

Zadaća 1 – Primijenjena vjerovatnoća i statistika na stvarnim podacima

Cilj

Cilj ove zadaće je primijeniti teorijske koncepte iz poglavlja 1 do 5 iz udžbenika *Probability and Statistics for Engineers and Scientists* na stvarne podatke. Svaka grupa će dobiti različit dataset i nezavisno analizirati statistička svojstva i odnose. Krajnji rezultat treba biti Jupyter Notebook koji je jasno strukturiran i objašnjen.

Pokrivena poglavlja

- Poglavlje 1: Deskriptivna statistika
- Poglavlje 2: Vjerovatnoća
- Poglavlje 3: Diskretne nasumične varijable i distribucije
- Poglavlje 4: Kontinuirane nasumične varijable i distribucije
- Poglavlje 5: Zajedničke distribucije

Instrukcije

Koristeći dodijeljeni dataset, sprovedite sveobuhvatnu statističku analizu koja pokazuje razumijevanje i primjenu koncepata iz prvih pet poglavlja. Strukturajte vaš rad slično datom primjeru zadaće.

Struktura zadaće i podjela zadataka

1. Uvod – Opis dataseta

- Opišite dataset: izvor, struktura, ključne varijable.
- Definišite ciljeve i hipoteze.
- **Primjer:** Podaci o klijentima sa varijablama poput godina, spol, godišnji prihod itd.

2. Deskriptivna analiza (Poglavlje 1)

- Identifikujte vrste varijabli (kategorijske/numeričke).
- Izračunajte osnovne statistike i koristite vizualizacije.
- **Primjer:** `.describe()`, `.info()`, boxplot prihoda po spolu.

3. Vjerovatnoća (Poglavlje 2)

- Definišite i analizirajte vjerovatnoće događaja.
- Izračunajte empirijske i uslovne vjerovatnoće.
- **Primjer:** $P(\text{Spol} = \text{Ženski})$, $P(\text{Score} \geq 80 \mid \text{Spol} = \text{Ženski})$.

4. Diskretne distribucije (Poglavlje 3)

- Odaberite tri do četiri diskretne varijable i vizualizirajte njihove distribucije (npr. koristeći bar-plot ili box-plot).
- Izračunajte očekivanu vrijednost i varijansu odabranih nasumičnih varijabli.
- Na osnovu ovih vizualizacija i proračuna, opišite odabrane varijable (izgled distribucije, šta predstavlja očekivana vrijednost i varijansa u odnosu na prikazanu distribuciju).

5. Kontinualne distribucije (Poglavlje 3)

- Odaberite tri do četiri kontinualne varijable i vizualizirajte njihove distribucije (npr. koristeći histogram i aproksimaciju koristeći parametar *kind="kde"*).
- Izračunajte očekivanu vrijednost i varijansu odabranih nasumičnih varijabli.
- Na osnovu ovih vizualizacija i proračuna, opišite odabrane varijable (izgled distribucije, šta predstavlja očekivana vrijednost i varijansa u odnosu na prikazanu distribuciju).

6. Zajedničke distribucije i zavisnost (Poglavlje 5)

- Analizirajte odnose između varijabli koristeći scatter plotove, korelaciju itd.
- Procijenite zajedničke i uslovne distribucije.
- **Primjer:** Scatter plot između godina i prihoda, analiza korelacije.

7. Zaključak i refleksija

- Sumirajte glavne nalaze i razmislite o rezultatima.
- Navedite ograničenja i prijedloge za budući rad.
- **Primjer:** Prihod ne zavisi značajno od spola.

Kriteriji za ocjenu

- Potpunost i tačnost statističke analize
- Jasnoća objašnjenja i interpretacija
- Efektivna upotreba vizualizacija
- Čist, komentiran kod i struktura notebooka
- Kreativnost i originalnost u analizi podataka