Міністерство Освіти та Науки України
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені ІВАНА ФРАНКА
Факультет прикладної математики та інформатики
Кафедра дискретного аналізу та інтелектуальних систем

ЗВІТ ПРО ВИКОНАННЯ
ІНДИВІДУАЛЬНОГО ЗАВДАННЯ №2

З дисципліни
Теорія Імовірності та Математична Статистика

Виконав студент
Групи ПМІ-24
Кутко Остап

2025

Постановка задачі: Відповідно до номеру в журналі, обрати файл з вхідними даними (вибіркою). Зчитати дані з текстового файлу, опрацювати статистичний матеріал, представивши дані таблично та графічно, на основі графічного представлення сформулювати гіпотезу про закон розподілу досліджуваної ознаки генеральної сукупності; передбачити можливість користувачу задати параметри розподілу вручну або оцінити на основі даних вибірки; для заданого користувачем рівня значущості перевірити сформульовану гіпотезу за критерієм χ^2.

Короткі теоретичні відомості

Ряд незалежних спостережень над випадковою змінною називають **вибіркою з генеральної сукупності**.

Всяке твердження про генеральну сукупність на основі вибірки називаємо **гіпотезою**.

Виділяють два основні типи гіпотез: **про закон розподілу випадкової величини** і **про значення параметрів розподілу випадкової величини**.

Міркування, на основі яких приходимо до висновків про гіпотезу називаємо **статистичним доведенням**.

Критерій гіпотези – випадкова величина К, за допомогою якої проводять перевірку гіпотези.

При статистичному доведенні можливо допустити одну з двох похибок: або відкинути істинну гіпотезу (похибка 1-го типу), або прийняти хибну гіпотезу (похибка 2-го типу). Умовно допустити похибку 1-го, тобто ймовірність відкинути істинну похибку називаємо **рівнем значущості даного критерія** і позначаємо через α.

Інформацію про випадкову величину, яка міститься у гіпотезі, називають **гіпотетичною або теоретичною**, а інформацію про неї, яку отримують на підставі вибірки, називають **статистичною або емпіричною**.

Основна (нульова) гіпотеза Н0 – гіпотеза, сформульована при статистичному доведенні.

Альтернативна (конкуруюча) гіпотеза Н1 – гіпотеза, яка повністю або частково логічно заперечує нульову гіпотезу.

Емпіричне значення критерію гіпотези – значення випадкової величини К, яке обчислюється на підставі певної вибірки.

Критична область – сукупність значень критерію К, для яких нульову гіпотезу Н0 відхиляють.

Область прийняття гіпотези - сукупність значень критерію К, для яких нульову гіпотезу Н0 приймають.

Критична точка $\mathbf{k_{кр}}$ – точка, яка відділяє критичну область і область прийняття гіпотези.

Розділяють 3 види критичних областей:

- Правостороння критична область, $К > k_{кр}$;
- Лівостороння критична область, $К < k_{кр}$;
- Двостороння критична область, $|К| > k_{кр}$;

Потужність критерію – ймовірність того, що нульова гіпотеза буде відхилена, якщо правильна конкуруюча гіпотеза, тобто ймовірність не допустити помилку другого роду за вибраного критерію.

Для перевірки гіпотез про закон розподілу часто застосовують критерій χ^2 („хі-квадрат”, по-іншому Пірсона), який ґрунтується на визначенні відхилення емпіричних характеристик від гіпотетичних характеристик.

$$\chi^2(r,n,F) = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^{r} \frac{n_i^2}{np_i} - n,$$ r – кількість інтервалів

Для $n \to \infty$ статистика χ^2 має розподіл, який задається густиною:

$$p_r(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{\Gamma((r-1)/2)2^{(r-1)/2}} x^{((r-1)/2)-1} e^{-x/2}, & x > 0 \end{cases}$$

Розподіл з густиною $p_r(x)$ називають розподілом χ^2 з r ступенями свободи.

1. Біномний закон розподілу

   Випадкова величина ξ може набувати цілих значень 0, 1,..., N з імовірностями р, = Р(ξ=i) = $C_N^i p^i (1-p)^{N-i}$, де р - параметр розподілу ($0 < p < 1$), який, якщо він невідомий, можна оцінити на підставі даних вибірки $p = \bar{x}/N$.

2. Закон розподілу Пуассона

Випадкова величина ξ може набувати цілих значень 0, 1,..., m з імовірностями p, = P(ξ=i) = $e^{-\lambda}\frac{\lambda^i}{i!}$, де λ > 0 - параметр розподілу, який, якщо він невідомий, можна оцінити на підставі даних вибірки p = x̄.

3. Рівномірний закон розподілу

Функція розподілу набуває вигляду

$$F(x)=\begin{cases} 0, & x<a \\ \dfrac{x-a}{b-a}, & a\le x<b \\ 1, & x\ge b. \end{cases}$$

Тут a та b – параметри розподілу, можуть бути оцінені на підставі даних вибірки a = x̄ - √3σ, b = x̄ + √3σ.

4. Показниковий закон розподілу

Функція розподілу

$$F(x)=\begin{cases} 0, & x<0 \\ 1-e^{-\lambda x}, & x\ge 0 \end{cases}$$

λ > 0 – параметр розподілу, його точкова оцінка на підставі вибірки λ=1/ x̄.

5. Нормальний закон розподілу

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}}\,dt\,.$$

Параметр розподілу: α = x̄.

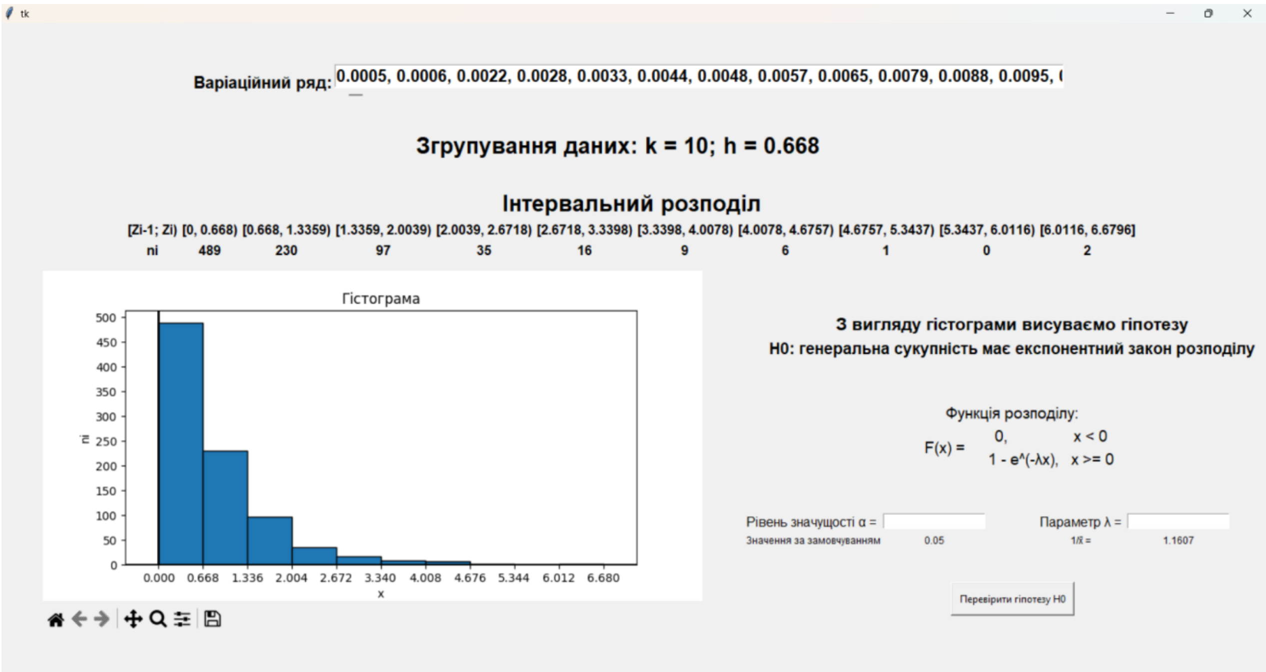
Програмна реалізація:

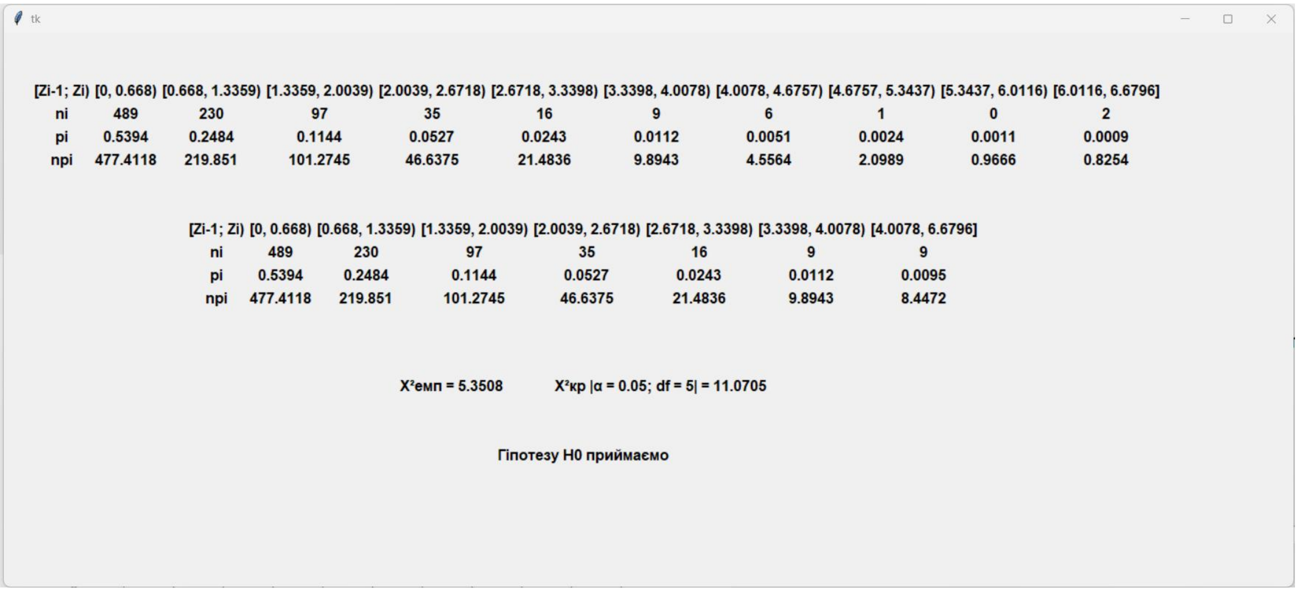
Програма, відповідно до варіанту, зчитує вибірку з файлу “sample6.txt”, утворює варіаційний ряд та ділить його на інтервали. Після цього за даним інтервальним розподілом будує гістограму частот. З вигляду цієї гістограми, я зробив припущення, що генеральна сукупність підпорядковується експонентному закону розподілу. У вікні програми є два поля для вводу рівня значущості та параметру λ. Якщо залишити ці поля пустими, то за замовчуванням рівень значущості буде 0.05, а λ – 1/ x̄, x̄ - середнє вибірки.

Після натискання на кнопку “Перевірити гіпотезу H0” відкривається нове вікно, в якому зображено повну інтервальну таблицю (вже з рядками pi (теоретичними імовірностями) та npi), таблицю з об’єднаними стовпцями (за умовою ni >= 5 та npi >=5), та значеннями хі квадрат емпіричне та критичне. Відповідно до цих значень і сформовано результат: гіпотезу H0 приймаємо або відхиляємо.

Програму реалізував на мові python, серед використаних структур списки та словники, використані бібліотеки: tkinter (інтерфейс), pandas (таблиці), matplotlib (гістограма), math (e та log10) та scipy (значення хі квадрат критичного).

Отримані результати:





tk

[Zi-1; Zi)	[0, 0.668)	[0.668, 1.3359)	[1.3359, 2.0039)	[2.0039, 2.6718)	[2.6718, 3.3398)	[3.3398, 4.0078)	[4.0078, 4.6757)	[4.6757, 5.3437)	[5.3437, 6.0116)	[6.0116, 6.6796]
ni	489	230	97	35	16	9	6	1	0	2
pi	0.5394	0.2484	0.1144	0.0527	0.0243	0.0112	0.0051	0.0024	0.0011	0.0009
npi	477.4118	219.851	101.2745	46.6375	21.4836	9.8943	4.5564	2.0989	0.9666	0.8254

[Zi-1; Zi)	[0, 0.668)	[0.668, 1.3359)	[1.3359, 2.0039)	[2.0039, 2.6718)	[2.6718, 3.3398)	[3.3398, 4.0078)	[4.0078, 6.6796]
ni	489	230	97	35	16	9	9
pi	0.5394	0.2484	0.1144	0.0527	0.0243	0.0112	0.0095
npi	477.4118	219.851	101.2745	46.6375	21.4836	9.8943	8.4472

X^2емп = 5.3508 X^2кр |α = 0.05; df = 5| = 11.0705

Гіпотезу H0 приймаємо

(тут взято значення α та λ за замовчуванням)

Висновок: Я написав програму, яка отримує вибірку з текстового файлу, групує її елементи в інтервальний розподіл та будує його гістограму частот. Пізніше перевіряє чи підпорядковується дана випадкова величина експонентному закону розподілу і виводить результат цієї перевірки.