

Seminarska naloga iz Statistike

Dejan Perić

5. december 2021

1 Prva naloga

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)

1.1 a)

Vzemite enostavni slučajni vzorec 200 družin in na njegovi podlagi ocenite povprečno število otrok na družino v Kibergradu.

1.2 b)

Ocenite standardno napako in postavite 95% interval zaupanja.

1.3 c)

Vzorčno povprečje in ocenjeno standardno napako primerjajte s populacijskim povprečjem in pravo standardno napako. Ali interval zaupanja iz prejšnje točke pokrije populacijsko povprečje?

1.4 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?

1.5 e)

Izračunajte standardni odklon vzorčnih povprečij za 100 prej dobljenih vzorcev. Primerjajte s pravo standardno napako za vzorec velikosti 200.

1.6 f)

Izvedite prejšnji dve točki še na 100 vzorcih po 800 družin. Primerjajte in razložite razlike s teorijo vzorčenja.

2 Druga naloga

2.1 a)

Ocenite povprečje in standardni odklon za telesno temperaturo posebej pri moških in posebej pri ženskah.

2.2 b)

Za povprečji iz prejšnje točke določite 95% intervala zaupanja.

2.3 c)

Preizkusite domnevo, da imajo moški in ženske v povprečju enako telesno temperaturo.

3 Tretja naloga

V datoteki Temp LJ se nahajajo izmerjene mesečne temperature v Ljubljani v letih od 1986 do 2020. Postavimo naslednja dva modela spreminjanja temperature s časom:

- Model A: vključuje linearni trend in sinusno nihanje s periodo eno leto.

- Model B: vključuje linearni trend in spreminjanje temperature za vsak mesec posebej.

Očitno je model B širši od modela A.

3.1 a)

Preizkusite model A znotraj modela B.

3.2 b)

Pri modeliranju je nevarno privzeti preširok model: lahko bi recimo postavili model, po katerem je temperatura vsak mesec drugačna, neidvisno od ostalih mesecev, a tak model bi bil neuporaben za napovedovanje. Akaikejeva informacija nam pomaga poiskati optimalni model – izberemo tistega, za katerega je le-ta najmanjša. Akaikejeva informacija je sicer definirana z verjetjem, a pri linearni regresiji in Gaussovem modelu je le-ta ekvivalentna naslednji modifikaciji:

$$AIC := 2m + n \ln \text{RSS},$$

kjer je m število parametrov, n pa je število opažanj. Kateri od zgornjih dveh modelov ima manjšo Akaikejevo informacijo?

4 Literatura