МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИ
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені ІВАНА ФРАНКА

Кафедра дискретного аналізу
та інтелектуальних систем

Індивідуальне завдання №3

з курсу ''Теорія ймовірності та математична статистика''

Виконав:
студент групи ПМі-21
Урдейчук Ростислав Ігорович

Оцінка

Перевірила:
доц. Квасниця Г.А.

Львів 2024

Постановка задачі:

1. 1. За даними кореляційної таблиці обчислити умовні середні yxi (i = 1,…,k).
2. 2. Побудувати поле кореляції, тобто нанести точки Mi(xi;yxi), i = 1,…,k , накоординатну площину. На основі цього зробити припущення про вигляд функціїрегресії (парабола, гіпербола і т.д.)
3. 3. В залежності від вигляду функції регресії ((1), (3), (6) чи (9)) скластивідповідну систему рівнянь ((2), (5), (8) чи (11)). Розв'язати її і знайти невідоміпараметри вибраної функції регресії.
4. 4. Записати рівняння кривої регресії Y на X : yx = f(x) (з конкретною знайденоюв пункті 3 функцією регресії f(x)) та побудувати її графік.
5. 5. Обчислити дисперсію (12) величини Y відносно кривої регресії Y на X .
6. 6. Визначити суму квадратів відхилень δ2 умовних середніх від значеньфункції регресії за формулою (13).

Варіант 17

17.

$Y\|X$	3	5	7	9	13	15	17
1	23						
1,5	2	19					
2		3	32	2			
3			8	23	5		
3,5				2	17	4	
4						20	3

Короткі теоретичні відомості:

Нехай вивчається генеральна сукупність, що характеризується системою кількісних ознак (X,Y). Для аналізу залежності між випадковими величинами X і Y зроблена вибірка, причому складова X набула значень $x_1,x_2,...,x_k$, складова Y – $y_1,y_2,...,y_l$, а подія $\{X = x_i,\ Y = y_j\}$ мала частоту появи n_{ij} $(i = 1,..,k\ ;\ j = 1,..,l)$. Результати цих спостережень записують у вигляді кореляційної таблиці:

$Y\|X$	x_1	x_2	…	x_i	…	x_k	m_j
y_1	n_{11}	n_{21}	…	n_{i1}	…	n_{k1}	m_1
y_2	n_{12}	n_{22}	…	n_{i2}	…	n_{k2}	m_2
…	…	…	…	…	…	…	…
y_j	n_{1j}	n_{2j}	…	n_{ij}	…	n_{kj}	m_j
…	…	…	…	…	…	…	…
y_l	n_{1l}	n_{2l}	…	n_{il}	…	n_{kl}	m_l
n_i	n_1	n_2	…	n_i	…	n_k	n

За даними кореляційної таблиці обчислюють умовні середні $\overline{y_{xi}}$ $(i = 1,...,k\ )$:

$$\overline{y_{x_1}} = \frac{y_1 n_{11} + y_2 n_{12} + ... + y_l n_{1l}}{n_1}, \quad \overline{y_{x_2}} = \frac{y_1 n_{21} + y_2 n_{22} + ... + y_l n_{2l}}{n_2}, \ ... ,$$

$$\overline{y_{x_i}} = \frac{y_1 n_{i1} + y_2 n_{i2} + ... + y_l n_{il}}{n_i}, \ ... , \overline{y_{x_k}} = \frac{y_1 n_{k1} + y_2 n_{k2} + ... + y_l n_{kl}}{n_k}.$$

Складають таблицю умовних середніх $\overline{y_x}$:

x	x_1	x_2	...	x_i	...	x_k
$\overline{y_x}$	$\overline{y_{x_1}}$	$\overline{y_{x_2}}$	...	$\overline{y_{x_i}}$	...	$\overline{y_{x_k}}$

Аналогічно можна скласти таблицю умовних середніх $\overline{x_{yj}}$:

y	y_1	y_2	...	y_j	...	y_l
$\overline{x_y}$	$\overline{x_{y_1}}$	$\overline{x_{y_2}}$	...	$\overline{x_{y_j}}$	...	$\overline{x_{y_l}}$

Для визначення вигляду функції регресії будують точки ($x;\overline{y}_x$) (або ($y;\overline{x}_y$)) і за їх розміщенням роблять висновок про приблизний вигляд функції регресії.

Якщо графік регресії $\overline{y}_x = f(x)$ або $\overline{x}_y = \phi(y)$ зображається кривою лінією, то кореляцію називають *нелінійною* (криволінійною).

Наприклад, функції регресії Y на X можуть мати вигляд:

$\overline{y}_x = ax^2 + bx + c$ (параболічна кореляція другого порядку);

$\overline{y}_x = ax^3 + bx^2 + cx + d$ (параболічна кореляція третього порядку);

$\overline{y}_x = \frac{a}{x} + b$ (гіперболічна кореляція);

$\overline{y}_x = ba^x$ (показникова кореляція).

Теорія криволінійної кореляції розв'язує ті самі задачі, що і теорія лінійної кореляції, а саме:

1) за даними кореляційної таблиці встановлюють форму кореляційного зв'язку, тобто визначають вигляд функції $\overline{y}_x = f(x)$ або $\overline{x}_y = \phi(y)$;

2) оцінюють щільність кореляційного зв'язку, тобто дають оцінку ступеню розсіювання значень випадкової величини Y навколо побудованої кривої регресії $\overline{y_x}$ (або значень випадкової величини X навколо $\overline{x}_y$).

4. **Коренева кореляція.** Припустимо, що аналіз залежності між змінними X і Y, вираженої кореляційною таблицею, приводить до вибору форми кореляційної залежності Y на X у вигляді рівняння

$$\overline{y}_x = a\sqrt{x} + b, \tag{9}$$

а у випадку регресії X на Y – рівняння

$$\overline{x}_y = c\sqrt{y} + d. \tag{10}$$

У цьому випадку невідомі параметри a і b будемо шукати з системи рівнянь

$$\begin{cases} a\sum_{i=1}^{k} n_i\sqrt{x_i} + bn = \sum_{i=1}^{k} \overline{y}_{x_i} n_i; \\ a\sum_{i=1}^{k} n_i x_i + b\sum_{i=1}^{k} n_i\sqrt{x_i} = \sum_{i=1}^{k} n_i \overline{y}_{x_i}\sqrt{x_i}. \end{cases} \tag{11}$$

Для відшукання параметрів c і d рівняння (10) складаємо аналогічну до (11) систему рівнянь, де змінні x і y міняються місцями.

5. **Оцінка щільності кореляційного зв'язку.** За побудованою кривою регресії $\overline{y}_x = f(x)$ (або $\overline{x}_y = \phi(y)$) можна оцінити відхилення значень випадкової величини Y від кривої регресії $\overline{y_x}$ (або значень випадкової величини X від кривої регресії x_y). Зокрема, обчислюють дисперсію величини Y відносно кривої регресії Y на X :

$$\sigma^2(y, \overline{y}_x) = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{l} (y_j - f(x_i))^2 n_{ij} = \frac{\Delta}{n},$$

$$\Delta = n_{11}[y_1 - f(x_1)]^2 + n_{21}[y_1 - f(x_2)]^2 + ... + n_{k1}[y_1 - f(x_k)]^2 + \tag{12}$$

$$+n_{12}[y_2 - f(x_1)]^2 + n_{22}[y_2 - f(x_2)]^2 + ... + n_{k2}[y_2 - f(x_k)]^2 + ... +$$

$$+n_{1l}[y_l - f(x_1)]^2 + n_{2l}[y_l - f(x_2)]^2 + ... + n_{kl}[y_l - f(x_k)]^2.$$

За міру розсіяння значень випадкової величини Y від кривої регресії $\overline{y_x}$ можна також взяти, наприклад, суму квадратів відхилень δ^2 умовних середніх

$$\overline{y_{x_i}} = \frac{1}{n_i} \sum_{j=1} y_j n_{ij}$$

обчислених за даними кореляційної таблиці, від значень $f(x_i)$ функції регресії:

$$\delta^2 = \sum_{i=1}^{k} \delta_i^2 n_i = \sum_{i=1}^{k} |\overline{y_{x_i}} - f(x_i)|^2 n_i . \tag{13}$$

Програмна реалізація:

1. **Обчислення умовних середніх**:
    - Ми обчислюємо умовні середні (y_{xi}) для кожного значення (x_i), використовуючи дані кореляційної таблиці.

2. **Побудова поля кореляції**:
    - Побудова графіка, де точки ((x_i, y_{xi})) наносяться на координатну площину.

3. **Припущення про вигляд функції регресії**:
    - Припускаємо, що функція регресії має вигляд $a\sqrt{x} + b$

4. **Розв'язання системи рівнянь та знаходження параметрів**:
    - Використовуємо метод найменших квадратів для знаходження параметрів (a), \(b) функції регресії.

5. **Побудова графіка кривої регресії**:
    - Побудова графіка кривої регресії на основі знайдених параметрів.

6. **Обчислення дисперсії**:
    - Обчислення дисперсії величини (Y) відносно кривої регресії (Y) на (X).

7. **Визначення суми квадратів відхилень**:
    - Визначення суми квадратів відхилень умовних середніх від значень функції регресії.

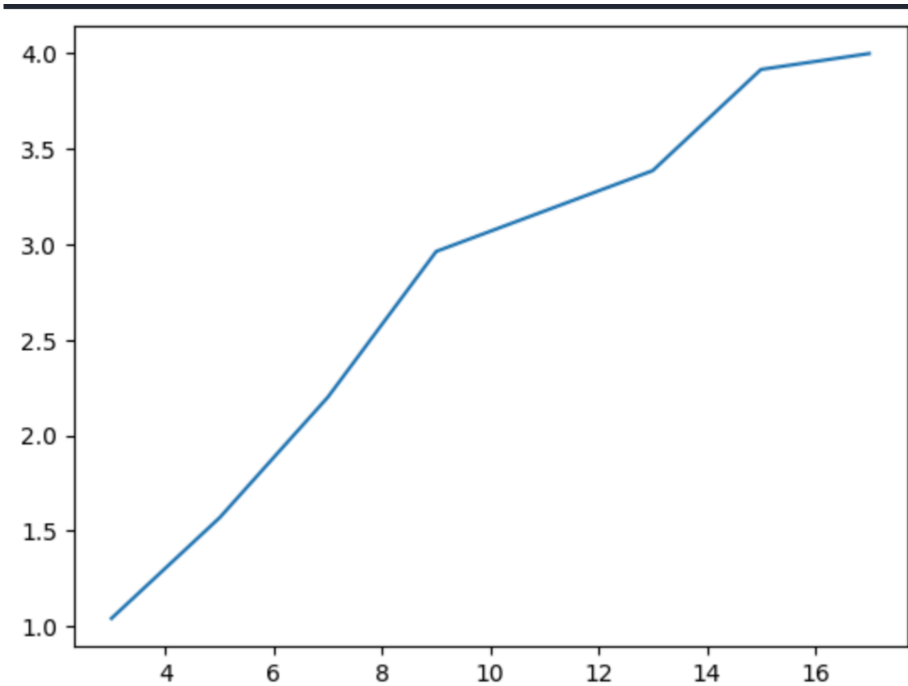
Аналіз отриманих результатів:

	3	5	7	9	13	15	17
1.0	23	0	0	0	0	0	0
1.5	2	19	0	0	0	0	0
2.0	0	3	32	2	0	0	0
3.0	0	0	8	23	5	0	0
3.5	0	0	0	2	17	4	0
4.0	0	0	0	0	0	20	3

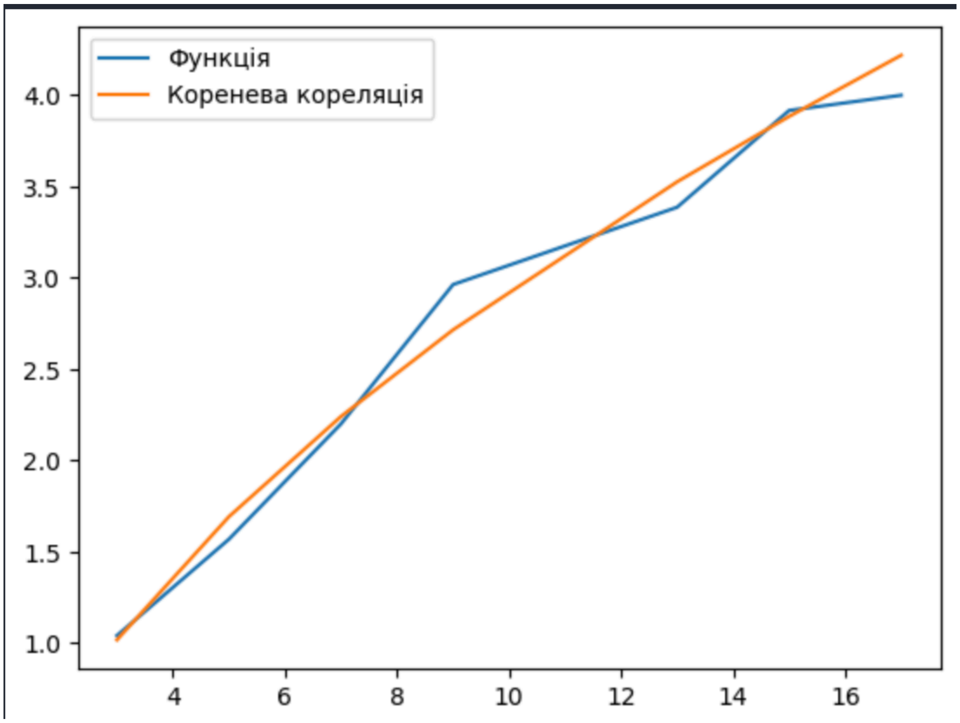
Умовні середні yxi (i = 1,…,k).

x	y
3.0	1.040000
5.0	1.568182
7.0	2.200000
9.0	2.962963
13.0	3.386364
15.0	3.916667
17.0	4.000000

Поле кореляції



```
(1.3392724090629646, -1.3029408473652275)
```



Рівняння кривої регресії

```
1.3392724090629646√x + -1.3029408473652275
```

```
Дисперсія:  0.08861522147592434
```

```
Сума квадратів відхилень δ2:  2.674348440643004
```

Висновок

Виконуючи дане індивідуальне завдання, проведений аналіз показав, що коренева функція регресії добре описує залежність між змінними (X) та (Y) у нашій вибірці даних. Крива регресії, побудована на основі отриманих параметрів, виявилася адекватною моделлю, що підтверджується низькими значеннями дисперсії та суми квадратів відхилень умовних середніх від значень функції регресії. Таким чином, модель регресії може бути використана для прогнозування значень (Y) для нових значень (X) у межах розглянутого інтервалу.