ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені ІВАНА ФРАНКА

Факультет прикладної математики та інформатики

Звіт

З курсу «Теорія ймовірності та математична статистика»

Індивідуальне завдання №1

Виконала

студентка групи ПМІ-23

Богданович Софія

2025

Умова завдання (варіант 2):

1. Згенерувати вибірку заданого об’єму (не менше 50) з проміжку [2; 12] 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.
   - згрупувавши дані, утворити інтервальний розподіл варіанти, побудувати гістограму та графік емпіричної функції розподілу, обчислити числові характеристики для згрупованих даних.
2. Згенерувати вибірку заданого об’єму (не менше 50) для неперервної статистичної змінної. На підставі отриманих вибіркових даних: • утворити інтервальний статистичний розподіл, побудувати гістограму та графік емпіричної функції розподілу, обчислити числові характеристики для згрупованих даних.

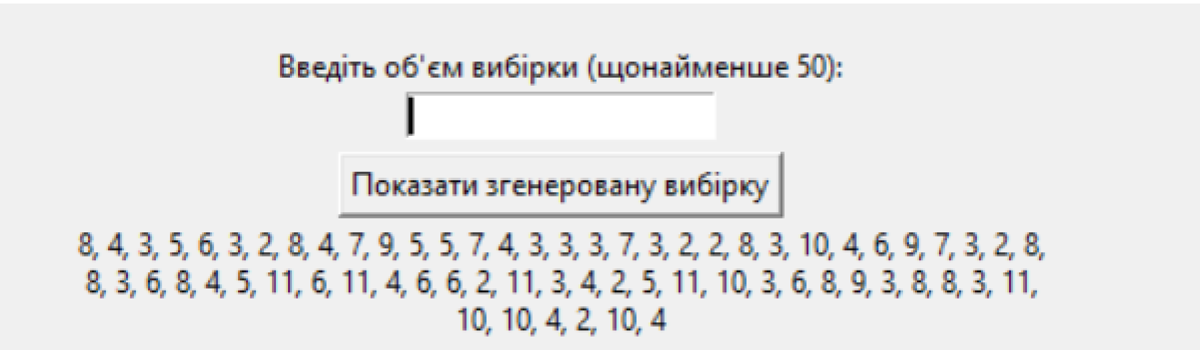
Хід роботи

Перша частина завдання.

1. Генеруємо вибірку заданого об’єму на проміжку [2; 12] для **дискретної статистичної змінної**.

Дискретна статистична змінна — це змінна, яка може набувати лише окремих значень із певного (часто скінченного або зліченного) набору.

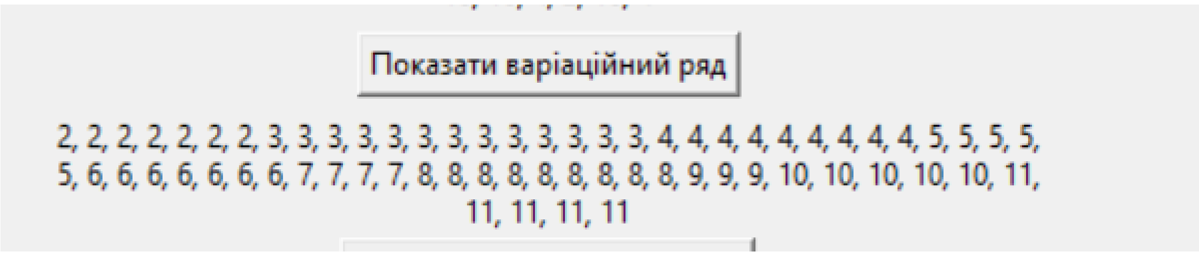
Наприклад, увівши число 67 отримуємо таку вибірку.



Приклад 1

2. Будуємо **варіаційний ряд** та **частотну таблицю.**

Варіаційний ряд — це впорядкований набір значень статистичної змінної (від найменшого до найбільшого)



Приклад 2

Частотна таблиця — це таблиця, що показує, як часто зустрічається кожне значення у вибірці.

Частотна таблиця називається ще **статистичним розподілом дискретної варіанти.**

Варіаційний ряд побудований на підставі згенерованих даних:

Частотна таблиця:

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

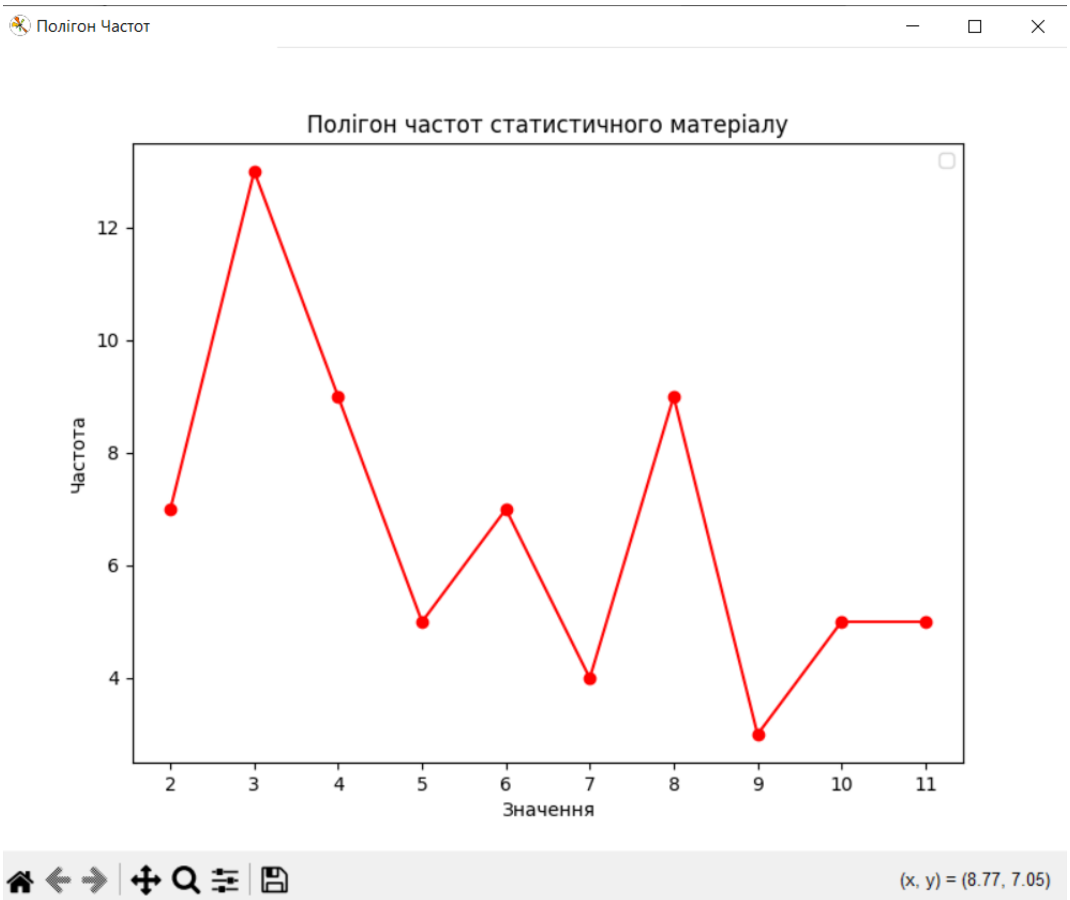
Приклад 3

3. Наступним кроком представляємо графічно статистичний матеріал.

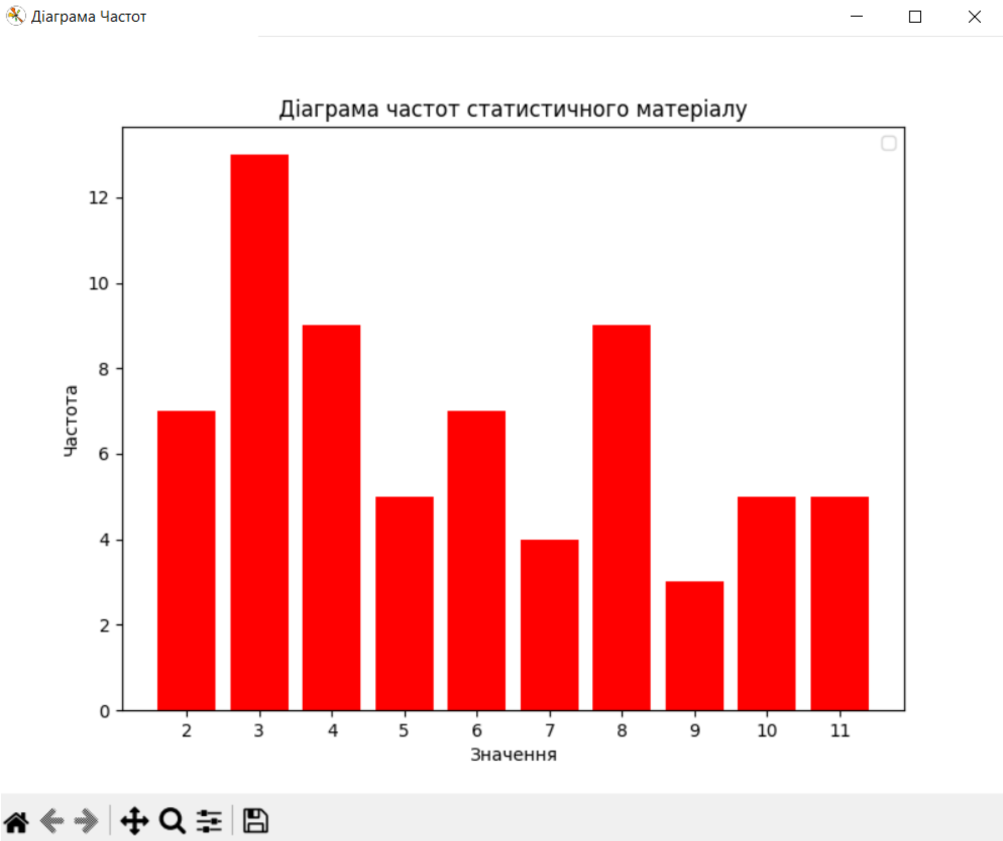
Діаграма частот — це графік (стовпчикова діаграма), де на **горизонтальній осі** відкладаються значення змінної, а на **вертикальній осі** – їхні частоти.

Полігон частот — це ламана лінія, яка з’єднує точки, що відповідають частотам значень.

Приклади діаграми та полігону частот для попередньо згенерованих даних:



Приклад 4 (полігон частот)



Приклад 5 (діаграма частот)

4. Далі будуємо графік **емпіричної функції розподілу**.

Для графічного представлення варіаційного ряду наносимо на вісь абсцис елементи варіаційного ряду $x_{(i)} (i = 1,2,\ldots, n)$ та пов'яжемо з кожною точкою $x_{(i)}$ масу $\frac{1}{n}$.

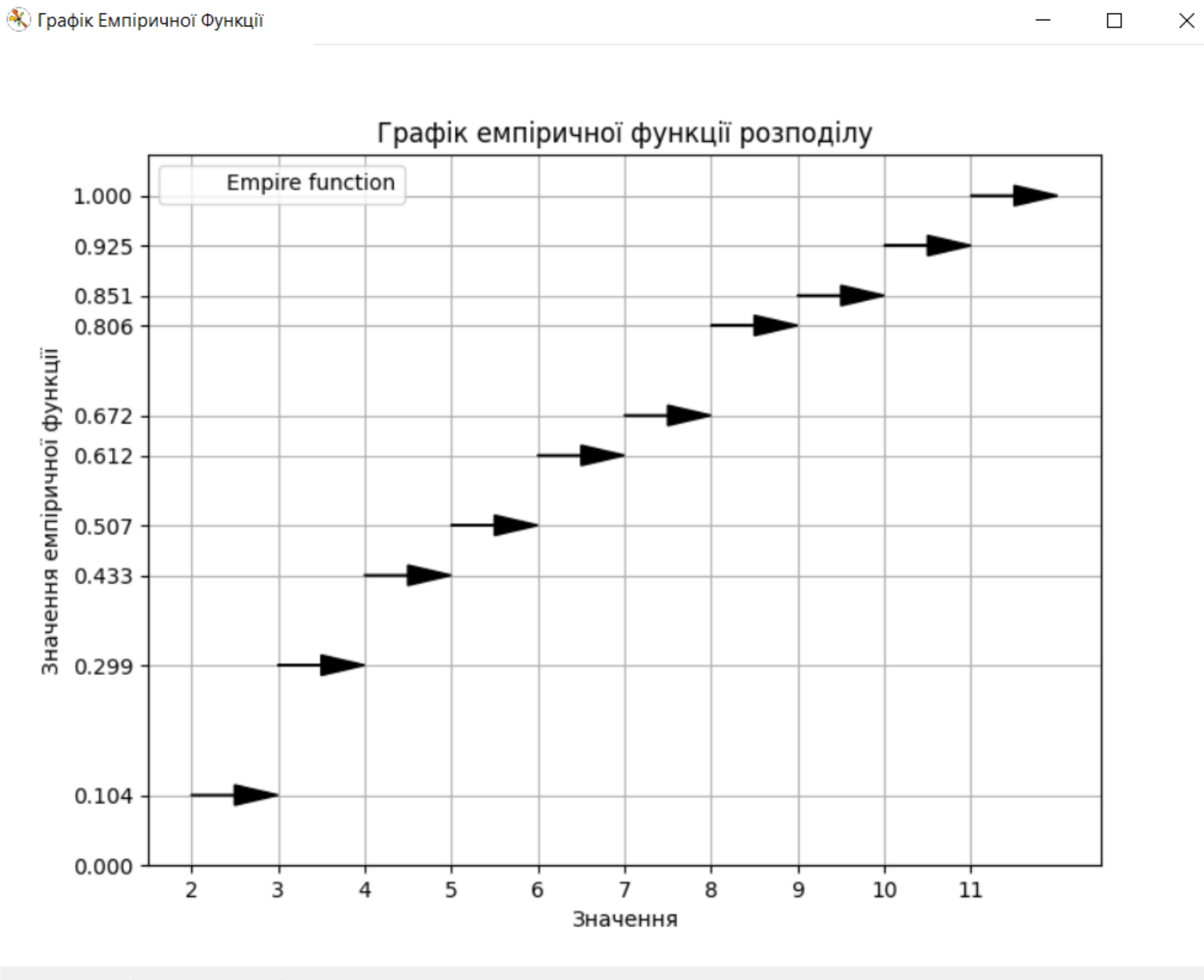
Нарисуємо східчасту лінію зі стрибками вверх у пунктах $x_{(i)}$ на $\frac{1}{n}$. Від $-\infty$ до $x_{(1)}$ маємо лінію на рівні нуль. У точці $x_{(1)}$ маємо стрибок на $\frac{1}{n}$ і відрізок на висоті $\frac{1}{n}$ до точки $x_{(2)}$ у точці $x_{(n)}$ останній стрибок на $\frac{1}{n}$ і лінія на висоті 1буде продовжуватися до безмежності. Якщо б зустрілося два однакові $x_{(i)}$, або більше, то в цій точці був би стрибок на $\frac{2}{n}$, або відповідно більше.

Одержане графічне представлення варіаційного ряду називаються **<u>емпіричною функцією розподілу</u>** або **<u>емпіричною кумулятою.</u>**

Таким чином, емпірична функція розподілу

$$F_n(x) = \begin{cases} 0, & x < x_{(1)} \\ \frac{k}{n}, & x_{(k)} \le x < x_{(k+1)} \quad (k = 1,\ldots,n-1) \\ 1, & x_{(n)} \le x \end{cases}$$

Графік емпіричної функції розподілу для нашого локального прикладу:



Приклад 6 (емпірична функція розподілу)

5. Наступним кроком обчислюємо **числові характеристики дискретного розподілу**.

Числові характеристики поділяють на 3 групи:

1. Числові характеристики **центральної тенденції (локації)**. До них відноситься:
   а) медіана (M_e)
   б) мода (M_o)
   в) середнє арифметичне ($\bar{x}$)
2. Числові характеристики **розсіяння**. До них відноситься:
   а) варіанса (s^2)
   б) стандарт (s)
   в) розмах (ρ)
   г) варація (v)
   д) інтерквантильність широт
3. Числові характеристики **форми**: До них відноситься:
   а) асиметрія (γ_1) *(Ас)*
   б) ексцес (γ_2) *(Ек)*

Кожна з перерахованих числових характеристик є деякою функцією від елементів статистичного матеріалу.

Означення. Медіаною називають цей елемент статистичного матеріалу, який ділить відповідний варіаційний ряд (3) на дві рівні за обсягом частини. Медіану позначаємо M_e.

Якщо обсяг статистичного матеріалу непарний, то медіана визначається однозначно. Наприклад, якщо варіаційний ряд статистичного матеріалу буде ($n = 2k+1$)

$$\underbrace{x_{(1)} \le x_{(2)} \le \cdots \le x_{(k)}}_{k} \le x_{(k+1)} \le \underbrace{\cdots \le x_{(2k+1)}}_{k}, \text{ то } M_e = x_{(k+1)}$$

Якщо обсяг статистичного матеріалу парний, то медіаною може бути інтервал.
Наприклад, якщо варіаційний ряд статистичного матеріалу буде ($n = 2k$)

$$\underbrace{x_{(1)} \le x_{(2)} \le \cdots \le}_{k-1} x_{(k)} \le x_{(k+1)} \le \underbrace{\cdots \le x_{(2k)}}_{k-2}$$

То $M_e = \lfloor x_{(k)}, x_{(k+1)} \rfloor$ $\quad M_e = \dfrac{x_{(k)} + x_{(k+1)}}{2}$

Означення. Модою називають цей елемент статистичного матеріалу, який найчастіше зустрічається. Моду позначаємо M_o.

Означення. Середнім арифметичним називається сума всіх елементів статистичного матеріалу, поділена на обсяг статистичного матеріалу, позначається $\bar{x}$:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Означення. **Варіансою** називається девіація (сума квадратів відхилень елементів статистичного матеріалу від середнього арифметичного) поділена на обсяг статистичного матеріалу без одного і позначається s^2.

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

Означення. Розмахом називається різниця між найбільшим і найменшим елементами статистичного матеріалу і позначається ρ .

$$\rho = \max(x_1, \cdots, x_n) = \min(x_1, \cdots, x_n) = x_{(n)} - x_{(1)}$$

Варіацією вибірки називається відношення стандарту цієї вибірки до середнього арифметичного

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

Означення. Квантилем порядку α , якщо він існує, називається цей елемент статистичного матеріалу (відповідного варіаційного ряду), до якого включно маємо $\alpha\%$ елементів статистичного матеріалу (відповідного варіаційного ряду).
Статистичний матеріал (1) має квантилі тільки порядків кратних $\cdot\frac{100}{n}$, інші квантилі не існують; елемент $x_{(i)}$ є кантилем порядку $i \cdot \frac{100}{n}$ $(i = 1, \cdots, n)$.
Означення. При $\alpha < \beta$, різницю між квантилем порядку β і квантилем порядку α називають **інтерквантильною широтою порядку** $\beta - \alpha$.
Для статистичного матеріалу (1) існують тільки інтерквантильні широти наступних порядків

$$q_{ij} = (j - i) \cdot \frac{100}{n}, \;\; j > i \;\; (i = 1, 2, \cdots n - 1; j = 2, 3, \cdots n)$$

Квантилі порядку 25, 50, 75 називаються **квартилями:** першим Q_1; другим Q_2; третім Q_3. Різниця між третім і першим квартилем $Q_3 - Q_1$ називається **інтерквартильною**
Квантилі порядку 10; 20;…, 90 називаються **децилями**: першим $Д_1$; другим $Д_2$; …дев'ятим $Д_9$. Різниця між дев'ятим і першим децилем $Д_9 - Д_1$ називається **інтердецильною широтою**. Очевидно, що

$$Д_1 = x_{\left(\frac{n}{10}\right)} \;\; ,\ldots, Д_9 = x_{\left(\frac{9n}{10}\right)}$$

Квантилі порядку 12,5; 25,0;…, 87,5 називаються **октилями**: першим O_1; другим O_2; …сьомим O_7. Різниця між сьомим і першим октилем $O_7 - O_1$ називається **інтероктильною широтою**. Очевидно, що

$$O_1 = x_{\left(\frac{n}{8}\right)} \;,\ldots,\; O_7 = x_{\left(\frac{7n}{8}\right)}$$

Обчисленні числові характеристики у нашому випадку:

Квантилі порядку 01; 02;…, 99 називаються **центилями**: першим C_{01}; другим C_{02}; …дев'яносто дев'ятим C_{99}. Різниця між дев'яносто дев'ятим і першим центилем $C_{99} - C_{01}$ називається **інтерцентильною широтою**.

Очевидно, що

$$C_{01} = x_{\left(\frac{n}{100}\right)}, \ldots, C_{99} = x_{\left(\frac{99n}{100}\right)}$$

Від C_{01} виключно до C_{99} включно розташовано 98% центральних елементів статистичного матеріалу.

Квантилі порядку 00,1; 00,2;…, 99,9 називаються **мілілями**: першим M_{001}; … дев'ятсот дев'яносто дев'ятим M_{999}. Різниця $M_{999} - M_{001}$ називається **інтермілільною широтою**.

Очевидно, що

$$M_{001} = x_{\left(\frac{n}{1000}\right)}, \ldots, M_{999} = x_{\left(\frac{999n}{1000}\right)}$$

Означення. Асиметрією або скошеністю статистичного матеріалу (1) називається відношення 3-го центрального моменту до 2-го центрального моменту в степені півтора

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

Означення. Ексцесом (крутістю, сплющеністю) статистичного матеріалу (1) називається відношення 4-го центрального моменту до 2-го центрального моменту в квадраті мінус три

$$E_k = \gamma_2 = \frac{\mu_4}{\mu_2^2} - 3.$$

Характеристика	Значення
Мода	3
Медіана	5
Середнє значення	6.0
Розмах	9
Варіанса	7.0459
Стандарт	2.6544
Варіація	0.4584
Дисперсія	8.0459
Середнє квадратичне відхилення	2.8365
Асиметрія	0.3692
Ексцес	-1.1278
Інтерквартильна широта	None
Інтероктильна широта	None
Інтердецильна широта	None
Інтерцентильна широта	None
Інтермілільна широта	None

Приклад 7 (числові характеристики дискретного розподілу)

None на скірішоті для інтерквантильних широт вказує на випадок, коли квартиль, октиль, дециль, центиль, міліль неможливо порахувати через кратність об’єму вибірки.

6. Згрупувавши дані, утворюємо **інтервальний розподіл варіанти**.

Інтервальний розподіл — це спосіб подання даних, коли значення змінної групуються в **інтервали**, а не розглядаються окремо.

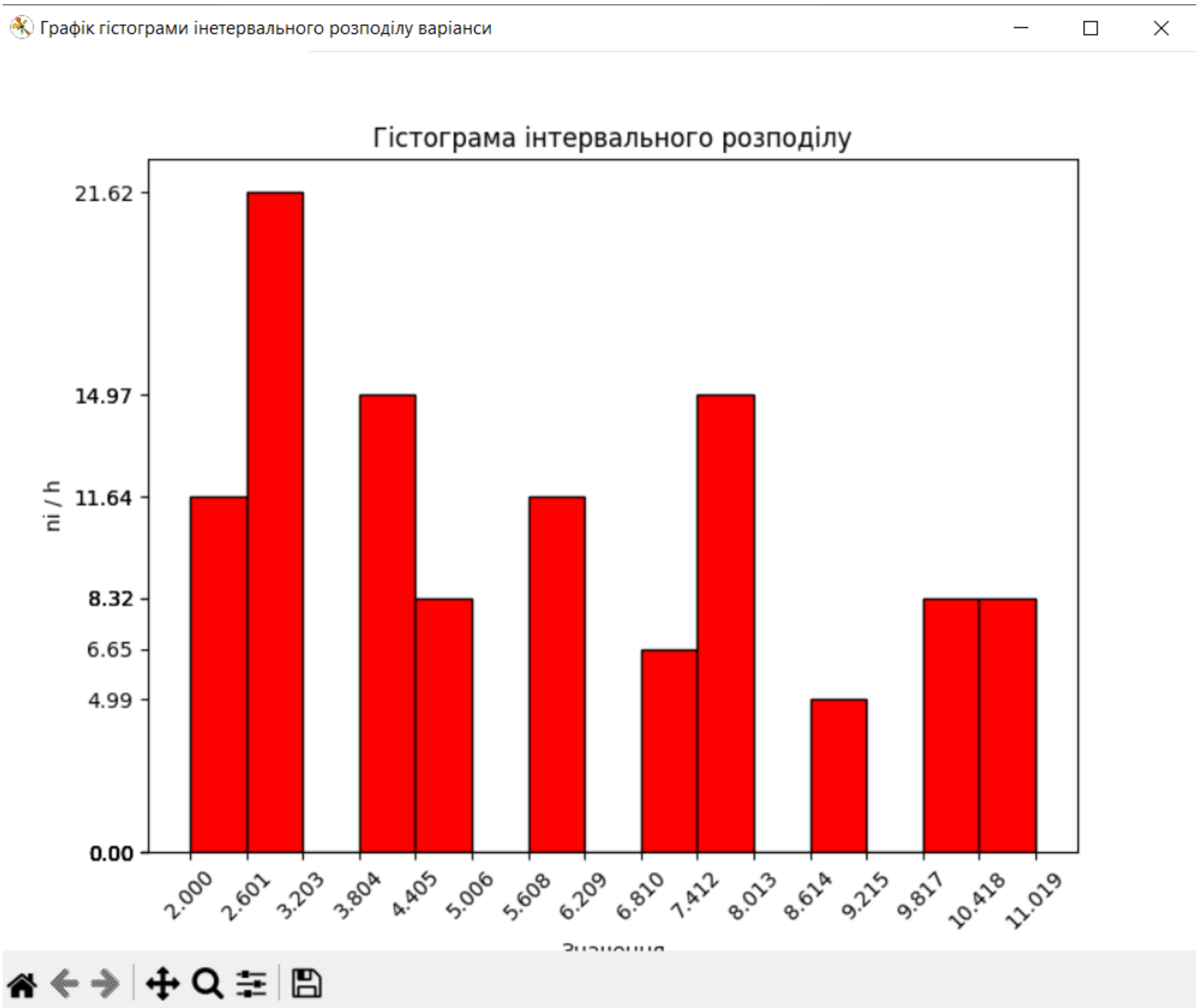
Інтервал	Частота
[2.0 - 2.6)	7
[2.6 - 3.2)	13
[3.2 - 3.8)	0
[3.8 - 4.41)	9
[4.41 - 5.01)	5
[5.01 - 5.61)	0
[5.61 - 6.21)	7
[6.21 - 6.81)	0
[6.81 - 7.41)	4
[7.41 - 8.01)	9
[8.01 - 8.61)	0
[8.61 - 9.22)	3
[9.22 - 9.82)	0
[9.82 - 10.42)	5
[10.42 - 11.02)	5

Приклад 8 (інтервальний розподіл)

7. Будуємо гістограму інтервального розподілу.

Гістограма інтервального розподілу — це стовпчикова діаграма, яка візуалізує, як дані розподілені за інтервалами.

В нашому випадку:



Приклад 9. Гістограма інтервального розподілу

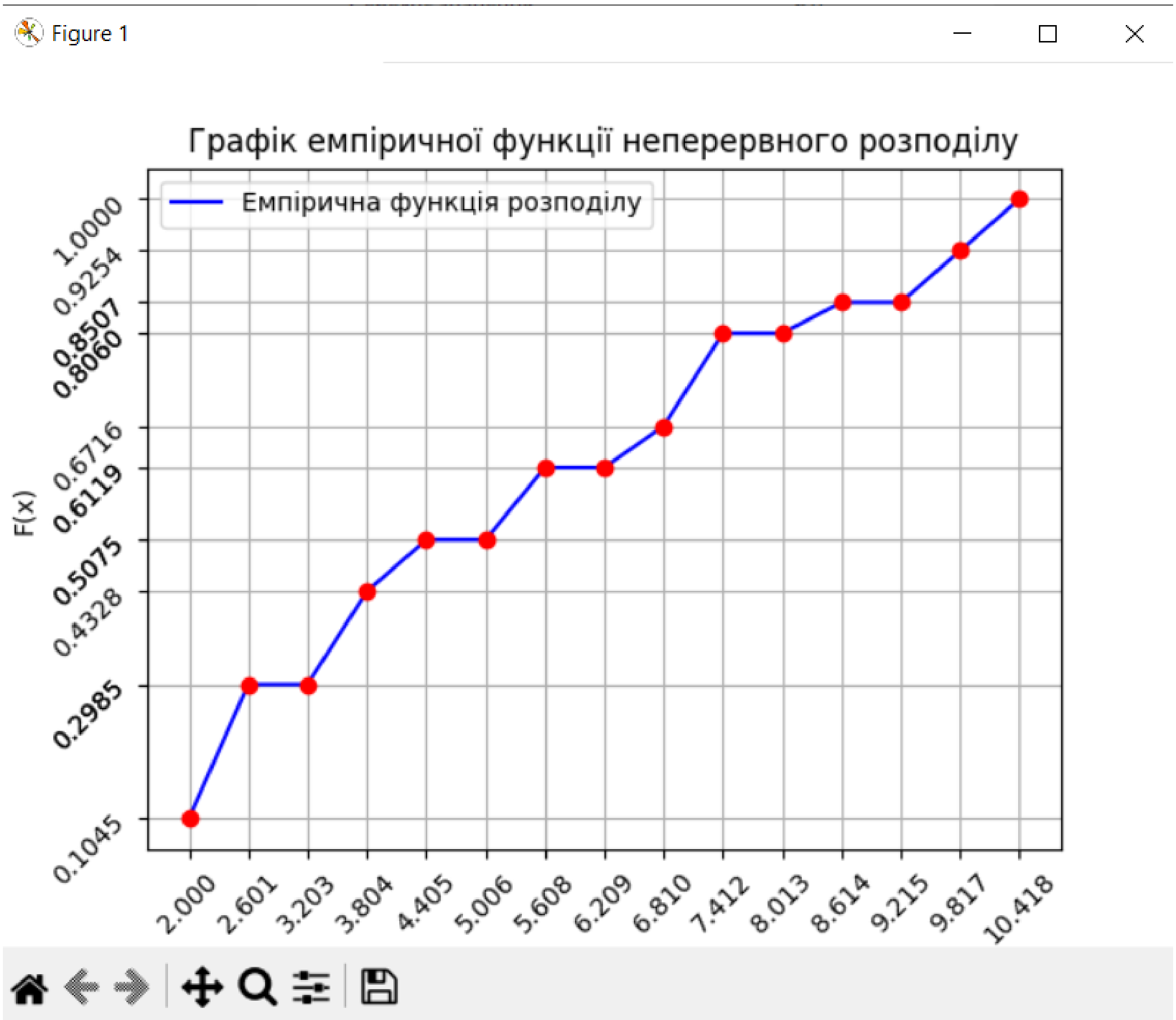
8. Будуємо графік емпіричної функції інтервального розподілу.

За формулою:

$$F^*(x)=\begin{cases} 0, & x<h_0, \\ F_1^*(x), & h_0\le x<h_1, \\ F_2^*(x), & h_1\le x<h_2, \\ \dots\dots\dots & \\ F_{r+1}^*(x), & h_r\le x<h_{r+1}, \\ 1, & x\ge h_{r+1}, \end{cases}$$

де $F_i^*(x)=\dfrac{w_i}{h_i-h_{i-1}}(x-h_{i-1})+\omega_{i-1},\quad i=\overline{1,r+1}.$

У нашому прикладі:



Приклад 10. Графік емпіричної функції інтервального розподілу

9. Обчислюємо числові характеристики для інтервального розподілу.

Формула для моди:

$$M_o(x) = h_{Mo-1} + \frac{n_{Mo} - n_{Mo-1}}{(n_{Mo} - n_{Mo-1}) + (n_{Mo} - n_{Mo+1})}(h_{Mo} - h_{Mo-1}).$$

Формула для медіани:

$$M_e = h_{M-1} + \frac{h_M - h_{M-1}}{n_M}\left(\frac{n}{2} - m_{M-1}\right).$$

Усі інші характеристики обчислюємо аналогічно до дискретного розподілу, але як значення беремо середнє значення кожного інтервалу.

Числові характеристики для інтервального розподілу у нашому прикладі:

Характеристика	Значення
Мода	2.6013
Медіана	4.9463
Середнє значення	5.7378
Розмах	8.418
Варіанса	7.7217
Стандарт	2.7788
Варіація	0.4843
Дисперсія	7.6065
Середнє квадратичне відхилення	2.758
Асиметрія	0.4185
Ексцес	-1.127

Приклад 11. Числові характеристики для інтервального розподілу

Друга частина завдання.

Згенеруємо вибірку об'єму 56 для **неперервної статистичної змінної**.

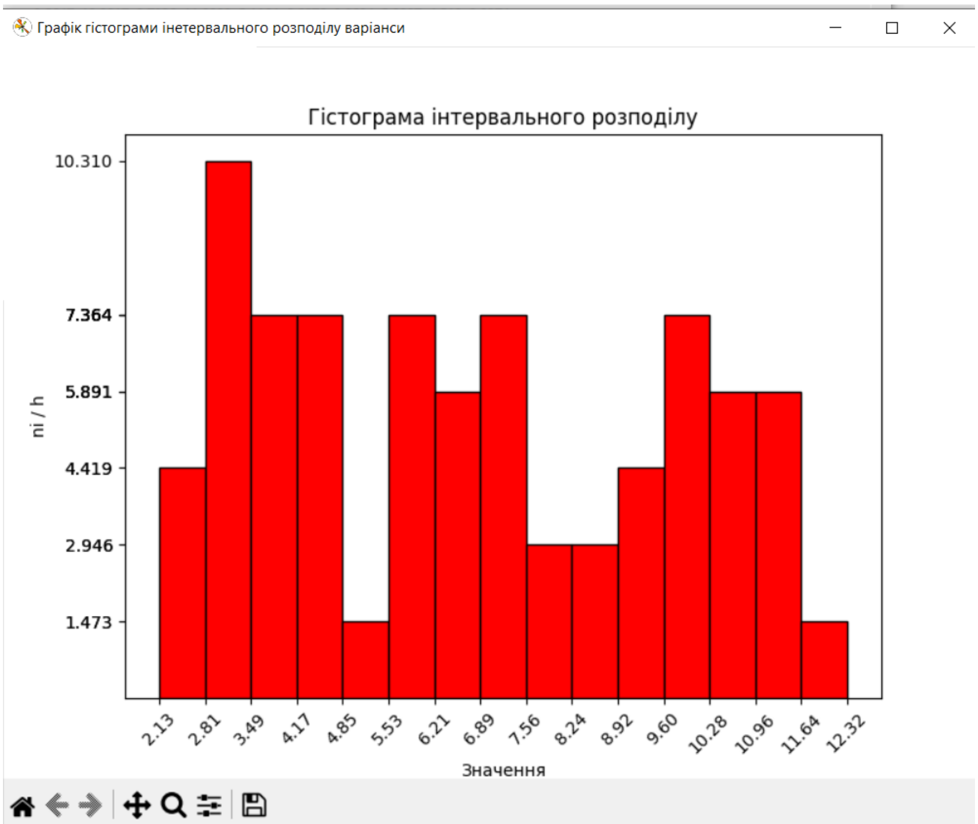
8.5647, 10.9017, 9.7026, 11.0998, 5.1164, 6.6576, 8.9694, 5.8897, 4.518, 8.9274, 11.8909, 2.8667, 7.5658, 5.536, 2.3951, 2.851, 10.8987, 7.2505, 4.3246, 3.9368, 7.2259, 5.6603, 2.1327, 6.413, 3.8669, 3.9223, 11.5503, 2.9765, 3.4023, 2.3498, 5.8316, 6.9068, 4.3735, 3.9049, 3.2177, 6.2655, 5.6854, 7.5206, 9.6207, 11.2361, 3.8974, 7.6169, 8.5681, 6.5223, 3.0893, 10.9596, 4.5798, 10.3218, 9.9301, 4.2855, 9.8702, 9.2431, 2.8211, 7.222, 10.7746, 9.7586

Приклад 12. Вибірка об'єму 56 для неперервної статистичної змінної

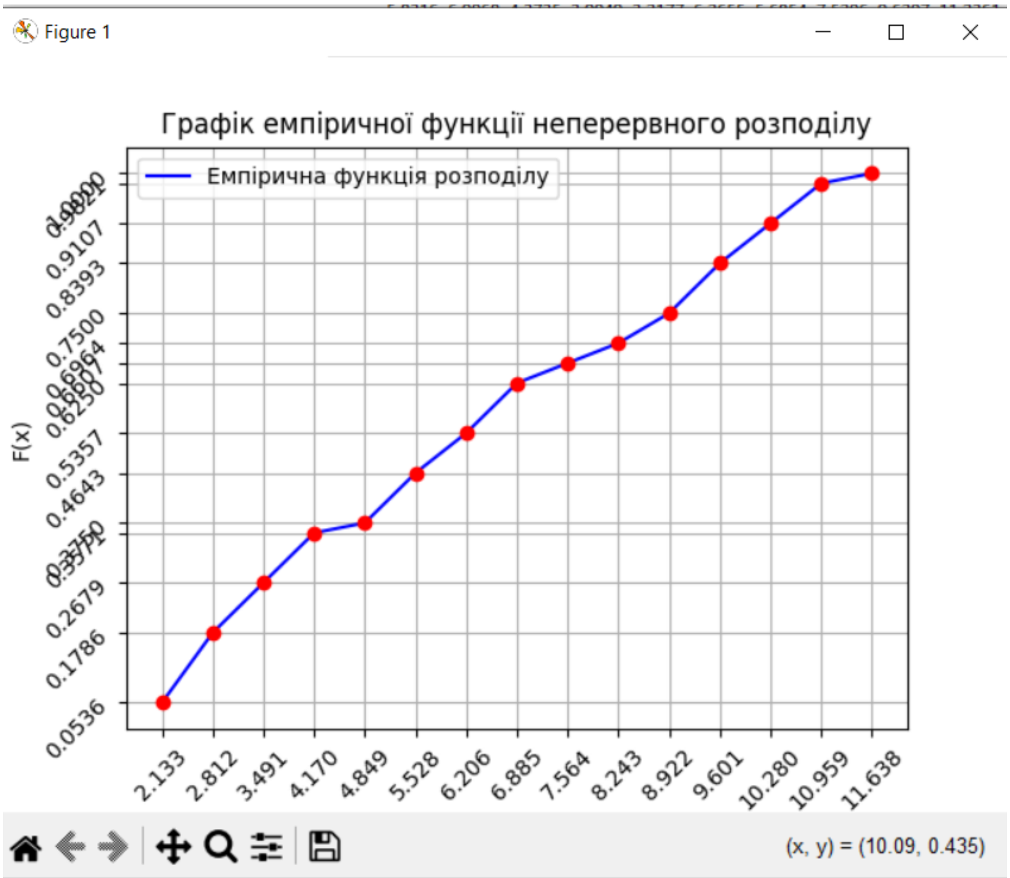
Утворимо інтервальний статистичний розподіл. Побудуємо гістограму та графік емпіричної функції розподілу, обчислимо числові характеристики для згрупованих даних.

Інтервал	Частота
[2.1327 - 2.8117)	3
[2.8117 - 3.4906)	7
[3.4906 - 4.1696)	5
[4.1696 - 4.8486)	5
[4.8486 - 5.5275)	1
[5.5275 - 6.2065)	5
[6.2065 - 6.8854)	4
[6.8854 - 7.5644)	5
[7.5644 - 8.2434)	2
[8.2434 - 8.9223)	2
[8.9223 - 9.6013)	3
[9.6013 - 10.2803)	5
[10.2803 - 10.9592	4
[10.9592 - 11.6382	4
[11.6382 - 12.3171	1

Приклад 13. Інтервальний статистичний розподіл



Приклад 14. Гістограма інтервального розподілу.



Приклад 15. Емпірична функція

Характеристика	Значення
Мода	2.8117
Медіана	6.546
Середнє значення	6.7521
Розмах	9.5055
Варіанса	8.5479
Стандарт	2.9237
Варіація	0.433
Дисперсія	8.3953
Середнє квадратичне відхилення	2.8975
Асиметрія	0.1651
Ексцес	-1.3035

Приклад 16. Числові характеристики для згрупованих даних