

ЛИНЕЙНАЯ РЕГРЕССИЯ, АДЕКВАТНОСТЬ МОДЕЛИ: РЕШЕНИЕ ЗАДАЧ

ЗАДАНИЕ.

Найти выборочные числовые характеристики и выборочное уравнение линейной регрессии $y_x = ax + b$. Построить прямую регрессии и изобразить на плоскости точки (x, y) из таблицы. Вычислить остаточную дисперсию. Проверить адекватность линейной регрессионной модели по коэффициенту детерминации.

x	1	2	4	6	10
y	2	6	11	16	24

РЕШЕНИЕ.

Здесь дана выборка объема $n = 5$ из двумерной генеральной совокупности (X, Y) . Для решения задачи требуется найти следующие величины:

1) выборочные средние:

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{1}{5} 23 = 4,6,$$

$$\bar{y} = \frac{1}{n} \sum y_i = \frac{1}{5} 59 = 11,8$$

2) выборочные дисперсии и среднеквадратичные отклонения:

$$D_x = \overline{x^2} - (\bar{x})^2 = \frac{1}{n} \sum x_i^2 - 4,6^2 = 31,4 - 4,6^2 = 10,24,$$

$$D_y = \overline{y^2} - (\bar{y})^2 = \frac{1}{n} \sum y_i^2 - 11,8^2 = 198,6 - 11,8^2 = 59,36$$

$$\sigma_x = \sqrt{D_x} = \sqrt{10,24} = 3,2, \quad \sigma_y = \sqrt{D_y} = \sqrt{59,36} \approx 7,705.$$

3) сумму $\overline{xy} = \frac{1}{n} \sum x_i y_i = \frac{1}{5} 394 = 78,8$

4) выборочный коэффициент корреляции

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y} = \frac{78,8 - 4,6 \cdot 11,8}{3,2 \cdot 7,705} \approx 0,994.$$

Коэффициент корреляции близок к единице, связь очень сильная (близка к функциональной), прямая.

Все расчеты в таблице ниже:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	2	1	4	2
2	6	4	36	12
4	11	16	121	44
6	16	36	256	96
10	24	100	576	240

Сумма	23	59	157	993	394
Среднее	4,6	11,8	31,4	198,6	78,8

Запишем выборочное уравнение линейной регрессии по формуле:

$$\frac{y_x - \bar{y}}{\sigma_y} = r_{xy} \frac{x - \bar{x}}{\sigma_x},$$

$$\frac{y_x - 11,8}{7,705} = 0,994 \frac{x - 4,6}{3,2},$$

$$y_x = 2,393x + 0,791.$$

По уравнению регрессии вычисляем y_x в заданных точках и заполняем таблицу.

x_i	y_i	$y_x(x_i)$	$(y_x(x_i) - y_i)^2$
1	2	3,184	1,402
2	6	5,577	0,179
4	11	10,363	0,406
6	16	15,149	0,724
10	24	24,721	0,520
23	59	58,994	3,231

Остаточная дисперсия $D_{ост} = \frac{1}{n} \sum (y_x(x_i) - y_i)^2 = \frac{1}{5} 3,231 = 0,646.$

Коэффициент детерминации $R^2 = 1 - \frac{D_{ост}}{D_y} = 1 - \frac{0,646}{59,36} = 0,989.$

Значение R^2 , близкое к единице, показывает, что выбранная регрессионная модель не противоречит экспериментальным данным.

По уравнению регрессии построим прямую в системе координат и на ту же систему координат нанесем точки (x, y) из таблицы.

Получим:

Задача скачана с сайта www.MatBuro.ru

©МатБюро - Решение задач по математике, статистике, экономике, программированию

Еще решения математической статистики: www.matburo.ru/ex_subject.php?p=ms

