

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені Тараса Шевченка
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
Кафедра програмних систем і технологій

Дисципліна
«Ймовірнісні основи програмної інженерії»

Лабораторна робота № 3
«Двовимірна статистика»

Виконала:	Саніна В.О.	Перевірив:	Марцафей А.С.
Група	ІПЗ-21	Дата перевірки	
Форма навчання	денна	Оцінка	
Спеціальність	121		
2022			

Тема: двовимірна статистика

Мета: навчитись використовувати на практиці набуті знання про міри в двовимірній статистиці.

Завдання:

1. Намалюйте діаграму розсіювання для даних. Укажіть, чи існує тренд у даних. Якщо так, то вкажіть, чи є це негативним трендом, чи позитивним.
2. Знайдіть центр ваги і коваріацію.
3. Знайти рівняння лінії регресії у від х.
4. Розрахуйте коефіцієнт кореляції між даними.
5. Зробити висновок про залежності.

Математична модель:

$$\text{Середнє } X \quad \bar{x} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

$$\text{Середнє } Y \quad \bar{y} = \frac{Y_1 + Y_2 + \dots + Y_n}{n}$$

$$\begin{aligned} \text{Лінія тренду} \quad m &= \frac{\sum (x - \bar{x}) * (y - \bar{y})}{\sum (x - \bar{x})^2} \\ b &= \bar{y} - m * \bar{x} \\ y &= mx + b \end{aligned}$$

Центр ваги $G(\bar{x}; \bar{y})$

$$\text{Коваріація } cov(x; y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}$$

$$\text{Дисперсія } Var(x) = \frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2$$

$$\text{Рівняння лінії регресії} \quad y = b_0 + b_1 x$$

$$b_1 = \frac{cov(x; y)}{var(x)}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$\text{Коефіцієнт кореляції} \quad s_x = \sqrt{Var(x)}$$

$$s_y = \sqrt{Var(y)}$$

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

Псевдокод алгоритмів:

Пошук середнього значення:

```
for i in range(0, lenght):  
    meanPrice += price[i]  
meanPrice /= len(price)
```

Знаходження лінії тренду:

```
for i in range(0, lenght):  
    first += (price[i] - meanX()) * (time[i] - meanY())  
    second += (price[i] - meanX()) ** 2  
m = first / second  
b = meanY() - m * meanX()  
y = m * x + b
```

Знаходження тренду:

if $y[\text{last}] > y[0]$: Тренд позитивний
else: Тренд негативний

Знаходження коваріації:

```
for i in range(0, lenght):  
    summ += price[i] * time[i]  
nsumm = (1 / lenght) * summ  
cov = nsumm - meanX() * meanY()
```

Знаходження дисперсії:

```
for i in range(0, lenght):  
    summa += price[i] ** 2  
summa = (1 / lenght) * summa  
dispersion = summa - meanX() ** 2
```

Знаходження коефіцієнту дисперсії:

```
for i in range(0, lenght):  
    first = (price[i] - meanX()) / math.sqrt(Dispersion())  
    second = (time[i] - meanY()) / math.sqrt(DispersionY())  
    summ += first * second  
cor = (1 / (lenght - 1)) * summ
```

Знаходження точки ваги:

(meanX; meanY)

Знаходження рівняння регресії:

$b1 = (\text{Covariation}()) / \text{Dispersion}()$

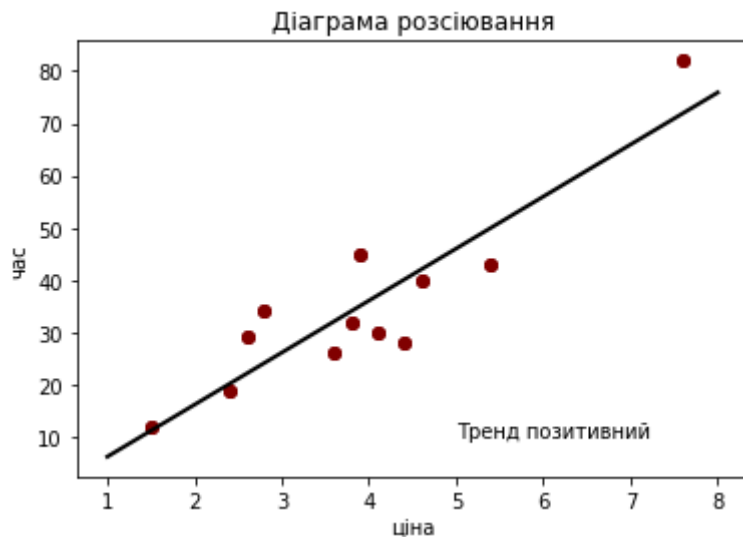
$b0 = \text{meanY}() - b1 * \text{meanX}()$

$y = b1 + b0 *$

Демонстрація роботи алгоритмів:

```
=====
Ціна      Час
=====
2.4      19.0
1.5      12.0
7.6      82.0
3.8      32.0
2.8      34.0
4.6      40.0
4.1      30.0
5.4      43.0
3.9      45.0
2.6      29.0
4.4      28.0
3.6      26.0
=====
Завдання 1
=====
Тренд позитивний
=====
Завдання 2
=====
Точка ваги:(3.89;35.0)
Коваріація: 23.0
=====
Завдання 3
=====
Рівняння регресії: y=9.96*x-3.76=0
=====
Завдання 4
=====
Коефіцієнт кореляції:0.98
=====
Завдання 5
=====
Коефіцієнт кореляції:0.98
Тренд негативний
Коефіцієнт кореляції наближається до 1, значить дані майже співпадають з лінією регресії
=====
```

```
=====
Ціна      Час
=====
2.4      19.0
1.5      12.0
7.6      82.0
3.8      32.0
2.8      34.0
4.6      40.0
4.1      30.0
5.4      43.0
3.9      45.0
2.6      29.0
4.4      28.0
3.6      26.0
2.4      19.0
1.5      12.0
7.6      82.0
3.8      32.0
2.8      34.0
4.6      40.0
4.1      30.0
5.4      43.0
3.9      45.0
2.6      29.0
4.4      28.0
3.6      26.0
2.4      19.0
1.5      12.0
7.6      82.0
3.8      32.0
2.8      34.0
4.6      40.0
4.1      30.0
5.4      43.0
3.9      45.0
2.6      29.0
4.4      28.0
3.6      26.0
2.4      19.0
1.5      12.0
7.6      82.0
3.8      32.0
2.8      34.0
4.6      40.0
4.1      30.0
5.4      43.0
3.9      45.0
2.6      29.0
4.4      28.0
3.6      26.0
2.4      19.0
1.5      12.0
7.6      82.0
3.8      32.0
2.8      34.0
4.6      40.0
4.1      30.0
5.4      43.0
3.9      45.0
2.6      29.0
4.4      28.0
3.6      26.0
2.4      19.0
1.5      12.0
7.6      82.0
3.8      32.0
2.8      34.0
4.6      40.0
4.1      30.0
5.4      43.0
3.9      45.0
2.6      29.0
4.4      28.0
3.6      26.0
2.4      19.0
=====
Завдання 1
=====
Тренд позитивний
=====
Завдання 2
=====
Точка ваги:(3.86;34.5)
Коваріація: 22.59
=====
Завдання 3
=====
Рівняння регресії: y=10.0*x-4.06=0
=====
Завдання 4
=====
Коефіцієнт кореляції:0.91
=====
Завдання 5
=====
Коефіцієнт кореляції:0.91
Тренд негативний
Коефіцієнт кореляції наближається до 1, значить дані майже співпадають з лінією регресії
=====
```



Висновок: Під час виконання лабораторної, було розроблено алгоритми на тему двовимірної статистики. Було намальовано графік розсіювання для даних, знайдено тренд у даних та визначено, що він позитивний (для тестових даних). Знайдено центр ваги даних та кореляцію за формулами обчислень. Знайдено рівняння лінії регресії у від x . Розраховано коефіцієнт кореляції між даними. В залежності від значення коефіцієнту, визначається залежність даних.