|  |
| --- |
| **Звіт**  **до лабораторної роботи № 1**  з дисципліни  **«Ймовірнісні основи програмної інженерії»**  **Студента групи ІПЗ-21**  **Павлюченка Артема Максимовича**  **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:**

Ми маємо три тестових файли (“input\_10.txt”,“input\_100.txt”,“input\_1000.txt”), в кожному з яких M+1 рядків з числами, М – кількість переглядів певного фільму. За допомогою методу open(input\_10.txt) і циклу for зчитуємо дані з файлу та записуємо в одновимірний масив, з я ким надалі працюватимемо. Через те що індексація масиву починається з 0, індекс першого фільму також 0, надалі так і буде використовуватися в усіх обрахунках. 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:** Побудувати таблицю частот та сукупних частот для переглянутих фільмів. Визначити фільм, який був переглянутий частіше за інші.

**Математична модель:**

Дані файлу я зчитав у одновимірний масив, і в таблиці частот дані одновимірного масиву являтимуть собою перший стовбець – Frequency. Для розрахунку відносної частоти була знайдена сума всіх частот та для конкретного фільму обрахована відносна частота методом ділення частоти цього фільму на суму всіх частот. Для знаходження кумулятивної частоти я створив окрему змінну, у якій накопичується сума частот перегляд фільмів. І до першого стовпця були занесення безпосередньо фільми наступним чином: Index of 0 movie. Для знаходження найбільш переглянутого фільму я скористався методом max для масиву з частотами переглядів фільмів.

**Задача №2:**

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

**Математична модель:**

Для знаходження моди потрібно знайти фільм з найбільшою частотою переглядів , це ми вже проробили наприкінці 1 завдання.

Для знаходження медіани я створив циклом for новий масив, розміром з суму всіх частот, у який передавав індекси всіх фільмів, починаючи з 0, а кількістю конкретного фільму у масиві виступала частота його переглядів. Потім для знаходження індексу медіани, я поділив довжину цього масиву навпіл та розглянув 2 випадки: парна та непарна довжина масиву.

**Задача №3:**

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

**Математична модель:**

Для знаходження дисперсії спочатку потрібно обрахувати середнє значення вибірки, тобто скласти всі фільми, помноживши їх на відповідну кількість переглядів і поділити на суму всіх переглядів. Далі, за формулою треба скласти всі частоти переглядів фільмів, помножені на відповідні квадрати індексів цих фільмів, поділити на суму всіх переглядів та відняти квадрат середнього значення вибірки.

А для знаходження середнього квадратичного відхилення потрібно взяти корінь з обрахованої дисперсії.

**Задача №4:**

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

**Математична модель:**

Для побудування гістограми частот розподілу треба скористуватися вбудованими методами, для яких треба створити масив індексів всіх фільмів і використати масив завдання 2 – з усіма індексами фільмів з відповідною їх кількістю.

**Задача №5:**

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

**Математична модель:**

Закон розподілу – це таке певне поняття, яке показує множину можливих подій з ймовірністю їх настанні. Проаналізувавши гістограму, можу сказати, що фільм з найбільшою кількістю переглядів(тобто мода вибірки) відповідає найвищому стовпцю гістограми, найменшу ймовірність переглядів представляють відповідно найнижчі стовпці гістограми із найменшою відносною частотою.

**4)Псевдокод алгоритму**

**Задача №1:**

a=[]

i=0

for element in file{

a[i]=element

i++

}

i=0

sum = 0

for element in range(a[0]+1):{

sum+=a[i]

i++

}

cum\_sum=0

i=0

for element in range(a[0]+1){

cum\_sum+=a[i]

i++

}

max = a.index(max(a))

**Задача №2:**

new\_arr = 0

k=0

for element in range(length of a){

i=0

for next\_element in range(a[i]){

new\_arr[i]=k

i++

}

k+=+

}

if length of new\_arr /2 =0

median = (new\_arr[length of new\_arr/2]+new\_arr[length of new\_arr/2-1])/2

else

median = new\_arr[length of new\_arr/2-1]

mode= a.index(max(a))

**Задача №3:**

sum3=0

i=0

for element in range(length of a){

sum3+=a[i]

i++

}

average = sum3/ cum\_sum

sum3\_1=0

i=0

for element in range(length of a){

sum3\_1+=a[i] \* (i\*i)

i++

}

disperia = sum3\_1/cum\_sum-(average\*average)

standard\_deviation = sqrt(dispersia)

**Задача №4:**

movie\_mas=[]

i=0

for element in range(length of a + 1){

movie\_mas[i]=i

i++

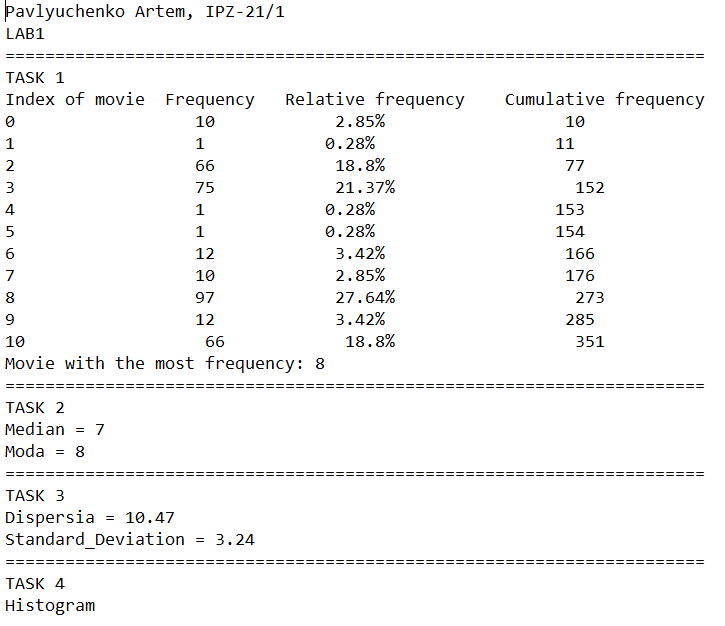
}

hist(new\_arr, movies\_mas)

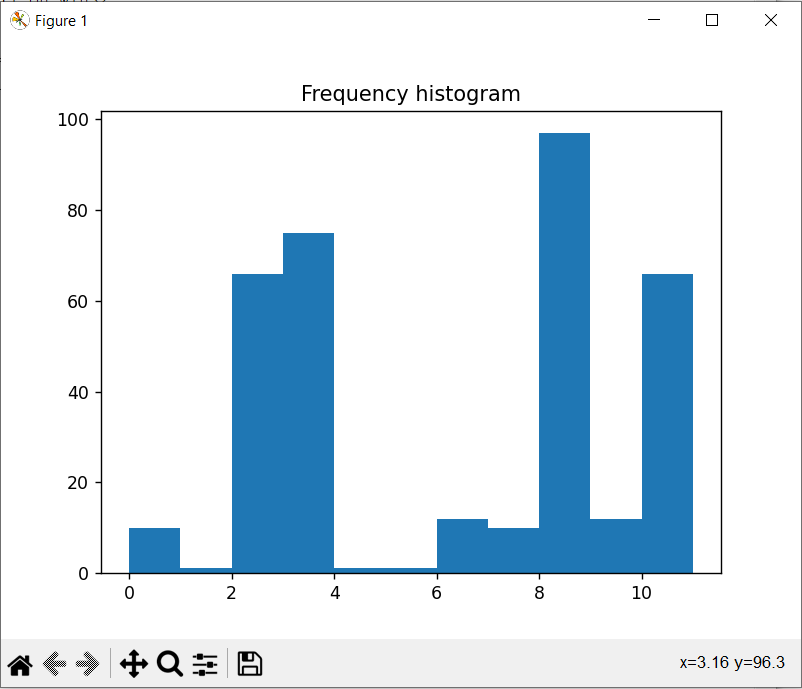
show()

**5) Випробування алгоритму**

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



Гістограма:



**6)Висновки:** протягом виконання даної практичної лабораторної роботи було обраховано на конкретному прикладі(на заданій вибірці) різні види частот – відносну та кумулятивну, моду та медіану заданої вибірки, дисперсію та середньоквадратичне відхилення цієї вибірки, побудовано гістограму на основі раніше обрахованих даних.