
Dodatak za §1.4. – Uzorak u Pythonu

Excel je alat prikladan za brzi pregled podataka i izradu grafova. Za zahtjevnuj obradu podataka često koristimo programske jezike. U praksi se kao alati za obradu podataka koriste R, Python, Matlab i neki drugi popularni alati. Kako se Python koristi u nastavi informatike (vidi primjerice knjigu *Izrada primjenskih programa u Pythonu : priručnik za prirodoslovno-matematičke gimnazije* ([18])), prikazat ćemo kao napraviti 1 000 000 simulacija reprezentativnog i slučajnog uzorka te provjeriti jesu li pojedina svojstva izvornih podataka ostala sačuvana u uzorku.

Idejno, algoritam izgleda ovako:

```
ulaz: podaci, veličina_uzorka, broj_eksperimenata;
za j = 1 ... broj_eksperimenata
    uzmi uzorak iz danih podataka veličine veličina_uzorka;
    provjeri i zabilježi svojstva na uzorku;
za svako svojstvo
    ispiši  $\frac{\text{koliko je puta svojstvo bilo zadovoljeno}}{\text{broj eksperimenata}}$ ;
```

Primjer implementacije u Pythonu / Jupyter bilježnici izgleda ovako:

```
[1]: # Uvoz paketa za generiranje slučajnih brojeva
import random

duljina_uzorka = 30
broj_eksperimenata = 100000

# Brojači
#Koliko puta je broj 50 bio najčešći broj u uzorku
najvise_broj_50 = 0
```

```
[2]: # Učitavanje podataka
with open('otkucaji.txt', encoding='utf-8') as podaci_izvor:
    podaci = [int(x.strip()) for x in podaci_izvor if x.strip().
               isdigit()]

duljina_podataka = len(podaci)
print(f"Duljina podataka {duljina_podataka}")
```

Duljina podataka 100

```
[3]: for _ in range(broj_eksperimenata):
    # Reprezentativni uzorak (bez ponavljanja)
    uzorak = random.sample(podaci, duljina_uzorka)
```

```
# Ako želiš slučajni uzorak s ponavljanjem, koristi:  
# uzorak = [random.choice(podaci) for _ in  
range(duljina_uzorka)]  
  
# Broj pojavljivanja svake vrijednosti u uzorku  
vrijednosti_pojavljivanja = {vrijednost: uzorak.  
count(vrijednost) for vrijednost in set(uzorak)}  
  
# Provjeri je li broj 50 najčešći u uzorku  
if vrijednosti_pojavljivanja.get(50, 0) ==  
max(vrijednosti_pojavljivanja.values()):  
    najvise_broj_50 += 1  
  
print("Broj eksperimenata gdje je broj 50 najčešći: {} {:.2f}%"  
.format(najvise_broj_50, najvise_broj_50 * 100/  
broj_eksperimenata))
```

Broj eksperimenata gdje je broj 50 najčešći: 59494 (59.49 %)