

**Міністерство освіти і науки України
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Факультет інформатики та обчислювальної техніки
Кафедра обчислювальної техніки**

Лабораторна робота №2
з дисципліни
«Аналіз даних з використанням мови Python»

Виконав:
студент групи ПІ-04
Пащенко Дмитро Олексійович

Перевірила:
Тимофєєва Ю. С.

Київ 2022

Код програми та результат виконання

```
In [159]: import numpy as np
import pandas as pd
import scipy.stats as stats
```

```
In [160]: data = pd.read_csv('Birthweight.csv')
data
```

29	752	49	3.32	36	40	1	27	12	152	48	37	12	25	170	0	0
30	1023	52	3.00	35	38	1	30	12	165	64	38	14	50	180	0	0
31	57	51	3.32	38	39	1	23	17	157	48	32	12	25	169	0	0
32	1522	50	2.74	33	39	1	21	17	156	53	24	12	7	179	0	0
33	223	50	3.87	33	45	1	28	25	163	54	30	16	0	183	0	0
34	272	52	3.86	36	39	1	30	25	170	78	40	16	50	178	0	0
35	27	53	3.55	37	41	1	37	25	161	66	46	16	0	175	0	1
36	365	52	3.53	37	40	1	26	25	170	62	30	10	25	181	0	0
37	619	52	3.41	33	39	1	23	25	181	69	23	16	2	181	0	0
38	1369	49	3.18	34	38	1	31	25	162	57	32	16	50	194	0	0
39	1262	53	3.19	34	41	1	27	35	163	51	31	16	25	185	0	0
40	516	47	2.66	33	35	1	20	35	170	57	23	12	50	186	1	0
41	1272	53	2.75	32	40	1	37	50	168	61	31	16	0	173	0	1

1. Знайти середній вік матерів і батьків і порівняти ці середні значення.

```
In [161]: mother_ages = np.array(data['mage'])
mother_ages
```

```
Out[161]: array([20, 19, 35, 20, 24, 29, 24, 21, 18, 20, 27, 24, 31, 21, 29, 28, 27,
                26, 19, 24, 20, 22, 19, 22, 20, 41, 20, 32, 31, 27, 30, 23, 21, 28,
                30, 37, 26, 23, 31, 27, 20, 37])
```

```
In [162]: father_ages = np.array(data['fage'])
father_ages
```

```
Out[162]: array([23, 19, 31, 26, 30, 31, 31, 21, 20, 22, 37, 29, 35, 25, 30, 39, 27,
                25, 23, 26, 24, 24, 20, 23, 20, 37, 20, 38, 41, 37, 38, 32, 24, 30,
                40, 46, 30, 23, 32, 31, 23, 31])
```

```
In [163]: mothers_avarage_age = np.mean(mother_ages)
mothers_avarage_age
```

```
Out[163]: 25.547619047619047
```

```
In [164]: fathers_avarage_age = np.mean(father_ages)
fathers_avarage_age
```

```
Out[164]: 28.904761904761905
```

```
In [165]: mothers_avarage_age < fathers_avarage_age
```

```
Out[165]: True
```

2. Перевірити чи нормально розподілена вага дітей.

```
In [166]: kid_weights = np.array(data['Birthweight'])
kid_weights
```

```
Out[166]: array([4.55, 4.32, 4.1 , 4.07, 3.94, 3.93, 3.77, 3.65, 3.63, 3.42, 3.35,
                3.27, 3.23, 3.2 , 3.15, 3.11, 3.03, 2.92, 2.9 , 2.65, 3.64, 3.14,
                2.78, 2.51, 2.37, 2.05, 1.92, 4.57, 3.59, 3.32, 3. , 3.32, 2.74,
                3.87, 3.86, 3.55, 3.53, 3.41, 3.18, 3.19, 2.66, 2.75])
```

```
In [167]: stat, p = stats.normaltest(kid_weights)
print('Stats = %.3f, p = %.3f' % (stat, p))
```

```
Stats = 0.105, p = 0.949
```

```
In [168]: "Вага нормально розподілена" if p > .05 else "Вага нормально не розподілена"
```

```
Out[168]: 'Вага нормально розподілена'
```

3. Перевірити за допомогою статистичних гіпотез чи у матерів, що палять, легші діти.

- **H0 (основна гіпотеза):** У матерів, що палять, діти мають однакову вагу, що й у матерів, що не палять
- **H1 (альтернативна гіпотеза):** У матерів, що палять, діти легші, ніж у матерів, що не палять

```
In [169]: kids_smoke = kid_weights[20:40]
kids_smoke
```

```
Out[169]: array([3.64, 3.14, 2.78, 2.51, 2.37, 2.05, 1.92, 4.57, 3.59, 3.32, 3.
3.32, 2.74, 3.87, 3.86, 3.55, 3.53, 3.41, 3.18, 3.19])
```

```
In [170]: kids_no_smoke = kid_weights[0:20]
kids_no_smoke
```

```
Out[170]: array([4.55, 4.32, 4.1 , 4.07, 3.94, 3.93, 3.77, 3.65, 3.63, 3.42, 3.35,
3.27, 3.23, 3.2 , 3.15, 3.11, 3.03, 2.92, 2.9 , 2.65])
```

```
In [171]: stat, p = stats.ttest_ind(kids_smoke, kids_no_smoke, alternative = "less")
print('Stats = %.3f, p = %.3f' % (stat, p))
```

```
Stats = -1.793, p = 0.040
```

```
In [172]: "H0 (основна гіпотеза)" if p > .05 else "H1 (альтернативна гіпотеза)"
```

```
Out[172]: 'H1 (альтернативна гіпотеза)'
```

4. Чи є зв'язок між зростом матері та дитини?

- **H0 (основна гіпотеза):** Зріст матері та дитини не корельовані
- **H1 (альтернативна гіпотеза):** Величини корельовані

```
In [173]: mother_heights = np.array(data['mheight'])
mother_heights
```

```
Out[173]: array([162, 171, 172, 174, 175, 165, 157, 165, 172, 157, 164, 168, 164,
155, 167, 158, 162, 167, 165, 149, 170, 160, 165, 159, 163, 166,
161, 173, 163, 152, 165, 157, 156, 163, 170, 161, 170, 181, 162,
163, 170, 168])
```

```
In [174]: kids_heights = np.array(data['Length'])
kids_heights
```

```
Out[174]: array([56, 53, 58, 53, 54, 51, 52, 53, 54, 50, 53, 51, 52, 53, 53, 48, 48,
48, 53, 43, 53, 51, 50, 50, 48, 46, 48, 58, 53, 49, 52, 51, 50, 50,
52, 53, 52, 52, 49, 53, 47, 53])
```

```
In [175]: stat, p = stats.pearsonr(mother_heights, kids_heights)
print('Stats = %.3f, p = %.3f' % (stat, p))
```

```
Stats = 0.485, p = 0.001
```

```
In [176]: "H0 (основна гіпотеза)" if p > .05 else "H1 (альтернативна гіпотеза)"
```

```
Out[176]: 'H1 (альтернативна гіпотеза)'
```