

STATISTIČKI PRAKTIKUM

PRVI KOLOKVIJ

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: [2+4+2=8 bodova]

Ivica nam je ponudio igrati igru u kojoj bacamo pravilan novčić 10 puta. Kada padne glava, osvajamo jednu kunu, a kada padne pismo plaćamo Ivanu jednu kunu. Pristali smo igrati i 8 puta je palo pismo. Nedugo zatim kod Ivana smo pronašli dva novčića, jedan pravilan, a drugi takav da pismo pada s vjerojatnošću 0.75. Sumnjamo da je Ivan koristio taj novčić, a ne pravilan novčić.

- Definirajte što je općenito statistički model te kako izgleda u ovom primjeru. Odredite hipoteze koje bi htjeli testirati u ovom primjeru.
- Želimo testirati hipoteze na osnovu test-statistike koja predstavlja broj pisama u igri i na osnovu kritičnog područja $\{8, 9, 10\}$. Definirajte što je općenito pogreška I. i II. tipa te za ovako definiran test izračunajte koliko one iznose.
- Što je p -vrijednost statističkog testa? Kolika je p -vrijednost u ovom primjeru?

ZADATAK 2: [6 bodova]

U pokusu bacanja novčića dobiven je 95%-tni interval za vjerojatnost pisma $[0.453794, 0.671206]$. Koliko je puta bačen novčić?

ZADATAK 3: [2+3+5=10 bodova]

Očekivanje i varijanca Poissonove distribucije s parametrom λ je upravo λ .

- Koristeći tu činjenicu definirajte dva procjenitelja parametra λ .
- Definirajte nepristranost procjenitelja. Jesu li procjenitelji definirani pod (a) nepristrani?
- Napravite simulaciju u kojoj ćete generirati 100 uzoraka veličine 100 iz Poissonove distribucije s parametrom 3. Na svakom uzorku izračunajte vrijednost dva procjenitelja. Na osnovu dobivenih rezultata, što možete reći o kvaliteti procjenitelja na temelju kriterija varijabilnosti?

ZADATAK 4: [5+4+5+4+5+3=26 bodova]

Instalirajte paket `catdata` naredbom `install.packages("catdata")` u konzoli (ne u skripti iz koje se generira izvješće). Učitajte tablicu podataka `heart` iz paketa `catdata` sljedećim naredbama (dio skripte za izvješće):

```
library("catdata")
data(heart)
heart <- as.data.frame(heart)
```

U tablici `heart` nalaze se podaci o 462 muškarca iz jedne regije u Južnoafričkoj republici poznatoj po visokoj stopi srčanih bolesti. Tablica između ostalog sadrži sljedeće varijable:

- varijabla `y` je 1 ako je ispitanik obolio od koronarne bolesti srca i 0 ako nije
 - varijabla `sbp` je sistolički tlak
 - varijabla `tobacco` je procjena ukupne količine duhana konzumirane u životu
 - varijabla `alcohol` je količina alkohola koju osoba konzumira
- (a) Odredite empirijsku distribuciju slučajne varijable koja modelira oboljenje od koronarne bolesti srca. Grafički prikazite procijenjenu distribuciju. Je li rizik od oboljenja veći za one koji su u obitelji imali slučajeva oboljenja ili za one koji nisu imali takvih slučajeva?
- (b) Je li prosječni tlak viši kod oboljelih ili kod zdravih? Odgovarajućim grafičkim prikazom usporedite podatke o tlaku za oboljele i zdrave. Što zaključujete iz tog prikaza?
- (c) Sistolički tlak viši od 140 smarta se povišenim. Možemo li na razini značajnosti 0.05 tvrditi da oboljeli imaju povišeni očekivani tlak? Procijenite vjerojatnost da ispitanik ima povišeni krvni tlak.
- (d) Na razini značajnosti 0.05, možemo li tvrditi da je za osobu koja nikad nije pušila i ne pije alkohol, vjerojatnost oboljenja manja od 0.2?
- (e) Je li na razini značajnosti 0.05 donji kvartil distribucije sistoličkog tlaka veći od 120? Što to govori o proporciji ispitanika s tlakom nižim od 120?
- (f) Napravite grafičke prikaze koji daju uvid u distribuciju varijable `alcohol`. Što možete reći o simetričnosti distribucije?