

POROČILO SEMINARSKE NALOGE

STATISTIKA

SARA BIZJAK

UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO
ODDELEK ZA MATEMATIKO

JULIJ 2020

1. NALOGA

Podatki so vzeti iz datoteke `Kibergard`, kjer se nahajajo informacije o 43.886 družinah, ki stanujejo v mestu Kibergard. Za vsako družino so zabeleženi naslednji podatki:

- 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 in vodje gospodinjstva (od 31 do 46: opisi v datoteki `z navodili`)

Nalogo sem reševala s pomočjo programa R. Koda, uporabljena za generiranje enostavnih slučajnih vzorcev in izračune, je dostopna v priloženi datoteki *naloga1.R*.

PRIMER A

Vzamemo enostavni slučajni vzorec 200 družin in na njegovi podlagi ocenimo delež družin v Kibergardu, v katerih vodja gospodinjstva nima srednješolske izobrazbe (niti poklicne niti splošne mature). Opisan delež znaša $p = 0.195$.

PRIMER B

Ocenimo standardno napako in postavimo 95% interval zaupanja. Standardno napako za delež izračunamo po formuli

$$\hat{se}(p) = \sqrt{\frac{p \cdot (1 - p)}{n - 1} \cdot \left(1 - \frac{n}{N}\right)},$$

kjer so $p = 0.195$, $n = 200$, $N = 43.886$.

Dobimo rezultat $\hat{se}(p) = 0.02802185$.

Interval zaupanja je enak: $[0.1400782, 0.2499218]$.

PRIMER C

Vzorčni delež in ocenjeno standardno napako primerjamo s populacijskim deležem in pravo standardno napako.

- Vzorčni delež: 0.195
- Populacijski delež: 0.2115025
- Razlika obeh deležev: 0.01650253
- Ocenjena standardna napaka (iz vzorca): 0,02802185
- Prava standardna napaka (iz celotne populacije): 0.02888282
- Razlika med ocenjeno in pravo standardno napako: 0.0008609634

Ker velja $0.2115025 \in [0.1400782, 0.2499218]$, interval zaupanja pokrije populacijski delež.

PRIMER D

GRAF

INTERVALI ZAUPANJA + KOLIKO JIH POKRIJE POPULACIJSKI DELEŽ

PRIMER E

Standardni odklon vzorčnih deležev za 100 prej dobljenih vzorcev je enak 0.02881085. Prava standardna napaka za vzorec velikosti 200 pa je 0.02888282. Razlikujeta se za $7.196778 \cdot 10^{-5}$.

PRIMER F

GRAF

INTERVALI ZAUPANJA + KOLIKO JIH POKRIJE POPULACIJSKI DELEŽ

2. NALOGA

3. NALOGA

4. NALOGA

5. NALOGA