**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**

**ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ІВАНА ФРАНКА**

Кафедра: теорії оптимальних процесів

**ІНДИВІДУАЛЬНЕ ЗАВДАННЯ №1**

**З «МАТЕМАТИЧНОЇ СТАТИСТИКИ»**

Роботу виконала:

**Трущак Анастасія Володимирівна**

студентка групи ПМА-23 Львівського національного університету імені Івана Франка

Роботу перевірила: **Николайчук Леся Василівна**

асистент кафедри теорії оптимальних процесів Львівського національного університету

імені Івана Франка

Львів – 2024

**Тема:** Методи подання статистичного матеріалу. Числові характеристики статистичного матеріалу.

**Мета:**

1. Згенерувати вибірку для дискретної і неперервної змінної за допомогою мови програмування Python.
2. Для дискретної змінної:
   * Побудувати варіаційний ряд.
   * Скласти частотну таблицю.
   * Розрахувати таблицю функції розподілу.
   * Обчислити числові характеристики, такі як мода, медіана, середнє арифметичне вибірки, варіанса, стандарт, дисперсія, середнє квадратичне відхилення, варіація, девіація, 1-4 початковий момент, 1-4 центральний момент, асиметрія, ексцес, квартилі, інтерквартильну широту, децилі, інтердецильну широту, центилі, інтерцентильну широту, мілілі, інтермілільну широту.
   * Побудувати графіки: діаграма частот, полігон частот, графік функції розподілу.
3. Для неперервної змінної:
   * Побудувати таблицю, поділену на інтервали, з вказанням середини інтервалу та кількістю входжень елементів в певний інтервал.
   * Обчислити числові характеристики: мода, медіана, середнє арифметичне вибірки, варіанса, стандарт, дисперсія, середнє квадратичне відхилення, варіація, девіація, 1-4 початковий момент, 1-4 центральний момент, асиметрія, ексцес, квартилі, інтерквартильну широту, децилі, інтердецильну широту, центилі, інтерцентильну широту, мілілі, інтермілільну широту.
   * Побудувати графік: гістограма.

**ТЕОРЕТИЧНІ ВІДОМОСТІ**

Сукупністьназивають **вибіркою**, а кількість **n** елементів у ній – обсягом вибірки.

Ряд називається **варіаційним рядом**.

Таблицю

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **X** |  |  |  |  |  | **…** |  |
| **v** | **n1** | **n2** | **n3** | **n4** | **n5** | **…** | **nk** |

Називають **частотною таблицею** або **статистичним розподілом** **дискретної змінної X**.

Якщо вибірка є вибіркою середнього або великого обсягу з неперервної генеральної сукупності, то таблиця утворюється дещо по-іншому. Ділимо інтервал зміни статистичних даних на r+1 інтервали можливої різної ширини, де , .

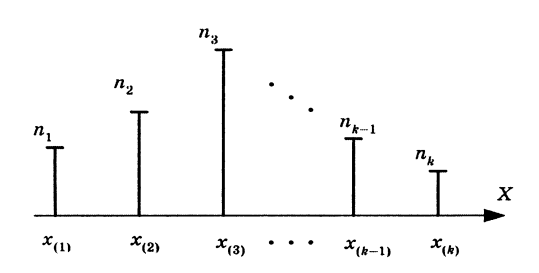
Число = називають **розмахом** **вибірки**.

Нехай ,,…, – середини побудованих інтервалів і нехай елементів варіаційного ряду вибірки потрапляють при такому розбитті на інтервал . Отримаємо згруповані дані, які можна записати таблично.

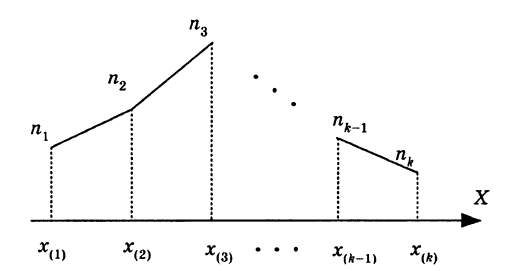
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | **…** |  |
| **Z** |  |  | **…** |  |
|  |  |  | **…** |  |

Таку частотну таблицю називають **статистичним розподілом дискретної варіанти Z, яка відповідає неперервній статистичній змінній X.**

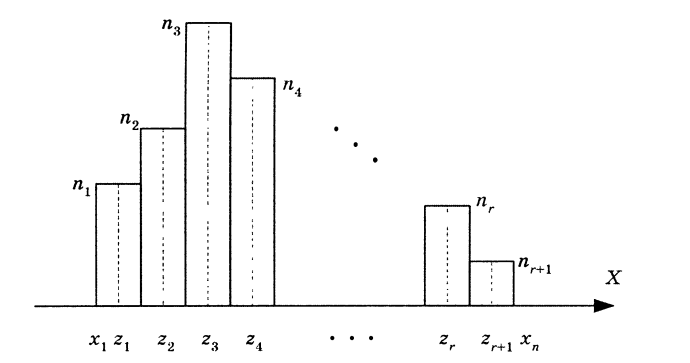
* **r+1** (кількість поділів) вираховують з нерівності: **,** де n – обсяг вибірки.

На осі OX нанесено значення , і в цих точках вверх відкладено вертикальні відрізки довжиною, яка дорівнює відповідним частотам . Цю сукупність відрізків називають **діаграмою частот**.

Ломану лінію, яка є сукупністю відрізків прямих, що з’єднують точки площини XOY з координатами (), , називають **полігоном частот**.

****

Якщо дані згруповано і отримано з неперервної генеральної сукупності, то ці дані можна зобразити за допомогою гістограми частот. Відкладено на осі OX інтервали , отримані в результаті поділу інтервалу . Прийнявши відрізки за основи прямокутників, будуються над віссю OX прямокутники, висоти яких дорівнюють , відповідно. Сукупність таких прямокутників називають **гістограмою частот.**

**

Функцію називають **емпіричною функцією розподілу або емпіричною кумулятою:**

,

де – кількість елементів вибірки, які не більші за *x*.

**ЧИСЛОВІ ХАРАКТЕРИСТИКИ СТАТИСТИЧНОГО МАТЕРІАЛУ**

**Статистики локації**

* **Медіана –** це елемент вибірки, який ділить відповідний варіаційний ряд на дві рівні за обсягом частини. Позначається Me.

Якщо обсяг вибірки ***n*** є непарним числом, тобто n=2\*k+1, де k-натуральне число, то медіана Me= :

k-елементів k-елементів

Якщо обсяг вибірки ***n*** є парним числом, тобто n=2\*k, де k-натуральне число, то медіана Me=:

k-елементів k-елементів

* **Мода** вибірки – це той елемент вибірки, який найчастіше трапляється у ній. Позначається Mo. Вибірка може мати декілька мод.
* **Середнє арифметичне вибірки** – це середнє арифметичне елементів даної вибірки, тобто сума всіх елементів вибірки поділена на кількість елементів у ній. Позначається і обчислюється:
* **Девіація** – це сума квадратів відхилень елементів вибірки від її середнього. Позначається dev і обчислюється:

dev=

**Числові характеристики розсіяння**

* **Розмах вибірки –** це різниця між найбільшим та найменшим елементом вибірки і позначають .

=-.

* **Варіанса статистичної змінної** – це сума квадратів відхилень елементів статистичного матеріалу від середнього цієї вибірки, поділену на обсяг вибірки без одного і позначають :
* **Стандарт** – це арифметичне значення кореня з варіанси. Позначають s.
* **Вибіркова дисперсія** – це міра розсіяння елементів вибірки. Позначають D.

D=

* **Варіація вибірки** – це відношення стандарту цієї вибірки до її середнього. Позначають v.
* **Квантилем** вибірки порядку α, якщо він існує, називають той елемент варіаційного ряду, до якого, включно з ним, є α відсотків елементів статистичного матеріалу.
* **Інтерквантильна широта** порядку β-α – це різниця між квантилем порядку β та квантилем порядку α (α< β).
* **Квартилі** – це квантилі порядків 25,50,75.
* **Інтерквартильна широта вибірки** – це різниця між третім і першим квартилями.
* **Децилі** – це квантилі порядків 10,20, … , 90.
* **Інтердецильна широта** – це різниця між дев’ятим і першим децилями.
* **Центилі** – це квантилі порядків 1,2, … , 99.
* **Інтерцентильна широта** – це різниця між дев’яносто дев’ятим і першим центилями.
* **Мілілі** – це квантилі порядків 0,1;0,2; … ; 99,9.
* **Інтермілільна широта** – це різниця між дев’ятсот дев’яносто дев’ятим і першим мілілями.
* **Моментом** порядку k відносно константи c вибірки називають середнє арифметичне k-их степенів усіх відхилень елементів цієї вибірки від константи c. Позначається :

Якщо c=0, то момент називають **початковим** і позначають :

Якщо с=, де – середнє арифметичне, то момент називають **центральним** і позначають :

**Статистики форми**

* **Асиметрія вибірки –** це відношення її третього центрального моменту до другого центрального моменту в степені півтора. Позначають :

Якщо **0**, то статистичний матеріал **скошений** **вліво**.

Якщо **<0**, то статистичний матеріал **скошений вправо**.

Якщо **=0**, то статистичний матеріал розташований **симетрично**.

* **Ексцес вибірки** – це відношення четвертого центрального моменту до другого центрального моменту в квадраті мінус три. Позначається :

Якщо **>0**, то статистичний матеріал **високовершинний**.

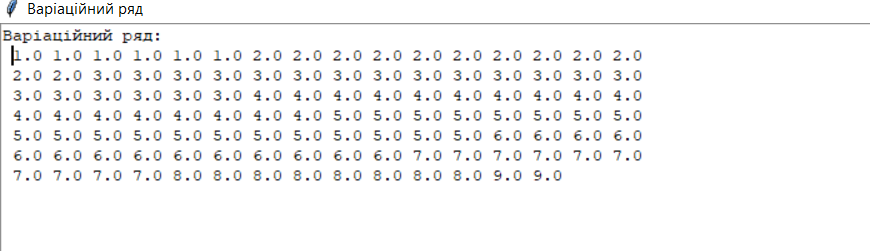
Якщо **0*,*** то статистичний матеріал **низьковершинний**.

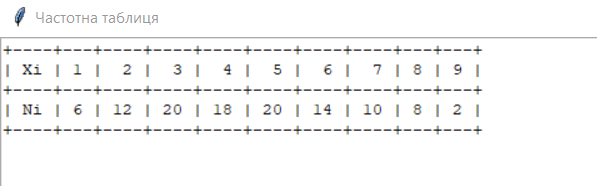
Якщо **=0**, то статистичний матеріал **нормальновершинний**.

**ОТРИМАНІ РЕЗУЛЬТАТИ ДЛЯ ДИСКРЕТНОЇ ЗМІННОЇ**

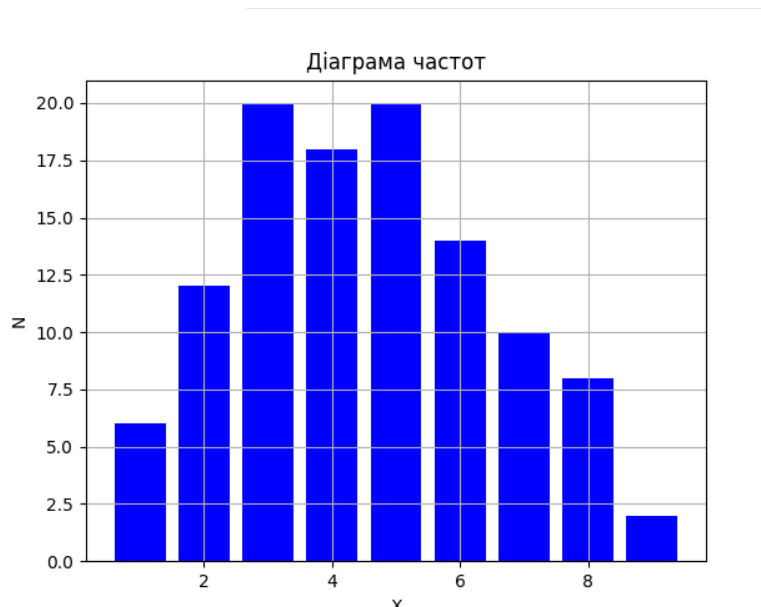
Я згенерувала вибірку з 110 елементів.

**Варіаційний ряд:**

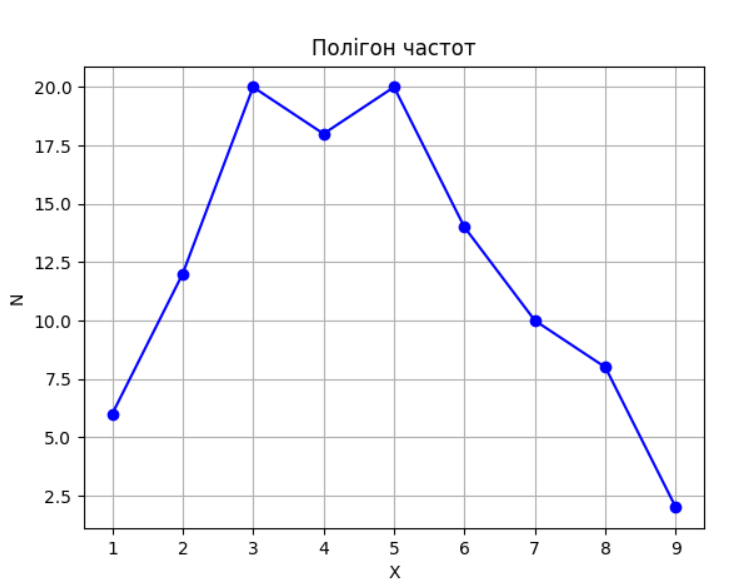
**Частотна таблиця для дискретної змінної**:



**Діаграма частот для дискретної змінної:**



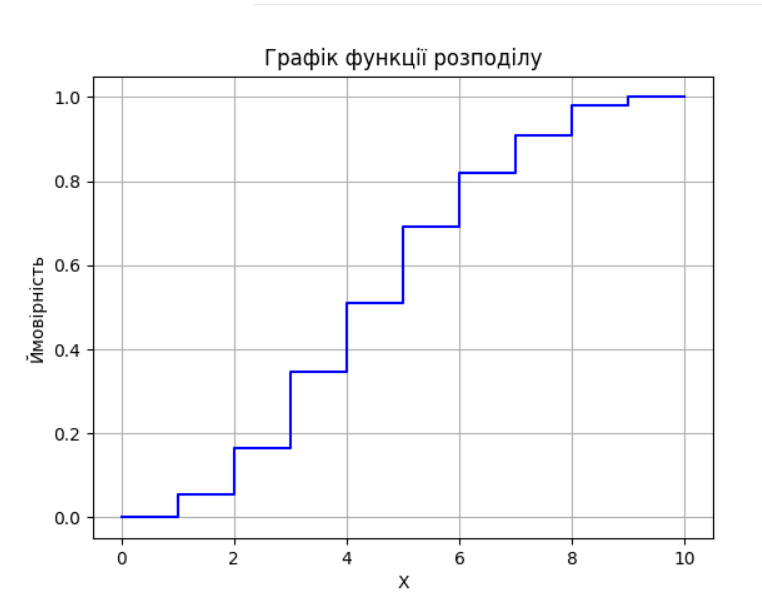
**Полігон частот для дискретної змінної:**

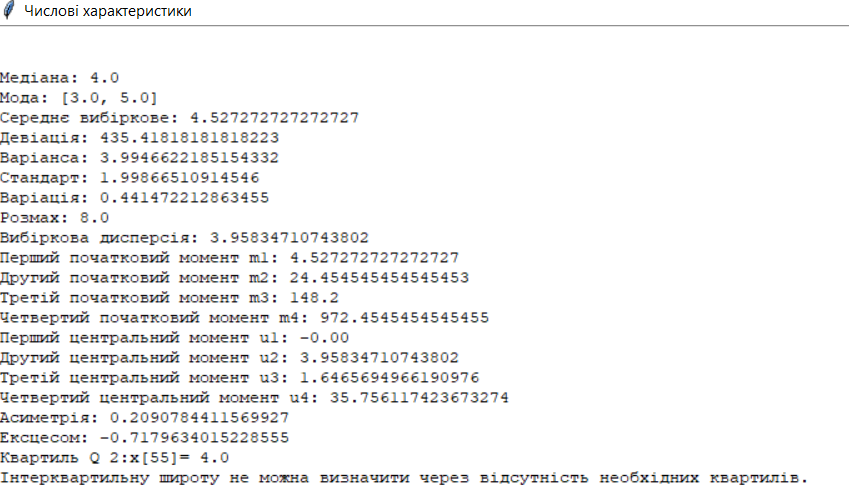
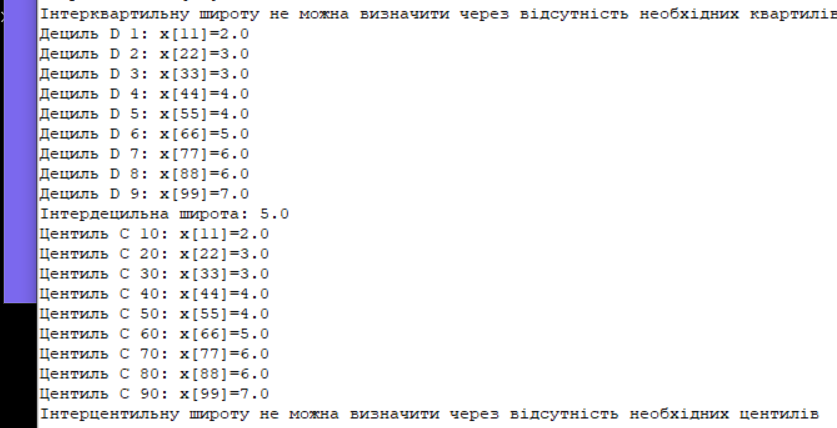
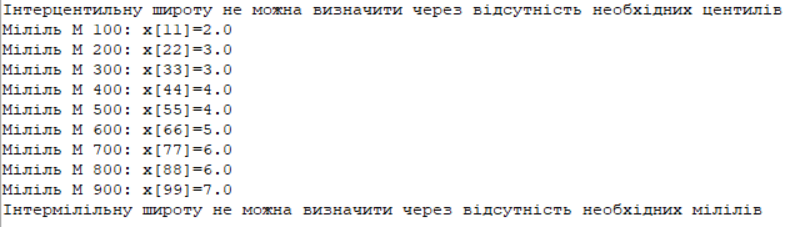


**Таблиця функції розподілу для дискретної змінної:**



**Графік функції розподілу для дискретної змінної:**

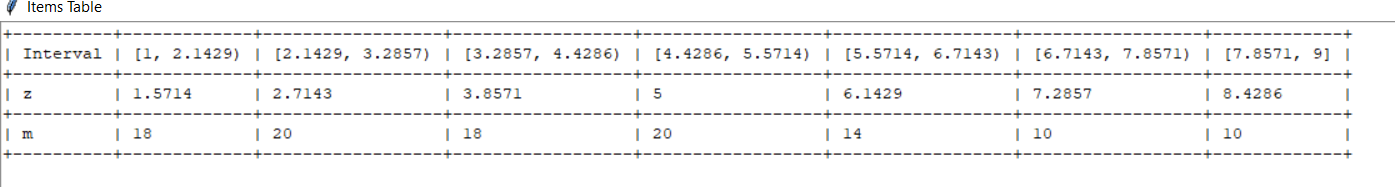


**Числові характеристики для дискретної змінної**

**ОТРИМАНІ РЕЗУЛЬТАТИ ДЛЯ НЕПЕРЕРВНОЇ ЗМІННОЇ**

Я використала ту ж саму вибірку з 110 елементів.

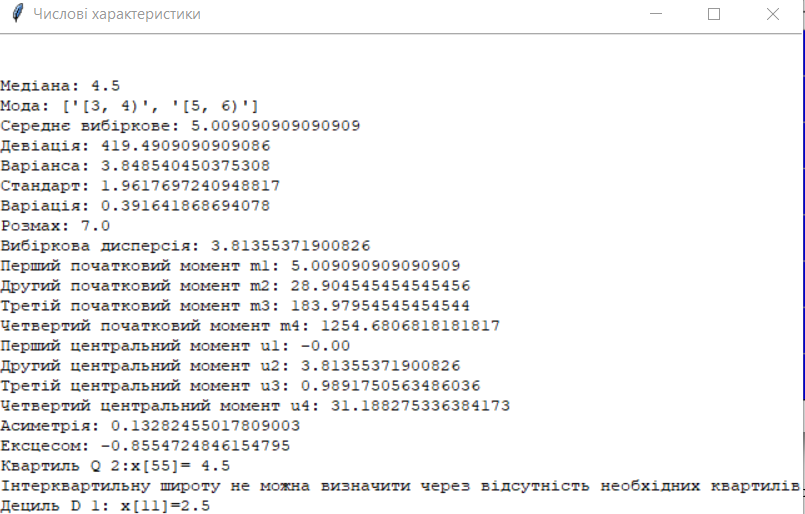
**Таблиця для неперервної змінної:**

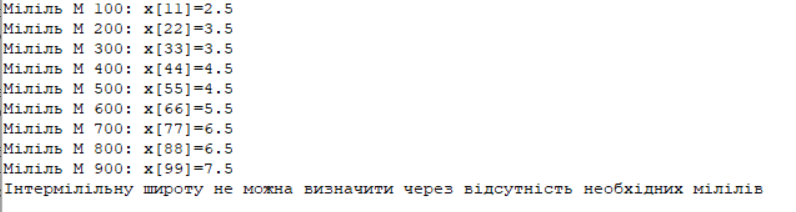
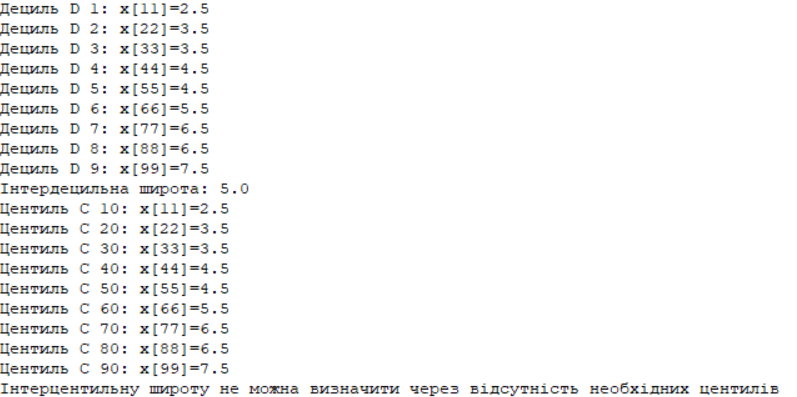


**Гістограма для неперервної змінної:**



**Числові характеристики для неперервної змінної:**

****



**ВИСНОВОК:** У ході виконання індивідуального завдання я успішно виконала поставлені завдання.

Спершу я згенерувала вибірку для дискретної змінної. Для дискретної змінної я побудувала варіаційний ряд, склала частотну таблицю та розрахувала таблицю функції розподілу. Після цього я обчислила числові характеристики, такі як мода, медіана, середнє арифметичне, варіанса, стандарт, дисперсія, середнє квадратичне відхилення, варіація, девіація, початкові та центральні моменти, а також асиметрія та ексцес. Я також обчислила квартилі (інтерквартильну широту не обчислила через відсутність необхідних квартилів), децилі, інтердецильну широту , центилі (інтерцентильну широту не обчислила через відсутність необхідних центилів), мілілі (інтермілільну широту не обчислила через відсутність необхідних мілілів). Також я побудувала графіки, такі як діаграма частот, полігон частот та графік функції розподілу, що допомогли мені візуалізувати розподіл.

Для неперервної змінної я використала ту ж саму вибірку. Побудувала таблицю з інтервалами, серединами інтервалів та кількістю входжень елементів у кожний інтервал. Обчислила різні числові характеристики, такі як мода, медіана, середнє арифметичне, варіанса, стандарт, дисперсія, середнє квадратичне відхилення, варіація, девіація, початкові та центральні моменти, а також асиметрія та ексцес. Я також обчислила квартилі (інтерквартильну широту не обчислила через відсутність необхідних квартилів), децилі, інтердецильну широту , центилі (інтерцентильну широту не обчислила через відсутність необхідних центилів), мілілі (інтермілільну широту не обчислила через відсутність необхідних мілілів). На завершення, я побудувала гістограму.

Отже, це індивідуальне завдання дозволило мені не лише створити вибірку для дискретної та неперервної змінних, але також дослідити їхні розподіли, отримати різноманітні числові характеристики та побудувати відповідні графіки.