

Gal Anton Gorše

PROJEKTNA NALOGA IZ STATISTIKE

UL FMF, Matematika — univerzitetni študij

2022/23

Pred vami je projektna 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.

Rešeno nalogo prosim oddajte v ustrezno rubriko na Učilnici v formatu ZIP. Tam naj bo zapakirana datoteka z imenom `Projektna_naloga.pdf`, v mapi `Priloge` pa naj bodo pomožne datoteke, npr. programi, s katerimi ste dobili rezultate. Toda v glavni datoteki morajo biti sproti vključeni vsi rezultati in grafikoni: imejte v mislih, naj, če je vse prav, pomožne datoteke ne bodo potrebne. Datoteke z besedili nalog ne oddajajte.

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

Rok oddaje je **ponedeljek, 4. september 2023**. Veliko uspeha pri reševanju!

NEKAJ NAPOTKOV ZA STAVLJENJE V T_EX-u oz. L^AT_EX-u

- Spremenljivke se dosledno stavijo ležeče, v T_EX-u torej med dolarji. Tako morate staviti, tudi če formula vsebuje en sam znak.
- Operatorji se stavijo pokončno, kar pa ne pomeni, da jih v T_EX-u postavimo kar izven dolarjev. Za najpogostejše operatorje so že naprogramirani ukazi.
- Če operator še ni definiran, ga sicer lahko stavimo recimo kot `\mathop{\mathrm{var}}` (ukaz `\mathop` je pomemben zaradi presledkov), a bistveno lažje je, če definiramo ukaz, recimo v preambuli:

```
\usepackage{amsmath}
\DeclareMathOperator{\var}{var}
```

- Dele formul je dostikrat smiselno ločiti z dodatnimi presledki. Temu so namenjeni ukazi `\,`, `\;`, `\>`, `\quad` in `\qquad`.
- Formule, ki so predolge za eno vrstico, je treba razlomiti. Najpogosteje se to naredi z uporabo okolij `array`, `align`, `align*`, `gather`, `gather*` in `split` (slednje znotraj okolja `equation` ali `equation*`). Za vse razen prvega potrebujemo knjižnico `amsmath`.
- Grafikone postavite **natančno** na mesto, kamor sodijo. Za to recimo v okolju `figure` uporabite določilo `H` (ne `h`), pri tem pa je treba v preambulo dati `\usepackage{float}`.
- Če boste decimalno vejico stavili kot običajno vejico, recimo `23,6`, vam bo T_EX naredil presledek, torej `23, 6`, ker bo mislil, da gre za naštevaje. Rešitev: `23{,}6`.

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 (od 31 do 46)

- (a) Vzemite enostavni slučajni vzorec 200 družin in na njegovi podlagi ocenite povprečno število otrok na družino v Kibergradu.
- (b) Ocenite standardno napako in postavite 95% interval zaupanja.
- (c) Vzorčno povprečje in ocenjeno standardno napako primerjajte s populacijskim povprečjem in pravo standardno napako. Ali interval zaupanja iz prejšnje točke pokrije populacijsko povprečje?
- (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 populacijsko povprečje?
- (e) Izračunajte standardni odklon vzorčnih povprečij 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 **Mangan** so podatki o deležu mangana v železu, pridobljenem v plavžu: skozi 24 dni so vsak dan analizirali pet odlitkov. Preučite normalnost dobljene empirične porazdelitve, tako da narišete:

- histogram z dorisano ustrezno normalno gostoto;
- viseči histogram razlik korenov frekvenc: glejte razdelek 9.7 v knjigi;
- primerjalni kvantilni (Q–Q) grafik: glejte razdelek 9.8 v knjigi.

Pri histogramu z dorisano normalno gostoto (ne pa tudi pri visečem histogramu) združite deleže mangana v razrede. Širino posameznega razreda določite v skladu z modificiranim Freedman–Diaconisovim pravilom.

Vir podatkov: I. Burr: *Applied Statistical Methods*. Academic Press, New York, 1974.

3. Odontoblasti so celice, iz katerih nastaja zobovina (dentin). V neki študiji so izmerili maksimalno dolžino odontoblastov pri 60 morskih prašičkih, ki so jim dodajali vitamin C v različnih količinah na dva različna načina: bodisi neposredno kot askorbinsko kislino (kar je zakodirano z VC) bodisi s pomarančnim sokom (kar je zakodirano z OJ).

- (a) Preizkusite, ali dodajanje vitamina C vpliva na dolžino odontoblastov.
- (b) Kateri način dodajanja je učinkovitejši? Preizkusite, ali je razlika statistično značilna.