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"

Виконав:
студент групи ПМі-23с
Гуменюк Станіслав

Оцінка

Перевірила:
доц. Квасниця Г.А.

Львів 2024

Постановка задачі:

на основі графічного представлення сформулювати гіпотезу про закон розподілу досліджуваної ознаки генеральної сукупності

для заданого користувачем рівня значущості перевірити сформульовану гіпотезу за критерієм χ^2.

ЗАДАЧА 2 Для контролю за готовою продукцією відібрано деталі, що виготовляються на однотипних верстатах-автоматах. Розподіл кількості n_i відібраних деталей залежно від їх контрольованого розміру X наведено в таблиці

X, мм	23,2-23,4	23,4-23,6	23,6-23,8	23,8-24,0	24,0-24,2	24,2-24,4	24,4-24,6	24,6-24,8	24,8-25,0	25,0-25,2
n_i(варіант7)	1	3	23	79	141	146	75	25	4	1

ЗАДАЧА 7 (варіанти 6-10). Для покращення обслуговування сільськогосподарських машин були зібрані дані про вихід з ладу техніки у господарствах району за період весняно-польових робіт. Розподіл кількості n_i перевірених одиниць залежно від кількості поломок X наведено у таблиці

X, к-ть поломок	0	1	2	3	4	5
n_i(варіант 7)	2483	1767	608	146	25	3

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 ni - емпіричні частоти, npi - теоретичні частоти, wi – емпіричні відносні частоти, pi – теоретичні ймовірності, n – обсяг вибірки.

Дана випадкова величина К має закон розподілу χ2, який описується густиною

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.

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

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

- статистичні дані вносятся у таблицю вигляду:де ni- число варіант вибірки що попадають в інтервал, [zi-1 , zi] - інтервал

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

- оскільки перевіряється гіпотеза про те, що генеральна сукупність задовольняє певному закону розподілу з густиною p(x), то для кожного інтервалу можна визначити теоретичні ймовірності pi 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 χ2 . (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 рітого, що випадкова величина приймає значення хі. Але тут є невелике зауваження: критерій Пірсона застосовують для великих обсягів вибірок, n ≥100. Також мають виконуватись умови ni5 , npi10 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 $p = \frac{\bar{x}}{N}$

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 $\lambda = \bar{x}$

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 $a = \bar{x} - \sqrt{3}s$, $b = \bar{x} + \sqrt{3}s$

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 $\lambda = \dfrac{1}{\bar{x}}$.

5) Нормальний закон розподілу має вигляд : $$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt,$$

параметри розподілу оцінюються на основі даних вибірки $a = \bar{x}$, $\sigma = s$

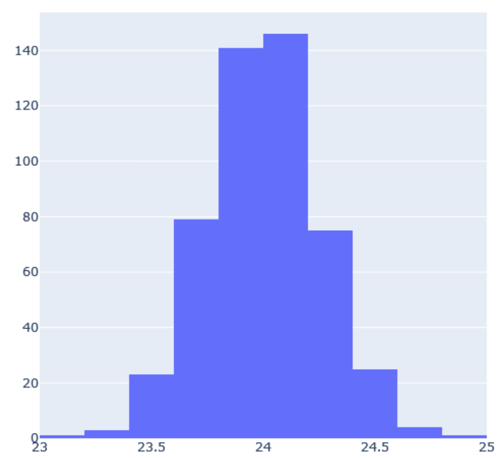
Програмна реалізація

Програму написано на платформі .NETз використанням мови C#. Було використано для графіків бібліотеку Plotly.NET та бібліотеку MathNet.Numerics для обрахунку деяких функцій

Отримані результати

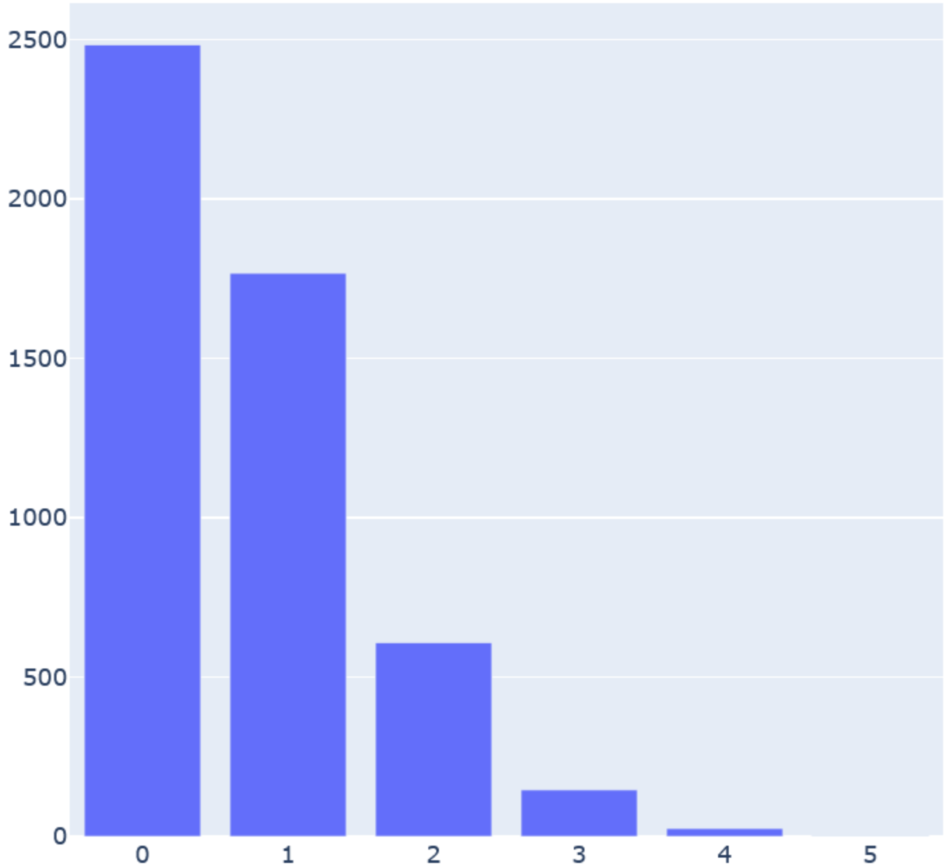
Task 1

```
H0: data is distributed normally

Data:
---------------------------------------------------------------------
| 23,2 | 23,4 | 23,6 | 23,8 | 24   | 24,2 | 24,4 | 24,6 | 24,8 | 25   |
---------------------------------------------------------------------
| 1    | 3    | 23   | 79   | 141  | 146  | 75   | 25   | 4    | 1    |
---------------------------------------------------------------------

| a = 24,002 | sigma = 0,26 |

Pi:
-------------------------------------------------------------------------------
| 0,001 | 0,009 | 0,051 | 0,158 | 0,278 | 0,28  | 0,16  | 0,052 | 0,01  | 0,001 |
-------------------------------------------------------------------------------
H0: data is distributed normally

Data:
---------------------------------------------------------------------
| 23,2 | 23,4 | 23,6 | 23,8 | 24   | 24,2 | 24,4 | 24,6 | 24,8 | 25   |
---------------------------------------------------------------------
| 1    | 3    | 23   | 79   | 141  | 146  | 75   | 25   | 4    | 1    |
---------------------------------------------------------------------

| a = 24,002 | sigma = 0,26 |

Pi:
-------------------------------------------------------------------------------
| 0,001 | 0,009 | 0,051 | 0,158 | 0,278 | 0,28  | 0,16  | 0,052 | 0,01  | 0,001 |
-------------------------------------------------------------------------------

Npi:
---------------------------------------------------------------------------------------------------
| 0,516   | 4,664   | 25,369  | 78,502  | 138,417 | 139,2   | 79,841  | 26,095  | 4,852   | 0,544   |
---------------------------------------------------------------------------------------------------

Merged data:
------------------------------------------
| 23,6 | 23,8 | 24   | 24,2 | 24,4 | 25   |
------------------------------------------
| 27   | 79   | 141  | 146  | 75   | 30   |
------------------------------------------

Merged Npi:
------------------------------------------------------------
| 30,549  | 78,502  | 138,417 | 139,2   | 79,841  | 31,491  |
------------------------------------------------------------

Enter alpha: 0.05

XKr = 7,814727903251172, Xemp = 1,16000894683962
H0 is accepted

Press any key to continue...
```



Task 2

```
H0: data is distributed by Poisson distribution

Data:
-------------------------------------------
| 0    | 1    | 2    | 3    | 4    | 5    |
-------------------------------------------
| 2483 | 1767 | 608  | 146  | 25   | 3    |
-------------------------------------------

| lambda = 0,7027027027027027 |

Pi:
-------------------------------------------------
| 0,495 | 0,348 | 0,122 | 0,029 | 0,005 | 0,001 |
-------------------------------------------------

Npi:
-------------------------------------------------------------------
| 2492,073 | 1751,186 | 615,282  | 144,12   | 25,318   | 3,558    |
-------------------------------------------------------------------

Merged data:
------------------------------------
| 0    | 1    | 2    | 3    | 5    |
------------------------------------
| 2483 | 1767 | 608  | 146  | 28   |
------------------------------------

Merged Npi:
--------------------------------------------------------
| 2492,073 | 1751,186 | 615,282  | 144,12   | 28,877   |
--------------------------------------------------------

Enter alpha: 0.05

XKr = 7,814727903251172, Xemp = 0,31314612479264164
H0 is accepted

Press any key to continue...
```



Під час написання цієї лабараторної роботи, я навчився користуватись деякими бібліотеками для роботи з статистикою, навчився писати програмні рішення для таких обрахунків. Застосував знання про розподіли та критерій Пірсона