

Projektna naloga iz statistike

Andraž Čepič

2. 6. 2022

V projektu ves čas uporabljamo Python s paketi Pandas, NumPy, Jupyter in Matplotlib.

Naloga 1

V namen obdelave podatkov smo napisali Jupyter zvezek `kibergrad.ipynb`. Za začetek naložimo podatke iz datoteke `Kibergrad.csv` v Pandas DataFrame objekt.

(a)

Izberemo enostavni slučajni vzorec velikosti 200 z vgrajeno funkcijo paketa Pandas, in sicer `pandas.DataFrame.sample`. Če so

$$X_1, \dots, X_{200}$$

števila otrok vsake od vzorčenih družin, je primerna ocena za povprečje enaka

$$\bar{X} = \frac{X_1 + \dots + X_{200}}{200}.$$

Za naš specifičen vzorec dobimo oceno za povprečno število otrok v mestu Kibergrad:

$$\bar{X} = 0,955$$

(b)

Ocena za standardno napako je podana s formulo

$$\widehat{\text{SE}}^2 = \frac{N-1}{N} \cdot \frac{1}{n(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2,$$

kjer je N velikost populacije in n velikost enostavnega slučajnega vzorca. V našem primeru je $N = 43.886$ in $n = 200$.