

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ІВАНА ФРАНКА
ФАКУЛЬТЕТ ПРИКЛАДНОЇ МАТЕМАТИКИ ТА ІНФОРМАТИКИ

Кафедра дискретного аналізу та інтелектуальних систем

Індивідуальне завдання №1
з курсу «Теорія ймовірності та математична статистика»

Виконав:

Студент групи ПМІ-23

Лук'янчук Денис

Викладач:

Доц. Пелюшкевич Ольга

Володимирівна

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 кроки:

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:

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**
- 6. Дисперсія вибірки**
- 7. Сигма**

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:

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

Вибірковою дисперсією статистичного розподілу вибірки називають середнє арифметичне квадратів відхилень його варіант від вибіркового середнього:

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

Вибірковим середнім квадратичним відхиленням називається число, яке обчислюється за формулою:

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

Характеристики форми:

1. Асиметрія
2. Екссес

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 (μ): $c = \bar{x}$. Початкового (m):
 $c = 0$

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

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

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

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

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 для першого завдання, а для другого що мінімальний та максимальний елемент визначається:

Статистичний аналіз

Номер варіанту (k):

14

Обсяг вибірки (n):

50

Розрахувати

Полігон частот

Емп. ф-ція (дискр.)

Гістограма

Емп. ф-ція (груп.)

Аналіз вибірки

Розмір вибірки (n):

Мінімальний елемент (k):

Максимальний елемент:

Виконати аналіз

Показати гістограму

Показати емпіричну функцію

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.

Завдання №1

Вибірка, варіаційний ряд, частотна таблиця

Вибірка: [20 17 24 21 18 20 23 16 20 24 24 21 18 17 21 21 16 19 18 15 21 19 15 18
14 23 19 22 14 24 24 23 16 20 17 22 16 18 16 20 18 22 20 15 17 22 15 23
22 23]

Варіаційний ряд: [14 14 15 15 15 15 16 16 16 16 16 17 17 17 17 18 18 18 18 18 18
19 19 19
20 20 20 20 20 20 21 21 21 21 21 22 22 22 22 22 23 23 23 23 23 24 24 24
24 24]

Частотна таблиця:

Значення | Частота | Відн. частота

14		2		0.040
15		4		0.080
16		5		0.100
17		4		0.080
18		6		0.120
19		3		0.060
20		6		0.120
21		5		0.100
22		5		0.100
23		5		0.100
24		5		0.100

Дискретний розподіл:

Числові характеристики (дискретний розподіл):

Середнє: 10.42

Медіана: 11.00

Мода: [np.int32(11), np.int32(9)]

Дисперсія (ген. сукупність): 8.96

Дисперсія вибірки: 9.15

Сигма (ген. сукупність): 2.99

Стандартне відхилення вибірки: 3.02

Розмах: 10.00

Середнє абсолютне відхилення: 2.60

Коефіцієнт варіації (%): 29.02

Квантілі:

Децилі: [6, 7, 8, 9, 11, 11, 12, 13, 14]

Міждецильний розмах: 8.00

Коефіцієнт асиметрії: -0.09

Коефіцієнт ексцесу: -1.16

Інтервальний розподіл:

Інтервальна частотна таблиця:			
Інтервал	Частота	Відн. частота	Середина
[14-16)	6	0.120	15.00
[16-18)	9	0.180	17.00
[18-20)	9	0.180	19.00
[20-22)	11	0.220	21.00
[22-24)	15	0.300	23.00

Числові характеристики (згруповані дані):

Середнє: 19.80

Медіана: 20.18

Мода: 22.42

Дисперсія вибірки: 7.84

Стандартне відхилення: 2.80

Розмах: 10

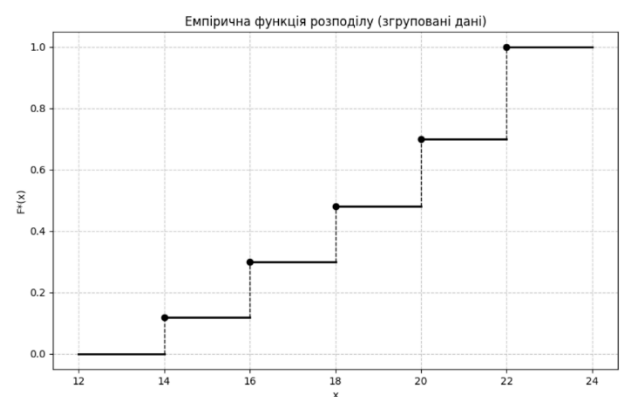
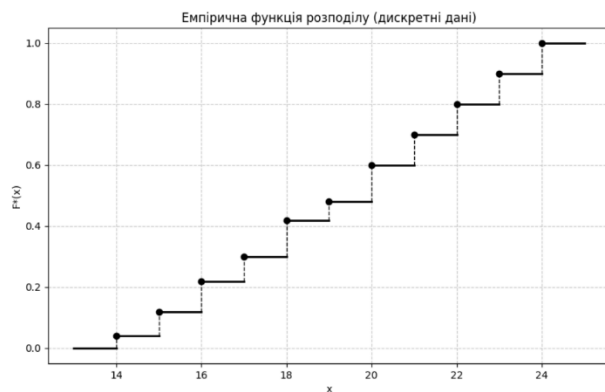
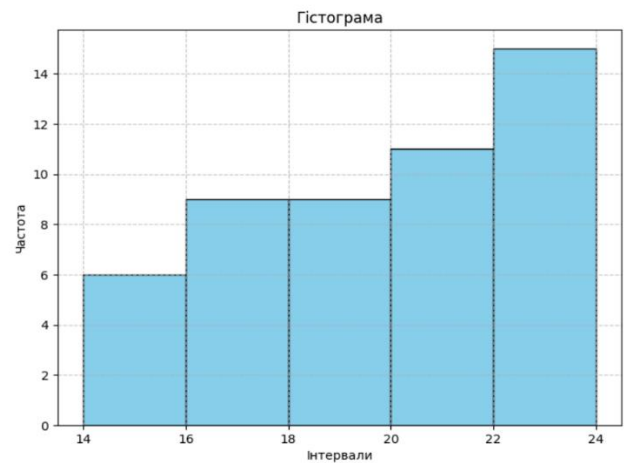
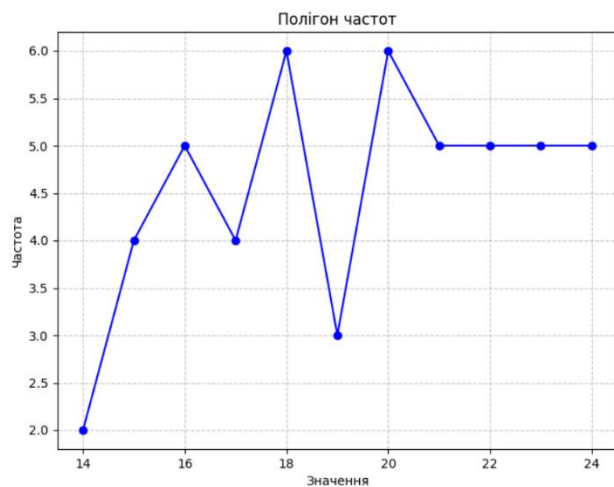
Середнє абсолютне відхилення: 2.45

Коефіцієнт варіації: 14.14

Коефіцієнт асиметрії: -0.32

Коефіцієнт ексцесу: -1.26

Графіки до 1 завдання:



Завдання №2

Вибірка, варіаційний ряд, частотна таблиця

Згенерована вибірка:

3.1, 13.0, 9.7, 2.7, 13.0, 13.1, 4.7, 6.8, 13.3, 2.6, 13.5, 8.1, 7.6, 7.8, 5.0, 9.6, 6.3, 8.6, 10.9, 11.6, 4.7, 2.4, 9.0, 1.0, 9.7, 5.5, 13.3, 12.1, 11.9, 14.9, 13.1, 3.3, 3.2, 1.5, 4.7, 6.1, 5.7, 10.6, 5.0, 9.7, 3.5, 12.9, 13.5, 14.7, 9.6, 4.6, 13.5, 15.0, 5.1, 1.5

Посортований варіаційний ряд:

1.0, 1.5, 1.5, 2.4, 2.6, 2.7, 3.1, 3.2, 3.3, 3.5, 4.6, 4.7, 4.7, 4.7, 5.0, 5.0, 5.1, 5.5, 5.7, 6.1, 6.3, 6.8, 7.6, 7.8, 8.1, 8.6, 9.0, 9.6, 9.6, 9.7, 9.7, 9.7, 10.6, 10.9, 11.6, 11.9, 12.1, 12.9, 13.0, 13.0, 13.1, 13.1, 13.3, 13.3, 13.5, 13.5, 13.5, 14.7, 14.9, 15.0

Частотна таблиця:

	Клас	Частота	Омега	Відносна частота	Кумулятивна частота
0	[1.0, 3.33)	9	1.5	9/50	0.18
1	[3.33, 5.67)	9	1.5	9/50	0.36
2	[5.67, 8.0)	6	1.0	6/50	0.48
3	[8.0, 10.33)	8	1.3	8/50	0.64
4	[10.33, 12.67)	5	0.8	5/50	0.74
5	[12.67, 15.0]	13	2.2	13/50	1.00

Статистичні характеристики:

Кількість класів за формулою Стерджеса: 6

Середнє: 8.23

Медіана: 8.29

Мода: 13.56

Девіація: 936.44

Розмах: 14.00

Варіанса: 19.11

Стандарт: 4.37

Коефіцієнт варіації: 0.53

Дисперсія: 18.73

Середнє квадратичне: 4.33

Початковий момент 1: 8.23

Центральний момент 2: 18.73

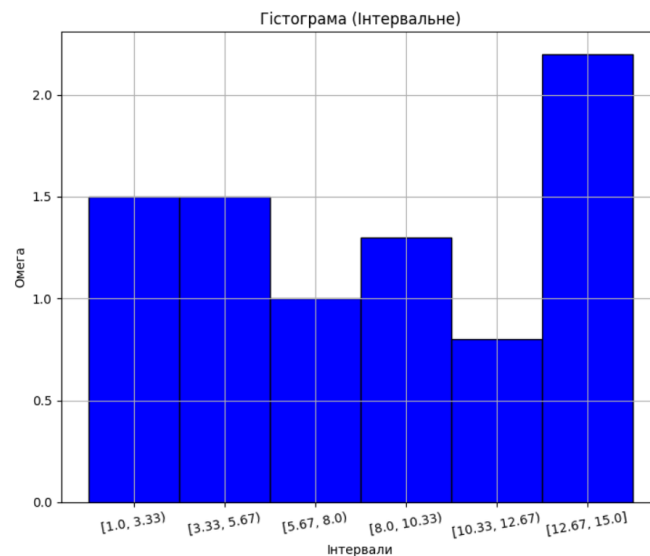
Центральний момент 3: -0.61

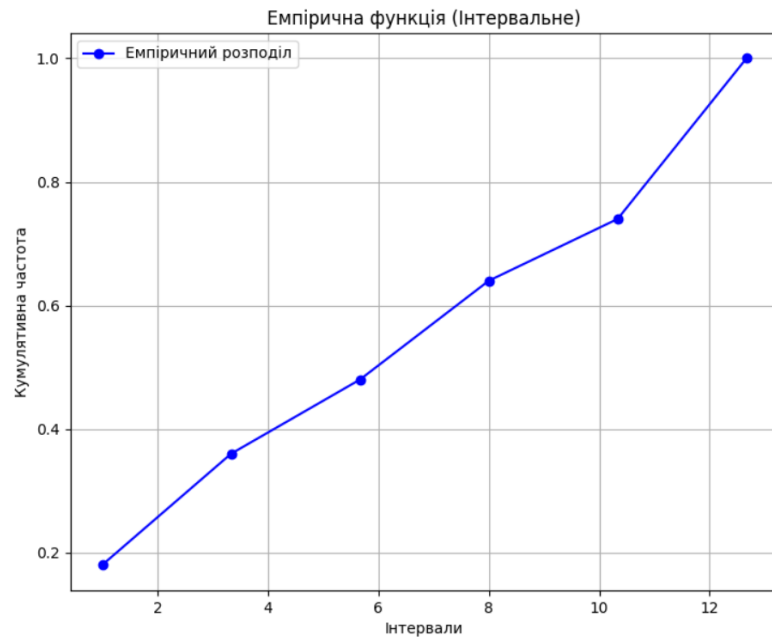
Центральний момент 4: 546.46

Асиметрія: -0.01

Екссес: -1.44

Графіки до 2 завдання:





Висновок: У цій лабораторній роботі я детально вивчив процес обрахунку та аналізу випадкових вибірок. Я розробив програму, яка може генерувати випадкові вибірки чисел за заданими критеріями, будувати варіаційний ряд і обчислювати числові характеристики, такі як центральна тенденція, розсіювання, квантилі, децилі, центилі та мілілі. Це дозволило мені краще розуміти взаємозв'язки між різними характеристиками та знаходити оптимальні методи їх обрахунку й візуалізації.