|  |  |  |  |
| --- | --- | --- | --- |
| МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені Тараса Шевченка  ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ  **Кафедра програмних систем і технологій**    Дисципліна  **«Ймовірносні основи програмної інженерії»**  **Лабораторна робота № 1**  **Центральні тенденції та міра дисперсії** | | | |
| **Виконав:** | Бойчук Максим Кирилович | **Перевірила**: | Вечерковська А.С. |
| Група | ІПЗ-23 | Дата перевірки |  |
| Форма навчання | денна | Оцінка |  |
| Спеціальність | 121 |
| 2022 | | | |

**Мета роботи :**

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

**Постановка задачі :**

* Для початку потрібно створити файл з вхідними данними. Потім запустимо середовище розробки та приступимо до першого завдання. Запишемо вхідні данні у одномірний масив та розраховуємо сукупну частоту. Для визначення фільма, який має найбільше переглядів використаємо функцію data.index(max(data)).

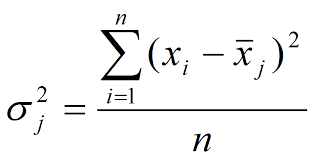
Перед тим як приступити до наступних завдань повторимо матеріали лекцій та згадаємо що таке мода, медіана, дисперсія та середнє квадратичне відхилення.

* Мода — значення випадкової величини, що трапляється найчастіше в сукупності спостережень.
* Медіана — величина ознаки, що розташована посередині ранжованого ряду вибірки
* Диспе́рсія — це міра розсіяння значень випадкової величини відносно середнього значення розподілу.
* Середнє квадратичне відхилення — показник розсіювання значень випадкової величини відносно її  центру розподілу.

Тепер, знайдемо інформацію про те, які існують алгоритми пошуку даних невідомих і почнемо будувати математичну модель.

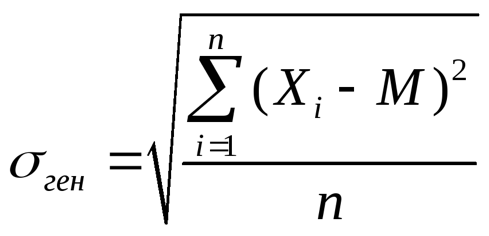
**Математична модель :**

* Для пошуку моди нам потрібно пробігтись по масиву та знайти ті елементи , які повторюються частіше за інші. Для цього використаємо функцію Counter, а потім за допомогою циклу та функції most\_common() порівняємо кількість повторень кожного зі значень у масиві.
* Щоб знайти медіану нам потрібно упорядкувати масив або по зростанню, або по спаданню та знайти елемент, що знаходиться по середині вибірки. Відсортовуємо по зростанню та створюємо алгоритм пошуку центрального елемента : ділимо кількість елементів у масиві на 2 та шукаємо елемент під таким індексом. Додатково беремо до уваги випадок при якому кількість елементів у вибірці не парна. У такому разі ми знайдемо індекс посередині між двома значеннями. Обрахуємо їх суму та розділимо на 2. Це і буде нашою медіаною.
* Розрахуємо дисперсію.



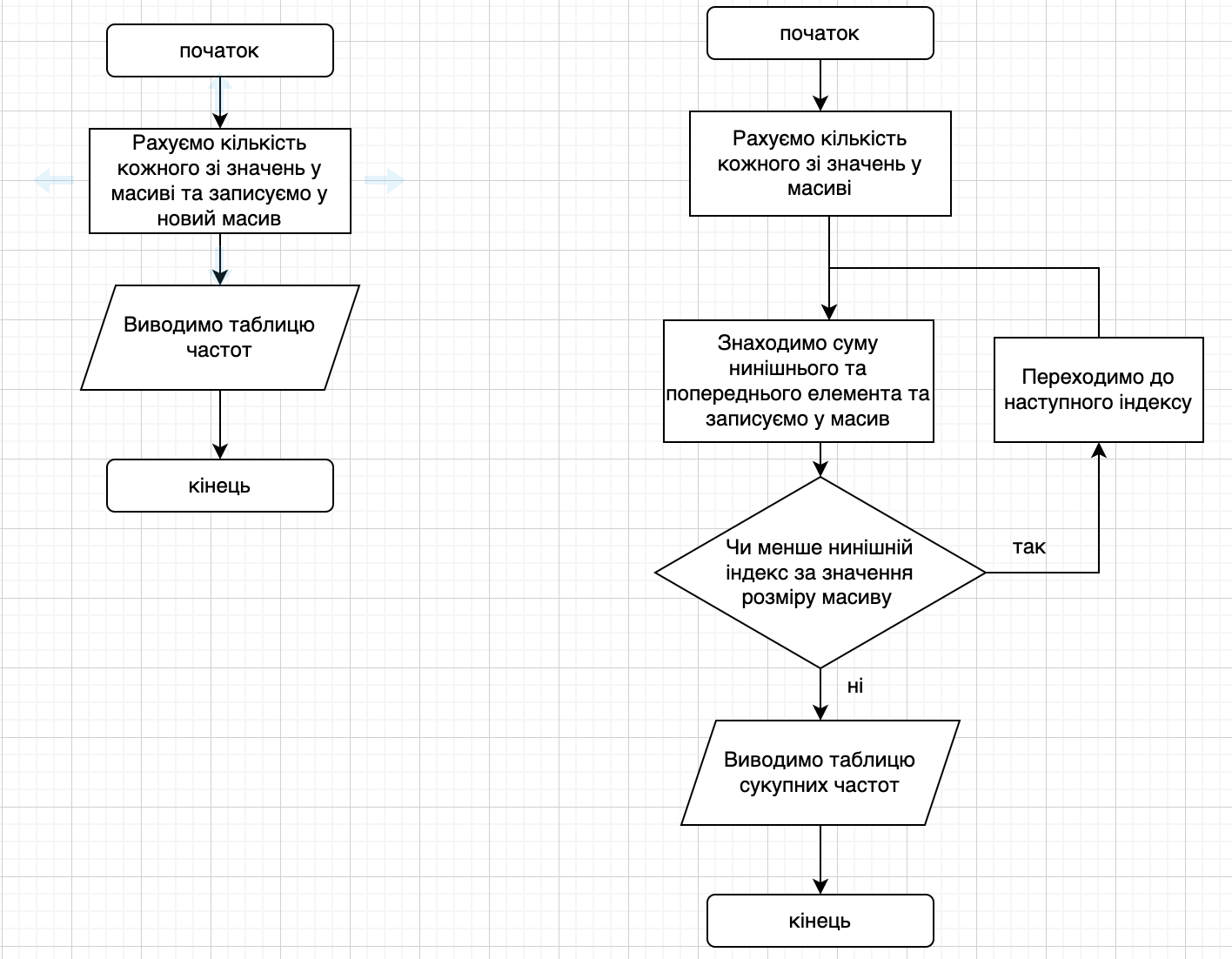
Спочатку знайдемо кількість данних у вибірці, потім знайдемо середнє значення поділивши суму елементів на їхню кількість. Тепер знайдемо квадратичне відхилення кожного з елементів за алгоритмом : число мінус середнє значення вибірки й зведемо кожен з результатів до квадрату. Тепер залишилось знайти суму кожного з відхилень у масиві та поділити її на кількість значень. Дисперсія розрахована.

* Для пошуку середнього квадратичного відхилення розподілу нам потрібно взяти корінь з дисперсії.

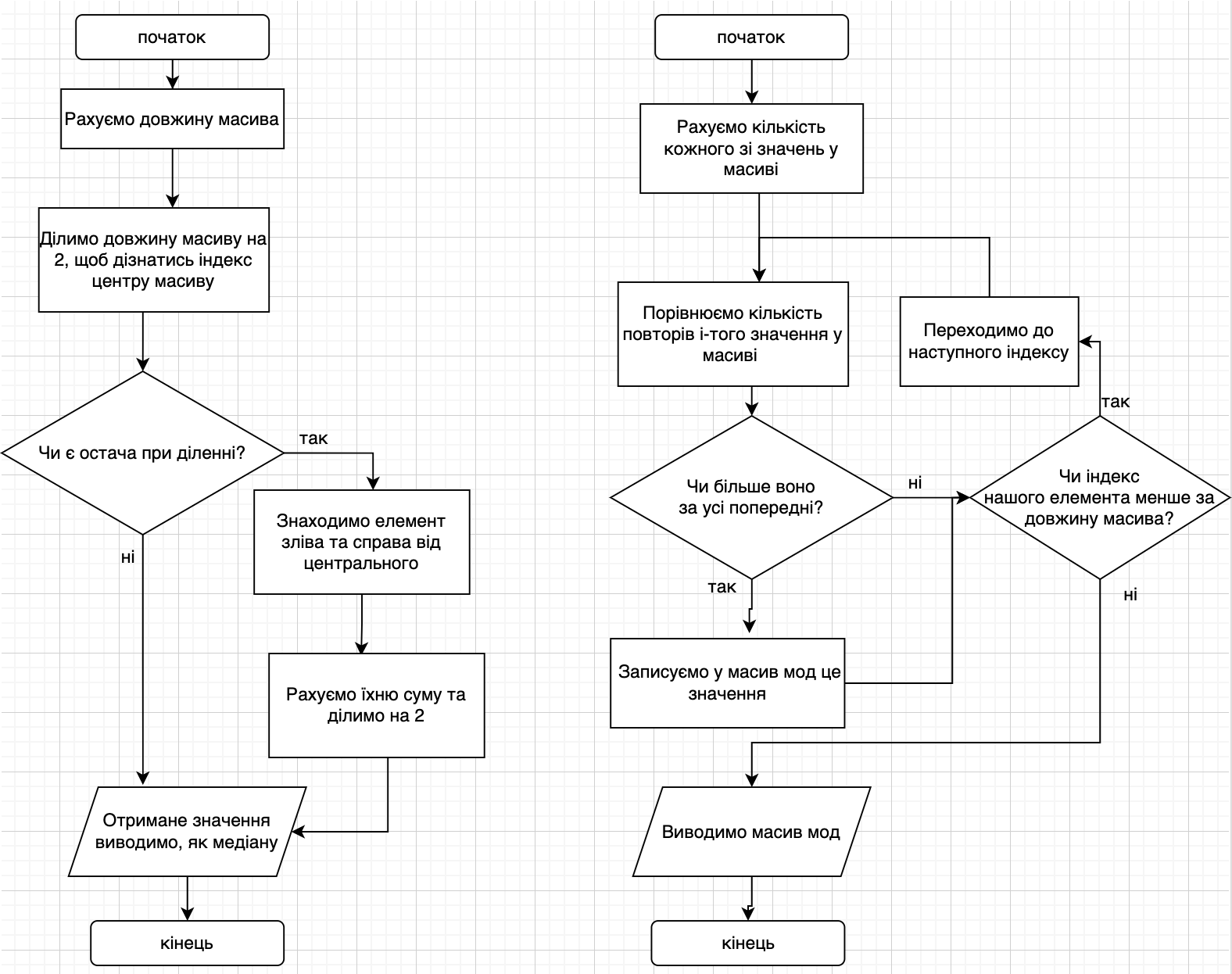


**Блок-схема алгоритму :**

**Завдання №1**

****

**Завдання №2**

****

**Завдання №3**

****

**Випробовування алгоритмів :**

**Завдання №1**

**Изображение выглядит как текст

Автоматически созданное описаниеИзображение выглядит как текст

Автоматически созданное описание**

**Завдання №2**

**Изображение выглядит как текст

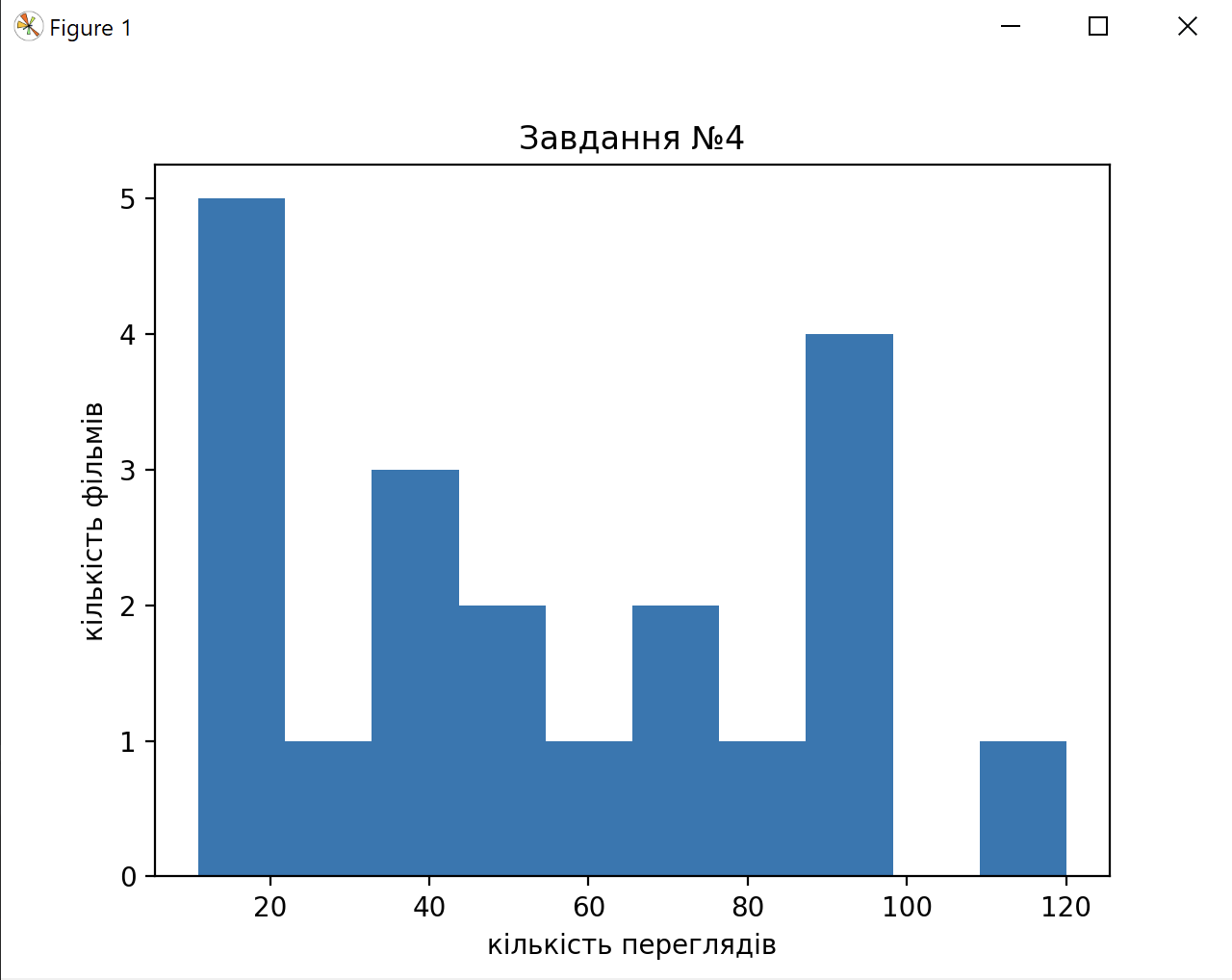
Автоматически созданное описание**

**Завдання №3**

**Изображение выглядит как текст

Автоматически созданное описание**

**Завдання №4**

****

**Завдання №5**

Гістограма каскадна й нахилена вліво. Найбільше y-значень знаходяться у діапазоні від 10 до 25 – 5. Найменше – {22;33} , {54;64} , {76;87} , {98;120} – по 1 випадку.

Є проміжок без значень - {98;119}

Закон розподілу ймовірностей — це поняття теорії ймовірностей, яке для дискретної випадкової величини показує множину можливих подій з ймовірностями їхнього настання.

Закон розподілу часто використовується для характеризування випадкової величини, яка має не дуже велику кількість реалізацій.