

**Žiga Gladek**

## 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 pomagata 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 prilagajte 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*. 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
- Mestna četrt, v kateri stanuje družina (od 1 do 4)
- Stopnja izobrazbe vodje gospodinjstva:
  - 31: Brez šolske izobrazbe
  - 32: Dokončan prvi, drugi, tretji ali četrti razred osnovne šole
  - 33: Nedokončana osnovna šola, a končanih vsaj pet razredov
  - 34: Dokončana osnovna šola
  - 35: Dokončan prvi letnik srednje šole
  - 36: Dokončan drugi letnik srednje šole
  - 37: Dokončan tretji letnik srednje šole
  - 38: Dokončan četrti letnik srednje šole, a brez mature
  - 39: Poklicna matura
  - 40: Splošna matura
  - 41: Dokončan višji strokovni študij
  - 42: Dokončan visoki strokovni študij
  - 43: Dokončan univerzitetni študij prve stopnje
  - 44: Dokončan univerzitetni študij druge stopnje (magisterij)
  - 45: Magisterij po starem programu
  - 46: Doktorat znanosti

- a) Vzemite enostavni slučajni vzorec 200 družin in na njegovi podlagi ocenite delež družin v Kibergradu, v katerih vodja gospodinjstva nima srednješolske izobrazbe, tj. niti poklicne niti splošne mature.
- b) Ocenite standardno napako in postavite 95% interval zaupanja.
- c) Vzorčni delež in ocenjeno standardno napako primerjajte s populacijskim deležem in pravo standardno napako. Ali interval zaupanja pokrije populacijski delež?
- d) Vzemite še 99 enostavnih slučajnih vzorcev in prav tako za vsakega določite 95% interval zaupanja. Narišite intervale zaupanja, ki pripadajo tem 100 vzorcem. Koliko jih pokrije populacijski delež?
- e) Izračunajte standardni odklon vzorčnih deležev za 100 prej dobljenih vzorcev. Primerjajte s pravo standardno napako za vzorec velikosti 200.
- f) Izvedite prejšnji dve točki še na 100 vzorcih po 800 družin. Primerjajte in razložite razlike s teorijo vzorčenja.

2. V datoteki **TempPulz** se nahajajo odčitki telesnih temperatur (v Fahrenheitovih stopinjah) ter pulzov 65 moških (kodiranih z 1) in 65 žensk (kodiranih z 2). Privzemite, da sta telesna temperatura in pulz tako pri moških kot pri ženskah porazdeljena normalno.
- Ocenite povprečje in standardni odklon za telesno temperaturo posebej pri moških in posebej pri ženskah.
  - Za povprečji iz prejšnje točke določite 95% intervala zaupanja.
  - Preizkusite domnevo, da imajo moški in ženske v povprečju enako telesno temperaturo.

*Pretvornik med Fahrenheitovimi in Celzijevimi stopinjami:*  $x^{\circ}\text{F} = y^{\circ}\text{C}$ , če je  $y = 5(x - 32)/9$ .

Vir podatkov: A. L. Shoemaker: What's normal? Temperature, gender, and heart rate. *J. Stat. Edu.* **3**, št. 2 (1996).

3. V neki raziskavi:

<http://www.statsci.org/data/oz/ms212.html>

so študentom merili pulz. Vsakemu študentu so pulz izmerili dvakrat. Določeni so imeli med obema meritvama fizično obremenitev (tek na mestu), določeni ne. Podatki so zbrani v tabeli **Pulz**, pri čemer imajo stolpci naslednje pomene:

VISINA	telesna višina
TEZA	telesna teža
STAROST	starost v letih
SPOL	1=moški, 2=ženski
KADI	1=kadilec, 2=nekadilec
ALKOHOL	1=pije, 2=ne pije
VADBA	1=vadi veliko, 2=vadi zmerno, 3=vadi malo ali pa sploh ne
OBREMENITEV	1=obremenitev, 2=brez obremenitve
PULZ1	prva meritev pulza
PULZ2	druga meritev pulza
LETO	leto meritve (1993–1998)

- Preizkusite, ali to, ali so bili študenti deležni obremenitve, vpliva na spremembo pulza.
- Se zdi, da so določeni študenti, ki so bili določeni za obremenitev, morda goljufali in sploh niso tekli? Ilustrirajte s primernim grafičnim prikazom.
- Pri študentih, ki so bili deležni obremenitve, preizkusite, ali vadba vpliva na spremembo pulza.