

STATISTIČKI PRAKTIKUM

PISMENI ISPIT 19.6.2019.

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]

- (a) Neka je X_1, \dots, X_n jednostavan slučajan uzorak takav da je $P(X_1 = -1) = 1 - p$ i $P(X = 1) = p$, $p \in (0, 1)$. Koristeći centralni granični teorem definirajte test statistiku koja će asimptotski imati standardnu normalnu distribuciju.
- (b) Definirajte test statistiku i objasnite način provođenja z -testa za testiranje jednakosti očekivanja normalne distribucije unaprijed zadanoj vrijednosti.
- (c) Navedite osnovna svojstva Spearmanovog koeficijenta korelacije ranga. Kakva je njegova veza s Pearsonovim koeficijentom korelacije? Neka je na nekim uzorcima Spearmanov koeficijent 0.5. Ako sve podatke pomnožimo s 2, koliki će biti Spearmanov koeficijent na tako dobivenim uzorcima?

ZADATAK 2: [8+4+7+7+10+12+4=52 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 `medcare` naredbom `data(medcare)`. Tablica sadrži sljedeće podatke za 4406 starijih osoba sa zdravstvenim osiguranjem:

- varijabla `ofp` označava broj posjeta liječniku
 - varijabla `hosp` je broj hospitalizacija
 - varijabla `healthpoor` je 1 ako je ispitanik lošeg zdravlja (u odnosu na prosječne)
 - varijabla `healthexcellent` je 1 ako je ispitanik izvrsnog zdravlja (u odnosu na prosječne)
 - varijabla `numchron` je broj kroničnih bolesti
 - varijabla `male` je 1 za muškarce i 0 za žene
 - varijabla `age` je dob od koje je oduzeto 60
 - varijabla `married` je 1 ako je u braku i 0 ako nije
 - varijabla `school` je trajanje obrazovanja.
- (a) Koristeći činjenicu da su očekivanje i varijanca Poissonove distribucije s parametrom λ jednaki λ , definirajte dva procjenitelja parametra λ . Jesu li ti procjenitelji nepristrani? Pretpostavimo li da broj kroničnih bolesti ima Poissonovu distribuciju, procijenite parametar te distribucije koristeći ta dva procjenitelja.
- (b) Odredite empirijsku distribuciju broja kroničnih bolesti te je grafički prikažite.
- (c) Je li očekivani broj posjeta liječniku veći za žene na razini značajnosti 0.05? Grafički usporedite dva skupa podataka.
- (d) Razlikuju li se na razini značajnosti 0.05 proporcije onih s lošim zdravljem kod ispitanika koji su u braku i onih koji nisu? Što ukazuju podaci, djeluje li brak pozitivno na zdravstveno stanje?

- (e) Jesu li na razini značajnosti 0.05 varijable dob (centrirana u odnosu na 60) i broj posjeta liječniku korelirane? Što možemo reći o zavisnosti? Možemo li tvrditi da postoji monotona veza između tih varijabli na razini značajnosti 0.05?
- (f) Je li na razini značajnosti 0.05 spol nezavisan od izvrsnosti zdravstvenog stanja (varijabla `healthexcellent`)? Procijenite distribuciju slučajnog vektora koji modelira ta dva obilježja. Odredite očekivane frekvencije pod pretpostavkom da su obilježja nezavisna. Procijenite uvjetne distribucije te na osnovu njih zaključite koji spol ima više natprosječno zdravih.
- (g) Procijenite očekivanu dob 95%-tnim pouzdanim intervalom.

ZADATAK 3: [10 bodova]

70 osoba testirano je može li napraviti 10 sklekova. Njih 30 je moglo dok ostali nisu uspjeli. Isti test ponovljen je i nakon dva mjeseca treninga. Svi koji su uspjeli prvi put, uspjeli su i u ponovljenom testu. Od onih koji nisu uspjeli prvi put, njih 6 je uspjelo na drugom testu dok ostali nisu ni tada. Povećava li trening sposobnosti na razini značajnosti 0.05?

ZADATAK 4: [10 bodova]

Novčić bacamo dok ne padne pismo, ali najviše četiri puta. Ponavljanjem ovog pokusa 100 puta utvrđeno je da je 49 puta novčić bačen jednom, 13 puta je bačen dva puta, 11 puta je bačen tri puta i 27 puta je bačen četiri puta. Možemo li tvrditi na razini značajnosti 0.05 da je novčić nepravilan?

ZADATAK 5: [8+5=13 bodova]

U R tablici podataka `airquality` u stupcu `Wind` nalaze se podaci o izmjerenim brzinama vjetra na jednoj lokaciji u New Yorku i to u miljama na sat.

- (a) Procijenite parametre linearnog modela temperature (varijabla `Temp`) u ovisnosti o brzini vjetra (varijabla `Wind`). Jesu li koeficijenti modela statistički značajno različiti od 0 na razini značajnosti 0.05? Nacrtajte procijenjeni regresijski pravac zajedno s točkama podataka. Prema dobivenom modelu, kako porast brzine vjetra za 3 milje na sat mijenja temperaturu?
- (b) Provjerite imamo li razloga sumnjati u nezavisnost grešaka dobivenog modela. Komentirajte.