

**STATISTIČKI PRAKTIKUM**

## PRVI KOLOKVIJ

## B grupa

**Napomena:**

- Za sve zadatke popratni tekst piše se na papir ili u R skripti u obliku komentara. Konačni dokument rješenja zadataka u R-u generira se na "File → Knit Notebook → Word". Taj se dokument šalje na `statprak@mathos.hr` zajedno s R skriptom. U predmetu poruke treba napisati svoje ime i prezime.

ZADATAK 1: [2+6+5+7+6+6+4+6+6+6+6+4=64 bodova]

Instalirajte paket `carData` naredbom `install.packages("carData")` u konzoli (ne u skripti iz koje se generira izvješće). Učitajte tablicu podataka `Arrests` iz paketa `carData` naredbom `data(Arrests)`. Tablica sadrži podatke o osobama uhićenim u Torontu zbog posjedovanja malih količina marihuane. Varijable su:

- `released` označava je li uhićenju slijedio poziv na sud (**Yes**) ili ne (**No**)
- `colour` je rasa uhićene osobe
- `year` je godina uhićenja
- `age` je dob uhićenog
- `sex` je spol (**Male** - muški, **Female** - ženski)
- `employed` pokazuje je li osoba zaposlena (**Yes**) ili ne (**No**)
- `citizen` pokazuje je li osoba građanin Toronta
- `checks` pokazuje koliko je puta prije uhićena osoba bila evidentirana u policijskim bazama (prethodna uhićenja i sl.).

- Odredite tipove varijabli u tablici.
- Koliki je udio muškaraca u uzorku? Koliki je udio mlađih od 35 godina?
- Koje godine je bilo najviše uhićenih? Prikažite grafički odgovarajućim prikazom.
- Usporedite prethodne evidencije (varijabla `checks`) po spolu tako da možete reći kako je spol povezan s brojem policijskih evidencija (potkrijepite usporedivim brojkama). Prikažite odgovarajućim grafičkim prikazom.
- Koliko godina ima najmlađa uhićena osoba? Radi li se o muškarcu ili ženi?
- Usporedite prosječnu dob uhićenih koji do sada nisu bili evidentirani i onih koji su bili evidentirani barem jednom. Grafički prikazite usporedbu.
- Usporedite raspršenost dobi nezaposlenih muškaraca i zaposlenih žena.
- Procijenite je li vjerojatnije da je uhićena osoba nikad prije evidentirana žena starija od 30 godina ili muškarac mlađi od 30 godina koji je evidentiran barem jednom.
- Je li udio žena veći među zaposlenim ili nezaposlenim osobama?
- Odredite 95%-tni pouzdani interval za očekivanu dob muškaraca.
- Odredite 95%-tni pouzdani interval za proporciju zaposlenih žena.
- Grafički usporedite distribuciju dobi s normalnom distribucijom i komentirajte.

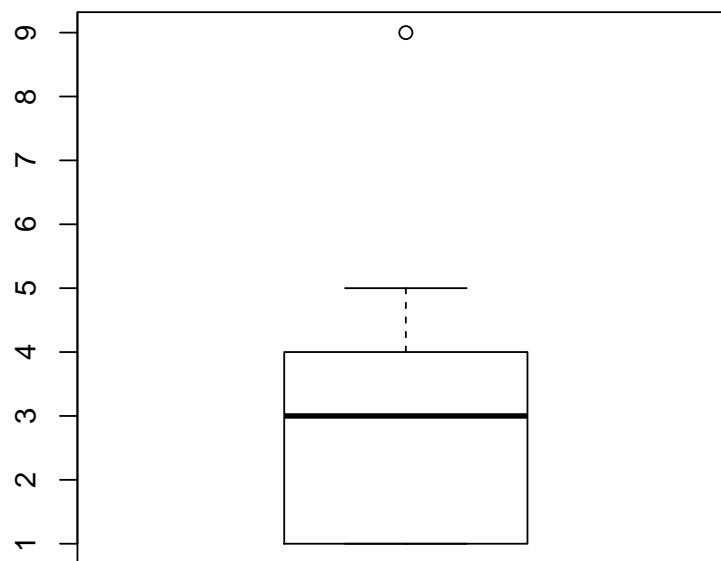
**ZADATAK 2:** [4+7+7+9=27 bodova]

Za igraču kockicu je nepoznata vjerojatnost  $p$  da padne šestica. Kockica se baca dva puta i bilježi se koliko šestica je palo u ta dva bacanja. Ponavljanjem dolazimo do slučajnog uzorka  $X_1, \dots, X_n$ .

- Definirajte statistički model za takav uzorak.
- Koristeći  $\bar{X}_n$  definirajte nepristran procjenitelj za vjerojatnost šestice  $p$ .
- Izračunajte varijancu tog procjenitelja.
- Pretpostavite da je kockica pravilna i simulirajte 100 uzoraka od kojih svaki sadrži  $n = 100$  ponavljanja promatranog pokusa. Izračunajte prosječnu vrijednost procjene definirane u (b). Procijenite varijancu procjenitelja i usporedite s rezultatom dobivenim u (c).

**ZADATAK 3:** [3+3+3=9 bodova]

Na slici je prikazan kutijasti dijagram nekog niza podataka. Upišite odgovarajući broj.



- Najveći podatak je \_\_\_\_\_.
- Barem 25% podataka je \_\_\_\_\_.
- Barem \_\_\_\_\_ % podataka je veće ili jednako 4.