|  |  |  |  |
| --- | --- | --- | --- |
| 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»** | | | |
| **Виконав:** | Хапланов Михайло Сергійович | **Перевірила**: | Марцафей Анна Сергіївна |
| Група | ІПЗ-24(2) | Дата перевірки |  |
| Форма навчання | денна | Оцінка |  |
| Спеціальність | 121 |
| 2022 | | | |

**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 коваріацію.

3. Знайти рівняння лінії регресії y від x.

4. Розрахуйте коефіцієнт кореляції між даними.

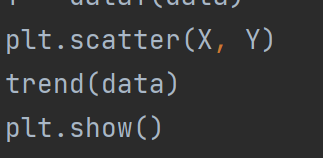
**Хід роботи**

**Завдання 1**

**1. Намалюйте діаграму розсіювання для даних. Укажіть, чи існує тренд у даних. Якщо так, то вкажіть, чи є це негативним трендом, чи позитивним.**

Діаграма розсіювання (точкова діаграма) — один з типів математичних діаграм, що використовує декартову систему координат для відображення значень двох змінних для набору даних.

print('Task 1')  
def dataX(data):  
 inputdatadata = []  
 for i in range(len(data)):  
 inputdatadata.append(data[i][0])  
 return inputdatadata  
  
def dataY(data):  
 inputdatadata = []  
 for i in range(len(data)):  
 inputdatadata.append(data[i][1])  
 return inputdatadata  
  
def trend(data):  
 if max(data)==data[len(data)-1]:  
 print("Тренд позитивний")  
 elif min(data)==data[len(data)-1]:  
 print("Тренд негативний") else:  
 print("Дані не мають терду")



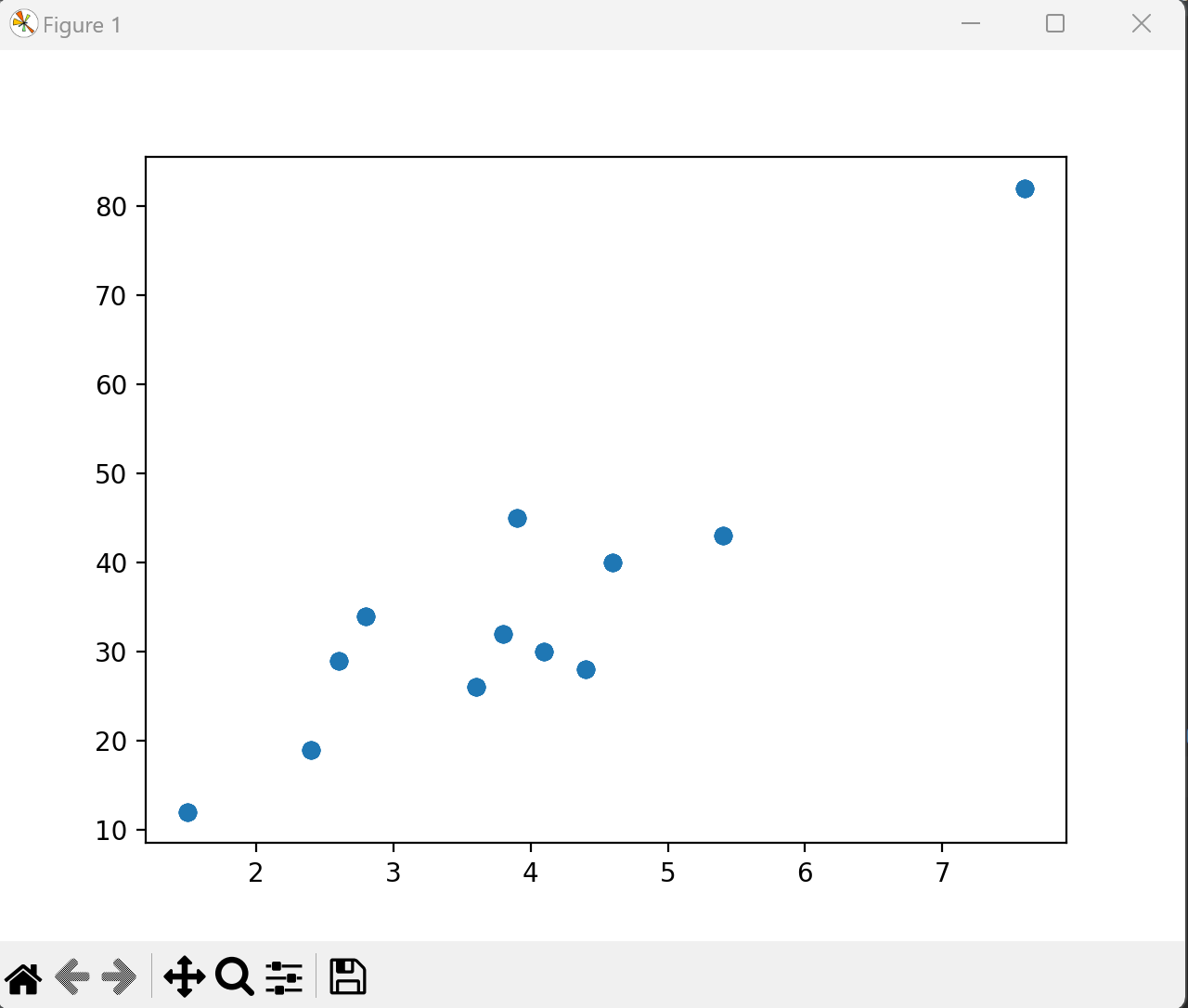
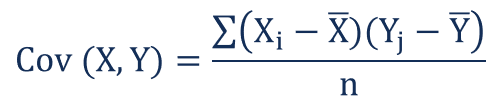


Рисунок 5 | Приклад виконання (діаграма розсіювання), input\_10.txt

**2. Знайдіть коваріацію.**

Коваріація — це міра спільної мінливості двох випадкових змінних. Якщо більші значення однієї змінної здебільшого відповідають більшим значенням іншої, й те саме виконується для менших значень, тобто змінні схильні демонструвати подібну поведінку, то коваріація є додатною. В протилежному випадку, коли більші значення однієї змінної здебільшого відповідають меншим значенням іншої, тобто змінні схильні демонструвати протилежну поведінку, коваріація є від'ємною. Отже, знак коваріації показує тенденцію в лінійному взаємозв'язку між цими змінними.

Коваріацію (рис. 8) було знайдено за допомогою наступної формули (рис. 6) та реалізовано відповідним кодом (рис. 7).



def task2(x, y):  
 global averagex, averagey  
 covarience = 0.0  
 for i in range(len(x)):  
 averagex += x[i]  
 averagey += y[i]  
 averagex = averagex / len(x)  
 averagey = averagey / len(y)  
 for i in range(len(x)):  
 covarience += (x[i] - averagex) \* (y[i] - averagey)  
 covarience = covarience / (len(x)-1)  
 print('Covarince: ', covarience)

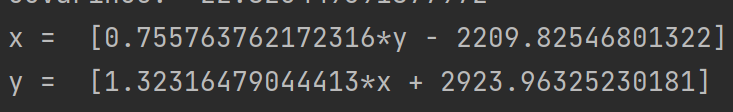


| Приклад виконання, input\_100.txt

**3. Знайти рівняння лінії регресії y від x.**

Рівняння, що відображує зміну середньої величини однієї ознаки (у) в залежності від другої (х), називається рівнянням регресії або рівнянням кореляційного зв'язку.

def task3(x, y ):  
 global avergex, avergey  
 byx, sumxy, sumx, sumy, sumx2, sumy2 = 0.0, 0.0, 0.0, 0.0, 0.0, 0.0  
  
 for i in range(len(x)):  
 sumx += x[i]  
 sumy += y[i]  
 sumxy += x[i] \* y[i]  
 sumx2 += x[i] \* x[i]  
 sumy2 += y[i] \* y[i]  
 byx = (len(x) \* sumxy - (sumx \* sumy)) / (len(x) \* sumx2 - sumx2)  
 x, y = sp.symbols("x, y")  
 line = sp.Eq(y-avergey, byx \* (x - avergex))  
 linex = sp.solve(line, y)  
 liney = sp.solve(line, x)  
  
 print("x = ", liney)  
 print("y = ", linex)

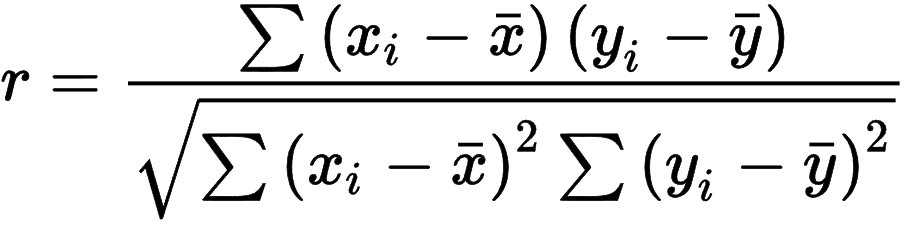


| Приклад виконання, input\_100.txt

**4. Розрахуйте коефіцієнт кореляції між даними.**

Коефіцієнт кореляції Пірсона — в статистиці, показник кореляції (лінійної залежності) між двома змінними X та Y, який набуває значень від −1 до +1 включно. Він широко використовується в науці для вимірювання ступеня лінійної залежності між двома змінними.

Коефіцієнт кореляції між даними (рис. 15) був знайдений за допомогою відповідної формули (рис. 13) та реалізований наступним кодом (рис. 14).



| Формула розрахунку коефіцієнту кореляції між даними

def task4(x, y):  
 global avergex, avergey  
 correlation, sum1, sum2, sum3 = 0.0, 0.0, 0.0, 0.0  
 for i in range(len(X)):  
 sum1 = (x[i] - avergex) \* (y[i] - avergey)  
 sum2 = (x[i] - avergex) \* (x[i] - avergex)  
 sum3 = (y[i] - avergey) \* (y[i] - avergey)  
 sum2 += sum2 \* sum3  
 correlation = sum1/math.sqrt(sum2)  
 print("Correlation: ", correlation)

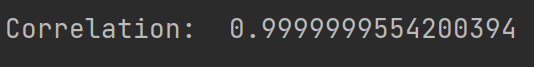


Рисунок 15 | Приклад виконання, input\_100.txt

**Висновок:** Під час виконання третьої лабораторної роботи було повторено операції з вхідними даними записаних у txt файл, реалізовано потрібні формули для знаходження інформації за завданнями, також побудовано діаграму розсіювання.

Також було проведено дослідження, де було порівняно коваріацію з кореляцією.

Коваріація та кореляція в основному оцінюють зв’язок між змінними. Найближчою аналогією зв'язку між ними є зв'язок між дисперсією та стандартним відхиленням.

Коваріація вимірює загальну варіацію двох випадкових змінних від їхніх очікуваних значень. Використовуючи коваріацію, ми можемо лише оцінити напрямок зв’язку (чи змінні мають тенденцію рухатися в тандемі чи показують зворотний зв’язок). Однак це не вказує ні на силу зв’язку, ні на залежність між змінними.

З іншого боку, кореляція вимірює силу зв’язку між змінними. Кореляція – це масштабна міра коваріації. Він безрозмірний. Іншими словами, коефіцієнт кореляції завжди є чистим значенням і не вимірюється в жодних одиницях.

Зв'язок між двома поняттями можна виразити за допомогою формули нижче:

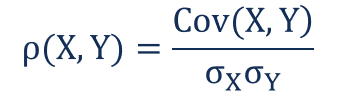


Рисунок 16 | Формула зображення зв'язку між коваріацією та кореляцією

Де:

* **ρ(X,Y)** – кореляція між змінними X і Y
* **Cov(X,Y)** – коваріація між змінними X і Y
* **σ X** – стандартне відхилення X-змінної
* **σ Y** – стандартне відхилення Y-змінної