|  |
| --- |
| **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. 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 – Movie. Для розрахунку частоти була знайдена був створений цикл перебору масиву, під час якого елементи у одиничній кількості занесені у один масив, їх частоти у інший. Для знаходження кумулятивної частоти я створив окрему змінну, у якій накопичується сума частот переглядів фільмів. Для знаходження найбільш переглянутого фільму я скористався методом max для масиву з частотами переглядів фільмів і циклом while, якщо таких декілька.

**Задача №2:**

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

**Математична модель:**

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

Для знаходження індексу медіани, я поділив довжину масиву навпіл та розглянув 2 випадки: парна та непарна довжина масиву. За індексом вже знайдено саму медіану.

**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, у які передати зчитаний масив з фільмами, і масив діапазону від 1 до першого елемента файлу.

**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. У нашому випадку, події-це фільми. Проаналізувавши гістограму, можу сказати, що фільм з найбільшою кількістю переглядів(тобто мода вибірки) відповідає найвищому стовпцю гістограми, відповідну найменшу ймовірність переглядів представляють найнижчі стовпці гістограми.

**4)Псевдокод алгоритму**

**Задача №1:**

a=[]

for element in file{

push element in a

}

Pop 0 of a

sort a

unique\_massive = []

frequency\_massive=[]

quantity =0

for element in a:{

if element not in unique\_massive then push quantity in frequency\_massive

quantity = 1

push element in unique\_massive

else quantity++

}

Push quantity in frequency\_massive

Pop 0 from frequency\_massive

cum\_sum=0

for element in frequency\_massive{

cum\_sum+=element

}

Moda\_massive = []

Maximum = max of frequency\_massive

Foundex\_maximum = maximum

Help\_massive = []

For element in frequency\_massive{

Push element in help\_massive

}

K=0

While founded\_maximum=maximum then

If k>0 then push unique\_massive[founded\_maximum index of help\_massive+k]in moda\_massive

Else push unique\_massive[founded\_maximum index of help\_massive]in moda\_massive

Pop founded\_maximum index of help\_massive from help\_massive

K++

**Задача №2:**

if length of a /2 =0

median = (a[length of a/2]+a[length of a/2-1])/2

else

median = a[length of a/2-1]

**Задача №3:**

sum\_a = 0

i=0

for element in range(length of unique\_massive){

sum\_a+=element\*frequency\_massive[i]

i++

}

average = sum\_a/ length of a

sum\_for\_var=0

i=0

for element in range(length of frequency\_massive){

sum\_for\_var +=element \* (unique\_massive[i]\*unique\_massive[i])

i++

}

Var = sum\_for\_var/len(a)-(average\*average)

Round var

Standard\_Deviation = sqrt(Var)

Round Standard\_Deviation

**Задача №4:**

NEW\_MAS=[]

For I in 100{

Push I in NEW\_MAS

}

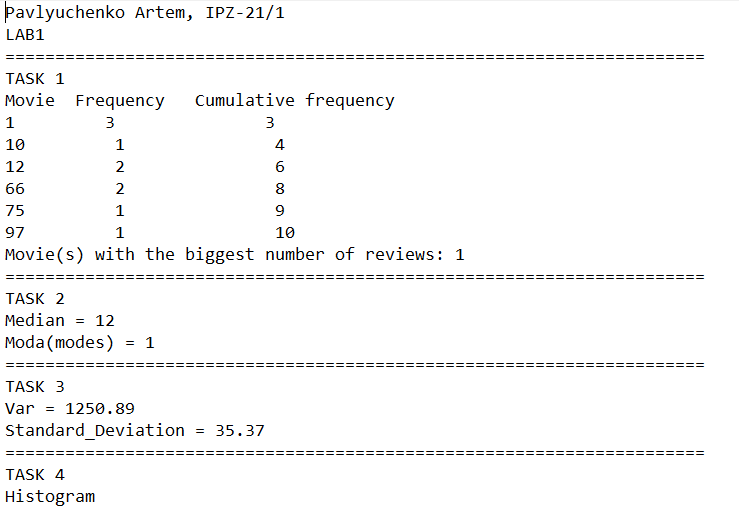
push (max of unique\_massive +a1) in unique\_massive

hist(a, NEW\_MAS)

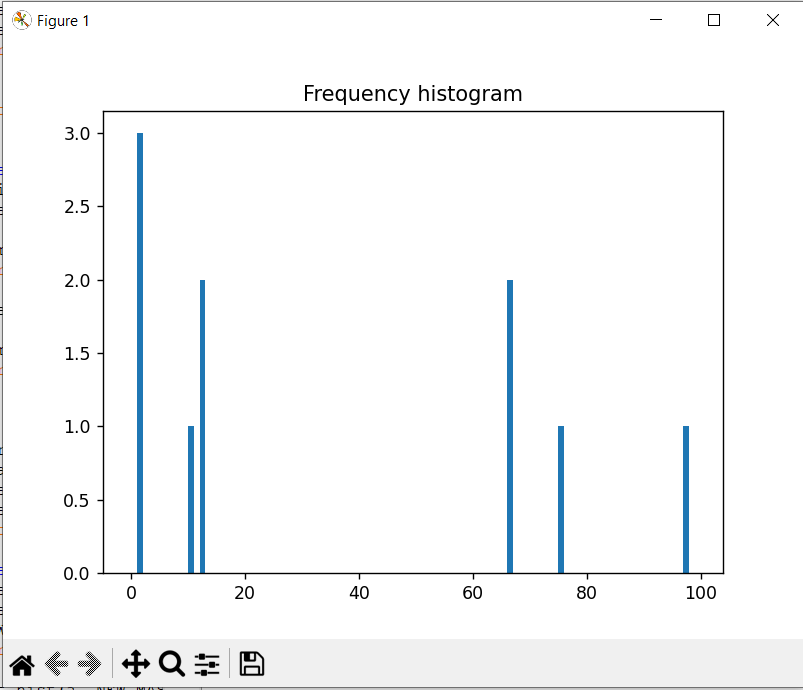
show()

**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) частоту, кумулятивну частоту, моду та медіану заданої вибірки, дисперсію та середньоквадратичне відхилення цієї вибірки, побудовано гістограму на основі раніше обрахованих даних.