

### Свойства коэффициентов множественной регрессии

Решить следующую задачу:

Заданы следующие данные опытов:

Хозяйства	Качество почвы (x), балл бонитета	Урожайность (Y), ц/га	$x_i y_i$	$x^2$
1	55	18,1+k		
2	50	21,1+k		
3	68	22,9+k		
4	48	18,9+k		
5	87	18,6+k		
6	60	30,5+k		
7	75	23,4+k		
8	80	27,6+k		
9	66	20,9+k		
10	58	18,2+k		
Сумма				
Среднее				

Здесь k – порядковый номер студента по журналу.

Выполнить следующее:

1. Построить уравнение регрессии.
2. Построить график.

3. Провести экономический анализ полученных результатов.

## Методические указания к выполнению заданий

Задача заключается в нахождении коэффициентов линейной зависимости, при которых функция двух переменных  $a$  и  $b$

$$F(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2$$

принимает наименьшее значение. То есть, при данных  $a$  и  $b$  сумма квадратов отклонений экспериментальных данных от найденной прямой будет наименьшей. В этом вся суть метода наименьших квадратов.

Таким образом, решение примера сводится к нахождению экстремума функции двух переменных.

### Вывод формул для нахождения коэффициентов.

Составляется и решается система из двух уравнений с двумя неизвестными.

$$F(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2$$

Находим частные производные функции по переменным  $a$  и  $b$ , приравниваем эти производные к нулю.

приравниваем эти производные к нулю.

$$\begin{cases} \frac{\partial F(a, b)}{\partial a} = 0 \\ \frac{\partial F(a, b)}{\partial b} = 0 \end{cases} \Leftrightarrow \begin{cases} -2 \sum_{i=1}^n (y_i - (ax_i + b))x_i = 0 \\ -2 \sum_{i=1}^n (y_i - (ax_i + b)) = 0 \end{cases} \Leftrightarrow \begin{cases} a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \\ a \sum_{i=1}^n x_i + nb = \sum_{i=1}^n y_i \end{cases}$$

Решаем полученную систему уравнений любым методом (например *методом подстановки* или *методом Крамера*) и получаем формулы для нахождения коэффициентов по методу наименьших квадратов (МНК).

$$\begin{cases} a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2} \\ b = \frac{\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i}{n} \end{cases}$$

$$F(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2$$

При данных  $a$  и  $b$  функция принимает наименьшее значение. Доказательство этого факта приведено ниже по тексту в конце страницы.

Вот и весь метод наименьших квадратов. Формула для нахождения

параметра  $a$  содержит суммы  $\sum_{i=1}^n x_i$ ,  $\sum_{i=1}^n y_i$ ,  $\sum_{i=1}^n x_i y_i$ ,  $\sum_{i=1}^n x_i^2$  и параметр  $n$  - количество экспериментальных данных. Значения этих сумм рекомендуем вычислять отдельно. Коэффициент  $b$  находится после вычисления  $a$ .

Пришло время вспомнить про исходный пример.

### Решение.

В нашем примере  $n=5$ . Заполняем таблицу для удобства вычисления сумм, которые входят в формулы искомых коэффициентов.

	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$\sum_{i=1}^5$
$x_i$	0	1	2	4	5	12
$y_i$	2,1	2,4	2,6	2,8	3	12,9
$x_i y_i$	0	2,4	5,2	11,2	15	33,8
$x_i^2$	0	1	4	16	25	46

Значения в четвертой строке таблицы получены умножением значений 2-ой строки на значения 3-ей строки для каждого номера  $i$ .

Значения в пятой строке таблицы получены возведением в квадрат значений 2-ой строки для каждого номера  $i$ .

Значения последнего столбца таблицы – это суммы значений по строкам.

Используем формулы метода наименьших квадратов для нахождения коэффициентов  $a$  и  $b$ . Подставляем в них соответствующие значения из

последнего столбца таблицы:

$$\begin{cases} a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2} \\ b = \frac{\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i}{n} \end{cases} \Rightarrow \begin{cases} a = \frac{5 \cdot 33,8 - 12 \cdot 12,9}{5 \cdot 46 - 12^2} \\ b = \frac{12,9 - a \cdot 12}{5} \end{cases} \Rightarrow \begin{cases} a \approx 0,165 \\ b \approx 2,184 \end{cases}$$

Следовательно,  $y = 0.165x + 2.184$  - искомая аппроксимирующая прямая.

Осталось выяснить какая из линий  $y =$

$0.165x + 2.184$  или  $g(x) = \sqrt[3]{x+1} + 1$  лучше аппроксимирует исходные данные, то есть произвести оценку методом наименьших квадратов.