|  |  |  |  |
| --- | --- | --- | --- |
| МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені Тараса Шевченка  ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ  **Кафедра програмних систем і технологій**  Дисципліна  **«Ймовірнісні основи програмної інженерії»**  **Лабораторна робота № 1**  **на тему:**  **«**Центральні тенденції та міра дисперсії**»** | | | |
| **Виконав:** | Литвинчук Владислав Валерійович | **Перевірила**: | Вечерковська  Анастасія  Сергіївна |
| Група | ІПЗ-21(1) | Дата перевірки |  |
| Форма навчання | денна | Оцінка |  |
| Спеціальність | 121 |
| 2022 | | | |

# **Мета:**

Навчитись використовувати на практиці набуті знання про центральні тенденції та міри дисперсії.

# **Постановка задачі:**

1. Побудувати таблицю частот та сукупних частот для переглянутих фільмів. Визначити фільм, який був переглянутий частіше за інші.
2. Знайти Моду та Медіану заданої вибірки.
3. Порахувати Дисперсію та Середнє квадратичне відхилення розподілу.
4. Побудувати гістограму частот для даного розподілу.
5. Зробити висновок з вигляду гістограми, про закон розподілу.

# **Математична модель:**

Для знаходження правильних відповідей для даної лабораторної роботи, необхідно використовувати формули статистики, наведені нижче:

***Сукупна частота****:*

де – сукупна частота елементу , k – кількість елементів у масиві частот, – частота появи цих елементів, i – індекс частоти j-го елементу.

***Мода****: ,* тобто мода – елемент масиву значень, який зустрічається в ньому найчастіше

***Медіана:***

***Середнє:***

де n – кількість елементів у масиві значень, – значення елементу.

***Середнє квадратичне відхилення:***

*де n – кількість елементів у масиві значень,*  – значення i-го елементу, x – середнє значення масиву елементів.

***Дисперсія:****,* де – квадрат середнього квадратичного відхилення розподілу

# **Псевдокод:**

1. Знаходження частоти

def getDictOfFrequencies(data\_set):  
 for номер фільму in масив значень:  
 if номер фільму in словнику частот:  
 додаємо до поточного фільму частота + 1

else:  
 додаємо поточний фільм в словник частот  
 return frequency\_dict

1. Знаходження кумулятивної частоти

def getDictOfCumulativeFrequencies(data\_set):  
 отримуємо словник частот  
 попередня кумулятивна частота = 0  
 for film\_number, frequency in словник частот:  
 поточна кумулятивна частота = частота + попередня кумулятивна частота

попередня кумулятивна частота = поточна кумулятивна частота  
 return cumulative\_frequency\_dict

1. Знаходження моди розподілу:

def getMode(data\_set):  
 Отримуємо словник частот  
 максимальна частота появи = 0  
 for фільм, частота in словник частот:  
 if частота > максимальна частота:  
 очистити modes  
 максимальна частота = частота  
 додаємо поточну частоту фільму в результуючий modes  
 elif частота = максимальна частота:  
 додаємо поточну частоту фільму в результуючий modes  
  
 if максимальна частота = 1:  
 мода = 'There is no mode in current distribution'  
 else:  
 мода = modes

1. Знаходження медіани розподілу:

def getMedian(data\_set, data\_length):  
 Сортуємо масив  
 Центральний індекс = ціла частина від довжина масиву / 2

if якщо довжина масиву – парне число:  
 медіана = елементу масиву з центральним індексом  
 else:  
 медіана = середнє значення між елементом з центральним індексом і наступним після нього

return median

1. Знаходження дисперсії:

def getVariance(data\_set):  
 Отримуємо середнє квадратичне відхилення  
 Дисперсія = квадрат середнього квадратичного відхилення  
 return variance

1. Знаходження середнього квадратичного відхилення:

def getStandardDeviation(data\_set, data\_length):  
 # getting average value of the data set  
 отримуємо середнє значення масиву  
 сума девіацій = 0  
 for value in data\_set:  
 сума девіацій += квадрат різниці поточного значення і середнього  
 середньо квадратичне відхилення = квадрат суми девіацій розділена на кількість елементів

return standard\_deviation

1. Знаходження середнього значення:

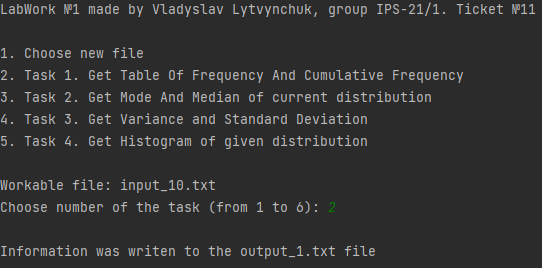
def getMean(data\_set, data\_length):

сума елементів = 0  
 for value in data\_set:

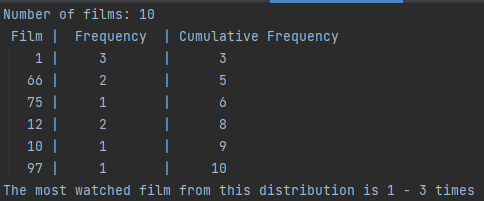
сума елементів += value   
 середнє = сума елементів / кількість елементів  
 return mean

# **Випробування алгоритму:**

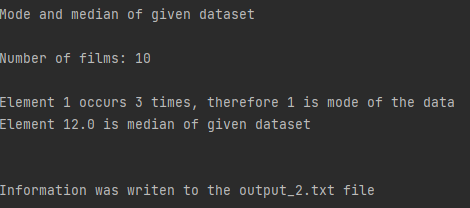
Консоль

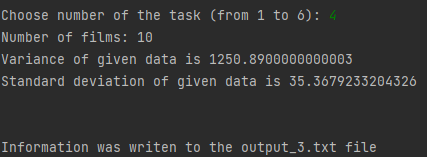


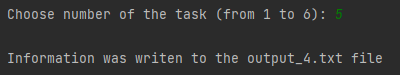
Файл output\_1.txt



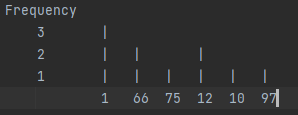
Консоль



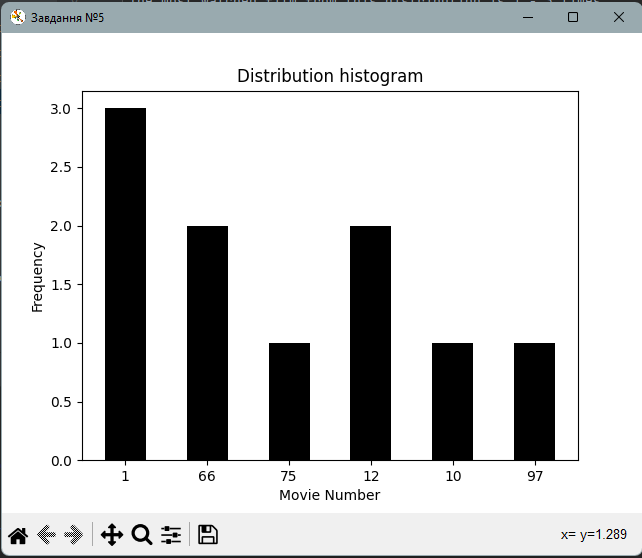




Файл output\_4.txt



Викликана програма для відображення гістограми



Гістограма не є рівномірною, має перекос в праву сторону, не має кластерів і пробілів.

# **Виcновок:**

Під час виконання даної лабораторної роботи було проведено аналіз алгоритмів і формул, необхідних для знаходження рішень. Також було розроблено алгоритми для роботи з центральними тенденціями та мірою дисперсії, а саме: розрахунок і побудова таблиці частот та сукупних частот, знаходження моди та медіани розподілу, розрахунок дисперсії та середньо квадратичного відхилення розподілу, а також побудова гістограми частот і визначення її виду.