

STATISTIČKI PRAKTIKUM

PRVI KOLOKVIJ

Napomena:

- Rješenja teorijskih zadataka (zadaci 1.a, 1.b, 1.c, 2.a i 2.b) pišu se na papir koji se predaje predmetnom nastavniku.
- Kod zadataka koji se rješavaju u R-u sav **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 → Compile Notebook → Word". Taj se dokument šalje na **statistika@mathos.hr - u predmetu poruke treba napisati svoje ime i prezime**.

ZADATAK 1: [3+4+4=11 bodova]

Odgovorite na sljedeća pitanja:

- a) Što je to p -kvantil distribucije, $p \in (0, 1)$? Diskutirajte problem jedinstvenosti kvantila diskretnih i neprekidnih distribucija te definirajte funkciju kvantila. (3 boda)
- b) Objasnite pojmove kritičnog područja, razine značajnosti i p -vrijednosti statističkog testa. Na primjeru z -testa, uz zadanu razinu značajnosti α , povežite princip donošenja odluke na temelju poznavanja kritičnog područja s principom donošenja odluke temeljenom na izračunatoj p -vrijednosti. (4 boda)
- c) Odredite kritično područje i p -vrijednost statističkog testa u čijoj alternativnoj hipotezi stoji da je p_0 -kvantil distribucije manji od neke unaprijed dane vrijednosti x_0 . (4 boda)

ZADATAK 2: [3+3+3=9 bodova]

Neka je (X_1, \dots, X_n) jednostavni slučajni uzorak (j.s.u.) iz uniformne distribucije na segmentu $[0, a]$, $a > 0$. Tada je $E[X_k] = a/2$ za svaki $k \in \{1, \dots, n\}$.

- a) Interpretirajte slabi zakon velikih brojeva u kontekstu zadanog j.s.u. (X_1, \dots, X_n) . (3 boda)
- b) Na temelju standardnog procjenitelja za očekivanje distribucije odredite procjenitelj za nepoznati parametar a uniformne distribucije na $[0, a]$. (3 boda)
- c) Simulirajte uzorak veličine 1000 iz uniformne distribucije na segmentu $[0, a]$ za $a = 1$ i na temelju procjenitelja za a određenog u zadatku (b) izračunajte njegovu procjenu. (3 boda)

ZADATAK 3: [2+5+3+5+5+5+5=30 bodova]

Datoteka `krave.csv` sadrži vrijednosti nekih financijskih parametara i rezultate praćenja proizvodnje mlijeka za uzorak krava s jedne velike farme u SAD-u. Varijable koje nas interesiraju su sljedeće:

- varijabla `laktacija` sadrži redni broj laktacije u kojoj se krava nalazi (laktacija je period u kojem krava proizvodi mlijeko)
- varijabla `ukupno_mlijeko` sadrži ukupnu količinu mlijeka u funtama (lb) kojeg je krava dala od početka svoje prve laktacije do završetka promatranja, tj. "ikada" ($1 \text{ lb} \approx 453.6 \text{ g}$)
- varijabla `hrana_trosak` sadrži ukupne troškove za hranu po kravi u promatranom razdoblju (u \$)
- varijabla `zdravlje_trosak` sadrži ukupne troškove veterinarskih usluga po kravi u promatranom razdoblju (u \$)
- varijabla `mlijeko_zarada` sadrži ukupnu zaradu od proizvedenog mlijeka po kravi u promatranom razdoblju (u \$).

Riješite sljedeće zadatke:

- (a) Kojim tipom distribucije biste modelirali varijablu **laktacija** i zašto? Procijenite njezinu distribuciju te izračunajte procjenu vjerojatnosti da je slučajno odabrana krava s promatrane farme u trećoj ili višoj laktaciji? (2 boda)
- (b) 95%-tnim pouzdanim intervalom procijenite vjerojatnost da se slučajno odabrana krava s promatrane farme nalazi u trećoj ili višoj laktaciji. Na temelju tog pouzdanog intervala odgovorite na sljedeće pitanje: možemo li na razini značajnosti $\alpha = 0.05$ tvrditi da se proporcija krava s promatrane farme koje su u trećoj ili višoj laktaciji razlikuje od 25%? (5 bodova)
- (c) Histogramom prikazite relativne frekvencije podataka iz varijable **ukupno_mlijeko** kategorizirane u 30 disjunktih intervala. Kojim tipom distribucije biste modelirali tu varijablu i zašto? Kojim funkcijama zadajemo taj tip distribucije? Grafički prikazite procjene tih funkcija za distribuciju ukupne količine mlijeka ikada proizvedenog od jedne krave s promatrane farme. (3 boda)
- (d) Procijenite očekivanje distribucije iz zadatka (c). Odredite i interpretirajte 99%-tni pouzdani interval za to očekivanje. (5 bodova)
- (e) Vlasnik farme tvrdi da je ukupno na ishranu i veterinarske usluge u promatranom razdoblju potrošio u prosjeku 3500 \$ po kravi. Neovisni nadzor poslovanja farme mu ne vjeruje i tvrdi da je taj prosječni trošak po kravi zapravo bio niži. Možete li na razini značajnosti $\alpha = 0.05$ podržati tvrdnju neovisnog nadzora? Koji ste test koristili i zašto? (5 bodova)
- (f) Neovisni nadzor na farmi utvrdio je da je standardna devijacija ukupne zarade od proizvedenog mlijeka po kravi 200\$. Vlasnik farme pak tvrdi da je varijabilnost te zarade veća. Postavite hipoteze i interpretirajte ih. Sugeriraju li vam histogram i qq-plot zadovoljenost pretpostavki statističkog testa za testiranje takve hipoteze ili ne? Možete li odlučiti tko je u pravu? (5 bodova)
- (g) Neovisni nadzor tvrdi da je medijan ukupne zarade od proizvedenog mlijeka po kravi veći od prosjeka uzorka takvih zarada. Možete li se na razini značajnosti $\alpha = 0.05$ složiti s tom tvrdnjom? Koji ste test koristili i zašto? (5 bodova)