|  |
| --- |
| **Звіт**  **до лабораторної роботи № 1**  з дисципліни  **«Ймовірнісні основи програмної інженерії»**  **Студента групи ІПЗ-21**  **Корсуна Максима Анатолійовича**  **на тему:**  **«Центральні тенденції та міра дисперсії»** |

**1)Назва:** Центральні тенденції та міра дисперсії

**Мета :** навчитись використовувати на практиці набуті знання про центральні тенденції та міри.

**2)Постановка задачі:**

У нас є три тестових файла (“input\_10.txt”,“input\_100.txt”,“input\_1000.txt”) у яких у кожному рядку записано індекс відповідного фільму. З цих файлів ми будемо брати потрібні дані ,обробляти їх відповідно до задач лабораторної роботи,та повертати результат до файлу “results.txt”.Для читання з файлу у нас буде функція GetDataFromFile(path),яка приймає назву файла та повертає словник у якого ключем буде індекс фільму,а значенням кількість переглядів фільму.Для побудови гістограми буде використано бібліотеку Python -matplotlib.

**3)Побудова математичної моделі**

**Задача №1:**

**Умова:** Побудувати таблицю частот та сукупних частот для переглянутих фільмів. Визначити фільм, який був переглянутий частіше за інші.

**Математична модель:**

Отримавши з файлу словник з переглядами всіх фільмів ми вже маємо масив з частотами фільмів.Щоб отримати масив з відносними частотами нам потрібно знайти загальну кількість фільмів та для кожного фільма поділити його частоту на загальну кількість фільмів.А для того,щоб отримати масив з кумулятивними частотами потрібно для кожного фільму додавати кумулятивну частоту попереднього фільму (кумулятивна частота нульового фільму дорівнює його частоті).

Для знаходження фільму,який переглянули найбільше,потрібно знайти найбільшу частоту серед усіх елементів.

**Задача №2:**

**Умова:**Знайти Моду та Медіану заданої вибірки.

**Математична модель:**

Щоб знайти моду потрібно знайти фільми з найбільшою частотою.Потрібно зауважити,що якщо у кожного фільму частота дорівнює 1,то моди не існує,а якщо всі елементи мають однакову частоту то всі вони є модою.

Для знаходження медіани потрібно знайти загальну кількість переглядів та знайти фільм ,який знаходиться за центральним індексом.Якщо кількість переглядів парне число,то ділимо результат суми двох центральних елементів на 2.

**Задача №3:**

**Умова:** Порахувати Дисперсію та Середнє квадратичне відхилення розподілу

**Математична модель:**

Щоб знайти дисперсію спочатку необхідно знайти середнє значення ,тобто просумувати значення індексів фільмів помноженних на їхню частоту та поділити на кількість переглядів.Потім ми знаходимо результат суми від множення частоти фільму на його індекс у квадраті.Поділивши цей результат на загальну кількість елементів та віднявши від результату середнє значення у квадраті ми отримаємо дисперсію.

Для знаходження середнього квадратичного відхилення необхідно просто внести результат дисперсії під корінь.

**Задача №4:**

**Умова:** Побудувати гістограму частот для даного розподілу

**Математична модель:**Для побудови гістограми нам необхідно знайти частоту кожного фільму та потім відшукати фільм з найбільшою частотою щоб визначити масштаб гістограми.

**Задача №5:**

**Умова:** Зробити висновок з вигляду гістограми, про закон розподілу.

**Математична модель:**З вигляду гістограми можна побачити наскільки частота одного фільму більша або менша за частоту будь-якого іншого.Тобто фільм з найменшою частотою є найменш ймовірним та навпаки.

**4)Псевдокод алгоритму**

**Задача №1:**

def GetCumulativeFrequency(movieViews):

cumulativeFrequencies={}

tmp=0

for elem in movieViews:

cumulativeFrequencies[elem]=movieViews[elem]+tmp

tmp=cumulativeFrequencies[elem]

return cumulativeFrequencies

**Задача №2:**

int Median(int[] movieViews) {

int middleInd = GetSum(movieViews) / 2 + 1

int i = 0

while middleInd>CumulativeFrequency(movieViews)[i] {

i++;

}

return i;

}

//функція яка ще використовується в програмі за допомогою якої ми шукаємо моду

int MostViewdMovie(int[] movieViews) {

int max = max(movieViews)

i=0

for elements in movieViews {

if movieViews[i] == max then

return i

}

}

**Задача №3:**

double GetDispersion(path) {

int[] movieViews = GetDataFromFile(path);

average = 0

var = 0

i=0

for elements in movieViews {

average += movieViews[i] \* i

i++

}

average /= GetSum(movieViews)

i=0

for elements in movieViews {

var += movieViews[i] \* i^2

}

var = (var / GetSum(movieViews) - average^2

return var

}

double GetStandartDeviation(path) {

return sqrt(GetDispersion(path))

}

**Задача №4(виконана на Python):**

for i in range(movieViews)

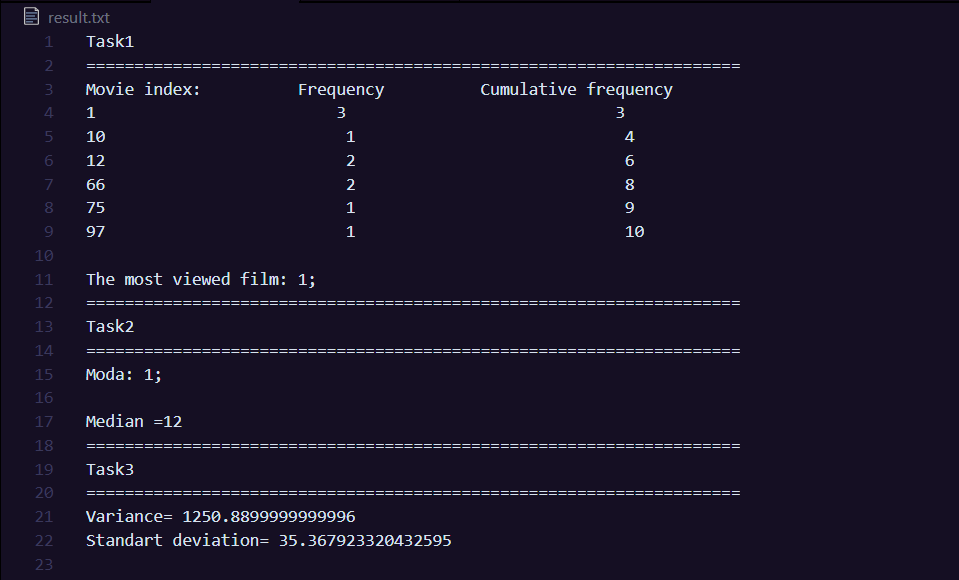
for j in range(movieViews[i])

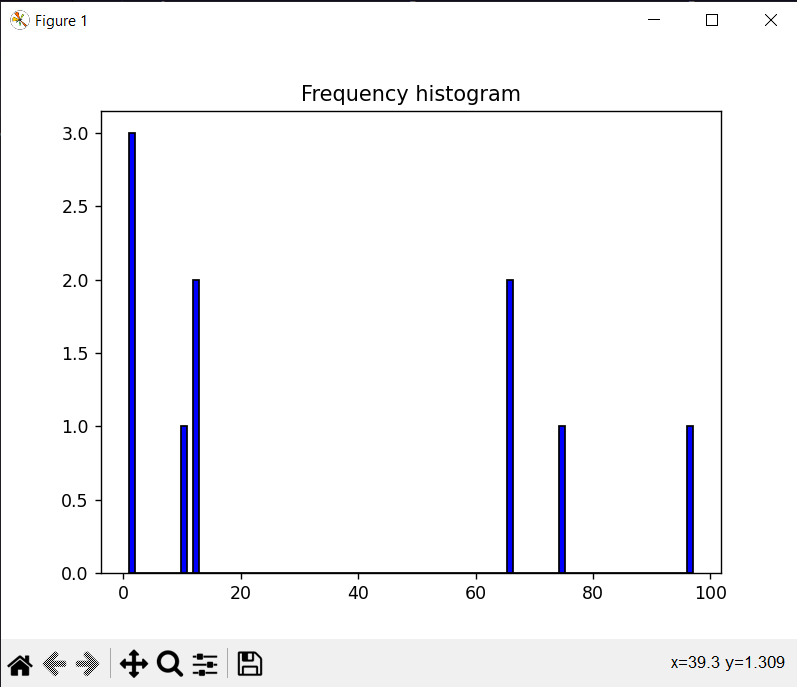
allViewsArr[i]=i+1

plt.hist(allViewsArr,bins=len(movieViews))

**5) Випробування алгоритму**

Результати для файлу “input\_10.txt”



****

**6)Висновки:**на цій лабораторній роботі було виконано практичні завдання на знаходження різних видів частот.Також були завдання на знахоження медіани, моди, дисперсії та середнього квадратичного відхилення вибірки.Ці дані є основними у статистиці та дають змогу точно аналізувати різні види інформації.Також було побудовано гістограму частот,яка графічно показує кількість перегляду кожного фільму.