

**Rok Strah**

## PROJEKTNA NALOGA IZ STATISTIKE

UL FMF, Matematika — univerzitetni študij

2021/22

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 PDF pod imenom `Projektna_naloga.pdf`.

Pri določenih nalogah si boste morali pomagati z računalnikom. Pri teh prosim priložite tako program ali datoteko kot tudi njegov izhod (numerične rezultate, grafikone ...). Vsaj izhode programov pa prosim še **sproti** prilagajte k rešitvam posameznih nalog v glavni datoteki. Na ta način prosim tudi priložite da izvozite izhod (še zlasti grafikone) programov za obdelavo preglednic (recimo excel, če ga boste že uporabili). Datoteke z besedili nalog ne oddajajte.

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

Veliko uspeha pri reševanju!

## NEKAJ NAPOTKOV ZA STAVLJENJE V T<sub>E</sub>X-u oz. L<sup>A</sup>T<sub>E</sub>X-u

- Spremenljivke se dosledno stavijo ležeče, v T<sub>E</sub>X-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<sub>E</sub>X-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<sub>E</sub>X naredil presledek, torej `23, 6`, ker bo mislil, da gre za naštevanje. Rešitev: `23{,}6`.

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)

Iz vsake četrti vzemite enostavni slučajni vzorec velikosti 100.

- a) Primerjajte dohodke po četrtih, tako da narišete vzporedne škatle z brki (glejte razdelek 10.6 v knjigi). Je videti, da so določenih četrtih dohodki višji?
- b) Iz severne četrti vzemite še štiri enostavne slučajne vzorce velikosti 100. Za vseh pet vzorcev iz severne četrti spet narišite vzporedne škatle z brki. Komentirajte!
- c) Za celotni Kibergrad izračunajte varianco dohodka, pojasnjeno s četrtmi, in preostalo (rezidualno) varianco. Kako se to ujema z opažanji od prej?

2. Pri najlonskih palicah so preizkušali lomljivost (Bennett in Franklin, 1954). V podobnih okoliščinah so ulili 280 palic in vsako od njih preizkusili na petih mestih. Rezultati poskusa so prikazani v tabeli na desni.

Če ima palica enakomerno strukturo, bi morale biti število mest, na katerih se je zlomila, porazdeljeno binomsko  $\text{Bin}(5, p)$  za določen neznan  $p$ . To naj bo naša osnovna ničelna domneva. Privzamemo tudi, da so palice med seboj neodvisne.

št. lomov	št. palic
0	157
1	69
2	35
3	17
4	1
5	1

- a) Ob predpostavki osnovne ničelne domneve ocenite  $p$  po metodi največjega verjetja.
- b) Združite zadnje tri vrednosti in s posplošenim Pearsonovim preizkusom hi kvadrat preizkusite osnovno ničelno domnevo proti alternativni domnevi, da ima število lomov katero drugo porazdelitev (glejte razdelek 9.5 v knjigi). Še vedno privzamemo, da ima število lomov na vseh palicah enako porazdelitev.

- c) Za  $i = 1, 2, \dots, n$  naj bodo dana neodvisna opažanja  $X_i \sim \text{Bin}(m_i, p_i)$ , kjer so parametri  $m_i$  znani, parametri  $p_i$  pa neznani. Razvijte preizkus na podlagi razmerja verjetij, ki bo preizkusil ničelno domnevo, da so vsi parametri  $p_i$  enaki, proti alternativni domnevi, da temu ni tako.
- d) Uporabite preizkus iz prejšnje točke na danih podatkih, vedite pa, da Wilksovega izreka ne morete uporabiti (premislite, zakaj pogoji niso izpolnjeni). Namesto tega uporabite metodo *bootstrap*: simulirajte 10.000 vrednosti preizkusne statistike pri ničelni domnevi, pri čemer za  $p$  vzemite oceno iz točke a). Nato pogledajte, koliko teh vrednosti presega vrednost preizkusne statistike, izračunano na konkretnih podatkih. Na podlagi tega ustrezno sklepajte, kaj storiti z ničelno domnevo.

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)

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