Національний технічний університет України «КПІ»

Кафедра автоматизованих систем обробки інформації і управління

Лабораторна робота №1

з дисципліни «ЙМОВІРНІСНІ МОДЕЛІ ТА СТАТИСТИЧНЕ ОЦІНЮВАННЯ В ІНФОРМАЦІЙНО-УПРАВЛЯЮЧИХ СИСТЕМАХ»

на тему: «Вибірковий контроль при конструюванні та виробництві виробів

електроніки за допомогою статистичних методів»

Виконав:

студент групи ІС-23

Шимків М.В.

Викладач:

Богданова Н.В

Київ 2024

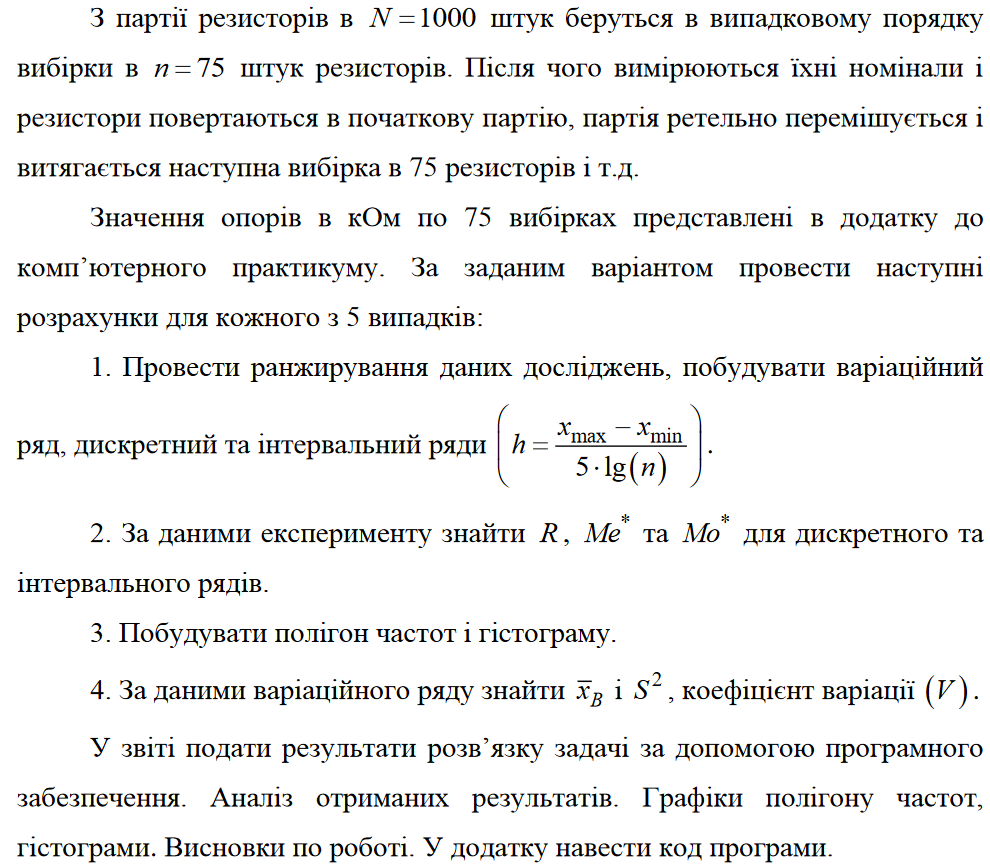
**Мета роботи**: ознайомлення з поняттям генеральної сукупності,

побудова дискретних та інтервальних статистичних рядів, статистичні оцінки

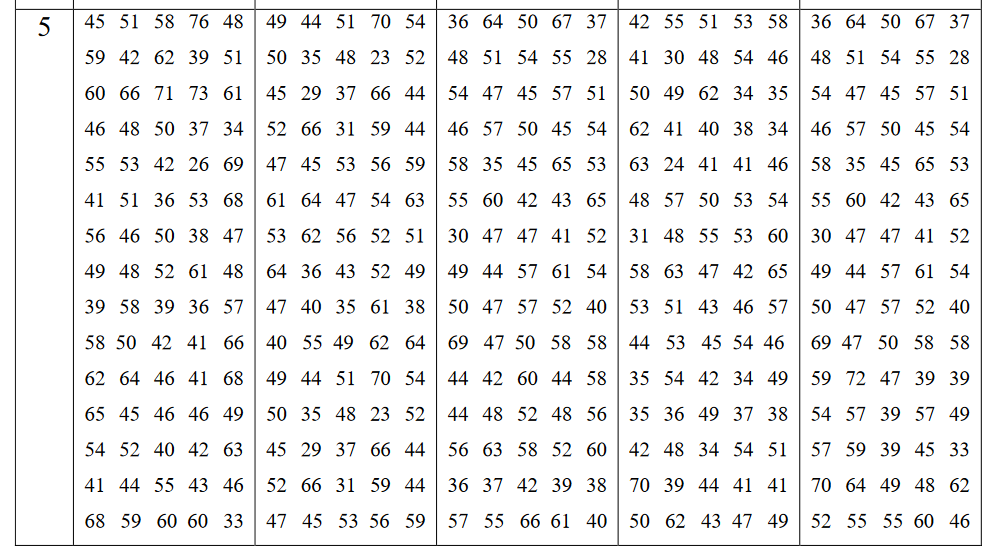
генеральної сукупності. Навчити студентів вмінню аналізувати варіаційні

ряди, визначати числові характеристики та будувати відповідні графіки.

**Завдання до роботи:**

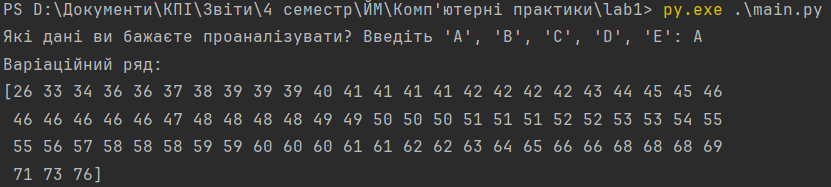


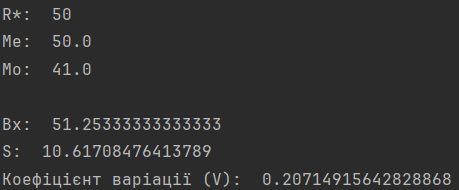
Варіант 25

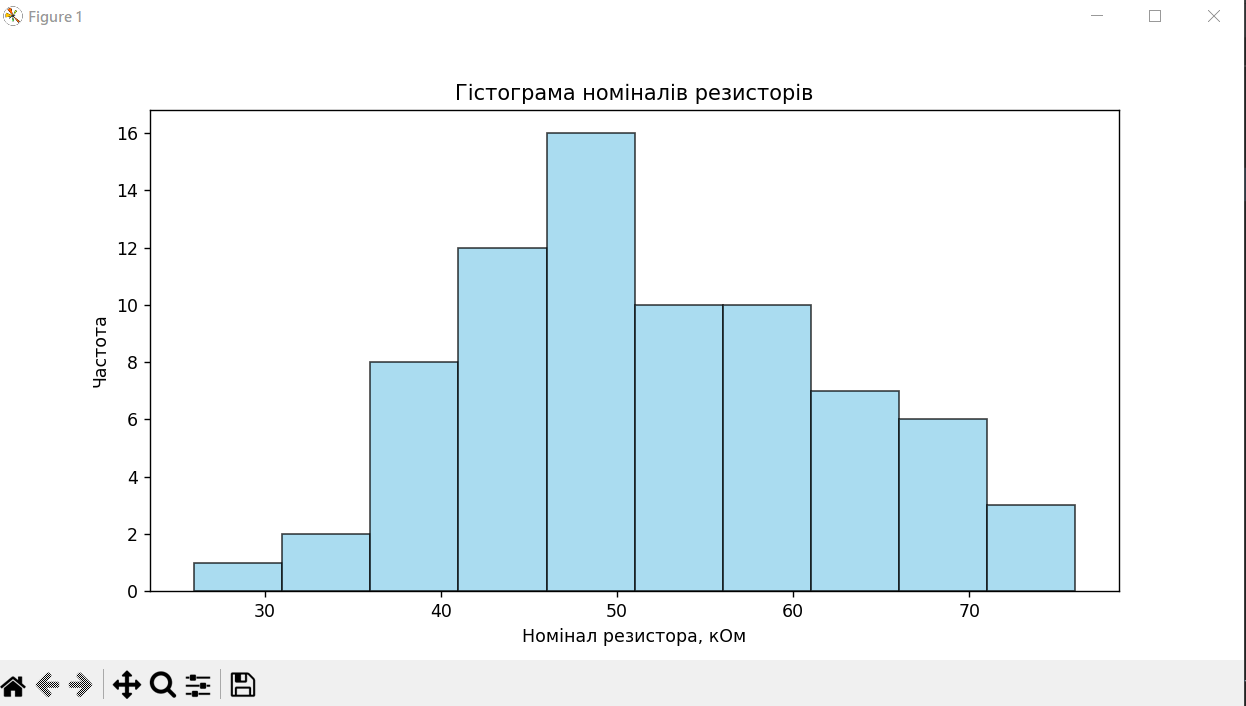


**Виконання:**

Результати для даних A:





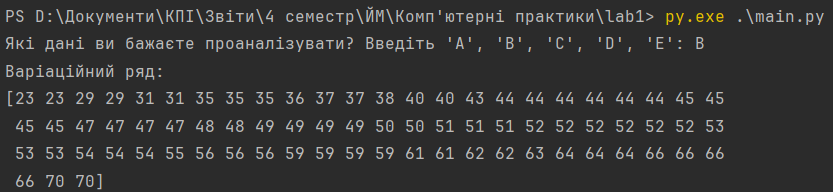


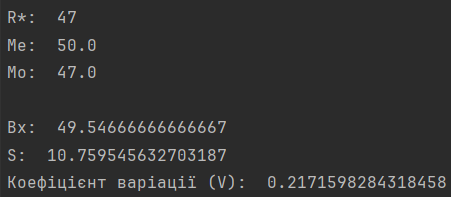
### Набір даних A:

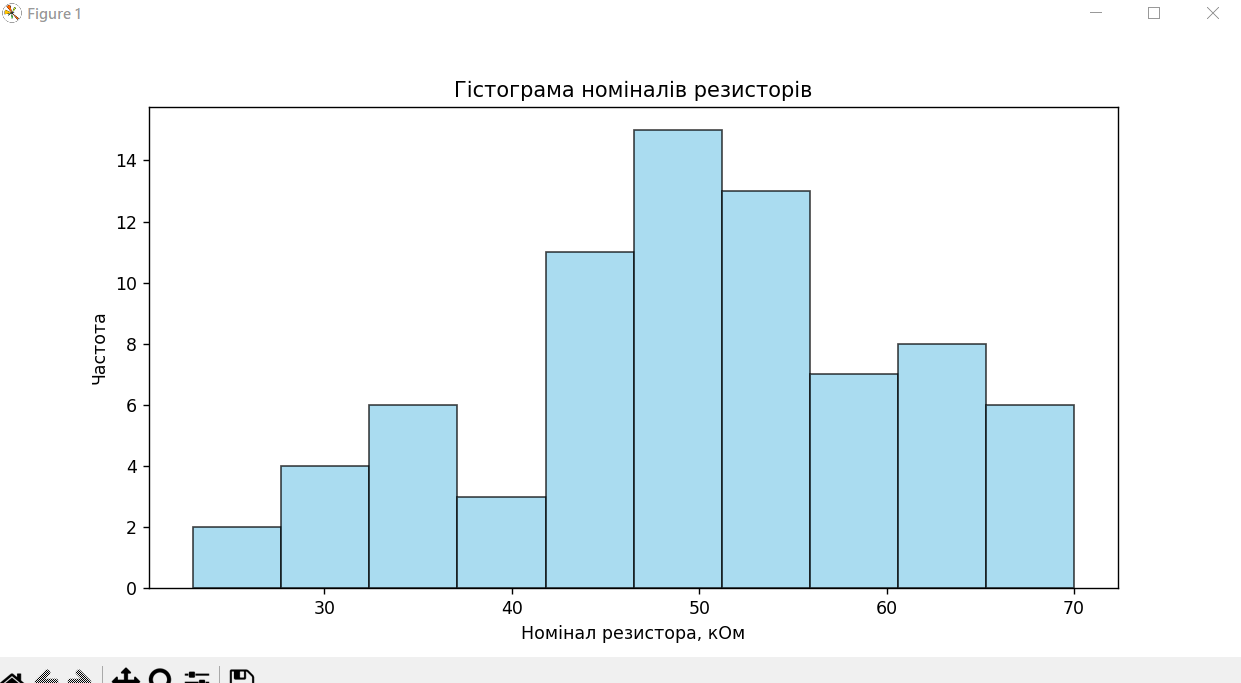
* \**Розмах (R*)\*\*: 50 - великий розмах вказує на значну різницю між найбільшим та найменшим значеннями в наборі.
* **Медіана (Me)**: 50.0 - це значення, яке розділяє набір даних на дві рівні частини. У цьому випадку, половина значень менше 50, а половина - більше.
* **Мода (Mo)**: 41.0 - це значення, яке зустрічається найчастіше. У цьому наборі даних це є найчастіше зустрічаючеся значення.
* **Середнє значення (Bx)**: 51.25 - це середнє арифметичне всіх значень. Це дає уявлення про середній номінал резисторів у цьому наборі.
* **Стандартне відхилення (S)**: 10.62 - велике значення стандартного відхилення вказує на значну варіацію в даних.
* **Коефіцієнт варіації (V)**: 0.207 - високе значення коефіцієнта варіації також вказує на значну варіацію в даних.

Отже, набір даних A має велику варіацію в номіналах резисторів, яка підтверджується великим розмахом та високим значенням стандартного відхилення. Мода близька до медіани, що може вказувати на деяку концентрацію резисторів навколо цієї точки. Однак, велика різниця між середнім значенням та модою може свідчити про наявність викидів або нерівномірного розподілу даних.

Результати для даних Б:





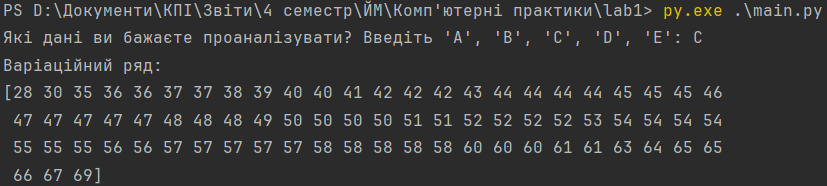


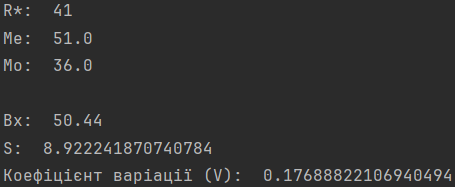
### Набір даних Б:

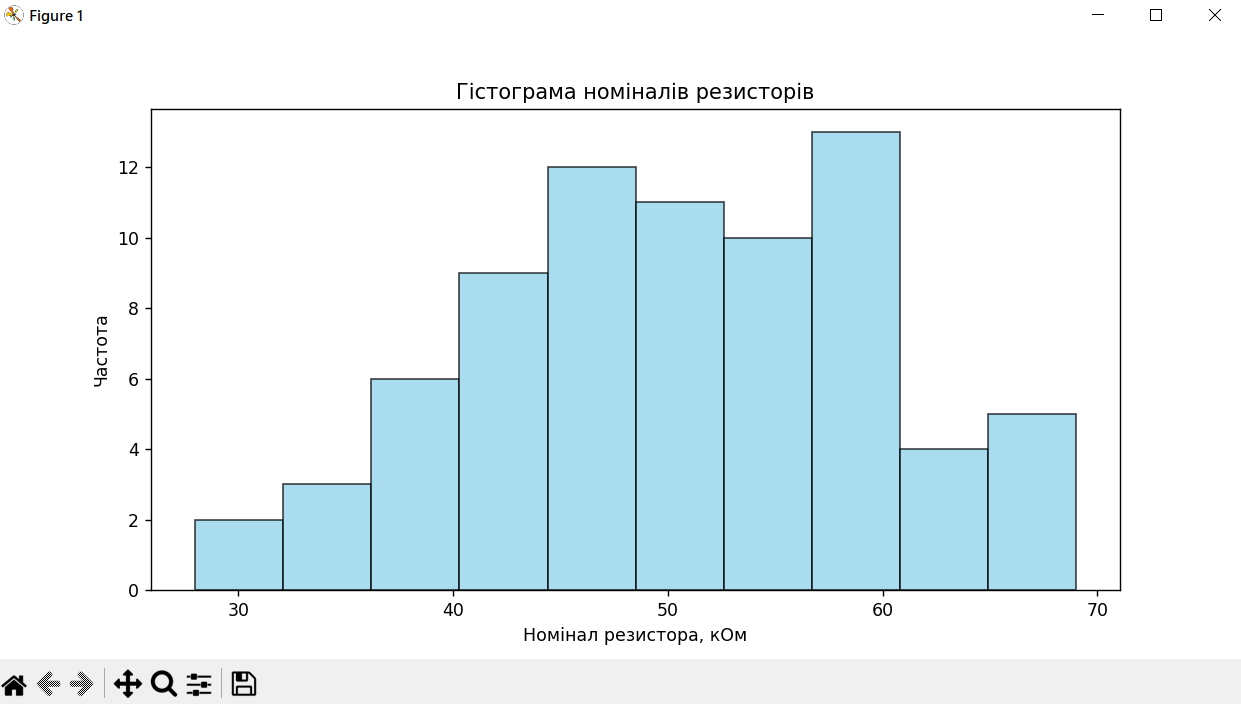
* \**Розмах (R*)\*\*: 47 - це великий розмах, що вказує на значну різницю між найбільшим та найменшим значеннями в наборі.
* **Медіана (Me)**: 50.0 - це значення, яке розділяє набір даних на дві рівні частини. У цьому випадку, половина значень менше 50, а половина - більше.
* **Мода (Mo)**: 47.0 - це значення, яке зустрічається найчастіше. У цьому наборі даних це є найчастіше зустрічаючеся значення.
* **Середнє значення (Bx)**: 49.55 - це середнє арифметичне всіх значень. Це дає уявлення про середній номінал резисторів у цьому наборі.
* **Стандартне відхилення (S)**: 10.76 - це міра розподілу даних відносно їх середнього значення. Велике значення стандартного відхилення вказує на значну варіацію в даних.
* **Коефіцієнт варіації (V)**: 0.217 - це відношення стандартного відхилення до середнього значення. Велике значення коефіцієнта варіації також вказує на значну варіацію в даних.

Отже, набір даних B має велику варіацію в номіналах резисторів, що підтверджується великим розмахом та високим значенням стандартного відхилення. Мода наближена до медіани, що може вказувати на деяку концентрацію резисторів навколо цієї точки. Однак, велика різниця між середнім значенням та модою може свідчити про наявність викидів або нерівномірного розподілу даних.

Результати для даних В:





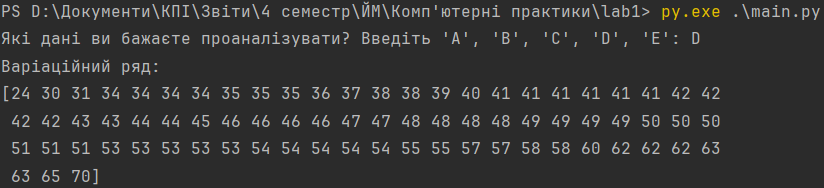


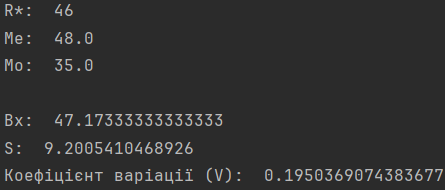
### Набір даних В:

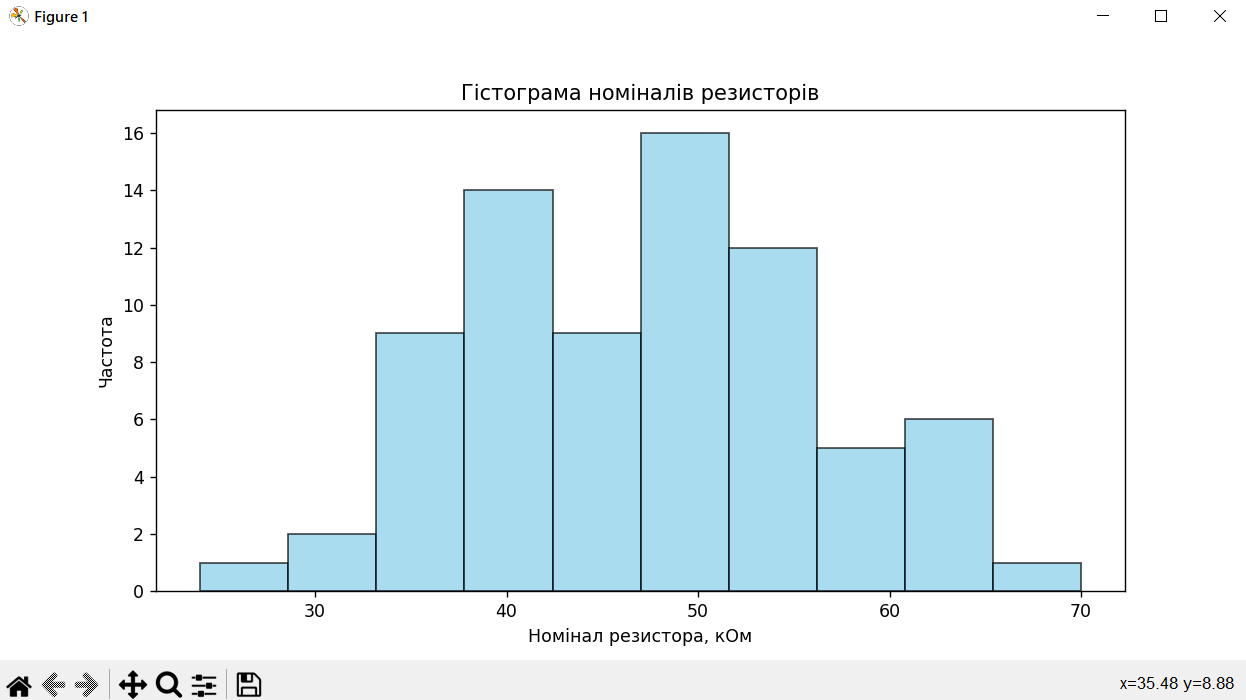
* \**Розмах (R*)\*\*: 41 - це великий розмах, що вказує на значну різницю між найбільшим та найменшим значеннями в наборі.
* **Медіана (Me)**: 51.0 - це значення, яке розділяє набір даних на дві рівні частини. У цьому випадку, половина значень менше 51, а половина - більше.
* **Мода (Mo)**: 36.0 - це значення, яке зустрічається найчастіше. У цьому наборі даних це є найчастіше зустрічаючеся значення, що може вказувати на те, що значна кількість резисторів має номінал близький до 36 кОм.
* **Середнє значення (Bx)**: 50.44 - це середнє арифметичне всіх значень. Це дає уявлення про середній номінал резисторів у цьому наборі.
* **Стандартне відхилення (S)**: 8.92 - це міра розподілу даних відносно їх середнього значення. Значення стандартного відхилення вказує на те, що в наборі даних є помірна варіація.
* **Коефіцієнт варіації (V)**: 0.177 - це відношення стандартного відхилення до середнього значення. Чим менше значення коефіцієнта варіації, тим менше варіація в даних.

Отже, набір даних C має помірну варіацію в номіналах резисторів, але все ж має великий розмах. Мода наближена до мінімального значення, що може вказувати на деяку концентрацію резисторів з номіналом навколо цієї точки. Медіана та середнє значення також дозволяють отримати уявлення про центральні тенденції в наборі даних.

Результати для даних Г:





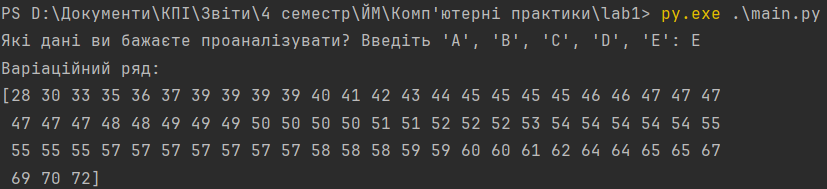


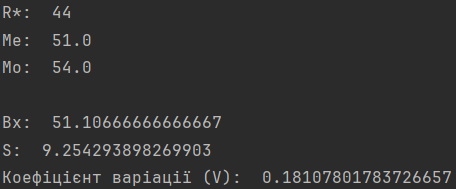
### Набір даних Г:

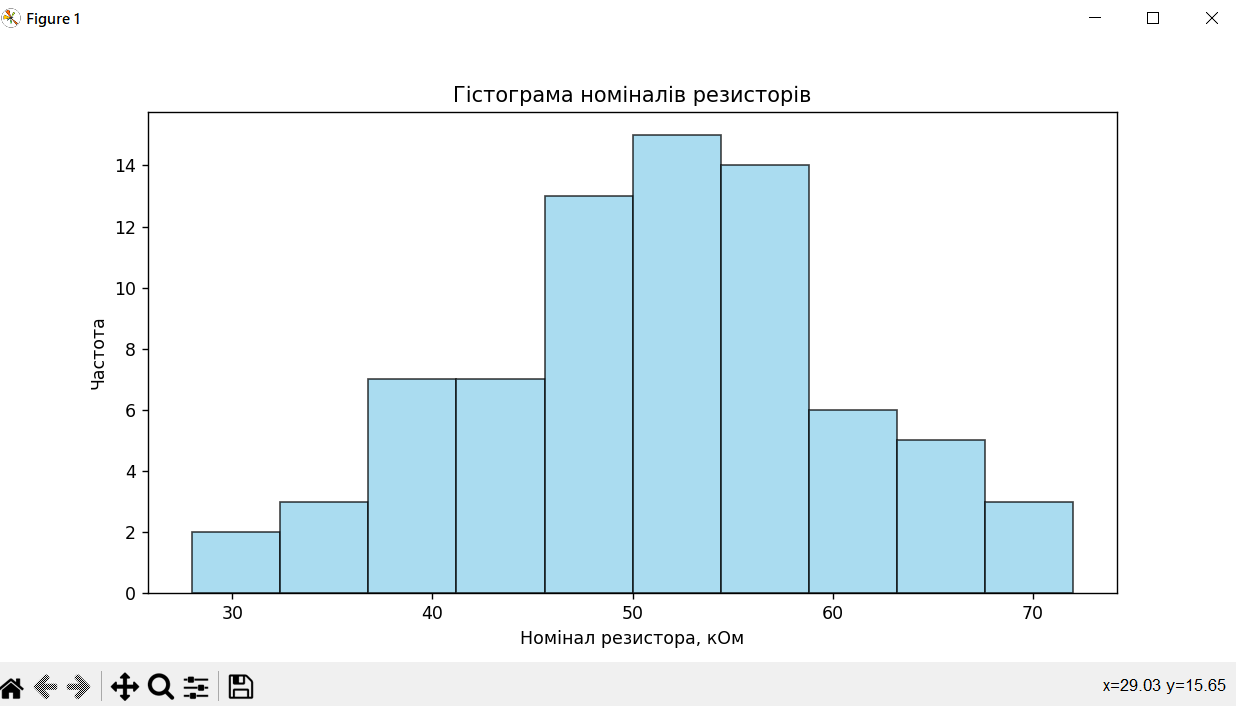
* \*Розмах (R)\*\*: 46 - це великий розмах, що вказує на значну різницю між найбільшим та найменшим значеннями в наборі.
* **Медіана (Me)**: 48.0 - це значення, яке розділяє набір даних на дві рівні частини. У цьому випадку, половина значень менше 48, а половина - більше.
* **Мода (Mo)**: 35.0 - це значення, яке зустрічається найчастіше. У цьому наборі даних це є найчастіше зустрічаючеся значення, що може вказувати на те, що значна кількість резисторів має номінал близький до 35 кОм.
* **Середнє значення (Bx)**: 47.17 - це середнє арифметичне всіх значень. Це дає уявлення про середній номінал резисторів у цьому наборі.
* **Стандартне відхилення (S)**: 9.20 - це міра розподілу даних відносно їх середнього значення. Значення стандартного відхилення вказує на те, що є помірна варіація в наборі даних.
* **Коефіцієнт варіації (V)**: 0.195 - це відношення стандартного відхилення до середнього значення. Чим менше значення коефіцієнта варіації, тим менше варіація в даних.

Отже, набір даних D має помірну варіацію в номіналах резисторів, але все ж має великий розмах. Мода наближена до мінімального значення, що може вказувати на деяку концентрацію резисторів з номіналом навколо цієї точки. Медіана та середнє значення також дозволяють отримати уявлення про центральні тенденції в наборі даних.

Результати для даних Д:







### Набір даних Д:

* \*Розмах (R)\*\*: 44 - великий розмах, що вказує на значну різницю між найбільшим та найменшим значеннями в наборі.
* **Медіана (Me)**: 51 - це значення, яке розділяє набір даних на дві рівні частини. У цьому випадку, половина значень менше 51, а половина - більше.
* **Мода (Mo)**: 54 - це значення, яке зустрічається найчастіше. У цьому наборі даних це є найчастіше зустрічаючеся значення, що вказує на те, що більшість резисторів мають номінал близький до 54 кОм.
* **Середнє значення (Bx)**: 51.11 - це середнє арифметичне всіх значень. Це дає уявлення про середній номінал резисторів у цьому наборі.
* **Стандартне відхилення (S)**: 9.25 - це міра розподілу даних відносно їх середнього значення. Значення стандартного відхилення вказує на те, що є помірна варіація в наборі даних.
* **Коефіцієнт варіації (V)**: 0.181 - це відношення стандартного відхилення до середнього значення. Значення коефіцієнта варіації вказує на помірний рівень варіації даних в порівнянні з їхнім середнім значенням.

Отже, набір даних E також має помірну варіацію в номіналах резисторів, але все ж має великий розмах. Мода наближена до максимального значення, що може вказувати на деяку концентрацію резисторів з номіналом навколо цієї точки. Медіана та середнє значення також дозволяють отримати уявлення про центральні тенденції в наборі даних.

**Висновки:**

Загальний аналіз показує, що всі набори даних мають високий розмах. Коефіцієнти варіації для всіх наборів даних відносно невеликі, що свідчить про відносну однорідність даних у кожному наборі. Варто також звернути увагу на значення моди та середнього значення, які можуть бути важливими для подальшого аналізу залежності вимірюваної характеристики від інших факторів.

Порівнюючи всі набори даних, можемо побачити, що вони мають схожі характеристики, але все ж таки є деякі різниці. Наприклад, набір даних С має найнижче середнє значення, а набір даних В - найвище стандартне відхилення. Такий аналіз допомагає ліпше зрозуміти розподіл даних та їхню варіабельність.

**Додаток з кодом:**

import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
  
def analyze\_data(data):  
 # Створення DataFrame з даних  
 df = pd.DataFrame(data, columns=[f'Resistor\_{i+1}' for i in range(len(data[0]))])  
  
 # 1. Ранжирування даних та побудова варіаційного ряду  
 ranked\_data = np.sort(df.values.flatten())  
 print("Варіаційний ряд:")  
 print(ranked\_data)  
  
 # 2. Знаходження R\*, Me та Mo  
 R\_star = ranked\_data[-1] - ranked\_data[0]  
 Me = np.median(ranked\_data)  
 Mo = df.mode().iloc[0][0]  
 print("\nR\*: ", R\_star)  
 print("Me: ", Me)  
 print("Mo: ", Mo)  
  
 # 3. Побудова полігону частот і гістограми  
 plt.figure(figsize=(10, 5))  
 plt.hist(ranked\_data, bins=10, color='skyblue', edgecolor='black', alpha=0.7)  
 plt.title('Гістограма номіналів резисторів')  
 plt.xlabel('Номінал резистора, кОм')  
 plt.ylabel('Частота')  
 plt.show()  
  
 # 4. Знаходження Bx, S та коефіцієнта варіації  
 Bx = np.mean(ranked\_data)  
 S = np.std(ranked\_data)  
 V = S / Bx  
 print("\nBx: ", Bx)  
 print("S: ", S)  
 print("Коефіцієнт варіації (V): ", V)  
  
# Задані дані  
data\_A = [  
 [45, 51, 58, 76, 48],  
 [59, 42, 62, 39, 51],  
 [60, 66, 71, 73, 61],  
 [46, 48, 50, 37, 34],  
 [55, 53, 42, 26, 69],  
 [41, 51, 36, 53, 68],  
 [56, 46, 50, 38, 47],  
 [49, 48, 52, 61, 48],  
 [39, 58, 39, 36, 57],  
 [58, 50, 42, 41, 66],  
 [62, 64, 46, 41, 68],  
 [65, 45, 46, 46, 49],  
 [54, 52, 40, 42, 63],  
 [41, 44, 55, 43, 46],  
 [68, 59, 60, 60, 33]  
]  
  
data\_B = [  
 [49, 44, 51, 70, 54],  
 [50, 35, 48, 23, 52],  
 [45, 29, 37, 66, 44],  
 [52, 66, 31, 59, 44],  
 [47, 45, 53, 56, 59],  
 [61, 64, 47, 54, 63],  
 [53, 62, 56, 52, 51],  
 [64, 36, 43, 52, 49],  
 [47, 40, 35, 61, 38],  
 [40, 55, 49, 62, 64],  
 [49, 44, 51, 70, 54],  
 [50, 35, 48, 23, 52],  
 [45, 29, 37, 66, 44],  
 [52, 66, 31, 59, 44],  
 [47, 45, 53, 56, 59]  
]  
  
data\_C = [  
 [36, 64, 50, 67, 37],  
 [48, 51, 54, 55, 28],  
 [54, 47, 45, 57, 51],  
 [46, 57, 50, 45, 54],  
 [58, 35, 45, 65, 53],  
 [55, 60, 42, 43, 65],  
 [30, 47, 47, 41, 52],  
 [49, 44, 57, 61, 54],  
 [50, 47, 57, 52, 40],  
 [69, 47, 50, 58, 58],  
 [44, 42, 60, 44, 58],  
 [44, 48, 52, 48, 56],  
 [56, 63, 58, 52, 60],  
 [36, 37, 42, 39, 38],  
 [57, 55, 66, 61, 40]  
]  
  
data\_D = [  
 [42, 55, 51, 53, 58],  
 [41, 30, 48, 54, 46],  
 [50, 49, 62, 34, 35],  
 [62, 41, 40, 38, 34],  
 [63, 24, 41, 41, 46],  
 [48, 57, 50, 53, 54],  
 [31, 48, 55, 53, 60],  
 [58, 63, 47, 42, 65],  
 [53, 51, 43, 46, 57],  
 [44, 53, 45, 54, 46],  
 [35, 54, 42, 34, 49],  
 [35, 36, 49, 37, 38],  
 [42, 48, 34, 54, 51],  
 [70, 39, 44, 41, 41],  
 [50, 62, 43, 47, 49]  
]  
  
data\_E = [  
 [36, 64, 50, 67, 37],  
 [48, 51, 54, 55, 28],  
 [54, 47, 45, 57, 51],  
 [46, 57, 50, 45, 54],  
 [58, 35, 45, 65, 53],  
 [55, 60, 42, 43, 65],  
 [30, 47, 47, 41, 52],  
 [49, 44, 57, 61, 54],  
 [50, 47, 57, 52, 40],  
 [69, 47, 50, 58, 58],  
 [59, 72, 47, 39, 39],  
 [54, 57, 39, 57, 49],  
 [57, 59, 39, 45, 33],  
 [70, 64, 49, 48, 62],  
 [52, 55, 55, 60, 46]  
]  
  
choice = input("Які дані ви бажаєте проаналізувати? Введіть 'A', 'B', 'C', 'D', 'E': ")  
  
if choice == 'A':  
 analyze\_data(data\_A)  
elif choice == 'B':  
 analyze\_data(data\_B)  
elif choice == 'C':  
 analyze\_data(data\_C)  
elif choice == 'D':  
 analyze\_data(data\_D)  
elif choice == 'E':  
 analyze\_data(data\_E)  
else:  
 print("Неправильний вибір.")