

Nejc Zajc

SEMINARSKA NALOGA IZ STATISTIKE

UL FMF, Matematika — univerzitetni študij

2020/21

Pred vami je seminarska naloga iz statistike, ki je sestavni del obveznosti pri tem predmetu. Predavatelj vam je na voljo, če potrebujete nasvet. Morda boste morali uporabiti kakšno različico statistične metode, ki je na predavanjih ali vajah nismo omenili. Lahko si pomagate z učbenikom:

John Rice: *Mathematical Statistics & Data Analysis*, Duxbury, 2007,

ali katero drugo knjigo. V primeru težav z dostopom do učbenika se oglasite pri predavatelju.

Pri določenih nalogah si boste morali pomagati z računalnikom. Pri teh prosim priložite tako program ali datoteko kot tudi izhod (numerične rezultate, grafikone ...). Vsaj izhode programov prosim sproti prilagajate k rešitvam posameznih nalog: vse skupaj sestavite v enotno PDF datoteko ali pa preprosto natisnite. Prosim tudi, da izvozite izhod (še zlasti grafikone) iz programov za obdelavo preglednic (recimo excel, če ga boste že uporabili). Datoteke z besedili nalog ne pošiljajte nazaj.

Če stopnja tveganja pri preizkusu ni navedena, morate preizkusiti tako pri $\alpha = 0.01$ kot tudi pri $\alpha = 0.05$.

Veliko uspeha pri reševanju!

1. V datoteki *Kibergrad* se nahajajo informacije o 43.886 družinah, ki stanujejo v mestu *Kibergrad*. Mesto ima štiri četrti: v severni četrti stanuje 10.149 družin, v vzhodni 10.390, v južni 13.457 in v zahodni 9.890. Za vsako družino so zabeleženi naslednji podatki (ne boste potrebovali vseh):

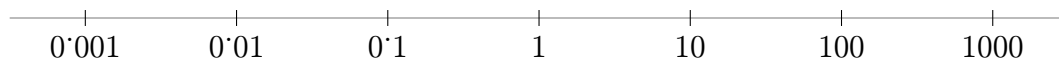
- Tip družine (od 1 do 3)
- Število članov družine
- Število otrok v družini
- Skupni dohodek družine
- Četrt, v kateri stanuje družina:
 - 1: Severna
 - 2: Vzhodna
 - 3: Južna
 - 4: Zahodna
- Stopnja izobrazbe vodje gospodinjstva (od 31 do 46)

Vzemite enostavni slučajni vzorec 400 enot.

- a) Na podlagi vzorca ocenite število otrok na družino. Ocenite še standardno napako vaše ocene in postavite 95% interval zaupanja.
 - b) Ali pri oceni povprečnega števila otrok pomaga, če stratificiramo po četrtih? Izvedite prejšnjo točko na stratificiranem vzorcu s proporcionalno alokacijo. Primerjajte!
2. V datoteki *ZarkiGama* se nahajajo podatki o časovnih razmikih med 3.935 zaznanimi fotoni, torej medprihodni časi (v sekundah).
- a) Naredite histogram medprihodnih časov. Se vam zdi, da je model s porazdelitvijo gama plavzibilen?
 - b) Ocenite parametra porazdelitve gama po metodi momentov in po metodi največjega verjetja. Primerjajte!
Namig: potrebovali boste funkcijo *digama*, ki je logaritemski odvod funkcije gama. Preberite kaj o njej recimo na wikipediji.
Namig: sistema enačb ne boste mogli rešiti eksaktno, to boste morali narediti numerično. Ena od učinkovitih možnosti je večrazsežna Newtonova metoda.
 - c) Ocenjeni porazdelitvi dorišite na histogram. Je videti razumno?
 - d) Histogram z dorisanimi gostotama narišite še na logaritemski lestvici. Lestvico transformirajte le na abscisni osi, vendar pa ustrezno transformirajte tudi dorisani gostoti.
 - e) Je porazdelitev medprihodnih časov videti konsistentna s Poissonovim modelom, po katerem so ti časi porazdeljeni eksponentno?

Pri histogramih združite čase oz. njihove logaritme v enako široke razrede. Širino posameznega razreda določite v skladu z modificiranim Freedman–Diaconisovim pravilom.

Logaritemska lestvica pomeni, da položaj ustreza logaritmu, oznaka pa izvirni vrednosti, npr.:



3. V datoteki **Zobje** se nahajajo podatki o dolžini zob morskih prašičkov, ki so jim dodajali vitamin C v različnih količinah na dva različna načina: bodisi neposredno (kar je zakodirano z VC) bodisi s pomarančnim sokom (kar je zakodirano z OJ).
 - a) Preizkusite, ali dodajanje vitamina C vpliva na rast zob.
 - b) Kateri način dodajanja je učinkovitejši? Preizkusite, ali je razlika statistično značilna.