nal1

August 17, 2022

```python
import numpy as np
import matplotlib.pyplot as plt
import csv
import random
```
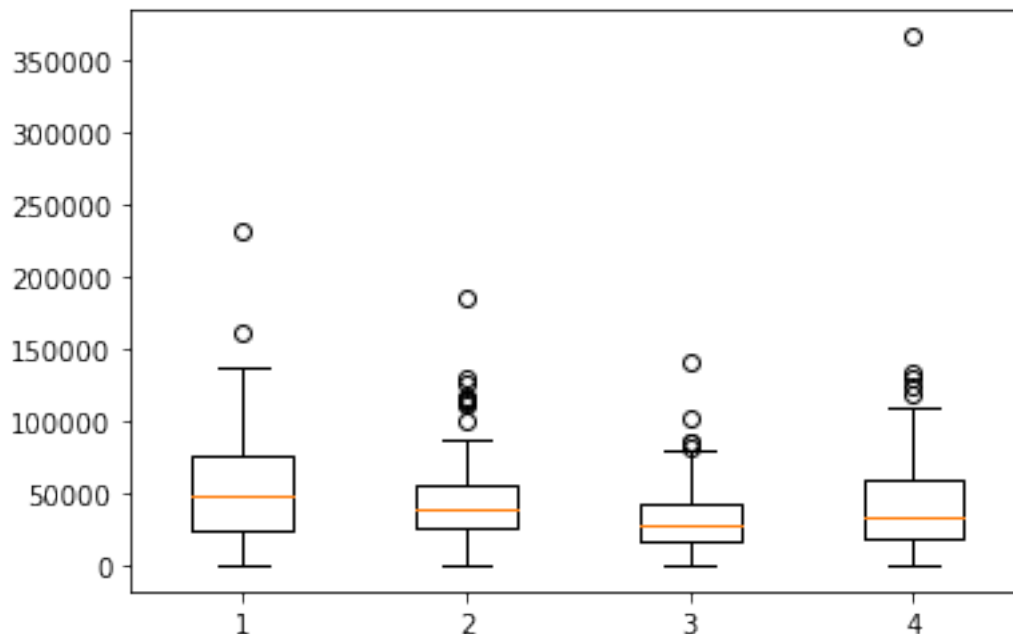
```python
with open('../data/Kibergrad.csv', 'r') as f:
    f.readline()
    reader = csv.reader(f)
    kibergrad = list(tuple(map(int, line)) for line in reader)

kibergrad_by_square = tuple(list(filter(lambda x: x[4] == i+1, kibergrad)) for
 ↪i in range(4))
```

```python
sample = tuple(random.choices(population=sq, k=100) for sq in
 ↪kibergrad_by_square)
```

```python
plt.boxplot(x=tuple(list(e[3] for e in sq) for sq in sample))
plt.show()
```

a) Sicer vidimo neke manjše razlike (dohodki so malce nižji v južni četrti in malce višji v severni), ampak menim, da četrti ne vplivajo bistveno na dohodek.

```python
sample2 = (sample[0], *(random.choices(population=kibergrad_by_square[0],
    k=100) for _ in range(4)))
```

```python
plt.boxplot(x=tuple(list(e[3] for e in sq) for sq in sample2))
plt.show()
```
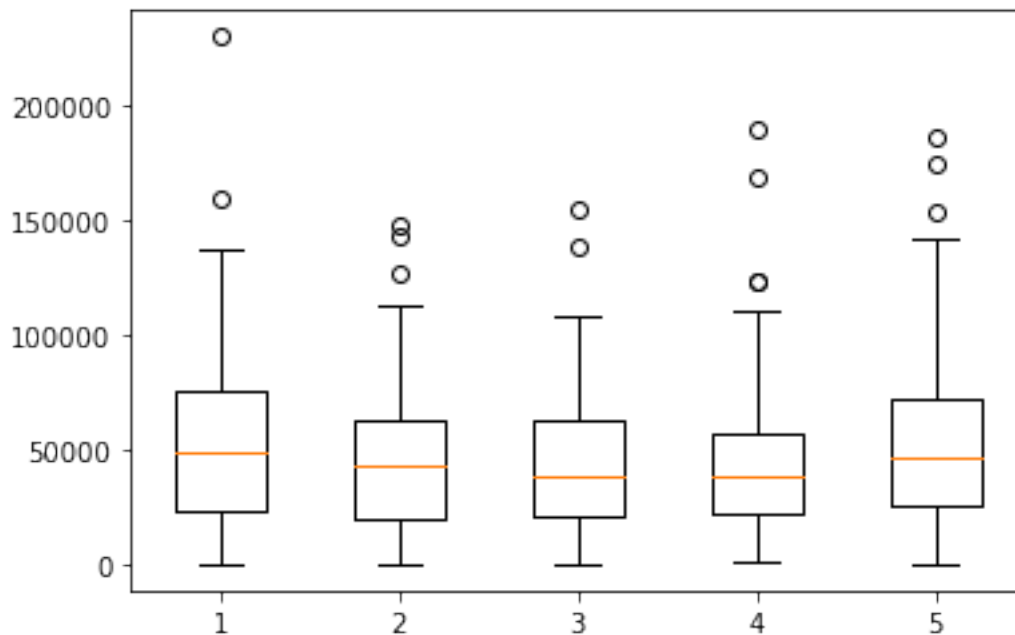

b) Graf zgleda približno kot zgornji, kar še bolj utemelji moje mnenje, da četrti ne vplivajo na dohodek.

```python
incomes_by_square = tuple(list(e[3] for e in sq) for sq in kibergrad_by_square)
explained_var = np.var(list(np.mean(sq) for sq in incomes_by_square))
residual_var = np.mean(list(np.var(sq) for sq in incomes_by_square))
total_var = np.var(list(e  for sq in incomes_by_square for e in sq))
print(f"explained variance: {explained_var}, residual variance: {residual_var}")
print(f"total variance: {total_var}, explained+residual var:
    {explained_var+residual_var}")
```

explained variance: 8688785.712368913, residual variance: 1027977659.8502393
total variance: 1026385669.8903824, explained+residual var: 1036666445.5626082

c) Kot vidimo, četrti res ne vplivajo bistveno na dohodek.