ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені ІВАНА ФРАНКА
Факультет прикладної математики та інформатики

Звіт із

Теорії ймовірності та математичної статистики

Індивідуальне завдання №2

Виконав:

Готюк Максим

Група ПМі-23

Викладач:

Квасниця Галина Андріївна

Оцінка:

2023

Постановка завдання:

У поданих нижче задачах наведено результати досліджень вибірок з деяких генеральних сукупностей.

- Зчитати дані з текстового файлу, побудувати полігон або гістограму частот;
- на основі графічного представлення сформулювати гіпотезу про закон розподілу досліджуваної ознаки генеральної сукупності (у задачах 1 – 5) рекомендуємо перевіряти вибірки на нормальний закон, а в задачах 6 -10 — на інші, наприклад, рівномірний, показниковий, біномний, закон розподілу Пуассона);
- передбачити можливість користувачу задати параметри розподілу вручну або оцінити на основі даних вибірки;
- для заданого користувачем рівня значущості перевірити сформульовану гіпотезу за критерієм $\chi 2$

Варіант **№5**

ЗАДАЧА 1 (варіанти 1-6). Для вивчення технічних властивостей нової марки бетону досліджувалися окремі його зразки. Розподіл кількості ni зразків бетону і відповідного їм стискування X (тобто такого, що спричиняє руйнування зразка) наведено в таблиці

X, кг/см^2	170-180	180-190	190-200	200-210	210-220	220-230	230-240	240-250	250-260	260-270
n_i(варіант 1)	4	9	32	54	72	65	50	25	12	7
n_i(варіант 2)	4	8	38	52	70	66	52	24	10	6
n_i(варіант 3)	4	8	26	52	76	66	48	24	11	6
n_i(варіант 4)	4	7	28	56	70	60	52	26	10	6
n_i(варіант 5)	4	9	28	48	70	72	52	22	10	6
n_i(варіант 6)	4	8	30	52	64	66	56	24	9	6

ЗАДАЧА 6 (варіант 1 - 5). Для вироблення рекомендацій щодо покращення роботи портів реєструвався час очікування кораблів на розвантаження в декількох портах країни. Розподіл за один рік кількості кораблів ni залежно від часу Т, який вони очікували на розвантаження, відображено в таблиці.

Т, год	0-6	6-12	12-18	18-24	24-30	30-36	36-42	42-48	48-54	>54
n_i(варіант 1)	518	384	284	211	156	116	86	63	44	34
n_i(варіант 2)	518	376	284	217	156	112	86	65	44	33
n_i(варіант 3)	502	384	292	211	150	116	90	63	42	35
n_i(варіант 4)	534	384	276	211	162	116	82	63	46	34
n_i(варіант 5)	518	392	284	205	156	120	86	61	44	34

Короткі теоретичні відомості:

Зважаючи на знання із попереднього індивідуального завдання теоретичні відомості в цьому індивідуальному завданні будуть наступні :

Однією з найбільш важливих задач математичної статистики є задача про визначення закону розподілу ймовірностей випадкової величини (ознаки генеральної сукупності) за даними вибірки. Якщо закон розподілу випадкової величини невідомий, то формулюють нульову гіпотезу про вигляд густини розподілу. Наприклад: „*Випадкова величина має густину нормального розподілу ймовірностей*". Для перевірки таких гіпотез часто застосовують критерій ***Пірсона :***

$$K = \sum_{i=1}^{m} \frac{(n_i - np_i)^2}{np_i} = n \sum_{i=1}^{m} \frac{(w_i - p_i)^2}{p_i}$$

де n_i - емпіричні частоти, np_i - теоретичні частоти, w_i – емпіричні відносні частоти, p_i – теоретичні ймовірності, n – обсяг вибірки.

Дана випадкова величина K має закон розподілу χ^2, який описується густиною

$$R(x,n) = \begin{cases} 0, & x \le 0; \\ \dfrac{x^{n-2} e^{-x^2/2}}{A_n}, & x > 0, \end{cases}$$

і він не залежить від невідомого закону розподілу ймовірностей досліджуваної випадкової величини, а залежить лише від $k = m - s - 1$ ступенів вільності, де m – число інтервалів інтервального розподілу статистичних ймовірностей, s – число параметрів теоретичного розподілу.

Перевірка гіпотези про вигляд густини розподілу ймовірностей неперервної випадкової величини за критерієм Пірсона має наступний вигляд:

- статистичні дані вносятся у таблицю вигляду:

$(z_{i-1}, z_i]$	$(z_0, z_1]$	$(z_1, z_2]$	...	$(z_{m-1}, z_m]$
n_i	n_1	n_2	...	n_m

де n_i- число варіант вибірки що попадають в інтервал, $[z_{i-1}, z_i]$ - інтервал

- оскільки перевіряється гіпотеза про те, що генеральна сукупність задовольняє певному закону розподілу з густиною p(x), то для кожного інтервалу можна визначити теоретичні ймовірності p_i попадання значень випадкової величини в цей інтервал; $p_i = P(z_{i-1} < Z \le z_i) = F(z_i) - F(z_{i-1})$,
- одержані результати обчислень записуємо у таблицю:

$(z_{i-1}, z_i]$	$(-\infty, z_1]$	$(z_1, z_2]$	...	$(z_{m-1}, +\infty)$
n_i	n_1	n_2	...	n_m
p_i	p_1	p_2	...	p_m

- обчислюється емпіричне значення критерію Пірсона $\chi^2_{емп} = \sum_{i=1}^{m} \frac{(n_i - np_i)^2}{np_i}$
- за даним рівнем значущості α і кількістю k = m − s −1 ступенів вільності знаходимо критичну точку $k_{кр} = \chi^2_{кр}(\alpha, k)$ за таблицею критичних значень розподілу χ^2 . (Додаток 5)
- Порівнюємо $\chi^2_{емп}$ і $\chi_{кр}$:
  - якщо $\chi^2_{емп} \ge \chi_{кр}$, то нашу гіпотезу H_0(про вигляд густини розподілу) відхиляють
  - якщо $\chi^2_{емп} < \chi_{кр}$, то гіпотезу приймають

Перевірка гіпотези про вигляд закону розподілу ймовірностей дискретної випадкової величини має практично ідентичний план, але з незначними відмінностями:

- статистичні дані вносятся у таблицю вигляду:

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

- на підставі гіпотетичного закону розподілу знаходимо теоретичні ймовірності p_iтого, що випадкова величина приймає значення x_i. Але тут є невелике зауваження: критерій Пірсона застосовують для великих обсягів вибірок, *n* ≥100.

  Також мають виконуватись умови $n_i \ge 5$, $np_i \ge 10$ в окремих групах. Якщо ці умови не виконуються, сусідні групи слід об’єднати.

Гіпотетичний закон розподілу може містити невідомі параметри, тоді за їх значення беруть їх точкові оцінки на основі даної вибірки.

1) Біномний закон розподілу.

Випадкова величина ξможе набувати цілих значень 0, 1, ..., N з ймовірностями $p_i = P(\xi = i) = C_N^i p^i (1-p)^{N-i}$, де p - параметр розподілу (0<p<1), який, якщо він відомий, можна оцінити на основі даних вибірки $p = \frac{\underline{x}}{N}$

2) Закон розподілу Пуассона
   Випадкова величина ξ може набувати цілих значень 0, 1, ..., m, ... з ймовірностями $p_i = P(\xi = i) = e^{-\lambda} \frac{\lambda^i}{i!}$, де λ>0 - параметр розподілу, який, якщо він відомий, можна оцінити на основі даних вибірки $\lambda = \underline{x}$
3) Рівномірний закон розподілу. Функція розподілу має вигляд:

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

   Тут a,b - параметри розподілу, можуть бути оцінені на основі даних вибірки $a = \underline{x} - \sqrt{3}s$, $b = \underline{x} + \sqrt{3}s$
4) Показниковий закон розподілу. Функція розподілу має вигляд:

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

$\lambda > 0$ - параметр розподілу, його точкова оцінка на основі вибірки $\lambda = \dfrac{1}{\overline{x}}$.

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

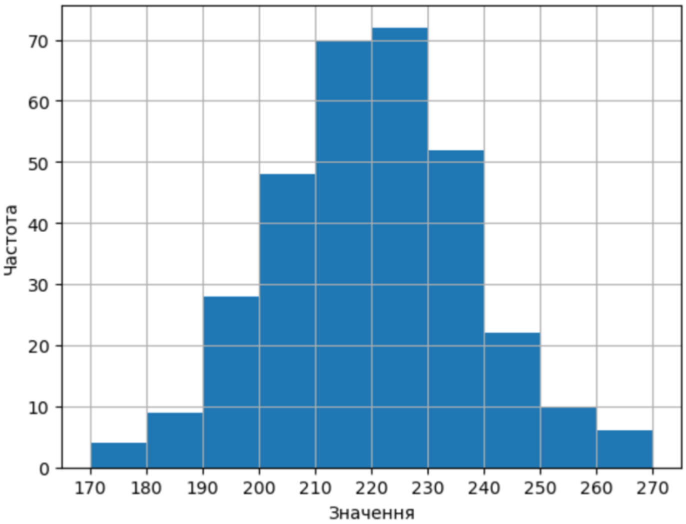
5) Нормальний закон розподілу має вигляд :
   параметри розподілу оцінюються на основі даних вибірки $a = \underline{x}$, $\sigma = s$

Програмна реалізація

Програма реалізована на допомогою мови python в середовищі Jupyter Notebook. Спочатку я ввів дані з файлу, тоді побудував за допомогою них гістограму, щоб зрозуміти, на функцію якого розподілу вона схожа. Тоді, згідно з результатом, формулював гіпотезу, рахував умовні ймовірності, якщо це було необхідно для можливості перевірки критерієм Пірсона, то об’єднував деякі дані вибірки. Тоді рахував $\chi 2$ критичне та емпіричне і на основі порівняння робив висновок щодо того, приймати гіпотезу, чи ні.

Отримані результати

Завдання 1



На основі графіку можна зробити висновок, що перевіряти потрібно на нормальний закон розподілу.

```
Хочете ввести параметри розподілу вручну або оцінити їх на основі вибіркових даних? (в/о) о
Введіть рівень значущості (альфа): 0.005
Інтервали: [170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270]
Значення: [4, 9, 28, 48, 70, 72, 52, 22, 10, 6]
Стандартне відхилення (сигма): 17.75
Рівень значущості (альфа): 0.005
```

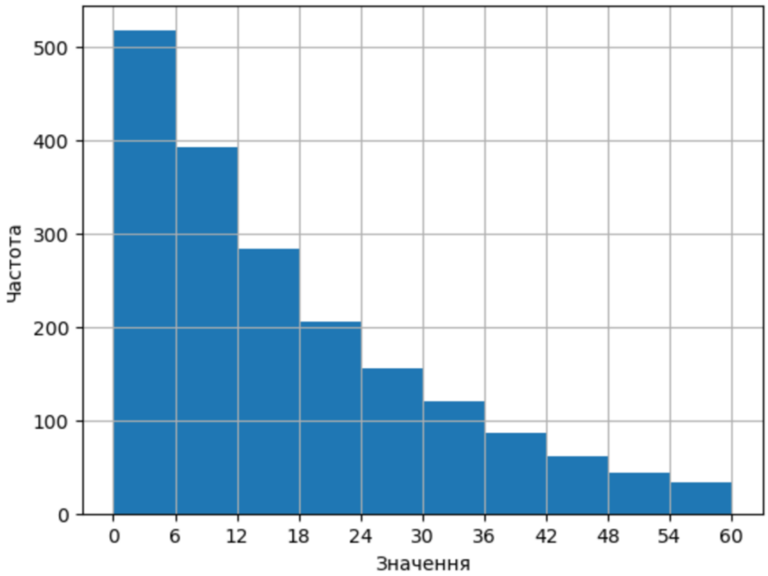
```
Теоретичні ймовірності: [0.011866208334873729, 0.03288298094112024, 0.08350850081496819, 0.1556466136902317 8, 0.212
9449204785182, 0.21386964487383, 0.15768332273832403, 0.08533800061895674, 0.033896200305875535, 0.0123636072033015
41]
```

Об’єднання інтервалів привело до такого розподілу:

```
[170, 190, 200, 210, 220, 230, 240, 260, 270]
[13, 28, 48, 70, 72, 52, 22, 16]
[0.04474918927599397, 0.08350850081496819, 0.15564661369023178, 0.2129449204785182, 0.21386964487383, 0.15768332273
832403, 0.08533800061895674, 0.046259807509177076]
```

```
X2 критичне: 9.99573227355399
X2 емпіричне: 1.6516000009137015
Гіпотеза приймається
```

Завдання 2



На основі графіку можна зробити висновок, що перевіряти потрібно на експонентний закон розподілу.

```
Хочете ввести параметри розподілу вручну або оцінити їх на основі вибіркових даних? (в/о) о
Введіть рівень значущості (альфа): 0.005
Інтервали: [0, 6, 12, 18, 24, 30, 36, 42, 48, 54, 60]
Значення: [518, 392, 284, 205, 156, 120, 86, 61, 44, 34]
Лямбда: 0.059
Рівень значущості (альфа): 0.005
```

```
Теоретичні ймовірності: [0.298924845244395, 0.2095687821400095, 0.1469234663707507, 0.10300439192310418, 0.07221382
000799714, 0.050627315037599985, 0.03549355272484622, 0.024883647969397793, 0.01744530735102956, 0.040914871230869
9]
```

Розподіл не потребував злиття.

```
X2 критичне: 12.298317366548037
X2 емпіричне: 50.73649296969306
Гіпотеза не приймається
```

Висновки

За час виконання лабораторної роботи я навчився працювати з критерієм Пірсона, а саме розуміти, як визначити тип розподілу, на який варто перевіряти, за допомогою графіку, обчислювати необхідні значення та робити висновок щодо правдивості гіпотези.