

STATISTIČKI PRAKTIKUM

PISMENI ISPIT 13.9.2016.

Napomena: Rješenja trebate kopirati u Word dokument. Komentare možete pisati u isti dokument ili na papir. Po završetku pisanja, trebate poslati dokument na `statistika@mathos.hr` i u predmetu poruke napisati svoje ime i prezime.

ZADATAK 1: [5+5+5=15 bodova]

- Neka je (X_1, \dots, X_n) jednostavan slučajan uzorak takav da je $P(X_1 = -1) = 1 - p$ i $P(X_1 = 1) = p$, $p \in (0, 1)$. Koristeći centralni granični teorem definirajte test statistiku koja će asimptotski imati standardnu normalnu distribuciju.
- 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?
- Objasnite test statistiku i način provođenja z -testa za testiranje jednakosti očekivanja normalne distribucije unaprijed zadanoj vrijednosti.

ZADATAK 2: [5+5+10+10+10+10+10=60 bodova]

U R bazi podataka `iris` nalaze se podaci o duljini listova (`Sepal.Length`), širini listova (`Sepal.Width`), duljini (`Petal.Length`) latica i širini (`Petal.Width`) latica cvijeta iris. U stupcu `Species` nalaze se nazivi vrste cvijeta iris "`setosa`", "`versicolor`" i "`virginica`". Sve testove treba provesti na razini značajnosti 0.05.

- Kakvim slučajnim varijablama možemo modelirati duljinu i širinu listova? Nacrtajte procijenjene funkcije distribucije tih varijabli. Procijenite vjerojatnost da je duljina listova manja ili jednaka 5.
- Testirajte možemo li širinu listova modelirati normalnom slučajnom varijablom. Izradite barem jedan grafički prikaz koji upućuje na to. Ukazuje li kutijasti dijagram na postojanje stršćih vrijednosti?
- Nađite 98%-tni interval pouzdanosti za očekivanu širinu listova. Možemo li tvrditi da je širina manja od 3.5?
- Razlikuje li se očekivane širine listova vrsti "`setosa`" i "`versicolor`" cvijeta iris? Možemo li tvrditi da postoji pozitivan pomak u distribuciji širine listova vrste "`setosa`" u odnosu na "`versicolor`"?
- Je li duljina listova korelirana s duljinom latica? Uz pretpostavku da su obilježja normalno distribuirana, je li očekivana duljina listova veća od očekivane duljine latica?
- Procijenite parametre linearnog modela koji opisuje ovisnost duljine latica o duljini listova. Jesu li koeficijenti modela statistički značajno različiti od nula? Nacrtajte procijenjeni regresijski pravac zajedno s parovima podataka. Prema dobivenom modelu, koliku će duljinu latica imati cvijet s duljinom lista 6?
- Razlikuju li se proporcije cvjetova koji imaju duljinu listova veću od 6.5 među vrstama "`versicolor`" i "`virginica`"?

ZADATAK 3: [13 bodova]

Četiri studije o raku pluća imale su sljedeću strukturu pacijenata: u prvoj studiji od 86 pacijenata 83 su bili pušači, u drugoj 90 od 93 su bili pušači, u trećoj 129 od 136 i u četvrtoj 70 od 82. Razlikuju li se proporcije pušača u sve četiri studije na razini značajnosti 0.05?

ZADATAK 4: [12 bodova]

U jednoj studiji promatrani su parovi 100 jednojajčanih blizanaca od kojih je jedan obolio od raka pluća. U 16 parova oba blizanca puše, dok u 59 parova ne puši niti jedan od blizanaca. Parova blizanaca kod kojih je oboljeli pušač, a zdravi nepušač ima 21. Parova kod kojih je oboljeli ne puši, a zdravi puši ima 4. Je li proporcija pušača među zdravim i oboljelim jednaka ($\alpha = 0.05$)?