

Міністерство освіти і науки України
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені ІВАНА ФРАНКА
Факультет прикладної математики та інформатики
Кафедра дискретного аналізу та інтелектуальних систем

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ № 2
з курсу “Математична статистика”

Виконала:
Студентка групи ПМІ-23
Богданович Софія

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 .

Теоретичні відомості

Сукупність всеможливих значень випадкової змінної називають **генеральною сукупністю або популяцією**.

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

Статистичною гіпотезою називають будь-яке твердження про властивості (ознаки) генеральної сукупності, що перевіряють на основі вибірки.

У математичній статистиці виділяють два основні типи статистичних гіпотез:

- гіпотези про закон розподілу ймовірностей випадкової величини (ознаки генеральної сукупності);
- гіпотези про значення параметрів розподілу випадкової величини (ознаки генеральної сукупності).

Сформульовану гіпотезу називають **основною** (нульовою) і позначають **H_0**

Альтернативною (конкуруючою) гіпотезою називають гіпотезу, яка повністю або частково логічно заперечує нульову гіпотезу, і позначають **H_1** .

На підставі статистичної перевірки гіпотези може бути прийняте одне з двох правильних **рішень**:

- гіпотезу приймають і вона правильна;
- гіпотезу відхиляють і вона неправильна.

На підставі статистичної перевірки статистичної гіпотези можуть бути допущені **помилки** (прийняті неправильні рішення) двох типів:

- гіпотезу відхиляють, але вона правильна (**помилка першого роду**);
- гіпотезу приймають, але вона неправильна (**помилка другого роду**).

Імовірність допустити помилку першого роду називають рівнем значущості і позначають через α .

Однією з найбільш важливих задач математичної статистики є **задача про визначення закону розподілу ймовірностей випадкової величини (ознаки генеральної сукупності) за даними вибірки.**

Для перевірки таких гіпотез часто застосовують критерій „хі-квадрат" Пірсона (критерій узгодження), який ґрунтується на визначенні відхилення емпіричних характеристик від гіпотетичних характеристик.

Схема перевірки гіпотези про вигляд закону розподілу ймовірностей дискретної випадкової величини за критерієм Пірсона:

- статистичні дані (результати вибірки) записують у вигляді дискретного статистичного розподілу:

x_i	x_1	x_2	...	x_m
n_i	n_1	n_2	...	n_m

- на підставі гіпотетичного закону розподілу знаходять теоретичні ймовірності p_i , того, що випадкова величина набуває значення x_i . Записуємо у таблицю:

x_i	x_1	x_2	...	x_m
n_i	n_1	n_2	...	n_m
p_i	p_1	p_2	...	p_m
np_i	np_1	np_2	...	np_m

- перевіряємо умови $n_i \geq 5$ і $np_i \geq 5$ для стовпців, якщо ці умови не виконуються, ці стовпці необхідно злити.
- Обчислюємо **емпіричне** значення критерію узгодження Пірсона

$$\chi^2_{\text{емп}} = \sum_{i=1}^{r+1} \frac{(n_i - np_i)^2}{np_i}$$

Випадкова величина K має відомий розподіл „хі-квадрат" з d.f.=r-s ступенями вільності, де r+1 - кількість інтервалів в інтервальному варіаційному ряді, s - кількість параметрів густини гіпотетичного розподілу:

- за цим рівнем значущості α і кількістю d.f.=r-s ступенів вільності знаходимо критичну точку $k_{\text{кр}}$ за таблицею критичних значень розподілу $\chi^2_{\text{емп}}$

- зіставляємо значення $\chi^2_{\text{емп}}$ і $k_{\text{кр}}$: якщо $\chi^2_{\text{емп}} \geq k_{\text{кр}}$, то гіпотезу H_0 про вигляд густини розподілу відхиляють; якщо ж $\chi^2_{\text{емп}} < k_{\text{кр}}$, то гіпотезу H_0 приймають.

Гіпотетичний закон розподілу може містити **невідомі параметри**, тоді за їх значення беруть їхні **точкові оцінки** на підставі цієї вибірки.

Для **біноміального розподілу** p обчислюємо як:

$$p = \frac{N}{\bar{x}}$$

Для **розподілу Пуассона** λ обчислюємо як:

$$\lambda = \bar{x}$$

Програмна реалізація та отримані результати

1. Зчитуємо дані, записуємо їх у вигляді дискретного статистичного розподілу.

Показати зчитану вибірку

3, 4, 5, 5, 2, 3, 3, 3, 2, 3, 3, 2, 4, 2, 3, 3, 4, 3, 4, 4, 1, 2, 7, 3, 3, 3, 4, ^
4, 3, 5, 3, 2, 3, 4, 1, 4, 2, 2, 5, 2, 4, 4, 1, 3, 4, 3, 3, 3, 3, 4, 3, 3, 1, 2, █
1, 7, 4, 1, 2, 3, 3, 4, 4, 5, 3, 1, 1, 3, 2, 3, 5, 7, 3, 3, 3, 2, 2, 3, 2, 1, 4, v
3, 5, 1, 5, 5, 3, 2, 5, 4, 2, 3, 3, 3, 4, 3, 1, 4, 3, 4, 4, 3, 3, 3, 3, 4, 3, 3, v
3, 5, 4, 4, 5, 2, 3, 1, 1, 1, 3, 2, 6, 1, 2, 1, 5, 3, 3, 3, 5, 2, 2, 3, 3, 5, 5, v

2, 2, 3, 5, 5, 3, 3, 3, 5, 3, 3, 1, 1, 1, 4, 2, 4, 6, 2, 1, 2, 1, 3, 2, 2, 3, 3, ^
3, 3, 0, 3, 5, 5, 3, 4, 2, 6, 3, 4, 5, 2, 4, 4, 5, 1, 2, 5, 5, 2, 3, 4, 2, 3, 1, █
4, 4, 3, 2, 5, 2, 5, 2, 3, 4, 1, 3, 5, 3, 1, 4, 3, 4, 3, 1, 1, 2, 6, 1, 3, 6, 2, v
4, 4, 6, 4, 3, 2, 4, 3, 2, 4, 1, 3, 5, 3, 3, 1, 3, 1, 5, 2, 4, 3, 3, 2, 4, 4, 2, v
3, 4, 4, 2, 3, 5, 4, 2, 5, 2, 4, 0, 4, 4, 6, 3, 5, 6, 3, 1, 2, 3, 2, 2, 5, 4, 3, v

2, 0, 2, 4, 3, 2, 3, 3, 4, 4, 4, 2, 2, 2, 1, 1, 3, 2, 5, 5, 3, 3, 5, 3, 5, 3, 6, ^
5, 2, 3, 3, 0, 4, 3, 2, 1, 2, 1, 3, 3, 2, 3, 2, 2, 2, 5, 6, 2, 3, 4, 5, 7, 3, 2, █
4, 5, 2, 3, 3, 4, 4, 6, 3, 2, 5, 2, 3, 4, 5, 3, 5, 3, 2, 3, 3, 3, 5, 4, 4, 2, 3, v
3, 3, 3, 3, 3, 4, 4, 6, 2, 1, 2, 2, 5, 3, 4, 4, 4, 4, 3, 2, 5, 2, 5, 2, 4, 2, 4, v
4, 2, 3, 3, 5, 2, 3, 3, 5, 5, 4, 3, 5, 1, 2, 2, 1, 2, 3, 3, 4, 1, 6, 3, 5, 2, 4, v

2, 3, 2, 4, 0, 3, 3, 1, 5, 2, 4, 2, 5, 5, 2, 3, 4, 5, 3, 3, 5, 4, 2, 4, 4, 6, 2, ^
3, 3, 4, 3, 3, 2, 4, 6, 5, 3, 2, 0, 3, 3, 3, 3, 4, 4, 5, 2, 4, 4, 3, 3, 2, 4, 5, █
3, 4, 2, 5, 4, 1, 4, 2, 4, 4, 1, 4, 4, 3, 6, 2, 4, 4, 3, 4, 2, 3, 4, 4, 4, 4, 5, v
5, 3, 3, 4, 4, 2, 3, 4, 4, 2, 5, 2, 3, 4, 4, 5, 4, 3, 2, 7, 3, 2, 1, 3, 2, 3, 4, v
2, 4, 2, 6, 4, 2, 3, 2, 4, 2, 5, 4, 2, 4, 2, 1, 3, 2, 3, 1, 5, 4, 2, 4, 3, 4, 3, v

1, 1, 0, 3, 1, 2, 3, 3, 2, 3, 2, 3, 4, 1, 3, 3, 7, 5, 1, 4, 4, 5, 3, 3, 4, 2, 3, ^
3, 2, 2, 3, 3, 0, 2, 5, 1, 3, 3, 6, 0, 4, 2, 4, 3, 2, 0, 5, 2, 5, 5, 5, 3, 1, 4, █
4, 4, 4, 2, 3, 4, 4, 3, 5, 3, 1, 2, 3, 4, 4, 4, 4, 2, 4, 3, 2, 3, 3, 3, 4, 3, 6, v
1, 2, 2, 3, 3, 5, 2, 4, 3, 3, 5, 2, 4, 3, 2, 4, 4, 4, 3, 1, 6, 5, 2, 4, 4, 4, 3, v
4, 5, 3, 5, 3, 2, 4, 4, 4, 3, 6, 2, 7, 2, 2, 4, 3, 3, 3, 1, 3, 3, 6, 2, 4, 2, 3, v

3, 3, 3, 5, 4, 3, 3, 2, 4, 6, 3, 3, 3, 4, 2, 3, 2, 0, 4, 3, 2, 3, 2, 1, 2, 4, 3,
 3, 2, 2, 3, 3, 4, 5, 1, 3, 1, 3, 4, 3, 3, 4, 2, 1, 2, 3, 2, 3, 5, 4, 2, 2, 3, 4,
 2, 4, 3, 3, 4, 2, 3, 1, 0, 3, 4, 2, 2, 4, 4, 3, 4, 1, 6, 2, 4, 4, 1, 3, 3, 3, 2,
 3, 0, 3, 3, 2, 3, 2, 3, 4, 3, 6, 5, 4, 3, 2, 3, 4, 3, 2, 3, 4, 5, 4, 4, 2, 5, 4,
 1, 4, 4, 4, 5, 2, 5, 3, 1, 4, 4, 4, 3, 6, 2, 7, 2, 2, 3, 5, 5, 2

Рисунок 1. Зчитана вибірка

Показати частотну таблицю

Значення	Частота
0	14
1	72
2	182
3	281
4	199
5	103
6	27
7	8

Рисунок 2. Частотна таблиця

2. Графічне представлення даних

Діаграма Частот

— □ ×

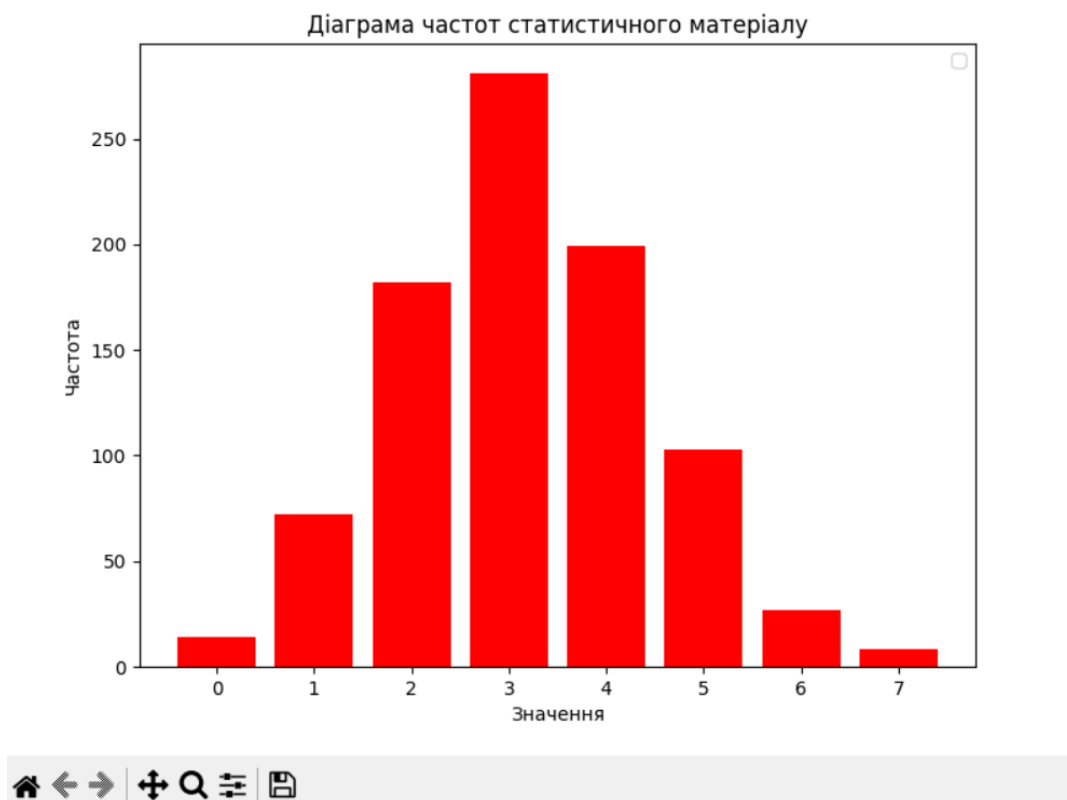


Рисунок 3. Діаграма частот

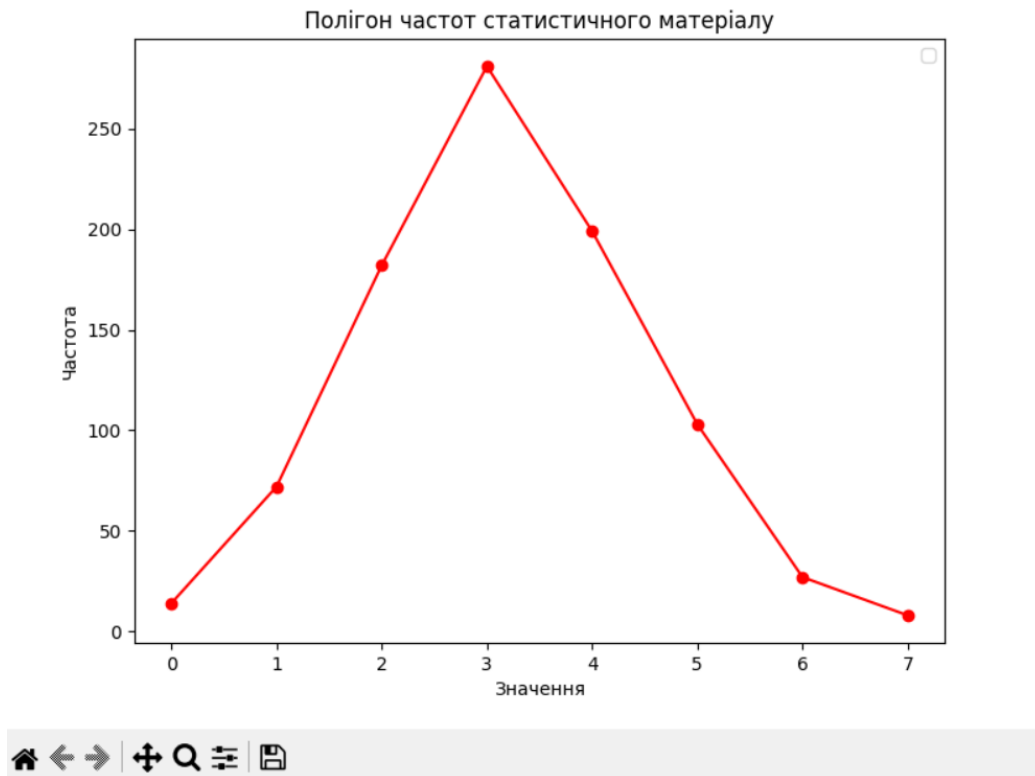


Рисунок 4. Полігон частот

3. Формулюємо гіпотезу про закон розподілу ознаки генеральної сукупності
Маємо можливість вибрати спосіб формулювання гіпотези:

Виберіть спосіб формулювання гіпотези

- Сформувати H_0 на основі графічного представлення
- H_0 : Розподіл Пуассона
- H_0 : Біноміальний розподіл

Рисунок 5. Формулювання гіпотези

4. Введення/розрахунок невідомих параметрів

На основі графічного представлення вибірки припускаємо, що ознака генеральної сукупності є біноміально розподілена

Введіть параметр p (або залиште порожнім для автоматичного розрахунку):

Submit

Рисунок 6. Невідомі параметри

5. Залишаємо введення порожнім для оцінки невідомого параметру p на основі вибірки. На підставі гіпотетичного закону розподілу знаходимо теоретичні ймовірності p_i того, що випадкова величина набуває значення x_i . Подаємо у вигляді таблиці.

Значення	Частота	p_i	$n \cdot p_i$
0.0	14.0	0.014698	13.022497
1.0	72.0	0.085122	75.418443
2.0	182.0	0.211276	187.190619
3.0	281.0	0.291329	258.117883
4.0	199.0	0.241029	213.551861
5.0	103.0	0.119648	106.008302
6.0	27.0	0.032997	29.235053
7.0	8.0	0.0039	3.455343

Рисунок 7. Теоретичні ймовірності

6. Перевіряємо умови $n_i \geq 5$ і $np_i \geq 5$ для записів у таблиці. Об'єднуємо записи.

Перевіряємо умови $n_i \geq 5$ та $n \cdot p_i \geq 5$. Виконуємо об'єднання записів у таблиці

Значення	Частота	p_i	$n \cdot p_i$
0.0	14.0	0.014698	13.022497
1.0	72.0	0.085122	75.418443
2.0	182.0	0.211276	187.190619
3.0	281.0	0.291329	258.117883
4.0	199.0	0.241029	213.551861
5.0	103.0	0.119648	106.008302
13.0	35.0	0.032997	32.690396

Рисунок 8. Таблиця після об'єднання

7. Обчислюємо емпіричне значення критерію узгодження Пірсона.

Обчислюємо χ^2 -квадрат емпіричне
 χ^2 -квадрат емпіричне = 3.640885804732945

Рисунок 9. Емпіричне значення

8. Обчислюємо χ^2 критичне, для чого спочатку вводимо рівень значущості α . Порівнюємо критичне та емпіричне значення, після чого робимо висновок про закон розподілу ознаки генеральної сукупності.

Обчислюємо χ^2 -квадрат критичне

Введіть рівень значущості α :

Submit

χ^2 -квадрат критичне = 15.09

Рішення: Приймаємо гіпотезу H_0

Рисунок 10. Критичне значення та рішення

В даному випадку гіпотеза про біноміальний закон розподілу прийматиметься для $\alpha < 0.9$.

У випадку припущення розподілу Пуассона, гіпотеза відхилятиметься для будь-яких α .

Висновок: Під час виконання цього індивідуального завдання було здійснено аналіз вибірки даних з текстового файлу, побудовано графічне представлення для формулювання гіпотези про закон розподілу досліджуваної ознаки генеральної сукупності. Оцінка параметрів розподілу була виконана як вручну, так і на основі вибірки, після чого була перевірена гіпотеза за допомогою критерію χ^2 для заданого рівня значущості. Результати показали відповідність вибірки певному статистичному розподілу.