ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені ІВАНА ФРАНКА

Факультет прикладної математики та інформатики

Кафедра дискретного аналізу

та інтелектуальних систем

Звіт з дисципліни

“Теорія ймовірності та математична статистика”

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ № 3

Виконала:

2021

Постановка задачі

1. За даними кореляційної таблиці обчислити умовні середні $y_{xi}(i = 1, \ldots, k)$.
2. Побудувати поле кореляції, тобто нанести на нього точки $M_i(x_i; y_{xi})$ на координатну площину. На основі цього зробити припущення про вигляд функції регресії.
3. В залежності від вигляду функції регресії скласти відповідну систему рівнянь. Розв'язати її і знайти невідомі параметри вибраної функції регресії.
4. Записати рівняння кривої регресії Y на X: $y_{xi} = f(x)$ та побудувати її графік.
5. Обчислити дисперсію величини Y відносно кривої регресії Y на X.
6. Визначити суму квадратів відхилень умовних середніх від значень функції регресії.

Короткі теоретичні відомості

Результати спостережень залежності між двома величинами X та Y, що мали частоти появи n_{ij} записують у вигляді кореляційної таблиці:

Y|X	x_1	x_2	...	x_i	...	x_k	m_j
y_1	n_{11}	n_{21}	...	n_{il}	...	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 цієї таблиці обчислюємо умовні середні $y_{xi}(i = 1, \ldots, k)$:

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

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

Для визначення вигляду функції регресії будують точки (x_i; y_x) і за їх розміщенням роблять висновок про приблизний вигляд функції регресії. Вона може мати вигляд наступних кореляцій.

1. Параболічна кореляція.

$$f(x) = ax^2 + bx + c,$$

де a, b та c - невідомі параметри, які можна знайти за системою рівнянь

$$\begin{cases} (\sum_{i=1}^{k} n_i x_i^4)a + (\sum_{i=1}^{k} n_i x_i^3)b + (\sum_{i=1}^{k} n_i x_i^2)c = \sum_{i=1}^{k} n_i \overline{y}_{x_i} x_i^2; \\ (\sum_{i=1}^{k} n_i x_i^3)a + (\sum_{i=1}^{k} n_i x_i^2)b + (\sum_{i=1}^{k} n_i x_i)c = \sum_{i=1}^{k} n_i \overline{y}_{x_i} x_i; \\ (\sum_{i=1}^{k} n_i x_i^2)a + (\sum_{i=1}^{k} n_i x_i)b + nc = \sum_{i=1}^{k} n_i \overline{y}_{x_i}. \end{cases}$$

2. Гіперболічна кореляція.

$$\overline{y}_x = \frac{a}{x} + b,$$

, де a та b - невідомі параметри, які можна знайти системою рівнянь

$$\begin{cases} a \sum_{i=1}^{k} \frac{1}{x_i} n_i + bn = \sum_{i=1}^{k} \overline{y}_{x_i} n_i; \\ a \sum_{i=1}^{k} \frac{1}{x_i^2} n_i + b \sum_{i=1}^{k} \frac{1}{x_i} n_i = \sum_{i=1}^{k} \frac{1}{x_i} \overline{y}_{x_i} n_i. \end{cases}$$

3. Показникова кореляція.

$\bar{y}_x = ba^x,$ де a та b - невідомі параметри, які можна знайти системою рівнянь

$$\begin{cases} \lg a \sum_{i=1}^{k} n_i x_i + n \lg b = \sum_{i=1}^{k} n_i \lg \overline{y}_{x_i}; \\ \lg a \sum_{i=1}^{k} n_i x_i^2 + \lg b \sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} n_i x_i \lg \overline{y}_{x_i}. \end{cases}$$

4. Коренева кореляція.

$\overline{y}_x = a\sqrt{x} + b,$ де a та b - невідомі параметри, які можна знайти системою рівнянь

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

За побудованою кривою регресії можна оцінити відхилення значень випадкової величини Y від кривої регресії y_x, а саме

дисперсію $\sigma^2 = \frac{\Delta}{n}$, де

$$\Delta = n_{11}[y_1 - f(x_1)]^2 + n_{21}[y_1 - f(x_2)]^2 + \ldots + n_{k1}[y_1 - f(x_k)]^2 +$$
$$+ n_{12}[y_2 - f(x_1)]^2 + n_{22}[y_2 - f(x_2)]^2 + \ldots + n_{k2}[y_2 - f(x_k)]^2 + \ldots +$$
$$+ n_{1l}[y_l - f(x_1)]^2 + n_{2l}[y_l - f(x_2)]^2 + \ldots + n_{kl}[y_l - f(x_k)]^2.$$

та суму квадратів відхилень

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

.

Програмна реалізація

Реалізувала програму мовою C# в середовищі Visual Studio. Для реалізації використовувала Windows Forms компоненти та бібліотеку Math.NET Numerics для розв’язання системи лінійних рівнянь.

Отримані результати

6.

Y|X	2	3	5	8	10	11	13
3						19	2
4				3	31	2	
6			1	16	3		
8		2	21	4			
10	3	31	5				
12	30	2					

Згідно з таблицею варіанту 6 знайшла середні y_{xi} :

y[2] = 11,818

y[3] = 10

y[5] = 8,296

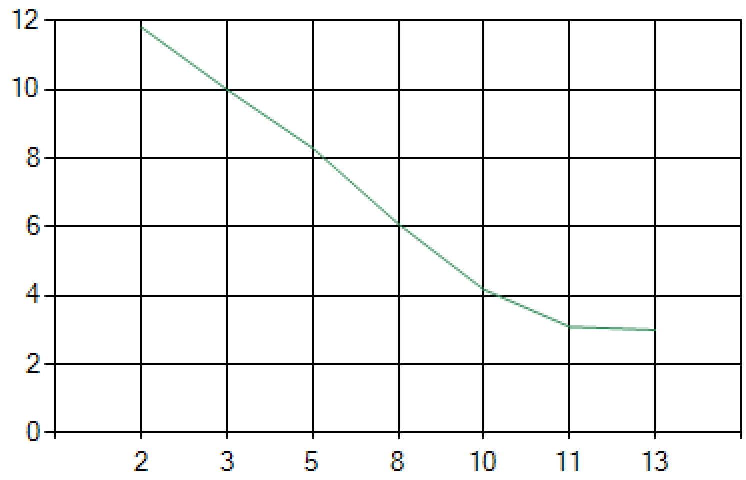
y[8] = 6,087

y[10] = 4,176

y[11] = 3,095

y[13] = 3

та графік 1 вигляду



Припустивши, що це гіперболічна кореляція, склала відповідну систему рівнянь

/41,905a + 175b = 1492

\14,104a + 41,905b = 436,443

Звідси, знаходимо a та b:

a = 19,4549701

b = 3,86708274

Рівняння кривої регресії Y на X: $y_x = \frac{19,4549701}{x} + 3,86708274$

З цього рівняння знайшла середні y_x:

y[2] = 13,595

y[3] = 10,352

y[5] = 7,758

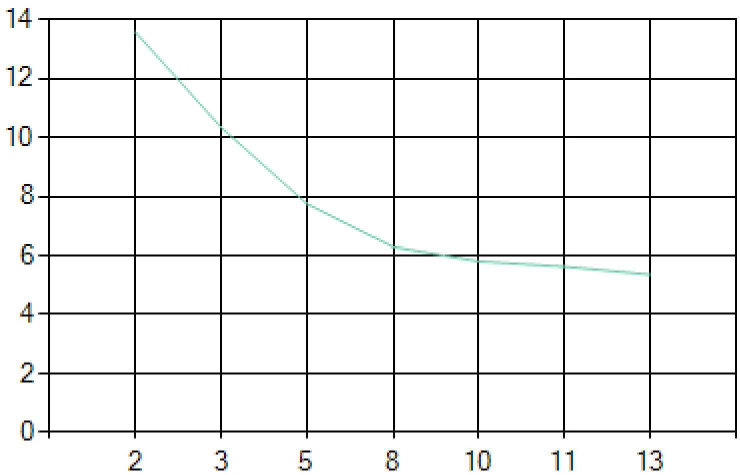
y[8] = 6,299

y[10] = 5,813

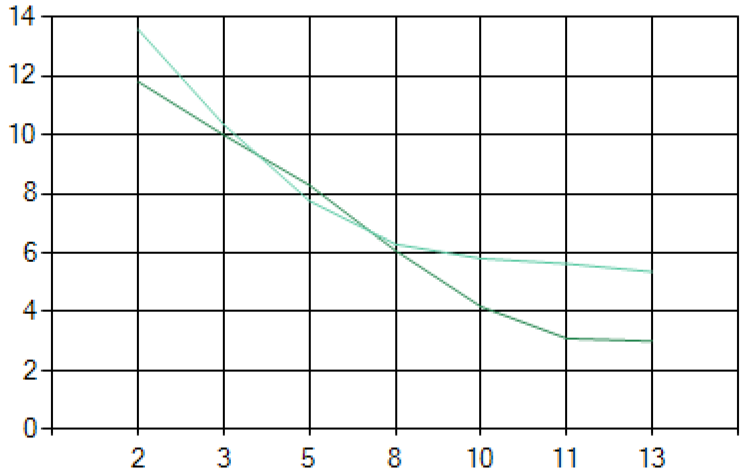
y[11] = 5,636

y[13] = 5,364

та графік 2 вигляду



Графіки 1 та 2 на одній координатній площині для порівняння



Обчислила дисперсію величини Y відносно кривої регресії Y на X.

σ^2 =2,61043286624728

Визначити суму квадратів відхилень умовних середніх від значень функції регресії.

$\delta^2 =$ 343,874180248822

Висновки

В ході виконання індивідуального завдання 3 я навчилась обчислювати умовні середні y_{xi} за даними кореляційної таблиці, будувати поле кореляції, на основі якого робити припущення про вигляд функції регресії, знаходити невідомі параметри вибраної функції, записувати рівняння кривої регресії, обчислювати дисперсію величини Y відносно кривої регресії Y на X та визначати суму квадратів відхилень умовних середніх від значень функції регресії.