

ПРАКТИЧНА РОБОТА №3

Статистичний аналіз та візуалізація даних з використанням python

Мета роботи: Навчитися проводити статистичний аналіз даних та візуалізувати результати, опанувати методи побудови варіаційних та інтервальних рядів, розподілів частот і відносних частот, емпіричних функцій розподілу, а також полігонів, гістограм та графіків. Освоїти автоматизацію цих процесів засобами Python для ефективного аналізу даних.

ПОРЯДОК ВИКОНАННЯ РОБОТИ

1. Підготовка до виконання роботи:
 - Ознайомитися з теоретичними відомостями.
 - Підготувати необхідні інструменти: зошит для ручних розрахунків, Excel та середовище Python.
2. Виконання завдань (ЗА ВАРИАНТОМ) вручну або засобами Excel:
 - Побудувати варіаційний ряд та розподіл частот для заданої вибірки.
 - Розрахувати ширину інтервалу за правилом Стреджеса та побудувати інтервальний ряд.
 - Побудувати гістограму та полігон частот.
 - Знайти емпіричну функцію розподілу та побудувати її графік.
 - Виконати аналіз групованих даних, побудувати накопичені частоти та графік щільності частот.
3. Написання програми на Python:
 - Створити програму, яка дозволяє вводити дані користувачем.
 - Реалізувати всі необхідні розрахунки (варіаційний ряд, частоти, інтервальний ряд, емпірична функція розподілу тощо).
 - Вивести на екран всі проміжкові обчислення.
 - Побудувати графіки (гістограму, полігон, емпіричну функцію розподілу).
4. Порівняння результатів:
 - Порівняти результати, отримані вручну/у Excel, з результатами, отриманими за допомогою програми на Python.
 - Проаналізувати розбіжності (якщо вони є) та пояснити їх причини.
5. Оформлення звіту:
 - Записати всі розрахунки, таблиці, графіки та висновки.
 - Додати код програми на Python та скріншоти результатів її роботи.

- Написати висновки щодо отриманих результатів.

Примітка:

- Якщо студент виконує завдання вручну, він повинен детально описувати всі розрахунки та будувати графіки на папері.
- Якщо студент виконує завдання у Excel, він повинен навести скріншоти таблиць та графіків з поясненнями.
- Програма на Python обов'язкова для всіх студентів.

Варіанти завдань

Варіант 1

Завдання 1: Варіаційний та інтервальний ряд, частоти та гістограма

Дані для обробки:

7, 4, 4, 8, 12, 12, 12, 7, 8, 12, 8, 12, 4, 12, 12, 4, 12, 4, 12, 12, 6, 9, 5, 3, 6, 6, 9, 3, 5, 6, 9, 5, 6, 6, 9, 6, 9, 6, 6, 6.

1. Побудуйте варіаційний ряд (впорядкуйте дані за зростанням).
2. Складіть розподіл частот та відносних частот.
3. Розрахуйте ширину інтервалу за правилом Стреджеса та побудуйте інтервальний ряд.
4. Створіть гістограму частот.

Завдання 2: Емпірична функція розподілу та полігон частот

Дані для обробки:

2,41; 2,62; 2,73; 2,52; 2,54; 2,41; 2,81; 2,53; 2,64; 2,61; 2,72; 2,45; 2,52; 2,1; 2,64; 2,52; 2,5; 2,33; 2,24; 2,4; 2,72; 2,51; 2,4; 2,61; 2,42; 2,43; 2,65; 2,54; 2,35; 2,54; 2,62; 2,9; 2,75; 2,24; 2,65; 2,45; 2,53; 2,32; 2,24; 2,55.

1. Розрахуйте ширину інтервалу за правилом Стреджеса та складіть інтервальний ряд.
2. Побудуйте полігон частот.
3. Знайдіть емпіричну функцію розподілу та створіть її графік.

Завдання 3: Групуваний розподіл і статистичний аналіз

Дані для обробки:

52, 54, 70, 73, 75, 52, 82, 93, 95, 82, 60, 62, 50, 79, 92, 91, 81, 53, 80, 70, 52, 82, 93, 95, 82, 55, 60, 62, 50, 90, 52, 54, 70, 73, 75, 52, 82, 93, 95, 82, 82, 82, 93, 55, 60, 62, 50, 79, 28, 100.

1. Запишіть впорядкований розподіл частот та відносних частот.
2. Розрахуйте ширину інтервалу за правилом Стреджеса та побудуйте згрупований розподіл частот та відносних частот.
3. Побудуйте згрупований полігон накопичених частот та графік щільності частот.

Варіант 2

Завдання 1: Варіаційний та інтервальний ряд, частоти та гістограма

Дані для обробки:

5, 3, 7, 8, 10, 5, 6, 7, 9, 10, 5, 6, 7, 8, 9, 10, 5, 6, 7, 8, 9, 10, 5, 6, 7, 8, 9, 10, 5, 6, 7, 8, 9, 10, 5, 6, 7, 8, 9, 10, 5, 6, 7, 8, 9, 10, 5, 6, 7, 8, 9, 10, 5, 6, 7, 8, 9, 10.

1. Побудуйте варіаційний ряд (впорядкуйте дані за зростанням).
2. Складіть розподіл частот та відносних частот.
3. Розрахуйте ширину інтервалу за правилом Стреджеса та побудуйте інтервальний ряд.
4. Створіть гістограму частот.

Завдання 2: Емпірична функція розподілу та полігон частот

Дані для обробки:

5.1, 5.3, 5.5, 5.8, 6.0, 6.2, 6.5, 6.7, 7.0, 7.3, 7.5, 7.8, 8.0, 8.2, 8.5, 8.7, 9.0, 9.3, 9.5, 9.8

1. Розрахуйте ширину інтервалу за правилом Стреджеса та складіть інтервальний ряд.
2. Побудуйте полігон частот.
3. Знайдіть емпіричну функцію розподілу та створіть її графік.

Завдання 3: Групований розподіл і статистичний аналіз

Дані для обробки:

45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100.

1. Запишіть впорядкований розподіл частот та відносних частот.
2. Розрахуйте ширину інтервалу за правилом Стреджеса та побудуйте згрупований розподіл частот та відносних частот.
3. Побудуйте згрупований полігон накопичених частот та графік щільності частот.

Варіант 3

Завдання 1: Варіаційний та інтервальний ряд, частоти та гістограма

Дані для обробки:

12, 15, 18, 20, 22, 25, 28, 30, 12, 15, 18, 20, 22, 25, 28, 30, 12, 15, 18, 20, 22, 25, 28, 30.

1. Побудуйте варіаційний ряд (впорядкуйте дані за зростанням).
2. Складіть розподіл частот та відносних частот.
3. Розрахуйте ширину інтервалу за правилом Стреджеса та побудуйте інтервальний ряд.
4. Створіть гістограму частот.

Завдання 2: Емпірична функція розподілу та полігон частот

Дані для обробки:

0,5; 0,7; 0,9; 1,1; 1,3; 1,5; 1,7; 1,9; 2,1; 2,3; 2,5; 2,5; 2,5; 2,7; 2,7; 2,7; 2,9;
3,1; 3,3; 3,5; 3,7; 3,9; 4,1; 4,3; 4,5; 4,7; 4,9; 5,1; 5,3; 5,5.

1. Розрахуйте ширину інтервалу за правилом Стреджеса та складіть інтервальний ряд.
2. Побудуйте полігон частот.
3. Знайдіть емпіричну функцію розподілу та створіть її графік.

Завдання 3: Групований розподіл і статистичний аналіз**Дані для обробки:**

30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 30, 35, 40, 45, 50,
55, 60, 65, 70, 75, 80, 85, 90, 95, 100.

1. Запишіть впорядкований розподіл частот та відносних частот.
2. Розрахуйте ширину інтервалу за правилом Стреджеса та побудуйте згрупований розподіл частот та відносних частот.
3. Побудуйте згрупований полігон накопичених частот та графік щільності частот.

Варіант 4**Завдання 1: Варіаційний та інтервальний ряд, частоти та гістограма****Дані для обробки:**

3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21.

1. Побудуйте варіаційний ряд (впорядкуйте дані за зростанням).
2. Складіть розподіл частот та відносних частот.
3. Розрахуйте ширину інтервалу за правилом Стреджеса та побудуйте інтервальний ряд.
4. Створіть гістограму частот.

Завдання 2: Емпірична функція розподілу та полігон частот**Дані для обробки:**

20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 20, 25, 30,
35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100.

1. Розрахуйте ширину інтервалу за правилом Стреджеса та складіть інтервальний ряд.
2. Побудуйте полігон частот.
3. Знайдіть емпіричну функцію розподілу та створіть її графік.

Завдання 3: Групований розподіл і статистичний аналіз**Дані для обробки:**

45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 45, 50, 55, 60, 65, 70, 75, 80,
85, 90, 95, 100, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100.

1. Запишіть впорядкований розподіл частот та відносних частот.

2. Розрахуйте ширину інтервалу за правилом Стреджеса та побудуйте згрупований розподіл частот та відносних частот.

3. Побудуйте згрупований полігон накопичених частот та графік щільності частот.

Варіант 5

Завдання 1: Варіаційний та інтервальний ряд, частоти та гістограма

Дані для обробки:

10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28.

1. Побудуйте варіаційний ряд (впорядкуйте дані за зростанням).
2. Складіть розподіл частот та відносних частот.
3. Розрахуйте ширину інтервалу за правилом Стреджеса та побудуйте інтервальний ряд.
4. Створіть гістограму частот.

Завдання 2: Емпірична функція розподілу та полігон частот

Дані для обробки:

1,0; 1,2; 1,4; 1,6; 1,8; 2,0; 2,2; 2,4; 2,6; 2,8; 3,0; 3,2; 3,4; 3,6; 3,8; 4,0; 4,2; 4,4; 4,6; 4,8; 5,0; 5,2; 5,4; 5,6; 5,8; 6,0; 6,2; 6,4; 6,6; 6,8.

1. Розрахуйте ширину інтервалу за правилом Стреджеса та складіть інтервальний ряд.
2. Побудуйте полігон частот.
3. Знайдіть емпіричну функцію розподілу та створіть її графік.

Завдання 3: Групований розподіл і статистичний аналіз

Дані для обробки:

40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100.

1. Запишіть впорядкований розподіл частот та відносних частот.
2. Розрахуйте ширину інтервалу за правилом Стреджеса та побудуйте згрупований розподіл частот та відносних частот.
3. Побудуйте згрупований полігон накопичених частот та графік щільності частот.

ТЕОРЕТИЧНІ ВІДОМОСТІ

1. Варіаційний ряд

Варіаційний ряд – це впорядкований за зростанням або спаданням набір значень вибірки. Він дозволяє оцінити розподіл даних та виявити їхні особливості.

Приклад:

Вибірka: 7, 4, 4, 8, 12, 12, 12, 7, 8, 12.

Варіаційний ряд: 4, 4, 7, 7, 8, 8, 12, 12, 12, 12.

Для побудови варіаційного ряду можна використовувати вбудовану у Python функцію **sorted()**, яка сортує список даних за зростанням.

```
data = [7, 4, 4, 8, 12, 12, 12, 7, 8, 12]
sorted_data = sorted(data)
print("Варіаційний ряд:", sorted_data)
```

2. Розподіл частот та відносних частот

Частота (n) – кількість повторень певного значення у вибірці.

Відносна частота (f) – частка певного значення у загальній кількості елементів вибірки, яка обчислюється за формулою:

$$f = \frac{n_i}{N}$$

де

n_i – частота значення,

N – загальна кількість елементів у вибірці.

Приклад:

Для значення 12 у вибірці з 10 елементів: 7, 4, 4, 8, 12, 12, 12, 7, 8, 12

Частота $n = 4$, відносна частота $f = 4/10 = 0,4$.

Для реалізації на Python можна використовувати модуль **collections.Counter** для підрахунку частот кожного значення у вибірці. Відносні частоти обчислюються шляхом ділення частоти кожного значення на загальну кількість елементів у вибірці.

```
from collections import Counter

data = [7, 4, 4, 8, 12, 12, 12, 7, 8, 12]
frequency = Counter(data)
N = len(data)
relative_frequency = {k: v / N for k, v in frequency.items()}

print("Частоти:", frequency)
print("Відносні частоти:", relative_frequency)
```

3. Інтервальний ряд

Інтервальний ряд використовується для групування даних, особливо коли значення є дробовими або кількість унікальних значень велика. Дані розбиваються на інтервали, і для кожного інтервалу вказується кількість елементів, що потрапили в нього.

Ширина інтервалу (h) може бути задана вручну або розрахована за правилом Стерджеса (Sturges):

$$k = 1 + 3,322 \cdot \log_{10}(N)$$

де

k – кількість інтервалів, N – обсяг вибірки.

Ширина інтервалу:

$$h = \frac{R}{k}$$

де R – розмах даних (різниця між максимальним і мінімальним значеннями).

Приклад:

Для вибірки: 7, 4, 4, 8, 12, 12, 12, 7, 8, 12, 8, 12, 4, 12, 12, 4, 12, 4, 12, 12, 6, 9, 5, 3, 6, 6, 9, 3, 5, 6, 9, 5, 6, 6, 9, 6, 9, 6, 6, 6

$\max = 12$ $\min = 3$ $N=50$ $R = \max - \min = 12 - 3 = 9$

$k = 1 + 3,322 \cdot \log_{10}(N) = 1 + 3,322 \cdot \log_{10}(50) \approx 6,6$ (округлюємо до 7).

Ширина інтервалу:

$h = \frac{R}{k} = \frac{9}{7} \approx 1,29$ (округлюємо до 1,3).

Для реалізації на Python можна використовувати модуль **numpy** для обчислення логарифму (**np.log10**), розмах даних R обчислюється як різниця між максимальним і мінімальним значеннями у вибірці.

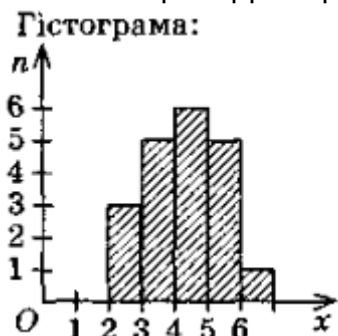
```
import numpy as np

data = [7, 4, 4, 8, 12, 12, 12, 7, 8, 12, 8, 12, 4, 12, 12, 4, 12, 4, 12,
12, 6, 9, 5, 3, 6, 6, 9, 3, 5, 6, 9, 5, 6, 6, 9, 6, 9, 6, 6, 6]
N = len(data)
R = max(data) - min(data)
k = 1 + 3.322 * np.log10(N)
h = R / k

print("Кількість інтервалів (k):", round(k))
print("Ширина інтервалу (h):", round(h, 2))
```

4. Гістограма

Гістограма – це графічне представлення інтервального ряду, де на осі X відкладаються інтервали, а на осі Y – частоти або відносні частоти. Висота стовпців відповідає кількості елементів у кожному інтервалі.



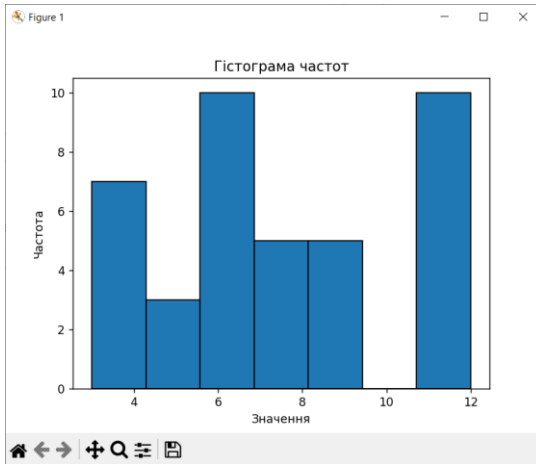
Для реалізації на Python можна використовувати модуль **matplotlib.pyplot** для побудови графіків.

Функція **plt.hist()** автоматично розбиває дані на інтервали та будує гістограму.

```
import matplotlib.pyplot as plt
import numpy as np

data = [7, 4, 4, 8, 12, 12, 12, 7, 8, 12, 8, 12, 4, 12, 12, 4, 12, 4, 12,
12, 6, 9, 5, 3, 6, 6, 9, 3, 5, 6, 9, 5, 6, 6, 9, 6, 9, 6, 6, 6]
```

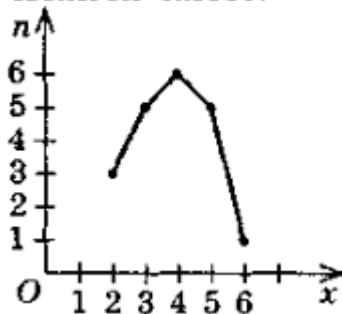
```
plt.hist(data, bins=7, edgecolor='black')
plt.title('Гістограма частот')
plt.xlabel('Значення')
plt.ylabel('Частота')
plt.show()
```



5. Полігон частот

Полігон частот – це графік, який з'єднує точки, де на осі X відкладаються середини інтервалів (або значення для дискретних даних), а на осі Y – частоти або відносні частоти.

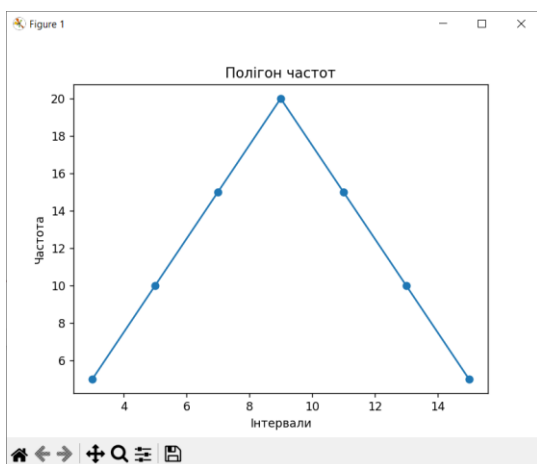
Полігон частот:



Для реалізації на Python можна використовувати функцію **plt.plot()** для побудови лінійного графіка модуля **matplotlib.pyplot**. Параметр **marker='o'** додає маркери до точок на графіку.

```
import matplotlib.pyplot as plt

intervals = [3, 5, 7, 9, 11, 13, 15]
frequencies = [5, 10, 15, 20, 15, 10, 5]
plt.plot(intervals, frequencies, marker='o')
plt.title('Полігон частот')
plt.xlabel('Інтервали')
plt.ylabel('Частота')
plt.show()
```

6. Емпірична функція розподілу

Емпірична функція розподілу $F(x)$ показує ймовірність того, що випадкова величина набуває значення, менше або рівне x . Вона обчислюється за формулою:

$$F(x) = \frac{\text{Кількість елементів} \leq x}{N}$$

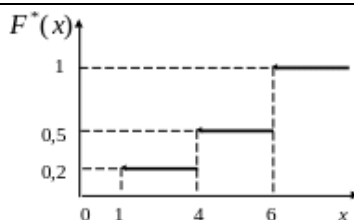
Графік емпіричної функції розподілу має ступінчастий вигляд.

Приклад:

Емпірична функція за даним розподілом вибірки:

варіанти	1	4	7
частоти	10	15	25

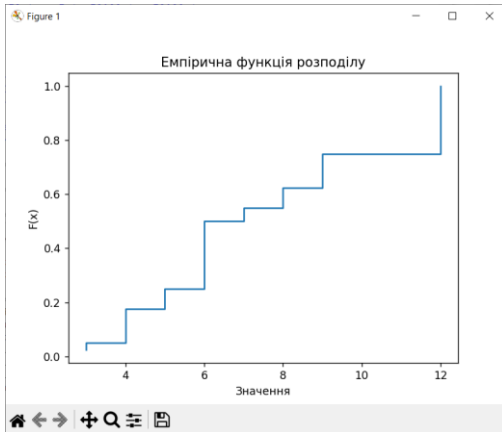
$$F^*(x) = \begin{cases} 0, & x \leq 1; \\ 0,2, & 1 < x \leq 4; \\ 0,5, & 4 < x \leq 7; \\ 1, & x > 7. \end{cases}$$



Для реалізації на Python можна використовувати функцію **np.sort()** модуля **numpy** для сортування даних, функція **plt.step()** модуля **matplotlib.pyplot** будує ступінчастий графік.

```
import matplotlib.pyplot as plt
import numpy as np

data = [7, 4, 4, 8, 12, 12, 12, 7, 8, 12, 8, 12, 4, 12, 12, 4, 12, 4, 12,
12, 6, 9, 5, 3, 6, 6, 9, 3, 5, 6, 9, 5, 6, 6, 9, 6, 9, 6, 6, 6]
sorted_data = np.sort(data)
y_values = np.arange(1, len(sorted_data) + 1) / len(sorted_data)
plt.step(sorted_data, y_values, where='post')
plt.title('Емпірична функція розподілу')
plt.xlabel('Значення')
plt.ylabel('F(x)')
plt.show()
```



7. Накопичені частоти та щільність частот

Накопичена частота – сума частот для всіх значень, менших або рівних поточному.

Щільність частоти – відношення частоти інтервалу до його ширини. Обчислюється за формулою:

$$\text{Щільність} = \frac{n_i}{h}$$

де

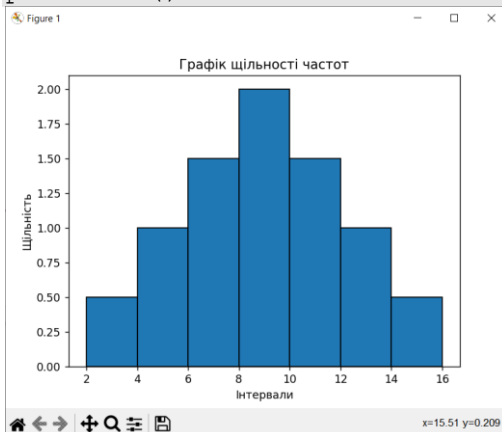
n_i – частота інтервалу,

h – ширина інтервалу.

Для реалізації на Python можна використовувати функцію **plt.bar()** модуля **matplotlib.pyplot** для побудови стовпчастої діаграми, параметр **width = 2** задає ширину стовпців.

```
import matplotlib.pyplot as plt

intervals = [3, 5, 7, 9, 11, 13, 15]
densities = [0.5, 1.0, 1.5, 2.0, 1.5, 1.0, 0.5]
plt.bar(intervals, densities, width=2, edgecolor='black')
plt.title('Графік щільності частот')
plt.xlabel('Інтервали')
plt.ylabel('Щільність')
plt.show()
```



ВМІСТ ЗВІТУ ДО ПРАКТИЧНОЇ РОБОТИ

Результатом виконання практичної роботи має бути:

1. **Необхідна кількість програм**, які написані мовою Python (файли *.py, що містить вихідний код з коментарями для вирішення завдань).
2. **Звіт** (в електронному вигляді), створений в MS Word (файл *.docx), що містить:
 - 1) титульну сторінку;
 - 2) пункт **ПОСТАНОВКА ЗАВДАННЯ** (вказати, що потрібно було зробити в практичній роботі);
 - 3) пункт **РЕЗУЛЬТАТ ВИКОНАННЯ РОБОТИ**:
 - для **кожного** завдання вказати *номер, загальне формулювання* (якщо є) і *формулювання підзадач* (якщо є);
 - після формулювання для кожного завдання (підзавдання) розмістити *програмний код*, який реалізує рішення (шрифт програмного коду: Courier New, 10, одинарний міжрядковий інтервал);
 - надати пояснення до програмного коду;
 - виконати тестування програм (навести скріни результату роботи).
 - навести розрахунки виконання вручну або в Excel.

Форматування звіту про виконання практичної роботи в електронному вигляді необхідно виконувати відповідно до вмісту файлу **Вимоги до оформлення звітів**

Програми (файли *.py) і звіт (файл *.docx) необхідно завантажити на сайт <https://mentor.khai.edu> в курсі «Основи статистики в Python» <https://mentor.khai.edu/course/view.php?id=8740>.