|  |
| --- |
| **Звіт**  **до лабораторної роботи № 3**  з дисципліни  **«Ймовірнісні основи програмної інженерії»**  **Студента групи ІПЗ-21**  **Павлюченка Артема Максимовича**  **на тему:**  **«ДВОВИМІРНА СТАТИСТИКА»** |

**1)Назва:** Двовимірна статистика

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

**2)Постановка задачі:**

Ми маємо два тестових файли (“input\_10.txt”,“input\_100.txt”), в кожному з яких 2 стовпці та М+1 рядками. Перший рядок містить число: M, де M — кількість значень у кожному з стовбців. Дослідження зроблено щодо суми покупки( стовбець 1) й часу, проведеного в супермаркеті(стовбець 2). За допомогою методу open(input\_10.txt) і методу read, replace та split вдалося розбити вхідний файл на 2 масиви- масив сум(float) та масив часу(int). Проробив всі дії за умовою і отримавши певні результати, за допомогою методу write() записуємо в вихідний файл наші результати. Через те, що метод приймає лише один параметр, довелося викликати його під час запису в файл кожного результату.

**3)Побудова математичної моделі**

**Задача №1:**

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

**Математична модель:**

Дані файлу зчитано у два одновимірних масиви, і для побудови діаграми розсіювання використано метод scatter модулю matplotlib.pyplot, у який передано масиви у якості аргументів. Для того, щоб краще проаналізувати діаграму, було побудовано за допомогою методу plot лінію регресії, у нього передано масив та змінну p, яка зберігає у собі результат методу poly1d бібліотеки numpy. Лінія тренду допомогла мені дійти до висновку, що тренду даних на цій діаграмі позитивний. Тренд є позитивним, коли зі збільшенням суми збільшується час перебування, негативним – коли навпаки. У нашому випадку, хоч і є невеликий розкид даних, але все ж таки він мінімальний, тренд позитивний.

**Задача №2:**

**Умова:** Знайдіть центр ваги і коваріацію.

**Математична модель:**

Центр ваги = ()

Коваріація = 1/n\*-

**Задача №3:**

**Умова:** Знайти рівняння лініїї регресії y від x.

**Математична модель:**

y=b1\*x+b0 –вид рівняння

b1 = cov(x,y)/var(x)

b0 = -b1\*

**Задача №4:**

**Умова:** Розрахуйте коефіцієнт кореляції між даними.

**Математична модель:**

Коефіцієнт кореляції = 1/(n-1)\*

**Задача №5:**

**Умова:** Зробити висновок про залежності.

**Математична модель:**

Проаналізувавши обрахований коефіцієнт кореляції між даними, можу зробити наступний висновок: залежність між даними майже лінійна, розкид між даними мінімальний.

**4)Псевдокод алгоритму**

GetMainMassive:

open(path) as file

create sum\_massive

create time\_massive

first\_massive = read file(replace "\t" on " ", "\n" on " ", "," on ".", split by " ")

if first\_massive(length of first\_massive-1)=="" then pop length of firsy\_massive-1 from first\_massive

index=0

for i in first\_massive:

if index mod 2 = 0 then push i in sum\_mas

else push i in time\_mas

index++

DrawScatterPlot:

scatter(sum\_mas, time\_mas)

z = polyfit of np by (sum\_mas, time\_mas ,1)

p = poly1d of np by z

plot(sum\_mas, time\_mas of p)

show()

SumElementsInMassive:

sum\_massive = 0

for el in massive:

sum\_massive+=el

GetAverageOfMassive:

average = sum\_massive/length of massive

round average

GetCumulativeSum:

cum\_sum=0

i=0

for el in mas1:

cum\_sum+=el\*mas2(i)

i++

GetCovariation:

cov = 1/length of sum\_mas\*cum\_sum - money\_average\*time\_average

round cov

GetSumDeviation:

var\_sum = 0

for el in mas:

var\_sum+=el\*el

GetDeviation:

var = var\_sum/length of mas - average\*average

round var

GetRegressionEquation:

b1 = cov/var\_money

round b1

b0=time\_average - b1\*money\_average

round b0

ControlSum:

CONTROL\_SUM=0

for el in mas1:

CONTROL\_SUM+=((el-average1)/sqrt(var1)\*(mas2(i)-average2)/sqrt(var2))

GetCorrelation:

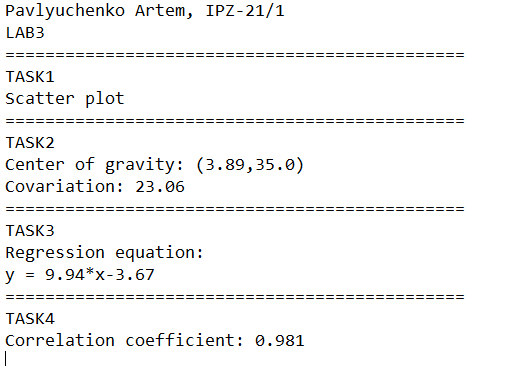
n = length of sum\_mas

CORRELATION = 1/(n-1)\*CONTROL\_SUM

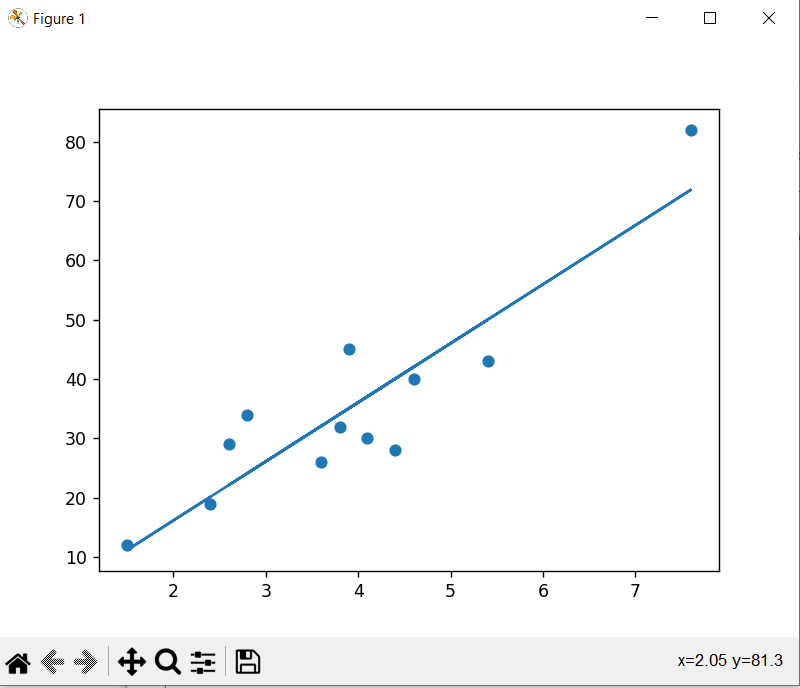
round CORRELATION

**5) Випробування алгоритму**

Результати для зчитаного файлу “input\_10.txt”:



Діаграма розсіювання:



**6)Висновки:** протягом виконання даної практичної лабораторної роботи було обраховано на конкретному прикладі(на заданих даних) центр ваги та коваріацію, обраховано рівняння регресії для набору даних, коефіцієнт кореляції між даними та побудовано діаграму розсіювання і проаналізовано її.