom na to da li su položili predmet ili ne.



Slika 1: Histogram broja izostanaka za studente koji su položili i one koji nisu.

- Osa x predstavlja broj izostanaka.

- Osa y predstavlja broj studenata koji su imali odgovarajući broj izostanaka.
- Zeleni stubići označavaju broj studenata koji su položili predmet (ostvarili bar 55% ukupnih bodova).
- Crveni stubići označavaju broj studenata koji nisu položili predmet.

Histogram je stacked, što omogućava direktnu komparaciju među grupama na svakoj vrijednosti broja izostanaka. Uočen je jasan trend da se većina studenata sa malim brojem izostanaka (posebno od 0 do 5) nalazi u grupi onih koji su položili ispit, dok se sa porastom broja izostanaka povećava udio onih koji nisu položili. Također, broj studenata generalno opada kako broj izostanaka raste, što znači da većina studenata ima relativno mali broj izostanaka. Ovaj histogram jasno sugerira da veća prisutnost na predavanjima korelira s većom vjerovatnoćom polaganja ispita.

Zaključak

Analizom podataka o broju izostanaka i konačnom uspjehu studenata iz predmeta Statistička teorija telekomunikacija pokazano je da postoji određena povezanost između prisutnosti na nastavi i polaganja predmeta. Nakon filtriranja ekstremnih vrijednosti i izračuna statističkih mjera, ustanovljeno je da studenti koji su izostali sa 15 ili više predavanja imaju značajno manju vjerovatnoću da polože predmet, koja u ovom uzorku iznosi samo 42.86%.

Također, histogram jasno pokazuje da se studenti sa manjim brojem izostanaka češće nalaze među onima koji su položili ispit, dok se pri većem broju izostanaka povećava udio onih koji nisu položili.

Kompletna kod se može pronaći na GitHub linku: <https://github.com/oeldar/Statisticka-teorija-telekomunikacija>