

STATISTIČKI PRAKTIKUM

DRUGI 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: [4+4=8 bodova]

- Definirajte koeficijent korelacije slučajnog vektora (X, Y) i navedite njegova svojstva.
- Kako procjenjujemo distribuciju dvodimenzionalnog diskretnog slučajnog vektora?

ZADATAK 2: [8+8+9+9+8+10=52 bodova]

Instalirajte paket `carData` naredbom `install.packages("carData")` u konzoli (ne u skripti iz koje se generira izvješće). Učitajte paket u skripti naredbom `library("carData")` i tablicu podataka `Mroz` naredbom `data(Mroz)`.

Tablica sadrži podatke o ženama na tržištu rada. Varijable su:

- `lfp` označava je li žena zaposlena (`yes`) ili ne (`no`)
 - `k5` je broj djece mlađe ili staro 5 godina
 - `k618` je broj djece stare 6 do 18 godina
 - `age` je dob
 - `wc` pokazuje je li žena studirala (`yes`) ili ne (`no`)
 - `hc` pokazuje je li suprug žene studirao (`yes`) ili ne (`no`)
 - `lwg` je logaritam plaće za žene koje su zaposlene, a za one koje nisu radi se o procjeni
 - `inc` je prihod ostatka obitelji bez ženine plaće.
- Je li na razini značajnosti 0.05 proporcija onih žena koje su studirale manja od proporcije onih koje nisu?
 - Jesu li na razini značajnosti 0.05 plaće žena koje su studirale (varijabla `lwg`) očekivano veće od onih koje nisu studirale?
 - Razlikuju li se na razini značajnosti 0.05 distribucije dohotka obitelji (varijabla `inc`) s obzirom na to je li žena studirala ili ne? Grafički usporedite procijenjene funkcije distribucija. Na osnovu toga zaključite kojim se još testom može testirati razlika u distribucijama u ovoj situaciji.
 - Je li na razini značajnosti 0.05 proporcija nezaposlenih veća među ženama koje nisu studirale nego među onima koje jesu? Procijenite te proporcije.
 - Ukazuje li na razini značajnosti 0.05 veća plaća žene (varijabla `lwg`) na veći dohodak ostatka obitelji?
 - Odredite zajedničku tablicu frekvenciju broja djece (bez obzira na dob) i obrazovanja supruga (varijabla `hc`). Odredite zajedničku tablicu frekvencija za iste varijable, ali u kojoj 5 ili više djece predstavlja jednu kategoriju. Na osnovu te tablice odgovorite je li broj djece nezavisan o obrazovanju supruga na razini značajnosti 0.05.

ZADATAK 3: [8+10+4=22 bodova]

Instalirajte paket `UsingR` naredbom `install.packages("UsingR")` u konzoli (ne u skripti iz koje se generira izvješće). Učitajte paket u skripti naredbom `library("UsingR")` i tablicu podataka `crime` naredbom `data(crime)`. Tablica sadrži podatke o stopi kriminala za 50 američkih država 1983. i 1993. godine.

- (a) Je li na razini značajnosti 0.05 očekivana stopa kriminala porasla?
- (b) Procijenite parametre linearnog modela koji modelira stopu kriminala 1993. godine na osnovu stope 1983. godine. Nacrtajte točke podataka i procijenjeni regresijski pravac. Za koliko je stopa kriminala narasla? Kako ocjenjujete jakost modela?
- (c) Ako pretpostavimo da će se ovakav trend nastaviti, procijenite stopu kriminala u Arizoni 2003. godine.

ZADATAK 4: [10 bodova]

Na prvom kolokviju nekog kolegija 52 studenta od ukupno 79 su prešla prag. Svi studenti izašli su na drugi kolokvij. Od onih koji su prešli prag na prvom kolokviju, njih 50 je i na drugom kolokviju prešlo prag. Od onih koji nisu prešli prag na prvom kolokviju, 5 je prešlo prag na drugom kolokviju. Jesu li na razini značajnosti 0.05 rezultati bolji na drugom kolokviju?

ZADATAK 5: [8 bodova]

U igri na sreću karta se izvlači iz špila od 52 karte tri puta s vraćanjem (karta se vraća u špil i onda se opet izvlači) i gleda se koliko puta je izvučen herc (jedna od četiri boje). Od 100 igara, 40 puta nije izvučen nijedan herc, 47 puta je izvučen jedan, 10 puta dva herca i 3 puta tri herca. Ima li razloga sumnjati na razini značajnosti 0.05 da u špilu nisu sve boje jednako zastupljene?

ZADATAK 6: BONUS [20 bodova]

50 studenata na prvom kolokviju ostvarilo je prosječni rezultat 20 bodova uz standardnu devijaciju 5. Na drugom kolokviju svaki student imao je točno 20% bolji rezultat. Možemo li na razini značajnosti 0.05 tvrditi da je očekivani rezultat bolji na drugom kolokviju?