

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:	Зарожевський Євген Сергійович	Перевірила:	Марцафей А.С.
Група	ІПЗ-22	Дата перевірки	
Форма навчання	денна	Оцінка	
Спеціальність	121		
2022			

Мета роботи:

Навчитись використовувати на практиці набуті знання про міри у 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.

Побудова математичної моделі:

Для побудови діаграми розсіювання та лінії регресії на графіку використовувалась бібліотека matplotlib мови програмування python.

Для визначення центру ваг використовувалась формула:

$$G(x, y) = G(\text{Mean}X, \text{Mean}Y)$$

Для визначення коваріації використовувалась формула:

$$C = \frac{\sum_{i=0}^n (x_i - \text{Mean}X) * (y_i - \text{Mean}Y)}{n}$$

Для визначення рівняння регресії використовувалась формула:

$$y = b_1 x + b_0,$$

$$\text{Де } b_1 = \frac{C}{\text{Var}(x)},$$

$$b_0 = \text{Mean}Y - b_1 * \text{Mean}X$$

Для визначення коефіцієнта кореляції використовувалась формула:

$$\text{Cor} = \frac{\sum_{i=0}^n \left(\frac{x_i - \text{Mean}X}{s_x} * \frac{y_i - \text{Mean}Y}{s_y} \right)}{n-1},$$

$$\text{Де } s_x = \sqrt{\text{Var}(x)},$$

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

Код алгоритму:

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

fileName = input("Enter name of file without extension: ")

f = open(f'{fileName}.txt', 'r')

length = int(f.readline())

x = []
y = []

for i in range(length):
    [item, item1] = f.readline().split()
    item = item.replace(',', '.')
    x.append(float(item))
    y.append(int(item1))

f.close()

totalX = 0
totalY = 0

for i in range(length):
    totalX += x[i]
    totalY += y[i]

meanX = round(totalX / len(x), 1)
meanY = totalY / len(y)

covariance = 0
varianceX = 0
varianceY = 0

for i in range(length):
    covariance += (x[i] - meanX) * (y[i] - meanY)
    varianceX += (x[i] - meanX) ** 2
    varianceY += (y[i] - meanY) ** 2

covariance /= length
varianceX /= length
varianceY /= length

b1 = covariance / varianceX
b0 = meanY - b1 * meanX

npX = np.array(x)
npY = b1 * npX + b0

standardDeviationX = math.sqrt(varianceX)
standardDeviationY = math.sqrt(varianceY)
correlation = 0

for i in range(length):
    correlation += ((x[i] - meanX) / standardDeviationX) * ((y[i] - meanY) /
```

```

standardDeviationY)

correlation /= (length - 1)

f = open('output.txt', 'w')
f.write(f'Center of gravity: ({meanX}, {meanY})\n')
f.write(f'Covariance: {covariance}\n')
f.write(f'Regression equation: y = {b1} + ({b0})*x\n')
f.write(f'Correlation coefficient: {correlation}\n')

if correlation > 0:
    f.write(f'Correlation is positive\n')
elif correlation < 0:
    f.write(f'Correlation is negative\n')

if correlation == 0:
    f.write('Conclusion: x and y are not linearly related\n')
elif math.fabs(correlation) > math.sqrt(3) / 2:
    f.write('Conclusion: x and y have a strong linear relation\n')
else:
    f.write('Conclusion: x and y have a weak linear relation\n')

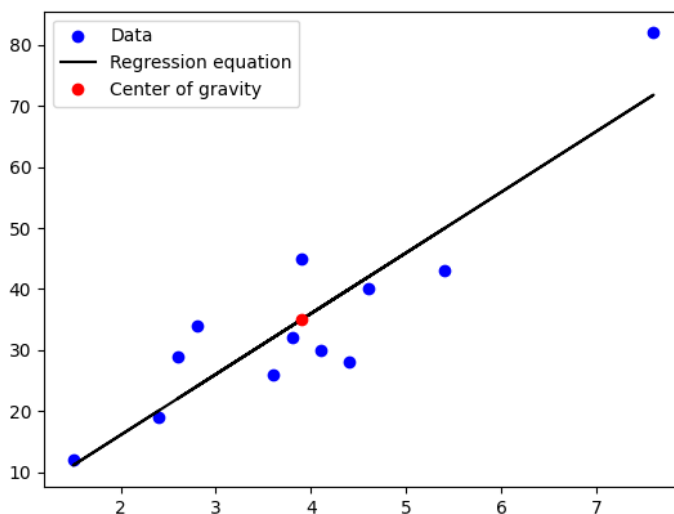
plt.scatter(x, y, c="blue", label="Data")
plt.plot(npX, npY, c="k", label="Regression equation")
plt.plot(meanX, meanY, 'ro', label="Center of gravity")

plt.legend()
plt.show()

```

Випробування алгоритму:

Результат роботи програми при вхідних даних розміром 10:

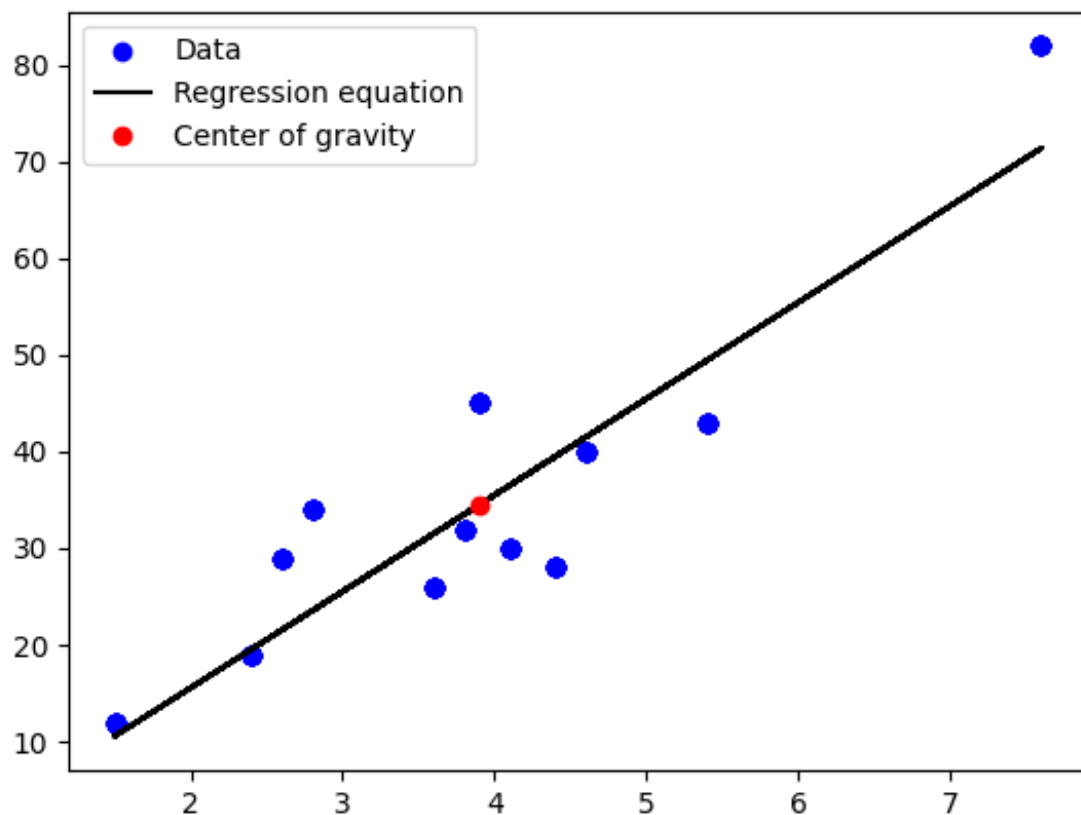


```

1 Center of gravity: (3.9, 35.0)
2 Covariance: 22.99999999999996
3 Regression equation:  $y = 9.953119365308332 + (-3.817165524702496)*x$ 
4 Correlation coefficient: 0.9828959169823492
5 Correlation is positive
6 Conclusion: x and y have a strong linear relation

```

Результат роботи програми при вхідних даних розміром 100:



```

1 Center of gravity: (3.9, 34.5)
2 Covariance: 22.592
3 Regression equation:  $y = 9.973512272647008 + (-4.3966978633233325)*x$ 
4 Correlation coefficient: 0.9106715347800048
5 Correlation is positive
6 Conclusion: x and y have a strong linear relation

```

Висновок:

Виконано завдання третьої лабораторної роботи. Опановані навички використання на практиці набутих знань про міри у двовимірній статистиці. Розроблена програма, що малює діаграму розсіювання даних, знаходить центр ваг, коваріацію та рівняння лінії регресії, розраховує коефіцієнт кореляції між даними та робить висновок про залежності. Програма коректно працює на різних об'ємах вхідних даних.