МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені ІВАНА ФРАНКА

Кафедра дискретного аналізу

та інтелектуальних систем

**Звіт із**

**Теорія імовірності та математичної статистики**

**Індивідуальне завдання №1**

з курсу ''Теорія ймовірності та математична статистика''

Виконав:

Готюк Максим

Група Пмі-23

Оцінка \_\_\_

Перевірила:

доц. Квасниця Г.А.

2024

## **Постановка завдання:**

1. Згенерувати вибірку заданого об’єму (не менше 50) з вказаного проміжку для дискретної статистичної змінної. На підставі отриманих вибіркових даних:

* побудувати варіаційний ряд та частотну таблицю;
* представити графічно статистичний матеріал;
* побудувати графік емпіричної функції розподілу;
* обчислити числові характеристики дискретного розподілу.

2. Згенерувати вибірку заданого об’єму (не менше 50) з вказаного проміжку для неперервної статистичної змінної. На підставі отриманих вибіркових даних:

* утворити інтервальний статистичний розподіл;
* побудувати гістограму та графік емпіричної функції розподілу;
* обчислити числові характеристики.

## **Короткі теоретичні відомості:**

Множина всіх можливих значень випадкової величини *X* називається **генеральною сукупністю** , а множина значень xi (i=1,2,...,k), яка одержана в результаті випробувань, –

**вибіркою з генеральної сукупності** або **статистичною сукупністю** .

Число елементів вибірки називається **обсягом вибірки**.

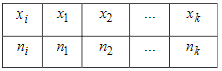
В результаті випробувань значення x1 випадкової величини X спостерігалося n1 разів, x2 – n2 разів, ..., xk nk разів. Спостережувані значення xi, (i =1, 2, ..., k) називаються **варіантами**, числа ni – **частотами**, відношення wi = ni/n – **відносними частотами**.

Послідовність варіант, записаних за зростанням, називається **варіаційним рядом**.

Статистичну сукупність характеризують статистичним розподілом вибірки,який може бути **дискретним** або **інтервальним**

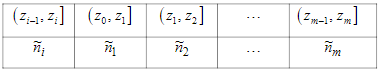
**Дискретним статистичним розподілом вибірки** називається відповідність між варіантами та їх частотами або відносними частотами.

Дискретний статистичний розподіл вибірки можна подати у вигляді таблиць:



**Інтервальним статистичним розподілом вибірки** називається відповідність між інтервалами варіаційного ряду та накопиченими частотами або накопиченими відносними частотами.

Інтервальний статистичний розподіл вибірки також записують у вигляді таблиць:



Статистичний розподіл вибірки можна задати графічно полігоном або гістограмою частот.

**Полігоном частот** називають ламану, відрізки якої послідовно з’єднують точки координатної площини

Гістограмою частот називається східчаста фігура , яка складена з прямокутників , основами яких є частинні інтервали  а їх висоти 

**Емпіричною функцією розподілу** F\*(x) випадкової величини *X* називається функція, яка для кожного дійсного числа *x* дорівнює відношенню  до обсягу вибірки *n*, тобто



**Числові характеристики поділяються на три групи:**

1. Числові характеристики центральної тенденції:медіана,мода,середнє арифметичне.

2.Числові характеристики розсіяння: варіанса, стандарт, розмах, варіація, інтерквантильні широти.

3. Числові характеристики форми: асиметрія,ексцес.

**Вибірковим середнім** x̄ статистичного розподілу вибірки називається середнє арифметичне її варіант xi з врахуванням їх частот, тобто



**Модою** Mо (x) дискретного статистичного розподілу вибірки називається те значення варіанти, xi якому відповідає найбільша частота ni. Іншими словами, мода - це варіанта, яка найчастіше зустрічається у вибірці.

**Медіаною** Me (x) дискретного статистичного розподілу вибірки називається варіанта, xi з обох сторін від якої міститься не більше ніж половина варіант, враховуючи їх частоту.

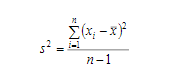
**Розмахом вибірки** називають різницю між найбільшим і найменшим значеннями її варіант, тобто



**Девіація** - сума квадратів відхилень елементів статистичного матеріалу від середнього арифметичного.



**Варіансою** s2 називається девіація, поділена на обсяг статистичного матеріалу без одного.



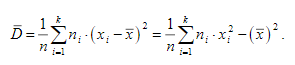
**Стандартом** (флуктуацією) s називається арифметичний корінь з варіанси.



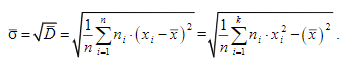
**Варіацією** v називається відношення стандарту до середнього арифметичного.



**Вибірковою дисперсією** статистичного розподілу вибірки називають середнє арифметичне квадратів відхилень його варіант xi від вибіркового середнього x̄, тобто



**Вибірковим середнім квадратичним відхиленням** називається число , яке є коренем з дисперсії, тобто



**Квантилем порядку **, якщо він існує, називається той елемент статистичного матеріалу (відповідного варіаційного ряду), до якого включно маємо % елементів статистичного матеріалу.

При  різницю між квантилем порядку і квантилем порядку називають **інтерквантильною широтою** порядку .

Квантилі порядку 25, 50, 75 називаються **квартилями**.

Квантилі порядку 12,5; 25,0;…, 87,5 називаються **октилями**.

Квантилі порядку 10; 20;…, 90 називаються **децилями**

Квантилі порядку 01; 02;…, 99 називаються **центилями**

Квантилі порядку 00,1; 00,2;…, 99,9 називаються **мілілями.**

Квартиль, октиль, дециль, центиль, міліль використовується тоді, коли об’єм статистичного матеріалу буде кратний відповідно 4, 8, 10, 100, 1000.

**Моментом** порядку k відносно сталої a називається вираз



При *а*=0 момент називається **початковим** і позначається через вираз



При *a*=x̄ момент називається **центральним**.

**Асиметрією** або скошеністю статистичного матеріалу називається відношення 3-го центрального моменту до 2-го центрального моменту в степені півтора.



**Ексцесом** ( крутістю, сплющеністю ) статистичного матеріалу називається відношення 4-го центрального моменту до 2-го центрального моменту в квадраті мінус три.



## **Програмна реалізація**

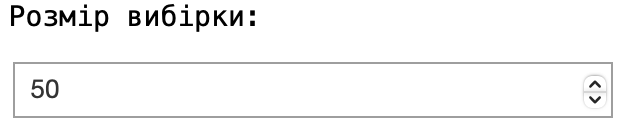
Для написання коду я використовував мову програмування Python, середовище Jupyter Notebook і наступні бібліотеки: numpy, matplotlib.pyplot, ipywidgets.

Для дискретного розподілу користувач вводить мінімальне і максимальне значення для елементів вибірки, а також її розмір розмір, після чого генерується масив рандомних цілих значень в заданих межах. Сортую масив і виводжу варіаційний ряд. Для виведення частотної таблиці створюю словник в якому ключами є унікальні елементи вибірки а значеннями – їхня кількість повторень. Рахую середнє вибіркове, медіану, моду, розмах вибірки, девіацію, варіансу, стандарт, варіацію, дисперсію, середньоквадратичне відхилення, а також моменти випадкової змінної для знаходження асиметрії та ексцесу. Тоді перевіряю, чи ділиться кількість елементів на 4, 8, 10 чи 100 та знаходжу квантилі, октилі, і т.д.

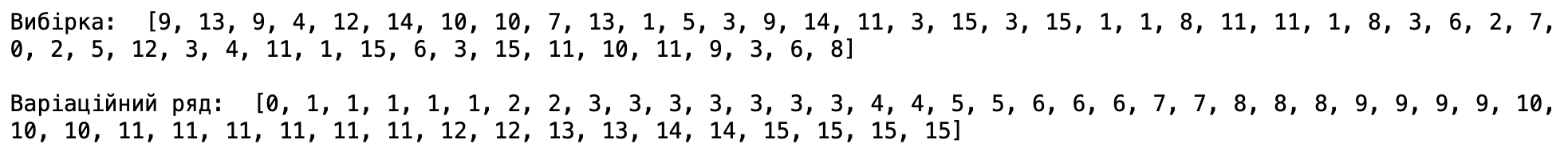
Для неперервниго розподілу користувач вводить мінімальне і максимальне значення для елементів вибірки, а також її розмір розмір, після чого генерується масив рандомних дійсних значень в заданих межах. Сортую масив та ділю на r + 1 відрізків, де r знаходиться за допомогою log2 розміру вибірки. Вивиджу масив інтервалів та відповідних частот. Рахую середнє вибіркове, медіану, моду, розмах вибірки, девіацію, варіансу, стандарт, варіацію, дисперсію, середньоквадратичне відхилення, а також моменти випадкової змінної для знаходження асиметрії та ексцесу.

## **Отримані результати**

**Дискретний розподіл:**

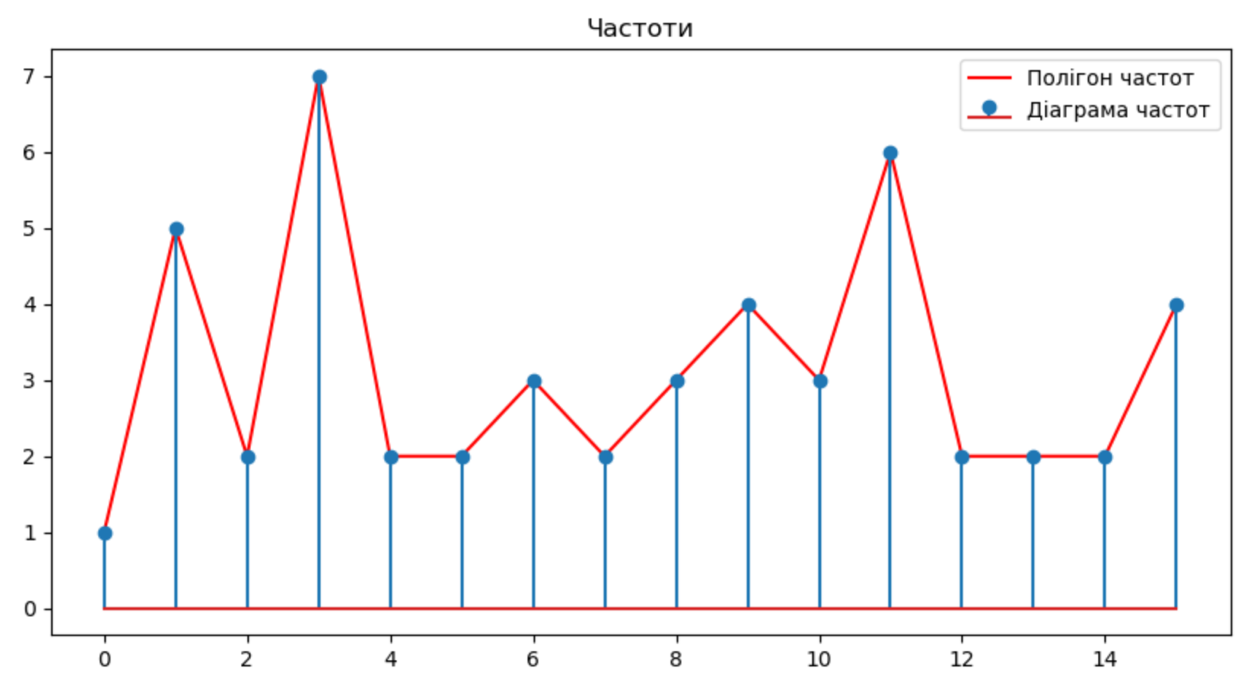


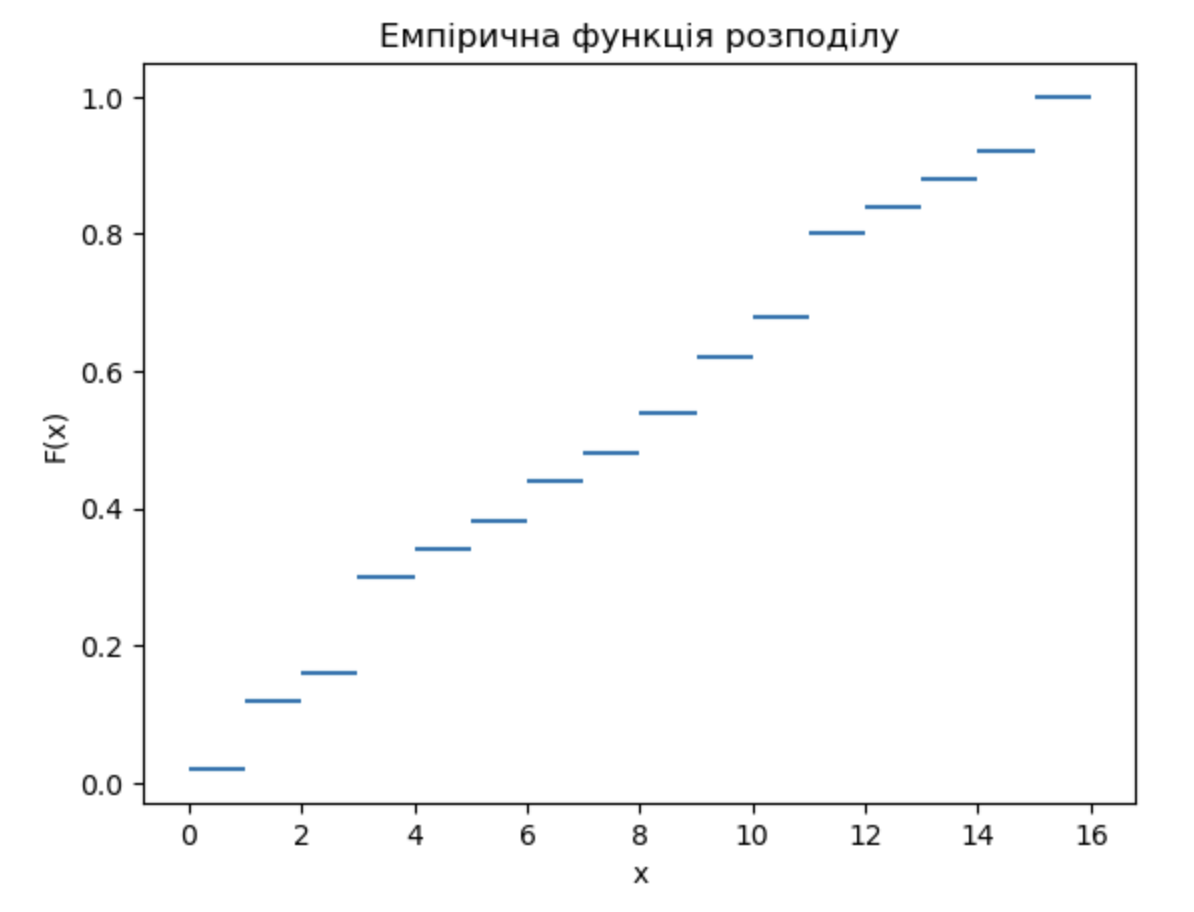


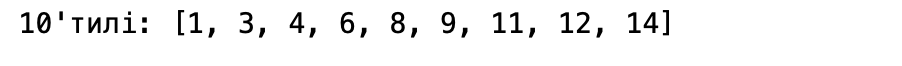
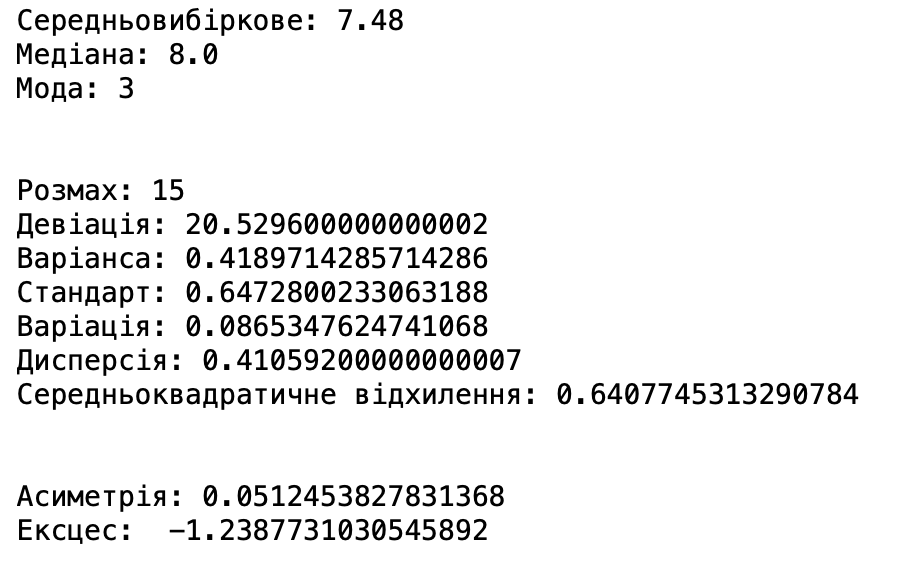


A number with numbers and symbols

Description automatically generated with medium confidence





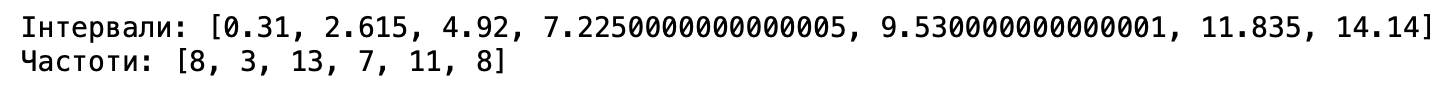


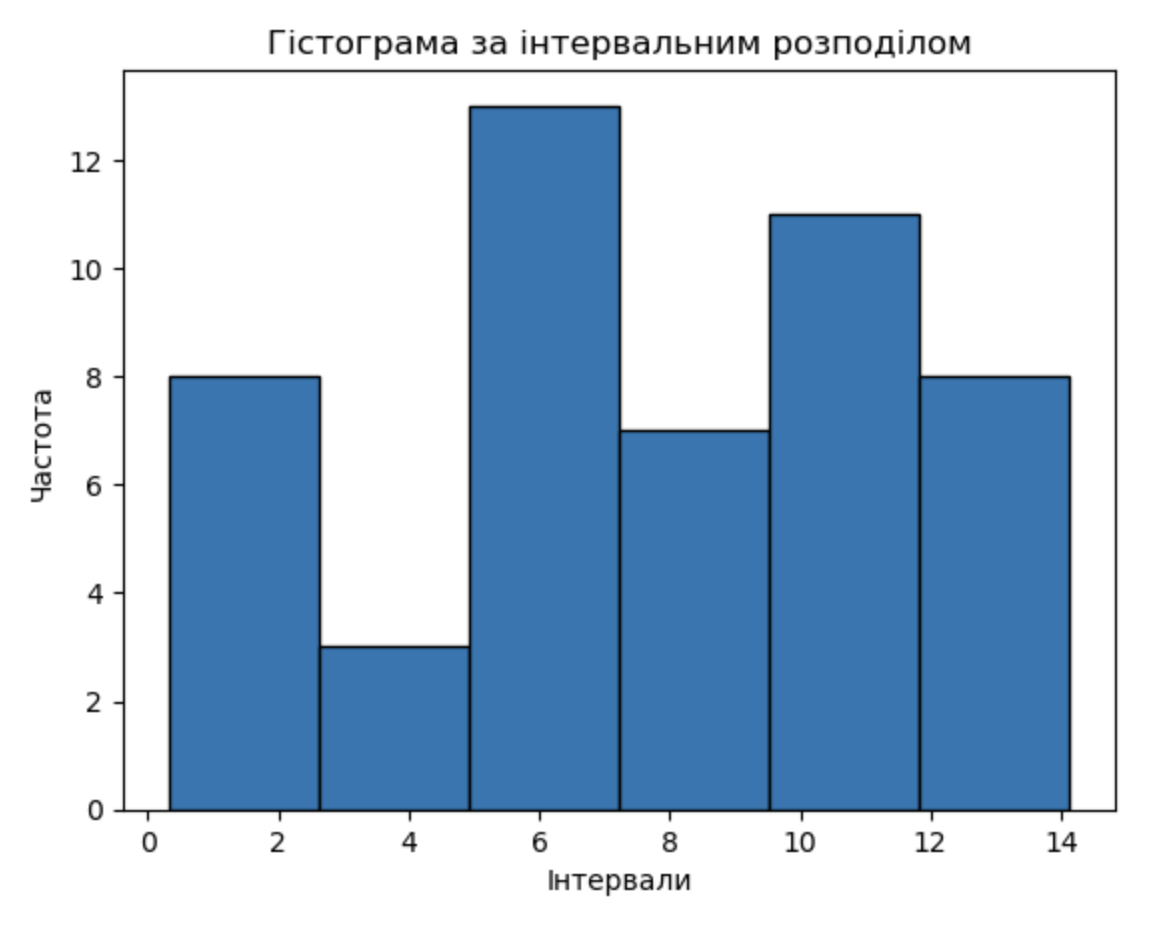
**Неперервний розподіл:**

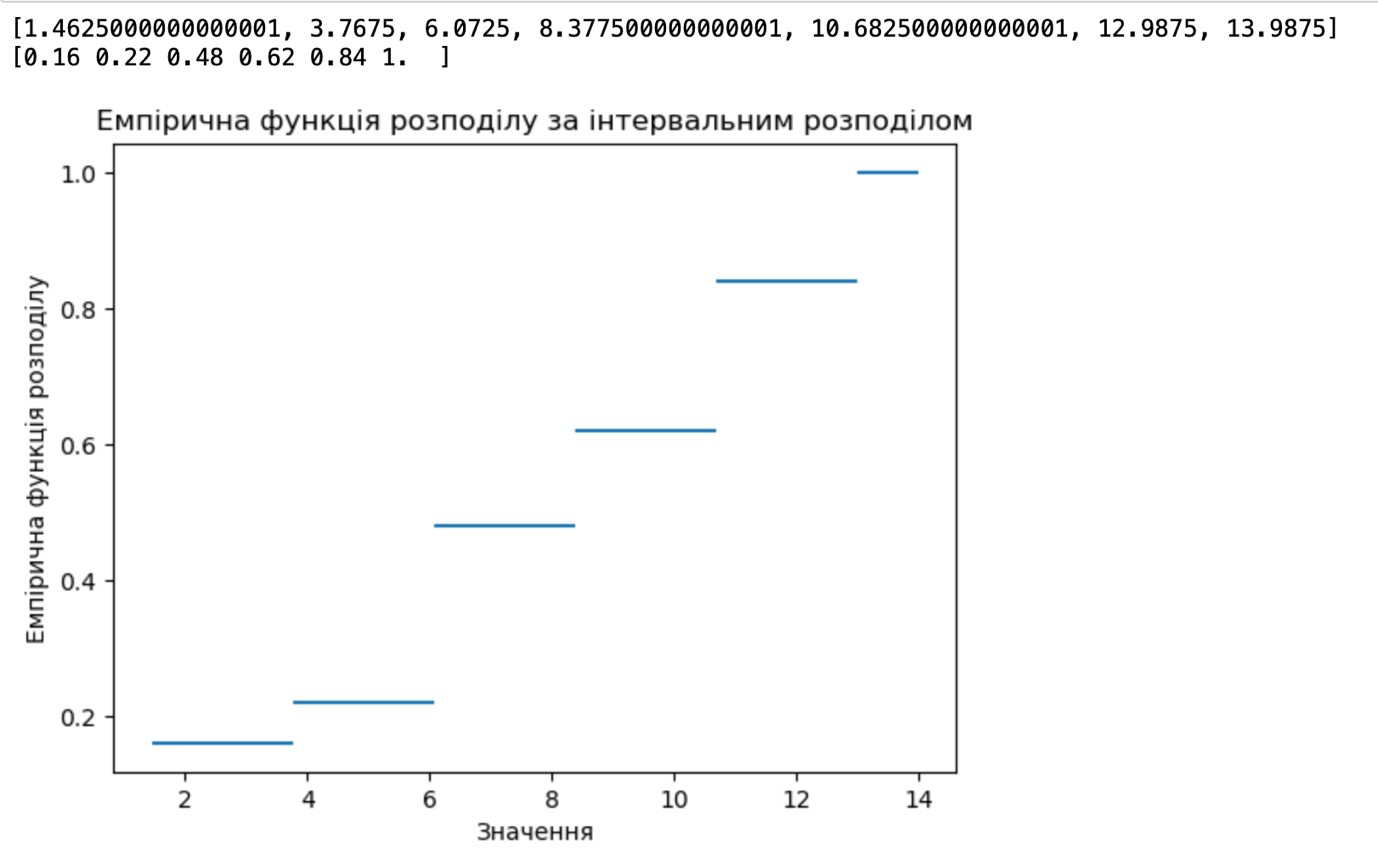
**A white rectangular object with black text

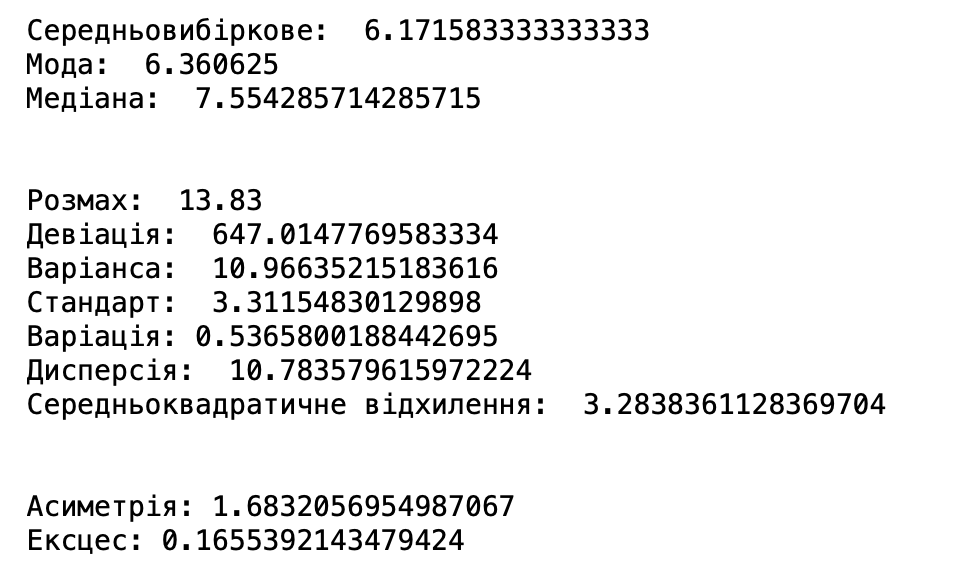
Description automatically generated**











**Висновок**

У даній лабораторній роботі я вирішив завдання з аналізу дискретної та неперервної статистичних змінних за допомогою програмування на мові Python та інструменту Jupyter Notebook, використовуючи бібліотеки numpy, matplotlib.pyplot та ipywidgets. Отримані дані дозволяють краще розуміти розподіл випадкових величин та його основні характеристики, що є важливим для подальшого аналізу та прийняття рішень у статистичних задачах.