Індивідуальне завдання №3

Постановка задачі

1) За даними кореляційної таблиці обчислити умовні середні yxi (i = 1,...,k)

2) Побудувати поле кореляції, тобто нанести точки Mi(xi;yxi), i = 1,...,k , на координатну площину. На основі цього зробити припущення про вигляд функції регресії (парабола, гіпербола і т.д.)

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 невідомі параметри вибраної функції регресії

4) Записати рівняння кривої регресії Y на X : yx = f(x) (з конкретною знайденою в пункті 3 функцією регресії f(x))та побудувати її графік

5) Обчислити дисперсію (12) величини Y відносно кривоїрегресії Y на X

6) Визначити суму квадратів відхилень δ2 умовних середніх від значень функції регресії за формулою (13).

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

Результати цих спостережень записують у вигляді кореляційної таблиці:

9.

$Y \mid X$	3	5	6	9	12	14	19
1,5	21						
2,5	4	31	3				
3		5	28	3	4		
3,5				25	4	3	
4					17	3	5
4,5						29	2

За даними кореляційної таблиці обчислюємо умовні середні за допомогою наступних формул:

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

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

Тепер потрібно скласти таблиці умовних середніх для X та У:

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

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

Після цього ми можемо побудувати графік регресії та оцінити її вигляд. В нашому випадку це – Коренева Кореляція. Це означає, що форма кореляційної залежності Y та X буде у вигляді наступного рів няння:

$$\overline{y}_x = a\sqrt{x} + b, \qquad \overline{x}_y = c\sqrt{y} + d.$$

Після знаходження всіх потрібних даних, можемо обчислити дисперсію величини Y відносно регресії Y на X за формулою:

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

Тепер можемо визначити суму квадратів відхилень δ2 умовних середніх від значень функції регресії за формулою:

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

Програмна реалізація

Програма зчитує введену в неї матрицю кореляцій, за допомогою спеціальних функцій. Далі формується таблиці для умовних середніх значень для подальшого їх використання в обчисленнях інших даних. Далі відбувається вивід графіка та умовний аналіз графіка на його вид. Оскільки в нашому випадку – це коренева кореляція, то обчислення параметрів a та b відбувається з розв’язанням системи двох рівнянь. Для обчислення дисперсії та суми квадратів відхилень δ2 умовних середніх від значень функції регресії, були створені дві окремі функції, які їх обраховують та записують в файл. Для кращої та компактнішої роботи програми, вона була завернута в exe файл. Аргументовано це тим, що програма не потребує вводу користувача як такого, адже сама матриця вже є задана в програмі, а отже і інтерфейсу програма не потребує. Таким чином після запуску exe файлу відкриється графік та буде створено txt файл, де й будуть зберігатися обчислені дані.