

**Завдання:** Згенерувати вибірку заданого об'єму з певного проміжку для дискретної та неперервної статистичної змінної. На підставі отриманих вибірових даних побудувати полігон частот, діаграму та гістограму. Також вивести всі числові характеристики

## Теорія:

**Статистичний розподіл** – таблиця, яка ставить у відповідність певному елементу вибірки те, скільки разів він у ній зустрічається. Для дискретної статистичної змінної ця таблиця складається природним чином. Для неперервної необхідно зробити додаткові крок:

1. Поділити вибірку на  $r+1$  рівних проміжків, де  $r + 1 = \lceil \log_2 n \rceil$ , де  $n$  – розмір вибірки.
2. Підрахувати для кожного проміжку кількість елементів, що входять у нього.
3. У таблиці як представлення кожного проміжку вважати середнє між початком і кінцем.

**Варіаційний ряд** – впорядкована за зростанням вибірка.

**Числові характеристики (статистики)** – характерні ознаки які притаманні для вибірок, показують такі ознаки як: точки скупчення елементів, розсіювання елементів по осі  $OX$ , щільність та ін. Поділяються на: характеристики центральної тенденції, характеристики розсіювання та характеристики форми.

### Характеристики центральної тенденції (локації):

1. Медіана.
2. Мода.
3. Середнє.

**Медіана (Me)** – елемент у центрі варіаційного ряду.

$$Me = \begin{cases} x_{k+1}, n = 2k + 1 \\ \frac{(x_k + x_{k+1})}{2}, n = 2k \end{cases}$$

Де  $n$  – кількість елементів вибірки,  $\{x_1, \dots, x_k, x_{k+1}, \dots, x_n\}$  – елементи варіаційного ряду.

**Мода (Mo)** – той елемент, який зустрічається найчастіше у вибірці (аналіз для неперервної статистичної змінної проводиться використовуючи статистичний розподіл.)

**Середнє** – середнє арифметичне вибірки :

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

### **Характеристики розсіювання:**

1. Розмах.
2. Девіація.
3. Варіанса.
4. Стандарт.
5. Варіація.

**Розмах** – різниця між найбільшим і найменшим елементами вибірки.

**Девіація** – сума відхилень елементів варіаційного ряду від середнього:

$$D = \sum_{i=1}^n (x_i - \bar{x})^2$$

**Варіанса** – девіація, поділена на кількість ступенів вільності:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{d. f.}$$

**Стандарт** – корінь від варіанси:

$$s = \left| \sqrt{s^2} \right|$$

**Варіація** – відношення стандарту до середнього:

$$v = \frac{s}{\bar{x}}$$

### **Характеристики форми:**

1. Асиметрія.
2. Екссесс.

**Момент статистичної змінної :**

$$M_k(c) = \sum_{i=1}^n (x_i - c)^k, k = (0, 1, 2 \dots)$$

Де  $c$  – деяка константа. Для центрального моменту ( $\mu$ ):  $c = \bar{x}$ . Початкового ( $m$ ):  
 $c = 0$

**Асиметрія** – відношення третього центрального моменту до другого в степені півтора:

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}$$

**Ексцес** – відношення четвертого центрального моменту до другого у квадраті, мінус три:

**Дециль** характеризує розподіл величин сукупності, при якій дев'ять значень дециля ділять її на десять рівних частин.

**Квантилем** вибірки порядку якщо він існує, називають той елемент варіаційного ряду, до якого включно з ним є елементів статистичного матеріалу.

## Хід роботи

Вибірка задається випадковим чином. Розмір вибірки та кількість елементів визначаються користувачем у відповідному вікні:

Введіть перше значення --> 2

Введіть останнє значення --> 14

Введіть кількість елементів --> 50

Кожна числова характеристика має відповідну написану функцію, для її обчислення за допомогою тих чи інших формул. Також усі характеристики та сама вибірка виводиться за допомогою спеціальних віджетів

Таблиця частот:

Значення	Частота
2	6
3	3
4	2
5	9
6	6
7	0
8	3
9	6
10	0
11	6
12	0
13	2
14	7

Таблиця частот:

Значення	Частота
[2, 4.0]	9
[4.0, 6.0]	11
[6.0, 8.0]	6
[8.0, 10.0]	9
[10.0, 12.0]	6
[12.0, 14.0]	9

Емпірична функція:

0,  $x < 2$

6/50,  $2 \leq x < 3$

9/50,  $3 \leq x < 4$

11/50,  $4 \leq x < 5$

20/50,  $5 \leq x < 6$

26/50,  $6 \leq x < 7$

26/50,  $7 \leq x < 8$

29/50,  $8 \leq x < 9$

35/50,  $9 \leq x < 10$

35/50,  $10 \leq x < 11$

41/50,  $11 \leq x < 12$

41/50,  $12 \leq x < 13$

43/50,  $13 \leq x < 14$

1,  $x \geq 14$

Середнє значення: 7.56

Медіана: 6.0

Мода: 5

Девіація: 3.99060120272418

Розмах: 12

Варіанса: 15.924897959183673

Стандарт: 3.99060120272418

Виправлена дисперсія: 0.0798120240544836

Варіація: 0.5278573019476429

Асиметрія: 0.28361371796542784

Ексцес: -1.1579514241860958

Окремо від дискретних, також виведені усі характеристики неперервної вибірки

Середнє значення: 7.76

Медіана: 7.666666666666667

Мода: 4.571428571428571

Девіація: 615.12

Розмах: 12.0

Варіанса: 12.553469387755102

Стандарт: 3.543087550111499

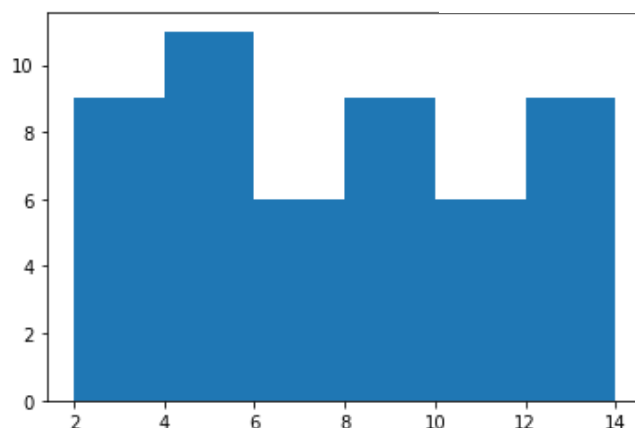
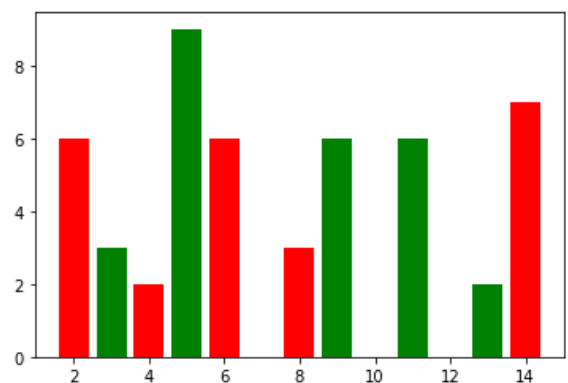
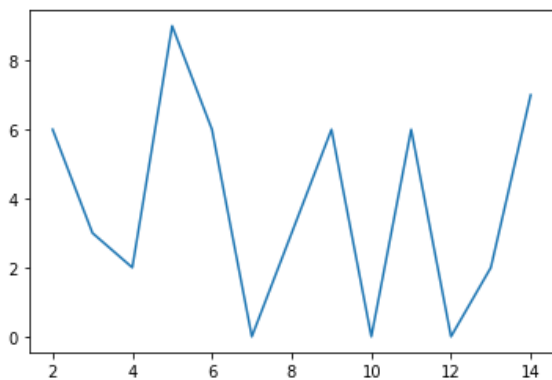
Виправлена дисперсія: 12.3024

Варіація: 0.45658344717931687

Асиметрія: 0.02042332535199287

Ексцес: -2.9663602421222164

Далі видно на відповідних малюнках: полігон, діаграму та гістограму



Квантили також виводяться в форматі таблиці, де None означає, що заданий квантиль не існує. Ось приклад виводу:

```
Квантили:  
Q1 = 3.0  
Q2 = 5.0  
Q3 = 9.0  
Інтерквартильна широта: 6.0
```

## Висновок

На даній лабораторній роботі я досконало вивчив, як обраховувати та аналізувати випадкові вибірки та їх характеристики. Та за допомогою цього створив програму, яка може згенерувати за критеріями випадкову вибірку чисел будувати варіаційний ряд, обчислювати числові характеристики (центральної тенденції, розсіювання, форми та квантиль, дециль, центиль та міліль). Завдяки цьому, я зміг краще зрозуміти відношення різних характеристик між собою та зміг знайти найкращі способи для їх обрахування та візуалізації

