

Міністерство Освіти та Науки України
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені ІВАНА ФРАНКА
Факультет прикладної математики та інформатики

Теорія Імовірності та Математична Статистика

ІНДИВІДУВАЛЬНЕ ЗАВДАННЯ № 1

Виконав студент
Групи ПМІ-24
Кутко Остап

Постановка задачі:

1. Згенерувати вибірку заданого об'єму (не менше 50) з проміжку $[k; k+10]$ (кномер варіанту) для дискретної статистичної змінної. На підставі отриманих вибіркових даних:

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

1.2. 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:

2.1. утворити інтервальний статистичний розподіл, побудувати гістограму та графік емпіричної функції розподілу, обчислити числові характеристики для згрупованих даних.

Короткі теоретичні відомості

Частотна таблиця або статистичний розподіл - таблиця з двома рядками у першому рядку виписуємо в зростаючому порядку можливі значення варіанти, а в другому – відповідні їм частоти.

Варіаційний ряд – упорядковані спостереження $x_{(i)}$.

Емпірична функція розподілу - це функція, яка вказує ймовірність, що дискретна випадкова змінна X менше або дорівнює певному значенню x . Зазвичай обчислюється як кумулятивна сума відносних частот для всіх значень менших або рівних x .

Різниця між найбільшим і найменшим елементами статистичного матеріалу називається **розмахом статистичного матеріалу**.

Для графічного представлення частотної таблиці на вісь абсцис наносимо можливі значення дискретної мінливої величини та відкладемо в цих точках відповідні частоти $n_i (i = 1, 2, \dots)$. Отримаємо **діаграму частот**.

Якщо з'єднати відрізками сусідні пункти $(x_{(i)}, n_i)$, то дістанемо **полігон частот**.

Емпірична функція розподілу для неперервної статистичної змінної - це функція, яка вказує ймовірність того, що випадкова змінна X буде менше або дорівнювати певному значенню x .

Гістограма частот для неперервного випадку - це графік, що показує розподіл частот для різних значень дискретної змінної. Кожна стовпчик гістограми відповідає діапазону значень, а висота стовпчика відображає частоту (або кількість випадків) у цьому діапазоні.

Також можна представити неперервну змінну у вигляді статистичного розподілу, який показує діапазон змінної та її відносні частоти або ймовірності у цьому діапазоні.

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. Вибіркове середнє статистичного розподілу** – середнє арифметичне її варіант $x_{(i)}$, з врахуванням їхніх частот.

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

- 2. Мода** – те значення варіанти $x_{(i)}$, якому відповідає найбільша частота. Для неперервного розподілу мода (моди) обчислюється так:
 $Mo = Z_{k-1} + (Z_k - Z_{k-1}) * (n_k - n_{k-1}) / (2 * n_k - n_{k-1} - n_{k+1})$, де $[Z_{k-1}, Z_k]$ – проміжок з найбільшою частотою, n_{k-1} – частота попереднього проміжку, n_k – частота обраного проміжку, n_{k+1} – частота наступного проміжку

- 3. Медіана** – варіанта, з обох сторін якої міститься не більше половини варіант, враховуючи їхню частоту.
 $Me = x_{k+1}, n = 2k + 1;$
 $Me = \{(x_k + x_{k+1}) / 2\}, n = 2k.$

Для неперервного розподілу:

$$Me = Z_{k-1} + (Z_k - Z_{k-1}) * (n / 2 - m_{k-1}) / n_k,$$

де $[Z_{k-1}, Z_k]$ – проміжок зліва і справа від якого кількість елементів не перевищує половини кількості елементів вибірки, m_{k-1} – накопичена частота попереднього інтервалу, n_k – кількість елементів у цьому проміжку.

Характеристики розсіювання

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

$$\rho = x_{max} - x_{min}$$

2. **Девіація** – сума квадратів відхилень елементів варіаційного ряду від середнього:

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

3. **Варіанса** – девіація, поділена на обсяг статистичного матеріалу крім одного:

$$s^2 = \frac{dev}{n - 1}$$

4. **Стандарт** – корінь від варіанси

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

5. **Варіація** – відношення стандарту до середнього:

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

6. **Вибіркова дисперсія** – середнє арифметичне квадратів відхилень варіант від вибіркового середнього

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

7. **Середньоквадратичне відхилення** – корінь з вибіркової дисперсії

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

Характеристики форми

Центральний момент статистичної змінної :

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

Де \bar{x} – середнє вибірки.

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

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

Якщо $A, > 0$, то статистичний матеріал скошений вправо.

Якщо $A, < 0$, то статистичний матеріал скошений вліво.

Якщо $A, = 0$, то статистичний матеріал симетричний стосовно середини проміжку.

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

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

У разі $E, > 0$ статистичний матеріал - високовершинний.

У разі $E, < 0$ статистичний матеріал - низьковершинний.

У разі $E, = 0$ статистичний матеріал - нормальновершинний.

Квантилем вибірки порядку a якщо він існує, називають той елемент варіаційного ряду, до якого маємо $a\%$ елементів статистичного матеріалу.

Квартилі – квантилі порядку 25, 50, 75.

Октилі – кванитлі порядку 12.5, 25, ..., 87.5.

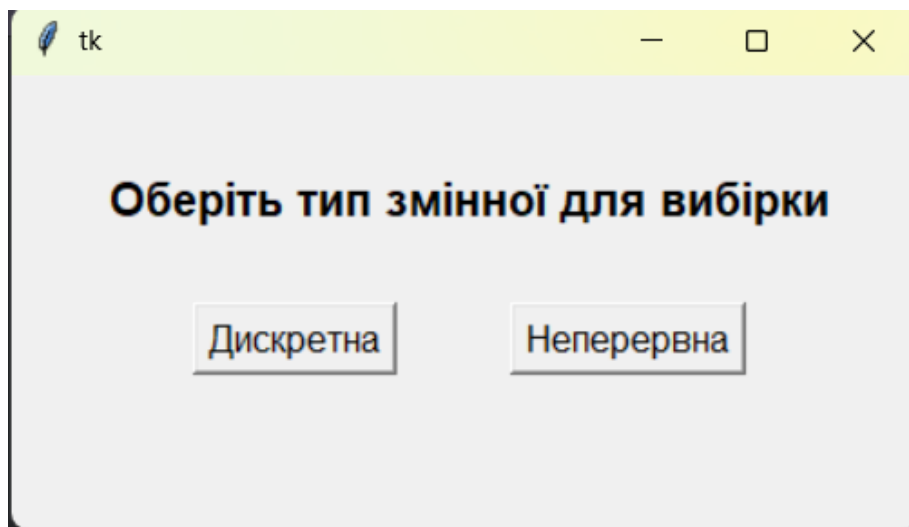
Децилі – кванитлі порядку 10, 20, 30, ..., 90.

Центилі – кванитлі порядку 1, 2, 3, ..., 99.

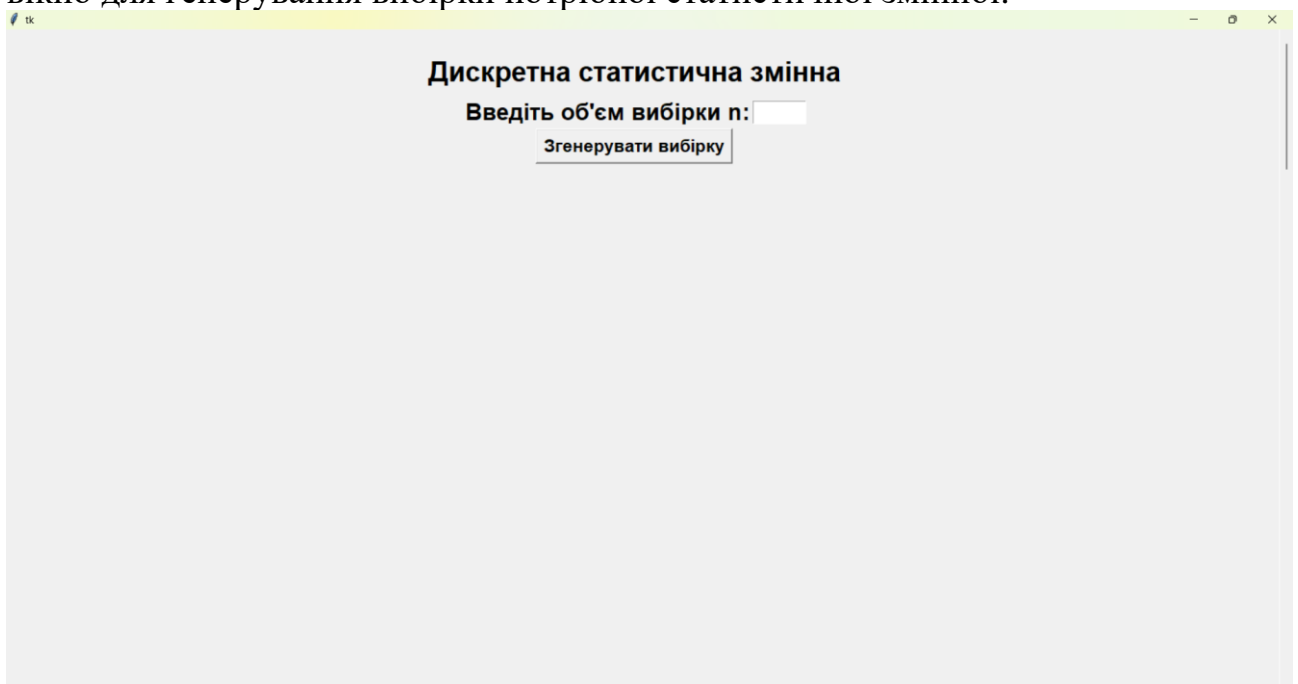
Мілілі – кванитлі порядку 1, 2, 3,..., 999.

Програмна реалізація:

Програма поділена на дві частини: вибірка для дискретної статистичної змінної та вибірка для неперервної статистичної змінної.



Після того, як користувач обрав один з варіантів, перед ним відкривається нове вікно для генерування вибірки потрібної статистичної змінної.



У відведеному полі вводимо потрібний об'єм вибірки і натискаємо на кнопку “Згенерувати вибірку”.

Після цього програма генерує вибірку заданого об'єму і з заданого проміжку (за умовами завдання і згідно з моїм номером варіанту це проміжок [6,16]), а також проводить всі необхідні обчислення і у вікні з'являється повний аналіз вибірки: варіаційний ряд, частотна таблиця, полігон та діаграма, емпірична функція та її графік, числові характеристики, кількість інтервалів та довжину кроку для групування даних, інтервальный розподіл, графік емпіричної функції та гістограма інтервального розподілу та числові характеристики інтервального розподілу. (у випадку неперервної статистичної змінної аналіз лише інтервального розподілу).



Програму реалізував на мові python, серед використаних структур списки та словники, використані бібліотеки: tkinter, numpy, pandas, matplotlib, random, math.

Отримані результати:

Дискретна статистична змінна:

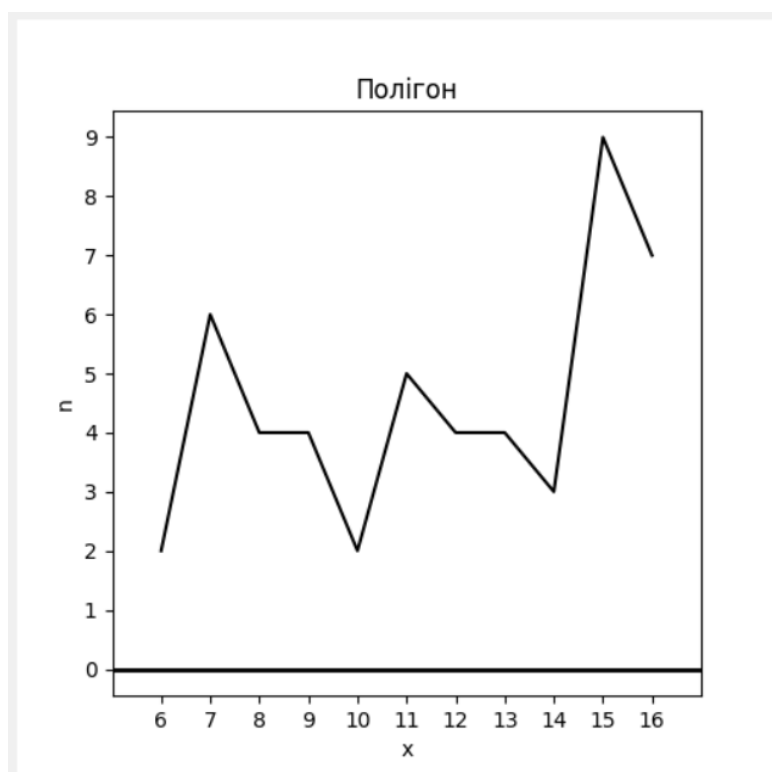
Ввід в текстове поле: 50

Згенерований варіаційний ряд: 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 9, 10, 10, 11, 11, 11, 11, 12, 12, 12, 12, 13, 13, 13, 13, 14, 14, 14, 15, 15, 15, 15, 15, 15, 15, 15, 16, 16, 16, 16, 16, 16, 16

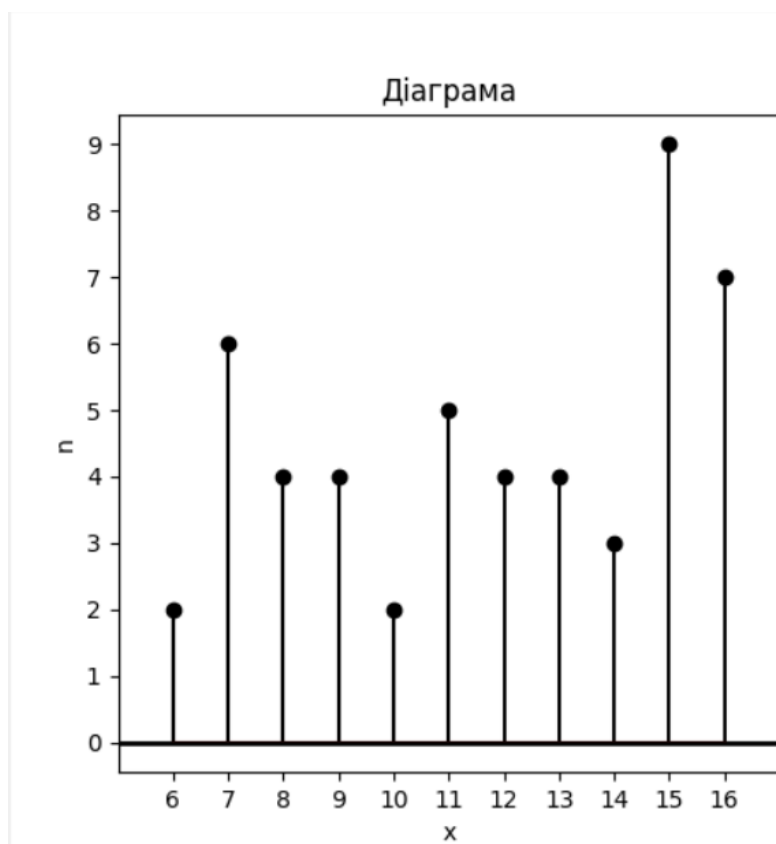
Частотна таблиця:

Частотна таблиця											
x_i	6	7	8	9	10	11	12	13	14	15	16
n_i	2	6	4	4	2	5	4	4	3	9	7

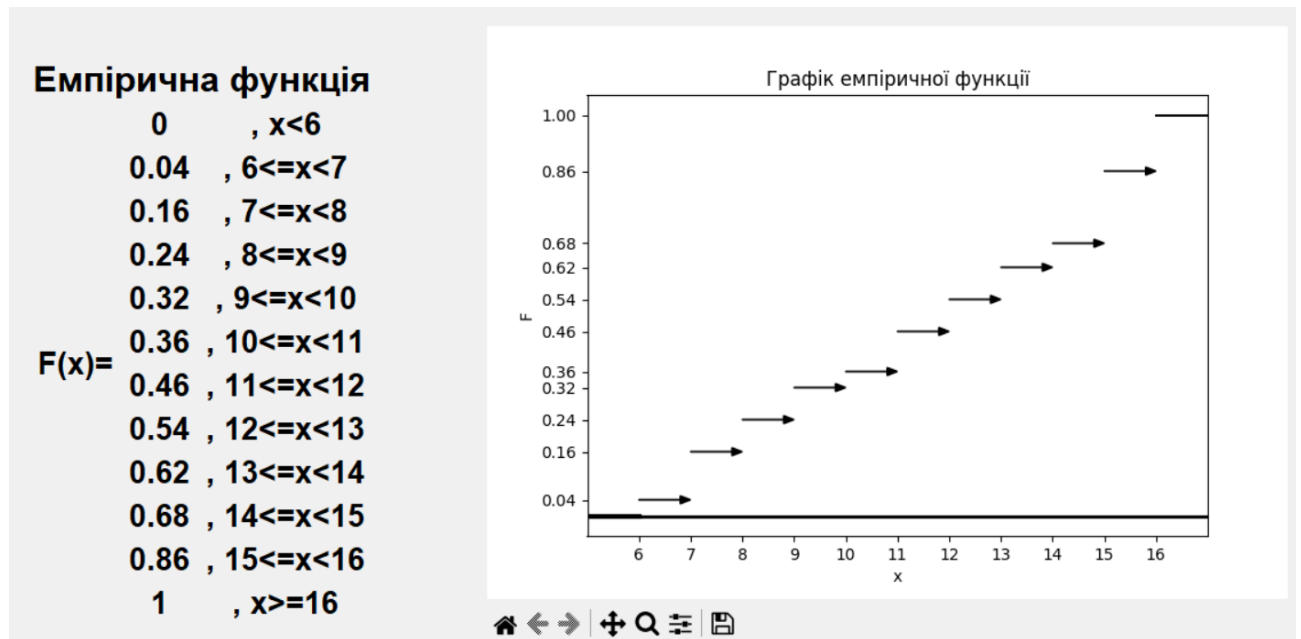
Полігон частот:



Діаграма частот:



Емпірична функція та її графік:



Числові характеристики:

Числові характеристики						
Медіана:	12.0	Стандарт:	3.3199	Квартилі:	-	Широти
Мода:	15	Виб. дисперсія:	10.8016	Октилі:	-	Інтерквартильна: -
Сер. арифм.:	11.72	Середньокв. відх.:	3.2866			Інтероктильна: -
Розмах:	10	Варіація:	0.2833	Децилі:	D1=7, D2=8, D3=9, D4=11,	Інтердецильна: 9
Девіація:	540.08	Асиметрія:	-0.2215		D5=12, D6=13, D7=15, D8=15, D9=16	Інтерцентильна: -
Варіанса:	11.022	Екссес:	-1.3535	Центилі:	-	Інтермілільна: -
				Мілілі:	-	

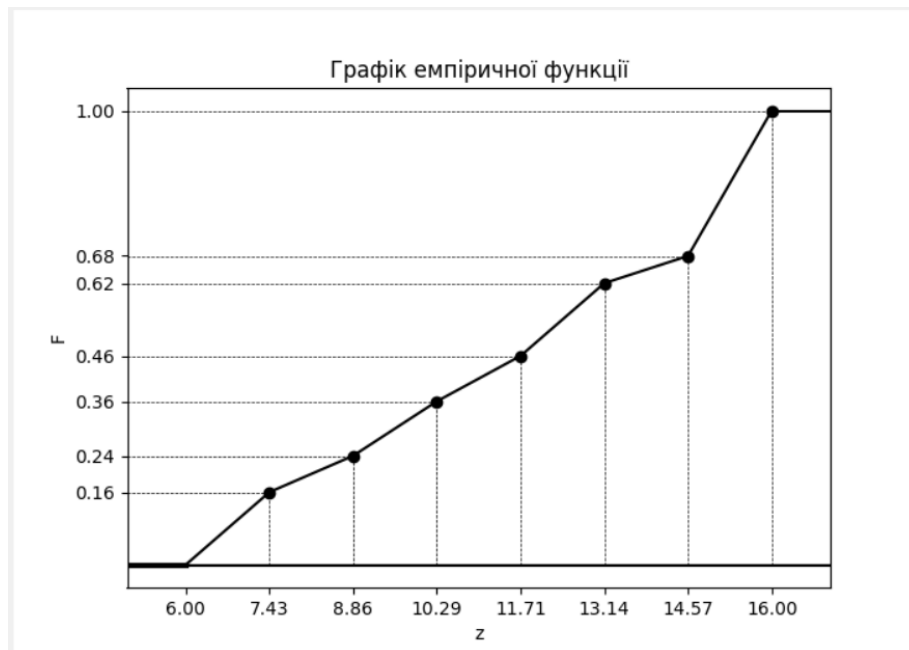
Кількість інтервалів і їхня довжина для групування даних:

Згрупування даних: $k = 7$; $h = 1.4286$

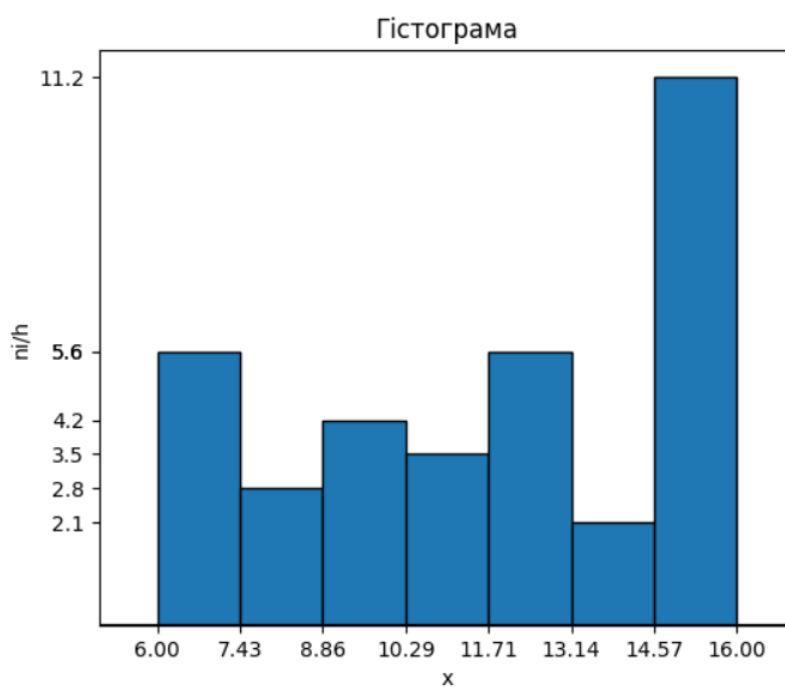
Інтервальний розподіл:

Інтервальний розподіл варіанти							
$[Z_{i-1}; Z_i]$	[6, 7.4286)	[7.4286, 8.8572)	[8.8572, 10.2858)	[10.2858, 11.7144)	[11.7144, 13.143)	[13.143, 14.5716)	[14.5716, 16]
n_i	8	4	6	5	8	3	16
n_i/h	5.5999	2.7999	4.1999	3.4999	5.5999	2.1	11.1998
w_i	0.16	0.08	0.12	0.1	0.16	0.06	0.32

Графік емпіричної функції інтервального розподілу:



Гістограма:



Числові характеристики інтервального розподілу:

Числові характеристики інтервального розподілу

Медіана:	12.0716	Стандарт:	3.204
Мода:	15.2119	Виб. дисперсія:	10.0606
Сер. арифм.:	11.6858	Середньокв. відх.:	3.1718
Розмах:	10	Варіація:	0.2742
Девіація:	503.029	Асиметрія:	-0.275
Варіанса:	10.2659	Екссес:	-1.3412

Дискретна статистична змінна:

Ввід в текстове поле: 50

Згенерований варіаційний ряд: 6.0041, 6.2051, 6.3043, 6.3973, 6.4755, 6.7739, 6.8567, 6.9612, 7.3547, 7.376, 7.5525, 7.5535, 7.7638, 7.8622, 7.9568, 8.0013, 8.2988, 8.4663, 8.552, 8.6187, 8.7138, 9.0325, 9.3754, 9.773, 10.0981, 10.4575, 10.7577, 10.8532, 11.4187, 11.579, 11.7244, 11.795, 11.9113, 12.6917, 12.8834, 13.0912, 13.3894, 13.4228, 13.6998, 13.7201, 13.8505, 13.9859, 14.1682, 14.4932, 15.1546, 15.1833, 15.2658, 15.5481, 15.574, 15.6295

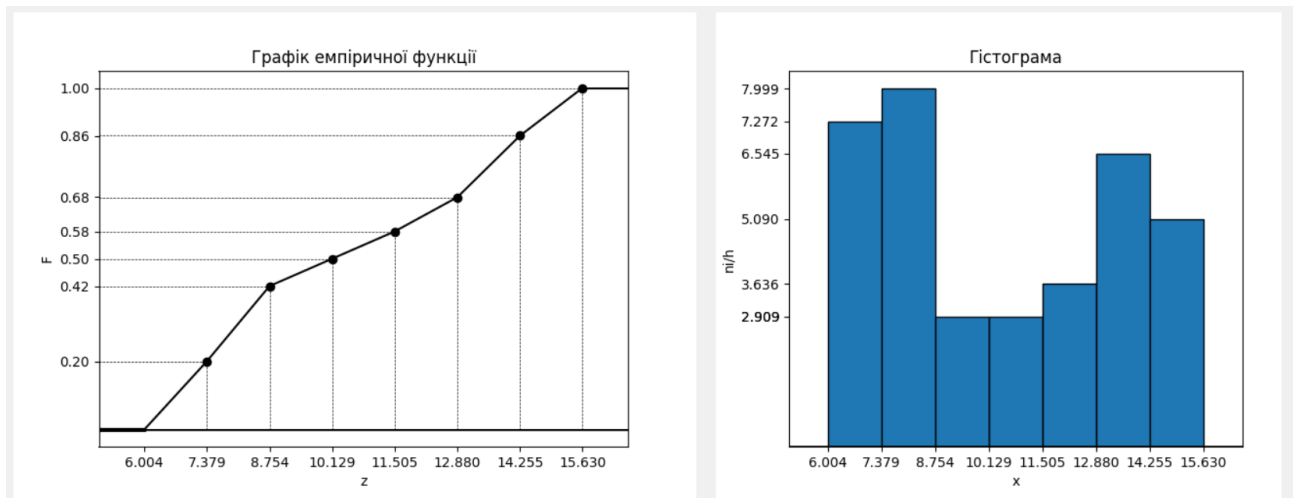
Кількість інтервалів, їхня довжина і інтервальний розподіл:

Згрупування даних: $k = 7$; $h = 1.3751$

Інтервальний розподіл варіанти

$[Z_{i-1}; Z_i]$	[6.0041, 7.3792)	[7.3792, 8.7543)	[8.7543, 10.1294)	[10.1294, 11.5045)	[11.5045, 12.8796)	[12.8796, 14.2547)	[14.2547, 15.6295]
n_i	10	11	4	4	5	9	7
n_i/h	7.2722	7.9994	2.9089	2.9089	3.6361	6.545	5.0905
w_i	0.2	0.22	0.08	0.08	0.1	0.18	0.14

Графік емпіричної функції та гістограма:



Числові характеристики:

Числові характеристики інтервального розподілу

Медіана:	10.1294	Стандарт:	3.0377
Мода:	7.5511	Виб. дисперсія:	9.0429
Сер. арифм.:	10.4869	Середньокв. відх.:	3.0071
Розмах:	9.6254	Варіація:	0.2897
Девіація:	452.1427	Асиметрія:	0.147
Варіанса:	9.2274	Екссес:	-1.5161

Висновок: Я написав програму, яка генерує випадкову вибірку заданого об'єму та на її основі будує варіаційний ряд, частотну таблицю, діаграму та полігон частот, емпіричну функцію та її графік а також обчислює числові характеристики. Крім того програма ділить вибірку на інтервали і працює вже з ними: утворює інтервальний розподіл, гістограму і графік емпіричної функції та обчислює числові характеристики.

Отже написана програма може успішно аналізувати вибірку дискретної і неперервної статистичної змінної, працювати як зі згрупованими, так і з незгрупованими даними, а також представити статистичний матеріал у вигляді таблиць і графіків.