

**STATISTIČKI PRAKTIKUM**

PISMENI ISPIT 3.7.2019.

**Napomena:**

- Za sve zadatke popratni tekst piše se ili 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:** [5+5+5=15 bodova]

- Navedite klasične pretpostavke linearnog regresijskog modela. Ako smo za neke podatke dobili linearni regresijski model s koeficijentom determinacije jednakim 1, što možemo reći o tim podacima?
- Definirajte jednostavni slučajni uzorak, parametarski statistički model te procjenitelja nepoznatog parametra distribucije iz koje dolazi jednostavni slučajni uzorak.
- Što je rang podatka  $x_i$  u nizu  $(x_1, \dots, x_n)$  u kojem su svi elementi međusobno različiti? Odredite rangove podataka u nizu (8.2, 5.4, 4.1, 2.2). Navedite test-statistiku barem jednog testa koji se temelji na rangovima podataka.

**ZADATAK 2:** [8+6+8+6+8+6+8+8=58 bodova]

Instalirajte paket `catdata` naredbom `install.packages("catdata")` u konzoli (ne u skripti iz koje se generira izvješće). Nakon učitavanja paketa (`library("catdata")`), učitajte tablicu `children` naredbom `data(children)`. Tablici sadrži sljedeće podatke za 1761 ispitanih žena:

- varijabla `child` označava broj djece
  - varijabla `age` je dob
  - varijabla `dur` je trajanje obrazovanja
  - varijabla `god` je slaganje s tvrdnjom vjeruje li u Boga, od 1 (potpuno se slaže) do 5 (ne slaže se uopće, i 6 - nije nikad razmišljala o tome)
  - varijabla `univ` je 1 ako je studirala i 0 ako nije studirala.
- Postoji li na razini značajnosti 0.05 razlika u zastupljenosti odgovora na pitanje o vjeri zabilježeno u varijabli `god`? Grafički prikažite empirijsku distribuciju te varijable.
  - Možemo li na razini značajnosti 0.05 tvrditi da su broj djece i trajanje obrazovanja povezani padajućom monotonom vezom? Prikažite prikladnim grafičkim prikazom.
  - Razlikuje li se na razini značajnosti 0.05 očekivana dob ispitanica s obzirom na to jesu li studirale ili ne? Grafički usporedite dva skupa i izračunajte prosječne vrijednosti.
  - Procijenite očekivanu dob 95%-tnim pouzdanim intervalom.
  - Je li na razini značajnosti 0.05 udio žena koje nemaju djecu veći kod onih koje su studirale nego kod onih koje nisu?
  - Je li proporcija žena bez djece veća od 20% na razini značajnosti 0.05?
  - Jesu li vjera (varijabla `god`) i studiranje nezavisni na razini značajnosti 0.05? Procijenite uvjetne distribucije za ta dva obilježja te analizirajte je li vjera jača kod onih žena koje su studirale ili kod onih koje nisu.
  - Razlikuju li se distribucije godina kod onih koje su na pitanje o vjeri odgovorile s 1 i s 5? Usporedno grafički prikažite procjene funkcija distribucija.

ZADATAK 3: [10 bodova]

Ivan tvrdi da se u kutiji nalazi jednak broj plavih, crvenih i zelenih kuglica. Od 20 izvučenih kuglica 11 je plavih, 6 crvenih i 3 zelene. Možemo li na razini značajnosti 0.05 tvrditi da Ivan ne govori istinu?

ZADATAK 4: [7+10 bodova]

Instalirajte paket `Flury` naredbom `install.packages("Flury")` u konzoli (ne u skripti iz koje se generira izvješće). Nakon učitavanja paketa (`library("Flury")`), učitajte tablicu `m.twins` naredbom `data(m.twins)`. Tablica sadrži neke tjelesne mjere za uzorak blizanaca.

- (a) Postoji li na razini značajnosti 0.05 razlika u očekivanoj visini (varijable `STA1` i `STA2`) između jednojajčanih blizanaca (varijabla `Type` jednaka `"monozygotic"`)?
- (b) Procijenite parametre linearnih modela koji opisuju visinu jednog blizanca u ovisnosti o visini drugoga i to posebno za jednojajčane blizance (varijabla `Type` jednaka `"monozygotic"`) i posebno za dvojajčane blizance (varijabla `Type` jednaka `"dizygotic"`). Usporedite parametre tih modela i komentirajte. Što možete reći o kvaliteti tih modela?