МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИ

ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені ІВАНА ФРАНКА

Кафедра дискретного аналізу

та інтелектуальних систем

**Індивідуальне завдання №2**

з курсу ''Теорія ймовірності та математична статистика''

Виконала:

студентка групи ПМі-21

Дудчак Валентина Юріївна

Оцінка

Перевірила:

доц. Квасниця Г.А.

Львів 2024

**Постановка задачі:**

1. Зчитати дані з текстового файлу, побудувати полігон або гістограму частот;
2. На основі графічного представлення сформулювати гіпотезу про закон розподілу досліджуваної ознаки генеральної cукупнocтi;
3. Передбачити можливість користувачу задати параметри розподілу вручну або оцінити на основі даних вибірки;
4. Для заданого користувачем рівня значущості перевірити сформульовану гіпотезу за критерієм 𝜒2.

**Варіант 4**

**Короткі теоретичні відомості:**

Перевірка гіпотез про закон розподілу випадкової величини — це процес, в ході якого статистично перевіряється відповідність спостережуваних даних певному теоретичному розподілу.

Він проходить у кілька етапів:

1. *Формулювання гіпотез*: висувається гіпотеза Н0 про закон розподілу, яким керується вибірка
2. *Визначення рівня значущості*: ймовірність відхилення нульової гіпотези, виражена у відсотках
3. *Обчислення тестової статистики*: цей процес залежить від того, який метод ми обираємо. Наприклад, за критерієм Пірсона, ми обчислюємо величину хі-квадратичне
4. *Прийняття рішення*: порівнюючи отримані значення з критичним значенням, ми приймаємо або відхиляємо початкову гіпотезу

Для критерія Пірсона ми обчислюємо величину хі-критичне, використовуючи рівень значущості та ступені свободи.

**Програмна реалізація:**

Для написання коду я використала мову програмування Python, середовище

Jupyter Notebook і наступні бібліотеки: pandas, numpy, matplotlib, scipy

Спочатку я зчитую дані вибірки з файлу за допомогою методів pandas, показую їх у вигляді таблиці та будую гістограму за допомогою бібліотеки matplotlib. Потім показую висунуту гіпотезу. Користувач може або ввести дані розподілу самостійно, або вибрати, щоб вони були оцінені. Він також вводить значення рівня значущості.

Потім обчислюю ймовірності для кожного проміжку в окремій функції. Для першої задачі, з неперервним випадком, я використовую функцію лапласа; для другої, з експонентним випадком, я використовую F(x) для експонентного закону розподілу.

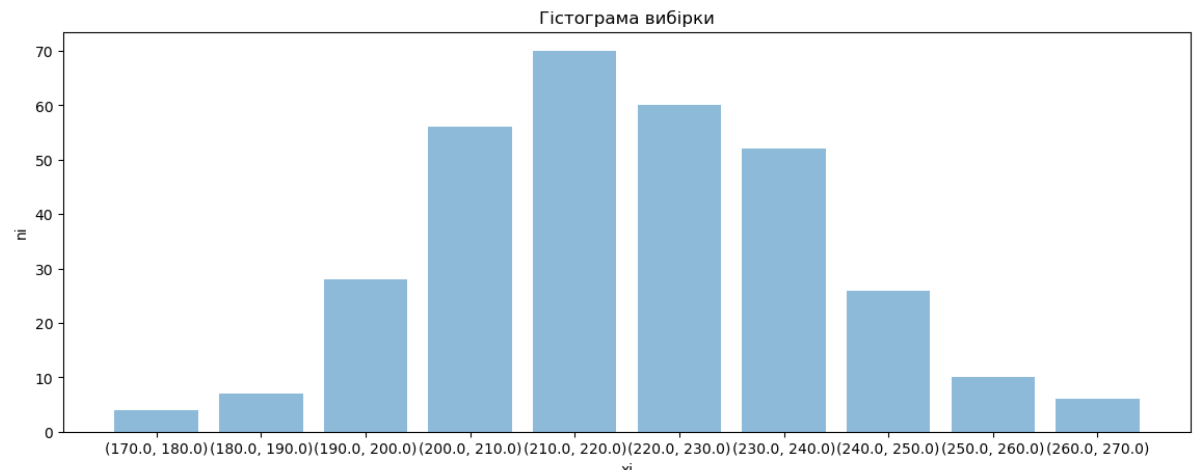
Після цього я перевіряю, чи для всіх рядків виконується критерій Пірсона, і якщо ні, то об’єдную ті з них, де це потрібно. Роблю це в окремій функції, використовуючи методи pandas — iloc та concat.

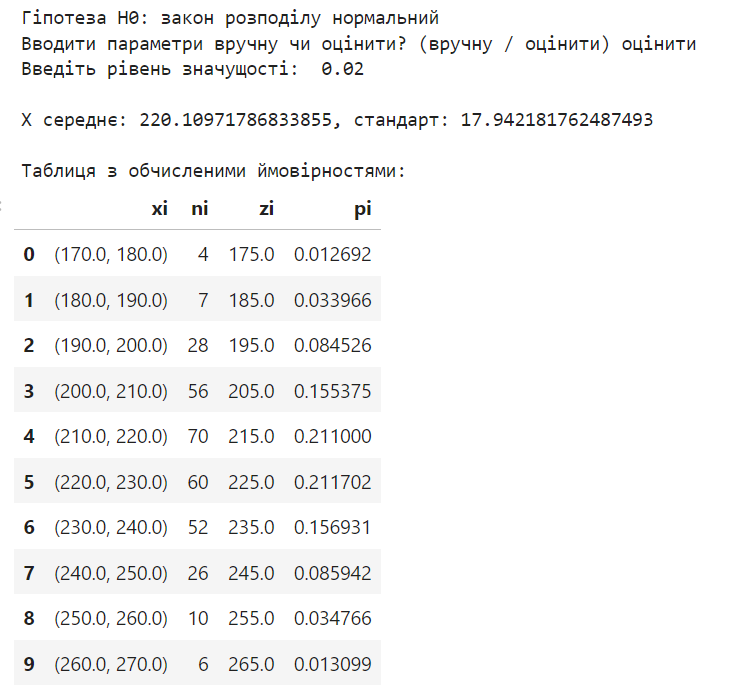
В кінці обчислюю значення хі квадрат емпіричного та критичного. Емпіричне рахую за формулою, а критичне — за допомогою функції chi2.ppf(). Якщо емпіричне значення менше за критичне, то гіпотеза приймається, якщо ні — відхиляється.

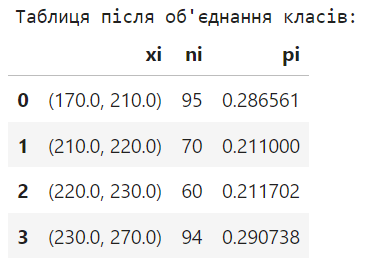
**Аналіз отриманих результатів:**

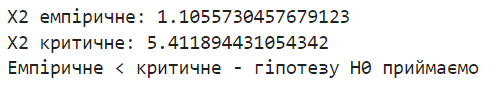
**Перша задача:**



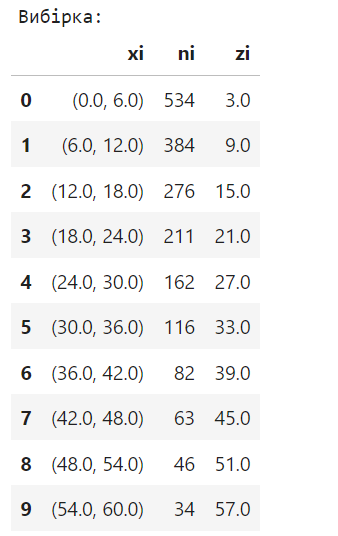


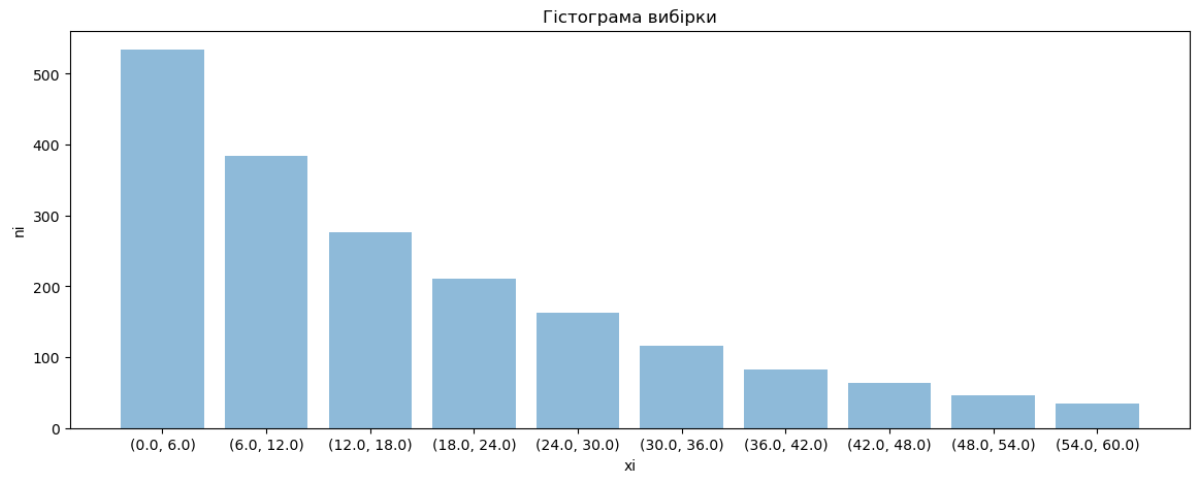


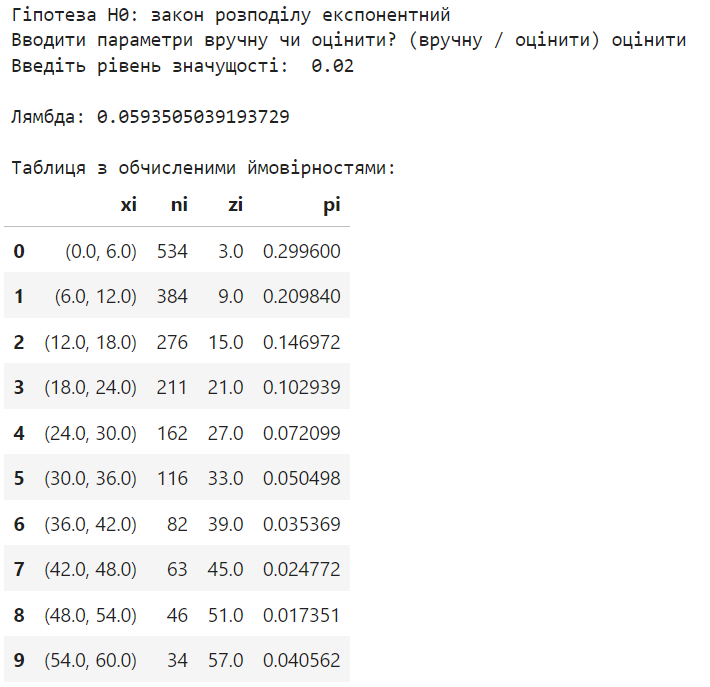


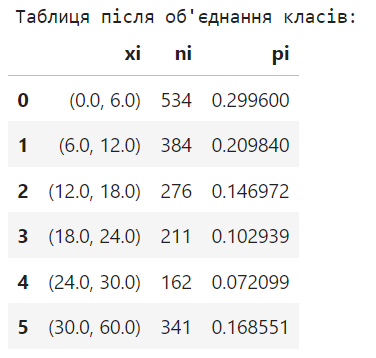


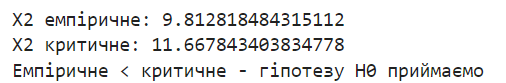
**Друга задача:**

****

****

****

****

****

**Висновок:**

Виконуючи дане індивідуальне завдання, я детальніше ознайомилась з темою перевірка гіпотез про закон розподілу випадкової величини методом Пірсона. Навчилась обчислювати хі квадрат емпіричне та критичне та пригадала, як обчислювати ймовірності для різних законів розподілу.