|  |  |  |  |
| --- | --- | --- | --- |
| 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**  «Центральні тенденції та міра дисперсії» | | | |
| **Виконав:** | Мичковський Богдан Ігорович | **Перевірила**: | Вечерклвська  А. С. |
| Група | ІПЗ-22 | Дата перевірки |  |
| Форма навчання | денна | Оцінка |  |
| Спеціальність | 121 |
| 2022 | | | |

**Мета:**

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

**Завдання:**

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.

**Формат вхідних/вихідних даних:**

Розроблена програма повинна зчитувати вхідні дані з файлу заданого формату та записувати дані у файл.

Вхідний файл представляє собою текстовий файл із М+1 рядків. Перший рядок містить число: M, де M —кількість переглядів і-того фільму.

До документа завдання також додаються приклад вхідних файлів різної

розмірності.

**Постановка задачі:**

Для того щоб побудувати таблицю частот та сукупних частот необхідно відсортувати значення кількості переглядів фільмів, прочитаних з наведених в завданні файлів. Фільм, який був переглянутий частіше за інші являє собою моду. Для виконання другого та завдання нам потрібно провести математичні обчислення для знаходження моди, медіани, дисперсії та середнє квадратичне. Використовуючи графічні можливості мови програмування, створити та відобразити гістограми.

**Побудову математичної моделі:**

Для виконання лабораторної роботи було використано формули з якими було ознайомлено під час прослуховування лекційного матеріалу та додаткового освоєння під час виконання даної роботи.

**Формули:**

Мода:

Для цього ми просто перебираємо всі елементи та записуємо той(ті) елемент(-ти), які мають найбільшу кількість повторів та якщо їх більше за 1.

Медіана:

Медіана - це центральний елемент з вибірки у якій непарна кількість елементів, та сума двох центральних елементів розділених на два, якщо парна кількість елементів.

Text

Description automatically generated

Дисперсія:

Logo

Description automatically generated

Середнє квадратичне відхилення:

A picture containing text

Description automatically generated

Серед

A picture containing text

Description automatically generated

**Псевдокод алгоритму:**

1. Зчитуємо значення вибірки з файлів. Зберігаємо в список.
2. Видаляємо пусті рядки у файлі.
3. Застосовуємо бульбашкове сортування для виставлення елементів в правильному порядку для подальших операцій.
4. (Завдання 1)

Створюємо список для обрахування частоти елементів:

Поки перебираєм всі елементи списку;

Якщо елемент співпадає, то ми додаємо 1 до його частоти.

Якщо ні, то ми зберігаємо це число в список та його частоту.

Зберігаємо дані в файл.

Повертаємо список відсортованих елементів та їх частот.

1. Виводимо табличку з результатами.
2. (Завдання 2)

(Мода)

Поки перебираєм елементи

Якщо попередній найбільший елемент менший поточного:

Максимальний = поточний

(Медіана)

Якщо довжина списку поділити на 2 парна;

То середнє = центральне ліве –центральне праве поділити на 2 по модулю.

Якщо довжина непарна;

То середнє = середнє.

1. (Завдання 3)

(Дисперсія, середнє квадратичне відхилення)

За формулою обрахуємо чисельник.

За формулою обрахуємо знаменник.

Підставимо отримані значення в формулу дисперсії наведеної вище.

Обрахуємо Середнє квадритичне значення за формулою наведеною вище.

1. (Завдання 4)

Використовуючи бібліотеку matplotlib.pyplot, побудуємо графік для трьох файлів, підписавши його осі.

**Випробовування алгоритму:**

Для виконання випробовуваних тестів було використано список з десяти значень, через те що інші списки відображають занадто багато елементів.

В даний список входять такі елементи: 1, 1, 1, 10, 12, 12, 66, 66, 75, 97

Text

Description automatically generated

З результату виконання програми видно, що модою з даного переліку елементів є число 1 так як його частота найбільша.

Медіаною є число 12.0 через те, що елемент 12 знаходиться по обидві сторони від центру парного списку.

Для знаходження дисперсії скористаємося онлайн калькулятором:

A picture containing text, screenshot, aquatic bird

Description automatically generated

Результат спів пав, отже обрахунки вірні.

За таким самим принципом знайдем середнє квадратичне:

Text

Description automatically generated with medium confidence

Результат спів пав, отже обрахунки вірні.

Гістаграма для даної вибірки матиме поданий нище вигляд.

Chart, histogram

Description automatically generated

Було проведено декілька спроб зобразити графік у вигляді “Кластерів та щілин”, який би ідеально описував потрібні проміжки на шрафіку, наприклад [0 – 10), [10 – 20), проте зіткнувся з тим що не знаю як це реалізувати у мові python

**Висновок:**

Під час даної лабораторної роботи було сформовано навички використання мови Python для виконання нескладних математичних операцій, набутих на лекціях знань про основні тенденції та міри. Проведено перше ознайомлення з побудовою гістаграм та графіків в середовищі PyCharm.