|  |  |  |  |
| --- | --- | --- | --- |
| МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені Тараса Шевченка  ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ  **Кафедра програмних систем і технологій**  **Звіт**  **Лабораторна робота № 3**  з дисципліни  **«Ймовірнісні основи програмної інженерії»** | | | |
| **Виконав:** | Дєєв Даніїл Олексійович | **Перевірила**: | Вечерковська А.С. |
| 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 центр ваги і коваріацію.

3. Знайти рівняння лініїї регресії y від x.

4. Розрахуйте коефіцієнт кореляції між даними.

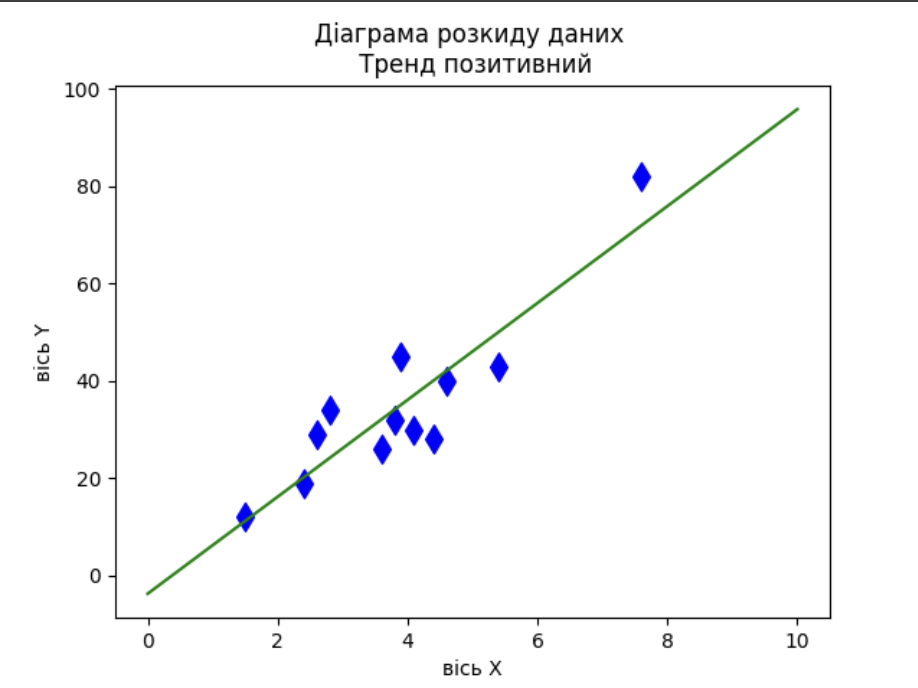
5. Зробити висновок про залежності.

**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 (розкиду) показує взаємозв’язок між двома видами зв’язаних даних і підтверджує їхню залежність. Такими двома видами даних можуть бути характеристика якості та чинник, що впливає на неї, дві різних характеристики якості, два чинники, що впливають на одну характеристику якості, і т. д.

Для побудови діаграми розсіювання потрібно не менше 30 пар даних (x, y). Осі x та y будують так, щоб довжини робочих частин були приблизно однакові. На діаграму наносять точки (x, y), назву діаграми, а *також* інтервал часу, число пар даних, назви осей, ПІБ, посада виконавця, і т. д. Точки, що далеко відстоять від основної *групи*, є викидами, і їх виключають.



Лінія регресії – це оцінка лінії, яка описує справжній, але невідомий лінійний зв’язок між двома змінними. Рівняння лінії регресії використовується для прогнозування (або оцінки) значення змінної відповіді на основі заданого значення пояснювальної змінної. Лінії регресії корисні в процедурах прогнозування. Його метою є опис взаємозв’язку залежної змінної (змінна y) з однією або багатьма незалежними змінними (змінна x).

**2. Знайдіть коваріацію**

**Коваріація - це величина, яка відображає, наскільки дві випадкові величини спільно змінюються щодо їхніх значень.**

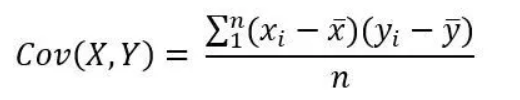
Це дозволяє нам знати, як поводиться змінна на основі того, що робить інша змінна. Тобто, коли X піднімається, як поводиться Y? Таким чином, коваріація може приймати наступні значення:

Коваріація (X, Y) менше нуля, коли “X” зростає, а “Y” падає. Існує негативний зв’язок.

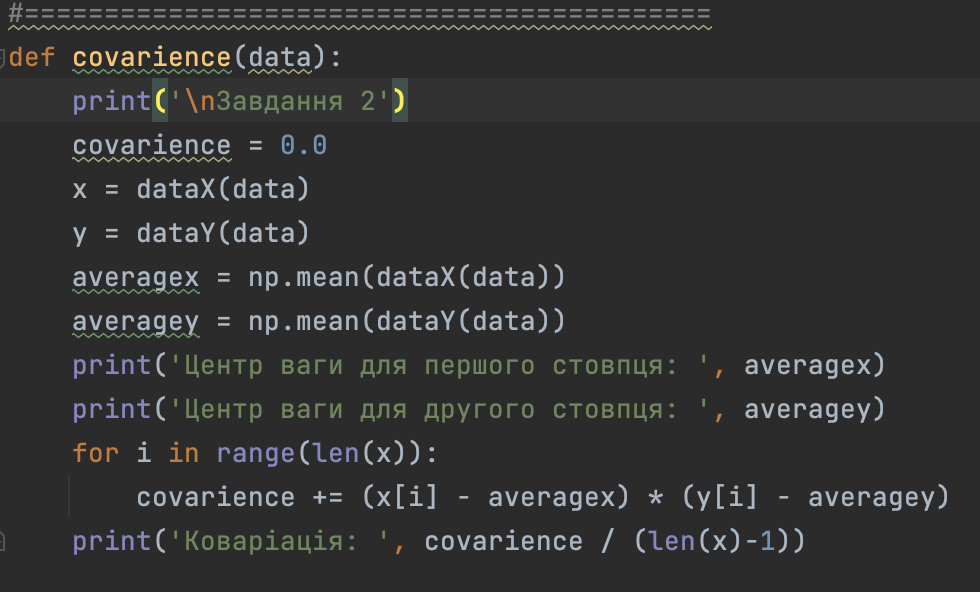
Коваріація (X, Y) перевищує нуль, коли "X" зростає, а "Y" зростає. Є позитивні стосунки.

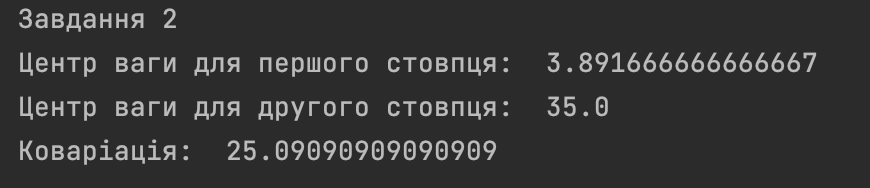
Коваріація (X, Y) дорівнює нулю, коли немає взаємозв'язку між змінними "X" і "Y".

**Розрахунок Коваріації**

Формула коваріації виражається таким чином:

Де y з наголосом є середнім значенням змінної Y, а x з наголосом є середнім значенням змінної X. "i" - положення спостереження, а "n" загальна кількість спостережень.





covarience = 0.0

x = dataX(data)

y = dataY(data)

averagex = np.mean(dataX(data))

averagey = np.mean(dataY(data))

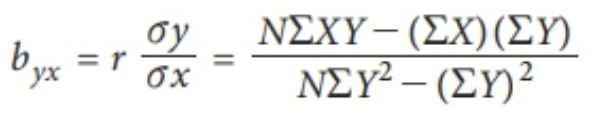
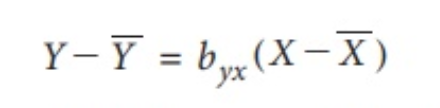
for i in range(len(x)):

covarience += (x[i] - averagex) \* (y[i] - averagey)

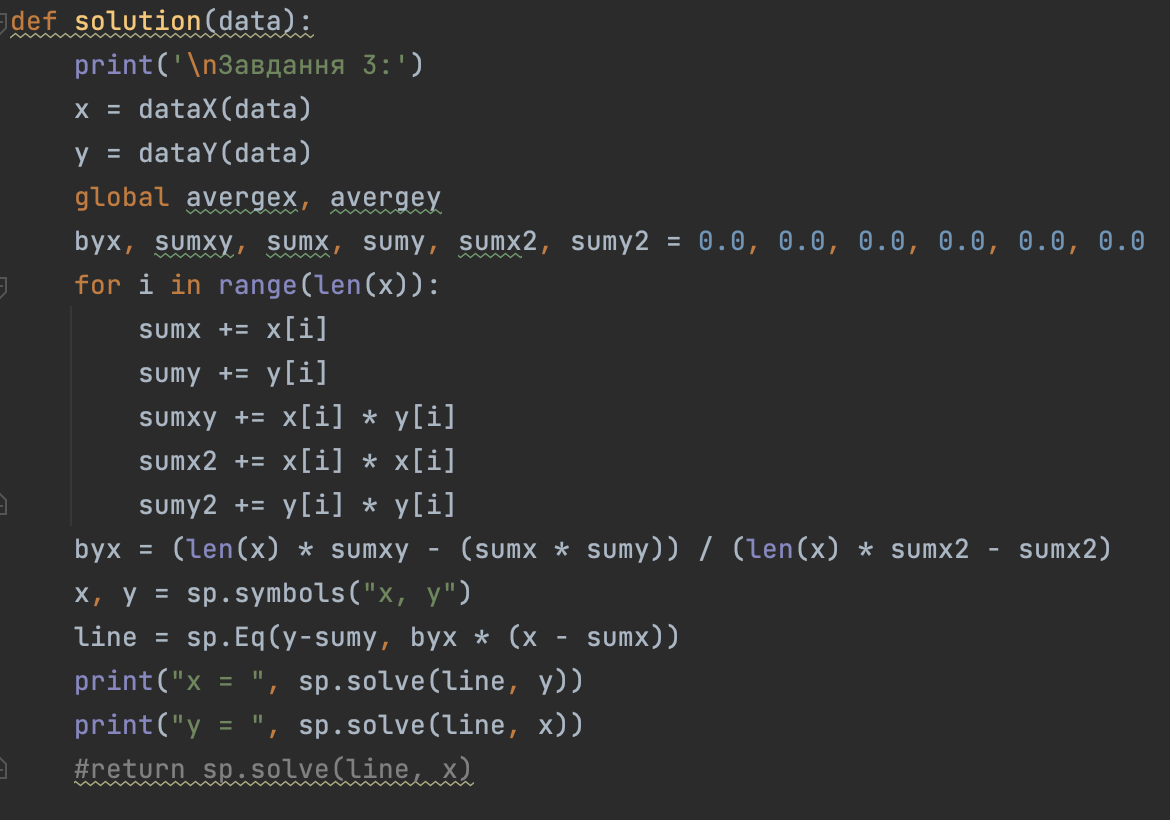
covarience = covarience / (len(x)-1))

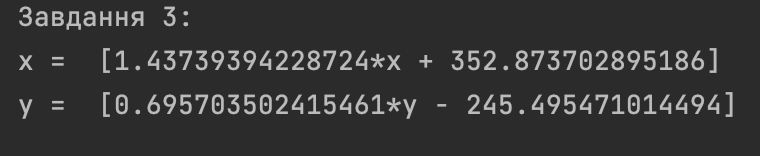
**3.** **Знайти рівняння лініїї регресії y від x.**

Регресійний аналіз між двома випадковими величинами дозволяє встановити чи існує між ними зв'язок та міру кореляції величин.  
Параметри рівняння регресії - це складові рівняння прямої регресії, якщо між випадковими величинами існує лінійний зв'язок.  
Щоб підтвердити гіпотезу про лінійність зв'язку між двома випадковими величинами обчислюють вибірковий коефіцієнт кореляції.

Рівняння регресії Y на X:

відомий як коефіцієнт регресії X на Y, і r є коефіцієнтом кореляції між X і Y, σx і σy є стандартними відхиленнями X і Y відповідно.





x = dataX(data)

y = dataY(data)

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")

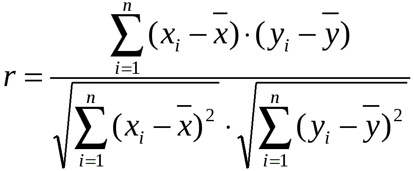
line = sp.Eq(y-sumy, byx \* (x - sumx))

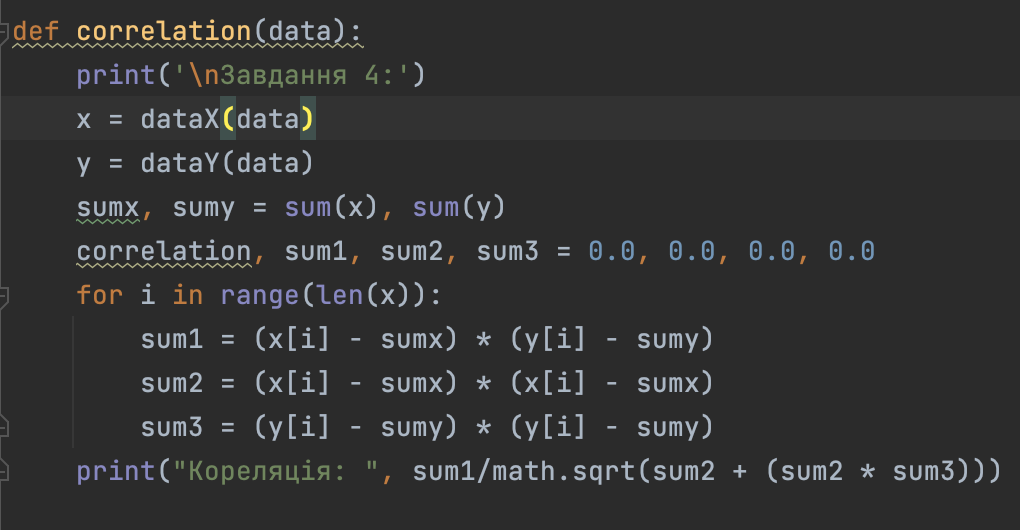
linex = sp.solve(line, y))

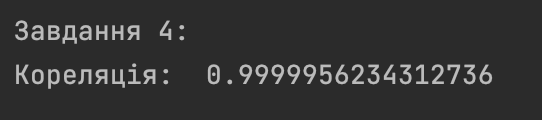
liney = sp.solve(line, x))

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

**Коефіцієнт кореляції** – показник, який використовують для вимірювання щільності зв'язку між результативними і факторними ознаками у кореляційно-регресійній моделі за лінійної залежності. За абсолютною величиною коефіцієнту кореляції коливається в межах від -1 до +1. Чим ближчий цей показник до 0, тим менший зв'язок, чим ближчий він до ±1 – тим зв'язок тісніший. Знак «плюс» при коефіцієнті кореляції означає прямий зв'язок між ознаками х і у, знак «мінус» – обернений.

**Формула коефіцієнту кореляції:**





x = dataX(data)  
y = dataY(data)  
sumx, sumy = sum(x), sum(y)  
correlation, sum1, sum2, sum3 = 0.0, 0.0, 0.0, 0.0  
for i in range(len(x)):  
 sum1 = (x[i] - sumx) \* (y[i] - sumy)  
 sum2 = (x[i] - sumx) \* (x[i] - sumx)  
 sum3 = (y[i] - sumy) \* (y[i] - sumy)  
correlation = sum1/math.sqrt(sum2 + (sum2 \* sum3))

**Висновок:** Під час виконання даної лабораторної роботи було використано на практиці набуті знання про міри в двовимірній статистиці та побудовано діаграму розкиду даних.

.