

МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИ
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені ІВАНА ФРАНКА

Кафедра дискретного аналізу
та інтелектуальних систем

Індивідуальне завдання №1
з курсу "Теорія ймовірності та математична статистика"

Виконав:
студент групи ПМі-21
Урбанський Максим

Оцінка

Перевірила:
доц. Квасниця Г.А.

Львів 2024

Постановка задачі:

1. Згенерувати вибірку заданого об'єму (не менше 50) з вказаного проміжку для дискретної статистичної змінної. На підставі отриманих вибірових 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 вибірових даних:
 - утворити інтервальний статистичний розподіл, побудувати гістограму та графік емпіричної функції розподілу, обчислити числові характеристики.

Короткі теоретичні відомості:

Генеральна сукупність – це вся сукупність об'єктів, які досліджуються.

Вибірка або статистичний матеріал – це об'єкти, довільно або випадково відібрані з генеральної сукупності для дослідження.

Обсяг (об'єм) сукупності – це кількість об'єктів цієї сукупності.

Коли реалізується вибірка, кількісна ознака, наприклад X , набуває конкретних числових значень $X = x_i$ які називають **варіантою**.

Варіаційним називають ряд, в якому варіанти упорядковані за зростанням.

Частинні інтервали варіант, які розміщені у зростаючій послідовності, утворюють **інтервальний варіаційний ряд**.

Додатне число n_i , яке показує, скільки разів варіанта x_i ($i=1, \dots, m$) трапляється в таблиці даних, називається **частотою варіанти**. Ряд n_1, n_2, \dots, n_m називається **рядом частот**.

Відношення частоти n_i варіанти x_i називають її **відносною частотою** і позначають через W_i , тобто

$$W_i = \frac{n_i}{n}$$

Статистичний розподіл вибірки встановлює зв'язок між рядом варіант, що зростає або спадає, і відповідними частотами. Він може бути зображений у вигляді

таблиці і називається **таблицею частот**:

x_i	x_1	x_2	\dots	x_m
n_i	n_1	n_2	\dots	n_m

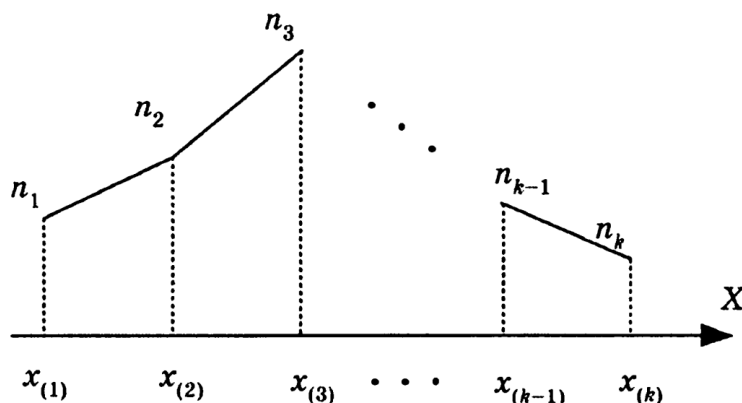
Залежність між упорядкованим рядом варіант і відповідними їм відносними частотами називається **статистичним розподілом відносних частот вибірки**:

x_i	x_1	x_2	\dots	x_m
W_i	$\frac{n_1}{n}$	$\frac{n_2}{n}$	\dots	$\frac{n_m}{n}$

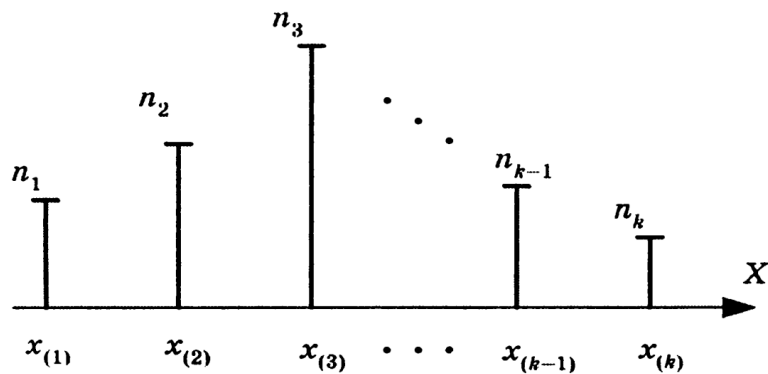
Перелік часткових інтервалів і відповідних їм частот, або відносних частот називають **інтервальним статистичним розподілом вибірки**:

h	$x_0 - x_1$	$x_1 - x_2$	\dots	$x_{k-1} - x_k$
n_i	n_1	n_2	\dots	n_k
W_i	W_1	W_2	\dots	W_k

Дискретний статистичний розподіл вибірки можна зобразити графічно у вигляді ламаної лінії, відрізки якої сполучають координати точок $(x_i; n_i)$, або $(x_i; W_i)$. У першому випадку ламану лінію називають **полігоном частот**, у другому - **полігоном відносних частот**.

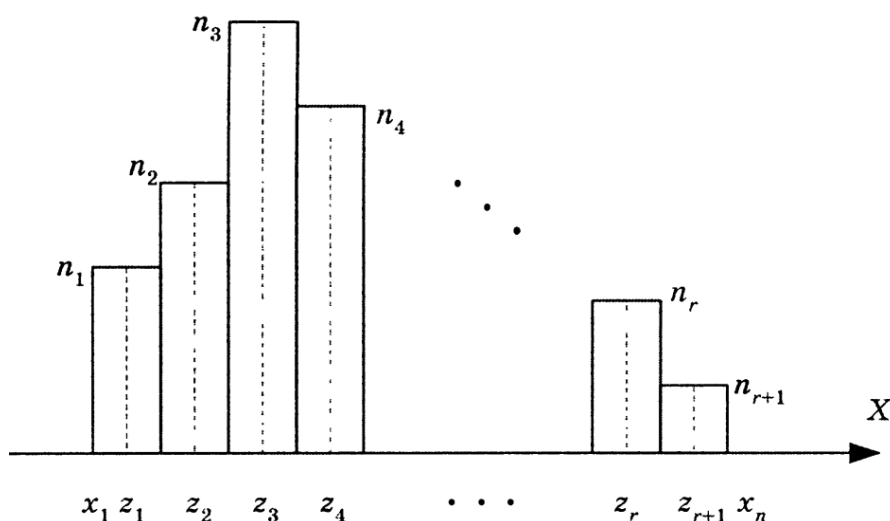


Діаграма частот - на вісі абсцис розташовані можливі значення дискретної величини і в цих точках відповідні частоти n_i



Ступінчасту фігуру, яка складається з прямокутників, основами яких є частинні

інтервали варіант довжиною $h = x_i - x_{i-1}$, а висоти дорівнюють $\frac{n_i}{h}$ або $\frac{W_i}{h}$ називають **гістограмою частот** або **гістограмою відносних частот** відповідно.



Емпіричною функцією розподілу називають функцію $F_n(x)$, яка визначає для кожного значення x відносну частоту події $X < x$.

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ \frac{m_i}{n}, & x_{(i)} \leq x < x_{(i+1)}, \quad (i = \overline{1, n-1}), \\ 1, & x_{(n)} \leq x, \end{cases}$$

де, m_i — кількість елементів вибірки, які не більші за x .

Числові характеристики дискретного розподілу:

- **Середнім арифметичним** називається сума всіх елементів статистичного матеріалу, поділена на обсяг статистичного матеріалу, позначається \tilde{x} .

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

Модою статистичного матеріалу називають той елемент цієї вибірки, який найчастіше трапляється у ній.

- Моду вибірки позначаємо Mo . Очевидно, що вибірка може мати декілька мод.
- **Медіаною** називають варіанту, яка ділить варіаційний ряд на дві частини, які рівні за числом варіант.
- **Девіація** - сума квадратів відхилень елементів статистичного матеріалу від середнього арифметичного.

$$dev = \sum_{i=1}^n n_i (x_i - \tilde{x})^2$$

- **Варіансою** називають середнє квадратичне відхилення значень у вибірці від

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

їх середнього значення, поділене на $n - 1$:

- **Вибірковою дисперсією** називають середнє значення квадратів відхилення варіант від дисперсії з урахуванням відповідних частотей.

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

- **Вибірковим середньоквадратичним відхиленням** називають квадратний корінь із вибіркової дисперсії:

$$\sigma = \sqrt{D}.$$

- **Стандартом** називається додатній арифметичний корінь з варіанси

$$S = +\sqrt{S^2}$$

- **Розмах** – це різниця між найбільшим і найменшим значенням варіаційного ряду (між крайніми елементами)

$$R = x_{max} - x_{min}$$

- **Асиметрією** (γ_1) або скошеністю статистичного матеріалу називається відношення третього центрального моменту до другого центрального моменту в степені півтора.

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

- **Ексцесом** (γ_2) (крутістю, сплюсненістю) статистичного матеріалу називається відношення четвертого центрального моменту до другого центрального моменту в квадраті мінус три.

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

Моментом або статистичним моментом порядку k відносно константи c вибірки x_1, x_2, \dots, x_n називають середнє арифметичне k -их степенів усіх відхилень елементів цієї вибірки від константи c .

Цю величину позначають $M_k(c)$. Отже,

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

•

Програмна реалізація:

Для виконання завдання я використовував мову програмування Python, середовище Jupyter Notebook і бібліотеку `matplotlib.pyplot` для графіків

Завдання 1

Користувач вводить мінімальне і максимальне значення для елементів вибірки, а також її розмір, після чого генерується масив рандомних значень в заданих межах.

Функцією `get_variant_row` сортую масив і виводжу варіаційний ряд.

Функції `get_frequency_table` і `show_frequency_table` створюють таблицю чистот і виводять її.

`freq_diagram` і `freq_polygon` - функції для представлення статистичного матеріалу у вигляді діаграми частот і полігону частот відповідно.

Функції, які обчислюють числові характеристики:

`get_average`, `get_mediana`, `get_moda`, `get_dev`, `get_variance`, `get_dyspersia`
`get_square_avg`, `get_standart`, `get_scope`, `get_moment`, `get_asymetry`, `get_exes`.

Функція `emp_func` створює емпіричну функцію а функція `plot_emp_func` рисує її.

Завдання 2

Користувач вводить число інтервалів на які хоче поділити вибірку.

Функція `create_intervals` формує інтервали з вибірки.

Функція `plot_histogram` подає дані у вигляді гістограми

`plot_interval_emp_func` - функція, яка рисує емпіричну функцію для інтервальної вибірки. Для знаходження цієї функції використовується раніше створена функція `emp_func`.

Функції, які обчислюють числові характеристики для інтервальної вибірки:

`get_interval_average`, `get_interval_median`, `get_interval_moda`, `get_interval_deviation`,
`get_interval_variance`, `get_interval_standart`, `get_interval_dyspersia`,
`get_interval_square_avg`, `get_interval_moment`, `get_interval_asymetry`,
`get_interval_exes`, `get_interval_scope`

Отримані результати:

Кількість: 50

Ліва границя: 3

Права границя: 15

Вибірка:

[7, 8, 8, 11, 12, 4, 14, 9, 8, 5, 10, 8, 13, 13, 7, 6, 14, 13, 13, 3, 15, 8, 8, 5, 14, 7, 11, 5, 7, 13, 4, 7, 6, 3, 6, 13, 8, 5, 12, 7, 13, 5, 14, 10, 14, 9, 9, 13, 15, 7]

Варіаційний ряд:

[3, 3, 4, 4, 5, 5, 5, 5, 5, 6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 8, 8, 9, 9, 9, 10, 10, 11, 11, 12, 12, 13, 13, 13, 13, 13, 13, 13, 13, 14, 14, 14, 14, 14, 15, 15]

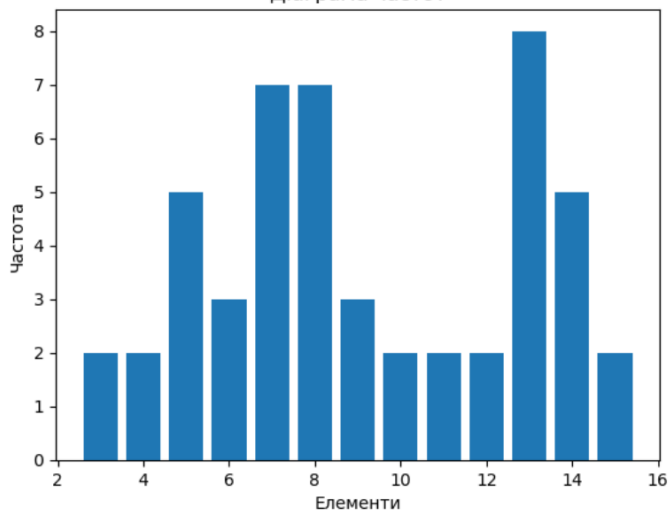
Частотна таблиця:

Частотна таблиця

X_i	n_i
3	2
4	2
5	5
6	3
7	7
8	7
9	3
10	2
11	2
12	2
13	8
14	5
15	2

Графічне представлення статистичного матеріалу:

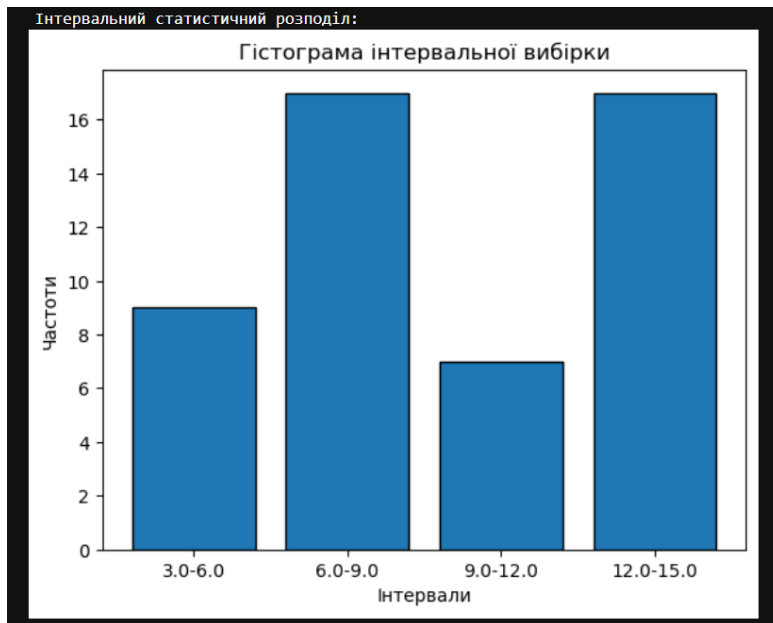
Діаграма частот



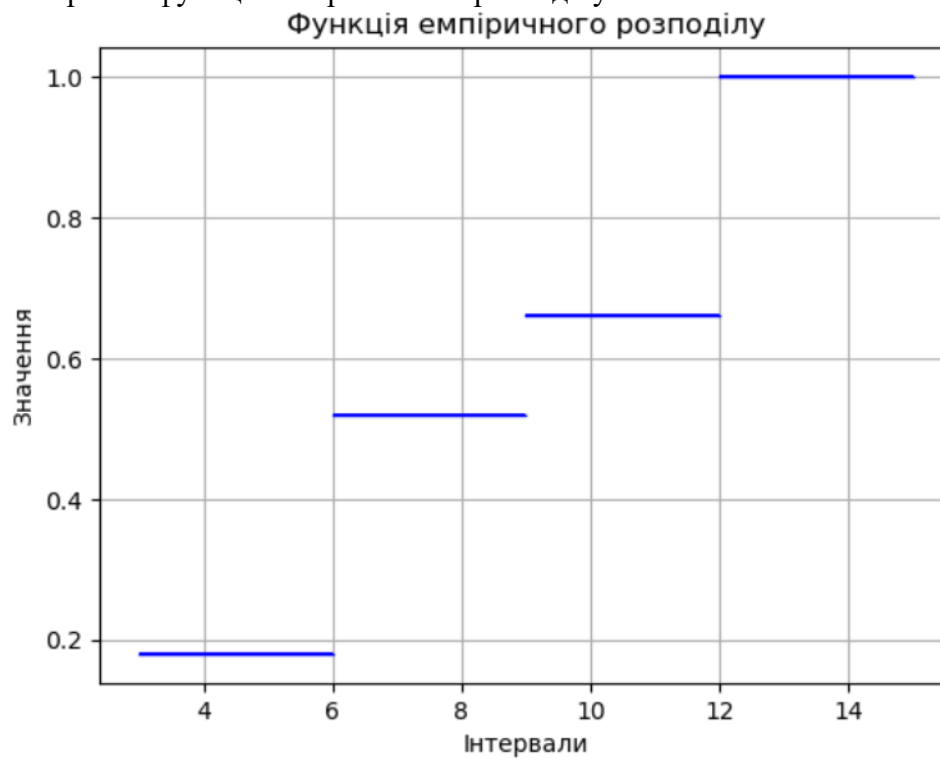
Інтервальний розподіл

Interval	n
(3.0, 6.0)	9
(6.0, 9.0)	17
(9.0, 12.0)	7
(12.0, 15.0)	17

Гістограма



Емпірична функція інтервального розподілу



Числові характеристики:
Середнє арифметичне: 9.42
Медіана: 3.2611764705882353
Мода: 7.333333333333333
Девіація: 571.68
Варіанса: 190.55999999999997
Вибіркова дисперсія: 142.92
Середньоквадратне відхилення вибірки: 11.954915307102764
Варіація: 1.4654296328439402
Стандарт: 13.804347141389917
Розмах: 12.0
Момент 2: 11.433599999999998
Момент 3: -0.5754239999999936
Момент 4: 204.49677312
Асиметрія: -0.014883788151693549
Ексес: -1.4356984689960592

Висновок:

Під час цієї індивідуальної роботи я навчився будувати варіаційний ряд, будувати частотну таблицю, діаграму частот, полігон частот та гістограму частот, будувати емпіричну функцію розподілу та обчислювати всі числові характеристики для інтервального та дискретного розподілів.

