5. РЕГРЕССИОННЫЙ АНАЛИЗ

Во многих задачах встречаются две или более переменных, между которыми существует характерная связь, причем необходимо исследовать природу этой связи. Связь между этими переменными характеризуется математической моделью, которая называется **уравнением регрессии**. Регрессионная модель должна аппроксимировать совокупность экспериментальных данных. Для аппроксимации широко используются полиномиальные модели. Регрессионные методы часто используются при анализе данных непланируемых экспериментов; такая ситуация может возникнуть при наблюдении неконтролируемых явлений. Однако планирование экспериментов для изучений регрессии дает больше преимуществ.

Регрессионный анализ как всякий статистический метод применим при определенных предположениях:

1) параметр оптимизаций y есть случайная величина с нормальным законом распределения;

2) дисперсия y не зависит от абсолютной величины y (однородность дисперсии в различных точках пространства);

3) значения факторов есть неслучайные величины (установление каждого фактора на заданном уровне и его поддержание существенно точнее, чем ошибка воспроизводимости).

5.1. Метод наименьших квадратов

В математической статистике разработано множество различных способов обработки результатов эксперимента. Среди всех этих методов выделяется своей простотой метод наименьших квадратов (МНК), разработанный более 150 лет назад Гауссом и Лежандром.

Пусть производится серия опытов, целью которых является исследование зависимости некоторой физической величины y от физической величины x. Предполагается, что величины x и y связаны функциональной зависимостью $y = f(x)$, вид которой необходимо определить из опытов. По МНК наилучшей считается та функция, для которой сумма квадрата разности отклонений наблюдаемых (экспериментальных) значений *f(x)* от вычисленных y является наименьшей:

$$\Phi = \sum_{i=1}^{N} \left[y_i - f(x_i)\right]^2 = \min, \tag{5.1}$$

$f(x_i)$ можно записать как $f(x_i, b_0, b_1 ...)$, тогда требуется выбрать коэффициенты $b_i (i = 0, 1, 2...)$, чтобы $\Phi = \sum_{i=1}^{N} \left[y_i - f(x_i, b_0, b_1 ...)\right]^2 = \min$.

Из курса математики: минимум некоторой функции, если он существует, достигает при одновременном равенстве нулю частных производных по всем неизвестным, то есть:

$$\frac{\partial \Phi}{\partial b_0} = \frac{\partial \Phi}{\partial b_1} = \frac{\partial \Phi}{\partial b_2} = ... = 0 \qquad \text{или}$$

$$\begin{cases} \sum_{i=1}^{N} 2\left[y_i - f(x_i, b_0, b_1, b_2 ...)\right] \dfrac{\partial f(x_i)}{\partial b_0} = 0, \\ \sum_{i=1}^{N} 2\left[y_i - f(x_i, b_0, b_1, b_2 ...)\right] \dfrac{\partial f(x_i)}{\partial b_1} = 0, \\ \sum_{i=1}^{N} 2\left[y_i - f(x_i, b_0, b_1, b_2 ...)\right] \dfrac{\partial f(x_i)}{\partial b_2} = 0, \end{cases} \tag{5.2}$$

....

Система уравнений содержит столько же уравнений, сколько неизвестных коэффициентов $b_0, b_1, b_2 \ldots$.

На практике более часто встречаются три случая: когда функция f линейна, когда она выражается уравнением второй степени (параболой) и в виде множественной регрессии.

5.2. Линейная регрессия

Требуется подобрать по МНК коэффициенты линейного уравнения регрессий $y = b_0 + b_1 x$ при выборке объемом N.

Система нормальных уравнений при этом имеет вид

$$\begin{cases} \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} (b_0 + b_i x_i) = 0; \\ \sum_{i=1}^{N} y_i x_i - \sum_{i=1}^{N} (b_0 + b_i x_i) x_i = 0 \end{cases} \tag{5.3}$$

или

$$\begin{cases} N b_0 + b_1 \sum_{i=1}^{N} x_{1i} = \sum_{i=1}^{N} y_i; \\ b_0 \sum_{i=1}^{N} x_{1i} + b_1 \sum_{i=1}^{N} x_{1i^2} = \sum_{i=1}^{N} x_{1i} y_i. \end{cases} \tag{5.4}$$

Коэффициенты b_0 и b_1 можно найти с помощью определителей:

$$b_0 = \frac{\begin{vmatrix} \sum_{i=1}^{N} y_i & \sum_{i=1}^{N} x_{1i} \\ \sum_{i=1}^{N} x_{1i} y_i & \sum_{i=1}^{N} x^2{}_{1i} \end{vmatrix}}{\begin{vmatrix} N & \sum_{i=1}^{N} x_{1i} \\ \sum_{i=1}^{N} x_{1i} & \sum_{i=1}^{N} x^2{}_{1i} \end{vmatrix}} = \frac{\sum_{i=1}^{N} y_i \sum_{i=1}^{N} x^2{}_{1i} - \sum_{i=1}^{N} x_{1i} y_i \sum_{i=1}^{N} x_{1i}}{N \sum_{i=1}^{N} x_{1i}^2 - (\sum_{i=1}^{N} x_{1i})^2}\ ; \tag{5.5}$$

$$b_1 = \frac{\begin{vmatrix} N & \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_{1i} & \sum_{i=1}^{N} x_{1i} y_i \end{vmatrix}}{\begin{vmatrix} N & \sum_{i=1}^{N} x_{1i} y_i \\ \sum_{i=1}^{N} x_{1i} & \sum_{i=1}^{N} x^2{}_{1i} \end{vmatrix}} = \frac{N\sum_{i=1}^{N} x_{1i} y_i - \sum_{i=1}^{N} y_i \sum_{i=1}^{N} x_{1i}}{N\sum_{i=1}^{N} x^2{}_{1i} \; - (\sum_{i=1}^{N} x_{1i})^2}\ . \tag{5.6}$$

Результаты эксперимента можно представить табл. 5.1.

Таблица 5.1

№ опыта	X_1	$У$
1	X_{11}	$У_1$
2	X_{12}	$У_2$
3	X_{13}	$У_3$
:	:	:
N	X_{1N}	$У_N$

Для выполнения вычислений ее расширяют (табл. 5.2).

Таблица 5.2

№ опыта	X_1	Y	X_1^2	YX_1	Y^2	$X_1 + Y$	$(X_1 + Y)^2$
1	X_{11}	Y_1	X_{11}^2	Y_1X_{11}	Y_1^2	$X_{11} + Y_1$	$(X_{11} + Y_1)^2$
2	X_{12}	Y_2	X_{12}^2	Y_2X_{12}	Y_2^2	$X_{12} + Y_2$	$(X_{12} + Y_2)^2$
3	X_{13}	Y_3	X_{13}^2	Y_3X_{13}	Y_3^2	$X_{13} + Y_3$	$(X_{13} + Y_3)^2$
i	X_{1i}	Y_i	X_{1i}^2	Y_iX_{1i}	Y_i^2	$X_{1i} + Y_i$	$(X_{1i} + Y_i)^2$
N	X_{1N}	Y_N	X_{1N}^2	Y_NX_{1N}	Y_N^2	$X_{1N} + Y_N$	$(X_{1N} + Y_N)^2$
Σ	$\sum_{i=1}^{N} X_{1i}$	$\sum_{i=1}^{N} Y_i$	$\sum_{i=1}^{N} X^2{}_{1i}$	$\sum_{i=1}^{N} Y_i X_{1i}$	$\sum_{i=1}^{N} Y_i^2$	–	$\sum_{i=1}^{N} (X_{1i} + Y_i)^2$
среднее	$\bar{X}_1$	$\bar{Y}$					

«Лишние» данные нужны для проверки правильности расчетов:

1 способ проверки: $$\sum_{i=1}^{N}(x_{1i}+y_i)^2=\sum_{i=1}^{N}x_{1i}^2+2\sum_{i=1}^{N}y_i x_{1i}+\sum_{i=1}^{N}y_i^2\,, \qquad (5.7)$$

2 способ проверки: $\bar{y}=b_0+b_1\bar{x}_i$. Подставляя в соотношение $\bar{y}$ и $\bar{x}_1$ и один из коэффициентов, можно найти другой коэффициент и сравнить его с расчетным.

Разделив все члены уравнений (5.5) и (5.6) на N^2, коэффициенты b_0 и b_1 определим через статистические моменты:

$$m_x=\frac{1}{N}\sum_{i=1}^{N}x_{1i}\,; \qquad m_y=\frac{1}{N}\sum_{i=1}^{N}y_i\,;$$

$$a_2=\frac{1}{N}\sum_{i=1}^{N}x_{1i}^2\,; \qquad a_{11}=\frac{1}{N}\sum_{i=1}^{N}x_{1i}y_i\,.$$

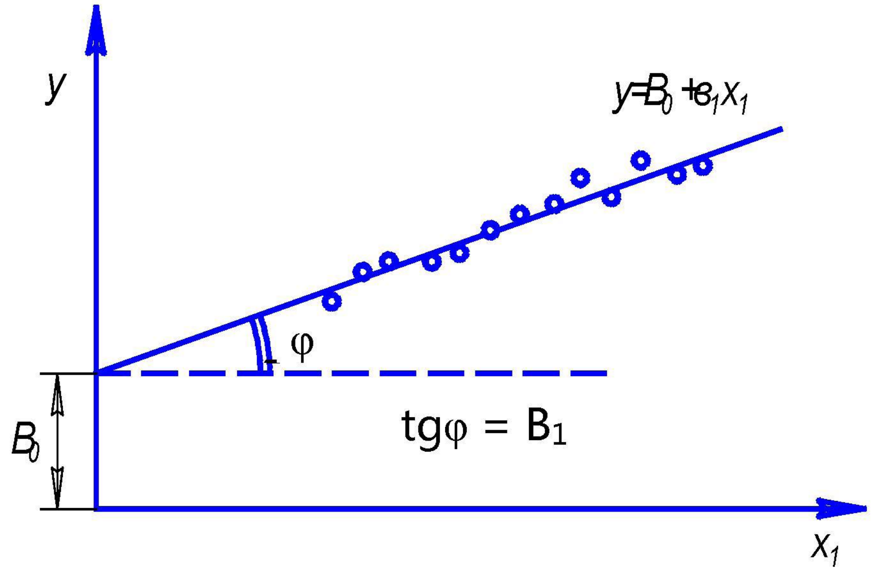
В результате получим:

$$b_0=\frac{m_y a_2-m_x a_{11}}{a_2-m_x^2}; \qquad b_1=\frac{a_{11}-m_x m_y}{a_2-m_x^2}. \qquad (5.8)$$

Таким образом,

$$y=\frac{m_y a_2-m_x a_{11}}{a_2-m_x^2}+x\frac{a_{11}-m_x m_y}{a_2-m_x^2}. \qquad (5.9)$$

На рис. 5.1 представлен график уравнения (5.9).



Если b_0=0, то
tgφ = y/x,
y= tg$\varphi\cdot x$

Рис. 5.1. Линейное уравнение регрессии

Геометрический смысл МНК рассмотрим на примере рис. 5.2.

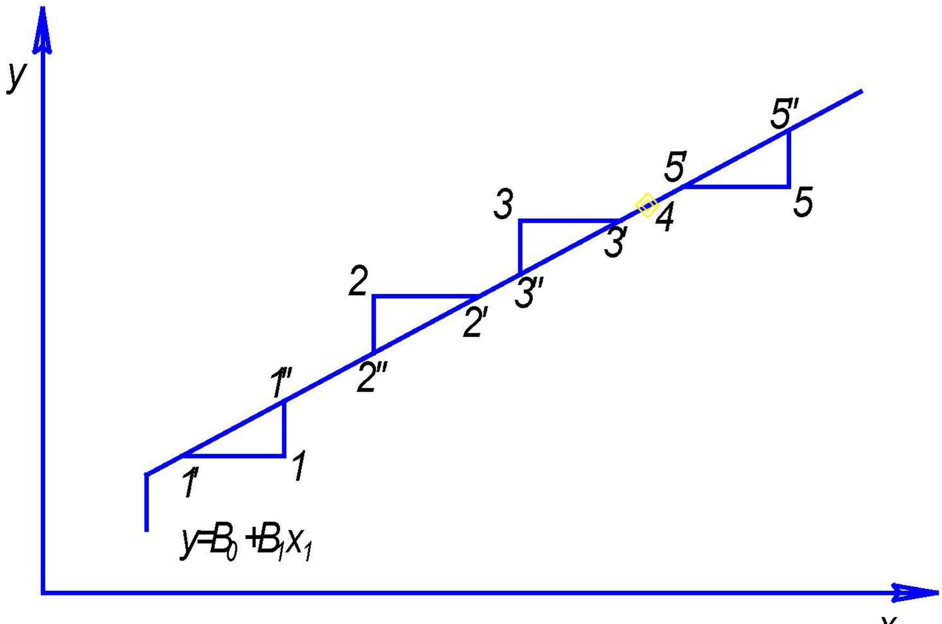


Рис. 5.2. Линейное уравнение регрессии (фрагмент)

Выберем пять экспериментальных точек. МНК состоит в том, чтобы минимизировать сумму квадратов отрезков, характеризующих расхождение между экспериментальными точками и полученным уравнением. Минимизировали отрезки, параллельные оси y. Из рисунка видно, что необходимо определить y в условиях каждого опыта. Будем называть такое значение предсказанным и обозначать $\hat{y}$. Затем находим все отрезки – разницу значений между теоретическим графиком и экспериментальными данными, возводим их в квадрат и складываем (табл. 5.3).

Таблица 5.3

№ опыта	Y	$\hat{Y}$	$\Delta Y = Y - \hat{Y}$	ΔY^2
1	Y_1	$\hat{Y}_1$	ΔY_1	$\Delta Y^2{}_1$
2	Y_2	$\hat{Y}_2$	ΔY_2	$\Delta Y^2{}_2$
3	Y_3	$\hat{Y}_3$	ΔY_3	$\Delta Y^2{}_3$
i	Y_i	$\hat{Y}_i$	ΔY_i	$\Delta Y^2{}_i$
N	Y_N	$\hat{Y}_N$	ΔY_N	$\Delta Y^2{}_N$
				$\sum_{i=1}^{N} \Delta Y_i^2$

Величина $\sum_{i=1}^{N} \Delta Y_i^2$ – остаточная сумма квадратов. МНК гарантирует, что эта величина минимально возможная.

5.3. Параболическая регрессия

Уравнение представляет собой полином второй степени $y = b_0 + b_1 x + b_2 x^2$, для определения b_0, b_1, b_2 необходимо решить систему уравнений:

$$\frac{\partial f(x)}{\partial b_0} = 1; \frac{\partial f(x)}{\partial b_1} = x; \frac{\partial f(x)}{\partial b_2} = x^2, \qquad (5.10)$$

или

$$\begin{cases} b_0 N + b_1 \sum_{i=1}^{N} x_i + b_2 \sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} y_i, \\ b_0 \sum_{i=1}^{N} x_i + b_1 \sum_{i=1}^{N} x_i^2 + b_2 \sum_{i=1}^{N} x_i^3 = \sum_{i=1}^{N} x_i y_i, \\ b_0 \sum_{i=1}^{N} x_i^2 + b_1 \sum_{i=1}^{N} x_i^3 + b_2 \sum_{i=1}^{N} x_i^4 = \sum_{i=1}^{N} x_i^2 y_i. \end{cases} \tag{5.11}$$

Коэффициенты этой системы также представляют собой статистические моменты системы двух величин x и y, а именно:

$$m_x = \frac{1}{N}\sum_{i=1}^{N} x_i\ ; \qquad m_y = \frac{1}{N}\sum_{i=1}^{N} y_i\ ;$$

$$a_2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2\ ; \qquad a_{11} = \frac{1}{N}\sum_{i=1}^{N} x_i y_i\ .$$

$$a_3 = \frac{1}{N}\sum_{i=1}^{N} x_i^3\ ; \qquad a_4 = \frac{1}{N}\sum_{i=1}^{N} x_i^4\ ;$$

$$a_{21} = \frac{1}{N}\sum_{i=1}^{N} x_i^2 y_i\ .$$

Пользуясь выражениями для коэффициентов, систему нормальных уравнений можно записать так:

$$\begin{cases} b_0 + b_1 m_x + b_2 a_2 = m_y; \\ b_0 m_x + b_1 a_2 + b_2 a_3 = a_{11}; \\ b_0 a_2 + b_1 a_3 + b_2 a_4 = a_{21}. \end{cases} \tag{5.12}$$

Решив систему уравнений относительно коэффициентов методом определителей, получим: $b_0 = \Delta b_0 / \Delta$, $b_1 = \Delta b_1 / \Delta$, $b_2 = \Delta b_2 / \Delta$,
где

$$\Delta b_0 = \begin{vmatrix} m_y & m_x & a_2 \\ a_{11} & a_2 & a_3 \\ a_{21} & a_3 & a_4 \end{vmatrix}; \qquad \Delta b_1 = \begin{vmatrix} 1 & m_x & m_y \\ m_x & a_{11} & a_3 \\ a_2 & a_{21} & a_4 \end{vmatrix};$$

$$\Delta b_2 = \begin{vmatrix} 1 & m_x & m_y \\ m_x & a_2 & a_3 \\ a_2 & a_3 & a_4 \end{vmatrix}; \qquad \Delta = \begin{vmatrix} 1 & m_y & a_2 \\ m_x & a_2 & a_3 \\ a_2 & a_3 & a_4 \end{vmatrix}.$$

После подстановки значений коэффициентов b_i квадратическая модель имеет вид

$$y = \frac{\Delta b_2}{\Delta} x^2 + \frac{\Delta b_1}{\Delta} x + \frac{\Delta b_0}{\Delta}. \qquad (5.13)$$

Аналогичными по структуре уравнениями будут определяться коэффициенты параболы любого порядка. Таким образом, когда экспериментальная зависимость определяется по методу наименьших квадратов для полинома некоторой степени, коэффициенты этого полинома находятся решением системы линейных уравнений. Коэффициенты этой системы линейных уравнений представляют собой статистические моменты различных порядков, характеризующие систему величин (x; y), если ее рассматривать как систему случайных величин.

5.4. Множественная регрессия

Если необходимо исследовать корреляционную связь между многими величинами, то пользуются уравнениями множественной регрессии:

$$y = b_0 + \sum_{i=1}^{\kappa} b_i x_i \ . \tag{5.14}$$

Однако решение системы нормальных уравнений с большим числом неизвестных b_i представляет собой достаточно трудоемкую задачу. В этом случае для проведения регрессионного анализа целесообразно использовать ЭВМ.

Перепишем уравнение (5.14) в виде

$$y = b_0 x_0 + \sum_{i=1}^{\kappa} b_i x_i \ , \tag{5.15}$$

где x_0 – фиксированная переменная, равная 1.

Статистические данные для выполнения вычислений представим в матричной форме:
матрицу независимых переменных как

$$x = \begin{vmatrix} x_{01} & x_{11} & \dots & x_{\kappa 1} \\ x_{02} & x_{12} & \dots & x_{\kappa 2} \\ \dots & \dots & \dots & \dots \\ x_{0N} & x_{1N} & \dots & x_{\kappa N} \end{vmatrix}; \tag{5.16}$$

вектор наблюдений y_i как

$$y = \begin{vmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{vmatrix}; \tag{5.17}$$

вектор коэффициентов b_i как

$$B=\begin{vmatrix} b_0 \\ b_1 \\ \dots \\ b_N \end{vmatrix}; \tag{5.18}$$

транспонированную матрицу X как

$$x^T=\begin{vmatrix} x_{01} & x_{02} & \dots & x_{0N} \\ x_{11} & x_{12} & \dots & x_{1N} \\ \dots & \dots & \dots & \dots \\ x_{k1} & x_{k2} & \dots & x_{kN} \end{vmatrix}. \tag{5.19}$$

Система нормальных уравнений для определения коэффициентов $b_0, b_1, ..., b_k$ имеет вид

$$\begin{cases} b_0\sum_{i=1}^{N}x_{01}^2+b_1\sum_{i=1}^{N}x_{01}x_{1i}+\dots+b_k\sum_{i=1}^{N}x_{0i}x_{ki}=\sum_{i=1}^{N}x_{0i}y_i; \\ b_0\sum_{i=1}^{N}x_{1i}x_{0i}+b_1\sum_{i=1}^{N}x_{1i}^2+\dots+b_k\sum_{i=1}^{N}x_{1i}x_{ki}=\sum_{i=1}^{N}x_{1i}y_i; \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ b_0\sum_{i=1}^{N}x_{ki}x_{0i}+b_1\sum_{i=1}^{N}x_{ki}x_{1i}+\dots+b_k\sum_{i=1}^{N}x_{ki}^2=\sum_{i=1}^{N}x_{ki}y_i. \end{cases} \tag{5.20}$$

Или в матричной форме

$$x^T xB = x^T y \;; \tag{5.21}$$

$$x^T x=\begin{vmatrix} \sum_{i=1}^{N}x_{0i}^2 & \sum_{i=1}^{N}x_{0i}x_{1i} & \dots & \sum_{i=1}^{N}x_{0i}x_{ki} \\ \sum_{i=1}^{N}x_{1i}x_{0i} & \sum_{i=1}^{N}x_{1i}^2 & \dots & \sum_{i=1}^{N}x_{1i}x_{ki} \\ \dots & \dots & \dots & \dots \\ \sum_{i=1}^{N}x_{ki}x_{0i} & \sum_{i=1}^{N}x_{ki}x_{1i} & \dots & \sum_{i=1}^{N}x_{ki}^2 \end{vmatrix};$$

$$x^T x\,\mathrm{B}=\begin{vmatrix} b_0\sum\limits_{i=1}^{N}x_{0i}^2 & b_1\sum\limits_{i=1}^{N}x_{0i}x_{1i} & \ldots & b_k\sum\limits_{i=1}^{N}x_{0i}x_{ki} \\ b_0\sum\limits_{i=1}^{N}x_{1i}x_{0i} & b_1\sum\limits_{i=1}^{N}x_{1i}^2 & \ldots & b_k\sum\limits_{i=1}^{N}x_{1i}x_{ki} \\ \ldots & \ldots & \ldots & \ldots \\ b_0\sum\limits_{i=1}^{N}x_{ki}x_{0i} & b_1\sum\limits_{i=1}^{N}x_{ki}x_{1i} & \ldots & b_k\sum\limits_{i=1}^{N}x_{ki}^2 \end{vmatrix};$$

$$x^T y=\begin{vmatrix} \sum\limits_{i=1}^{N}x_{oi}y_i \\ \sum\limits_{i=1}^{N}x_{1i}y_i \\ \ldots \\ \sum\limits_{i=1}^{N}x_{ki}y_i \end{vmatrix}.$$

Матрица-столбец коэффициентов B определяется следующим образом:

$$B=(x^T x)^{-1}x^T y, \tag{5.22}$$

где $(x^T x)^{-1}$ – матрица, обратная матрице $(x^T x)$.

$$(x^T x)^{-1}=\begin{vmatrix} d_{00} & d_{01} & \ldots & d_{0k} \\ d_{10} & d_{11} & \ldots & d_{1k} \\ \ldots & \ldots & \ldots & \ldots \\ d_{k0} & d_{k1} & \ldots & d_{kk} \end{vmatrix}. \tag{5.23}$$

Элемент обратной матрицы вычислим из соотношения

$$d_{jh}=\frac{1}{\Delta}\left(\sum_{i=1}^{N}x_{ui}x_{ji}\right),$$

где Δ – определитель матрицы $(x^T x)$; $\sum\limits_{i=1}^{N}x_{ui}x_{ji}$ – алгебраическое дополнение элемента в матрице $(x^T x)$.

Вычисления возможны, если обратная матрица (5.23) не вырождена, то есть переменные $x_1; x_2 \dots x_к$ линейно независимы.

Геометрический смысл метода наименьших квадратов можно интерпретировать, как показано на рис. 5.3. В идеальном случае Y = BX (Y = $\bar{Y}$). В реальности из-за наличия погрешности $\hat{Y}$ = BX + ε.

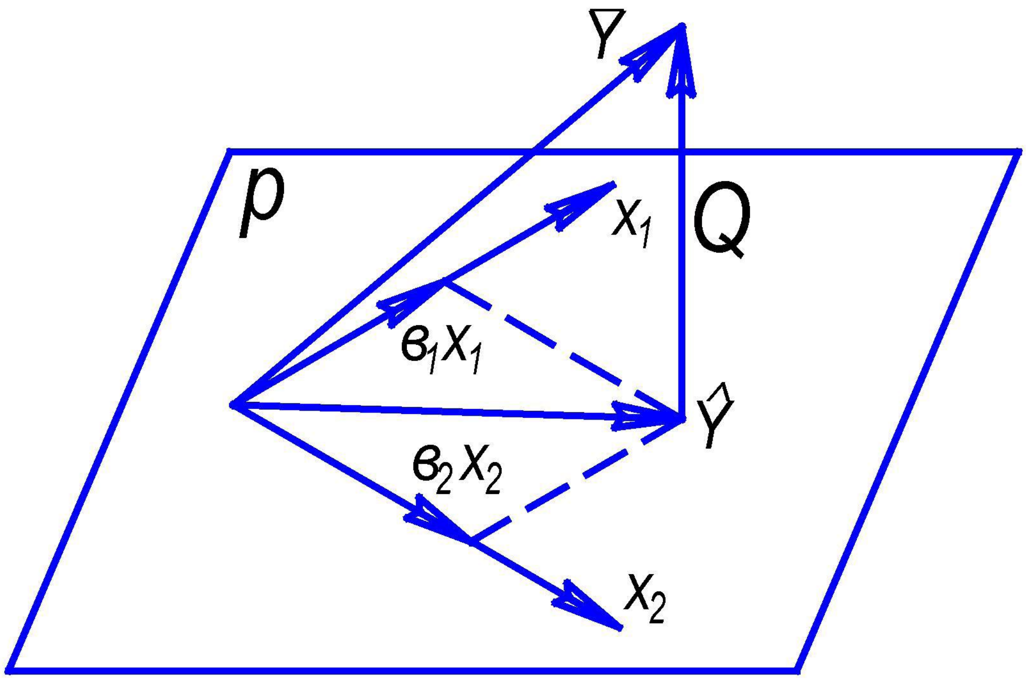


Рис. 5.3. Геометрическое представление метода наименьших квадратов

Минимальное расстояние между вектором наблюдений $\hat{Y}$ и гиперплоскостью P будет соответствовать длине перпендикуляра, опущенного из конца этого вектора на гиперплоскость (Q = | Ŷ – Y | будет минимальным).

По условию ортогональности вектор Ŷ – Y к гиперплоскости можно записать в виде $X^T(\hat{Y} - Y) = 0$ (X^T – транспонированная матрица). Приняв во внимание, что Y = BX, получим нормальное уравнение $X^T XB = X^T \hat{Y}$.

5.5. Примеры применения регрессионного анализа

Пример. Проводился некоторый эксперимент, результаты которого представлены в табл. 5.4. Необходимо определить теоретические значения по методу наименьших квадратов. Построить график экспериментальных и вычисленных (теоретических) зависимостей.

Таблица 5.4

№ опыта	X	$\bar{y}$
1	1	223
2	2	218
3	3	212
4	4	208
5	5	202
6	6	198
$\sum$	21	1261

Регрессионная зависимость имеет вид $\hat{y} = b_0 + b_1 x$.

По МНК суммарное квадратическое отклонение $Q = \sum_1^6 (\bar{y} - \hat{y})^2$ должно быть минимальным.

$$\sum_1^6 (\bar{y} - \hat{y})^2 = \sum_1^6 (\bar{y} - (b_0 + b_1 x))^2 = \sum_1^6 (\bar{y}^2 - 2\bar{y}(b_0 + b_1 x) + (b_0 + b_1 x)^2) =$$

$$= \sum_1^6 (\bar{y}^2 - 2\bar{y}\, b_0 - 2\bar{y}\, b_1 x) + b_0^2 + 2b_0 b_1 x + b_1^2 x^2.$$

Вычисляем частные производные по b_0 и b_1:

$$\begin{cases} \partial Q / \partial b_0 = -\sum_1^6 2\bar{y} + \sum_1^6 2b_0 + \sum_1^6 2b_1 x = 0; \\ \partial Q / \partial b_1 = -\sum_1^6 2\bar{y}x + \sum_1^6 2b_0 x + \sum_1^6 2b_1 x^2 = 0. \end{cases}$$

Вывод полученной системы уравнений аналогичен формуле (5.4).

Преобразуем систему к следующему виду:

$$\begin{cases} -\sum_{1}^{6}\bar{y} + 6b_0 + \sum_{1}^{6} b_1 x = 0; \\ -\sum_{1}^{6}\bar{y}x + \sum_{1}^{6} 2b_0 x + \sum_{1}^{6} b_1 x^2 = 0. \end{cases} \quad (5.24)$$

Подставляем в систему уравнений численные значения из табл. 5.4. Для удобства расчетов табл. 5.4 дополним новыми колонками, куда внесем значения, полученные при расчетах (табл. 5.5).

Таблица 5.5

№ опыта	X	$\bar{y}$	X^2	$X \cdot \bar{y}$	$\hat{y}$	ε, %
1	1	223	1	223	222,81	0,08
2	2	218	4	436	217,76	0,11
3	3	212	9	636	212,69	0,33
4	4	208	16	832	207,64	0,17
5	5	202	25	1010	202,59	0,28
6	6	198	36	1188	197,53	0,24
∑	21	1261	91	4325		

Подставляя значения из табл. 5.5 в систему уравнений (5.24), получим

$$\begin{cases} 1261 = 6b_0 + 21b_1; \\ 4325 = 21b_0 + 91b_1. \end{cases}$$

Решая систему уравнений, получим $b_0 = 227,87$ и $b_1 = -5,057$.

Подставляя коэффициенты b_0 и b_1 в исходное уравнение, получим $\hat{y}$ = 227,87 –5,057x.

Рассчитаем теоретические значения для каждой точки плана:

$$\hat{y}_1 = b_0 + b_1 x_1 = 227{,}87 - 5{,}057 \cdot 1 = 222{,}81;$$
$$\hat{y}_2 = b_0 + b_1 x_2 = 227{,}87 - 5{,}057 \cdot 2 = 217{,}76;$$
$$\hat{y}_3 = b_0 + b_1 x_3 = 227{,}87 - 5{,}057 \cdot 3 = 212{,}69;$$
$$\hat{y}_4 = b_0 + b_1 x_4 = 227{,}87 - 5{,}057 \cdot 4 = 207{,}64;$$
$$\hat{y}_5 = b_0 + b_1 x_5 = 227{,}87 - 5{,}057 \cdot 5 = 202{,}59;$$
$$\hat{y}_6 = b_0 + b_1 x_6 = 227{,}87 - 5{,}057 \cdot 6 = 197{,}53.$$

Относительная ошибка вычисляется по формуле $\varepsilon = (\hat{y} - \bar{y})/\hat{y} \cdot 100$ %.

По результатам обработки данных строим экспериментальные и вычисленные зависимости (рис. 5.4).

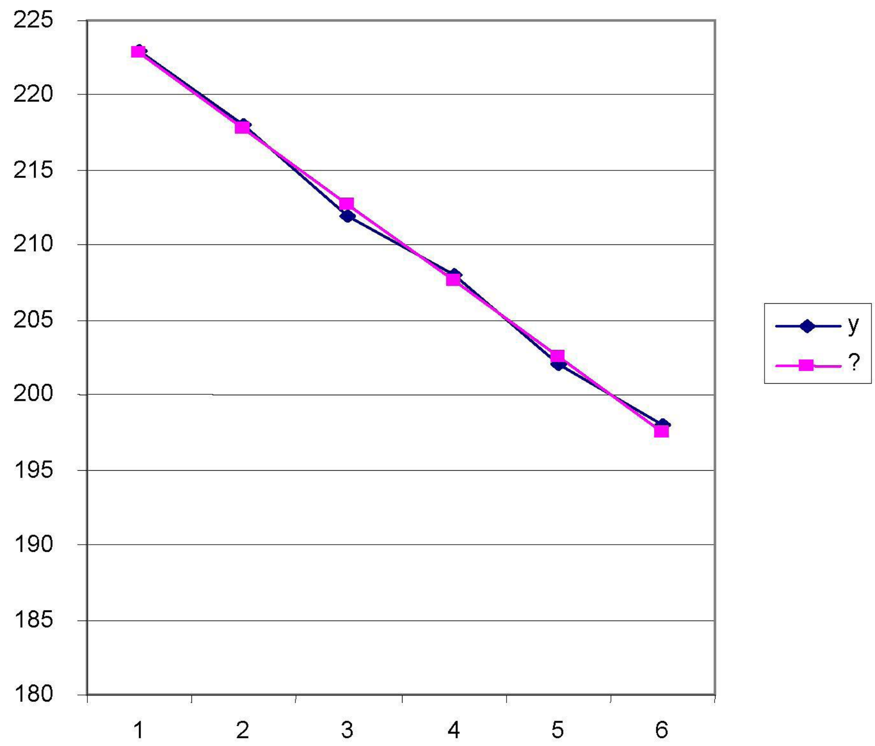


Рис. 5.4. Экспериментальные и вычисленные (теоретические) зависимости

Пример. Пусть отклик y линейно зависит от факторов X_1 и X_2. Исходные данные для проведения регрессионного анализа приведены в табл. 5.6.

Таблица 5.6

Фактор i / Точка j	X_0	X_1	X_2	$\bar{y}_i$
1	+1	2	1	10
2	+1	2	2	12
3	+1	8	10	17
4	+1	2	4	13
5	+1	6	8	15
6	+1	3	4	10
7	+1	5	7	14
8	+1	3	3	12
9	+1	9	10	16
10	+1	10	11	18

Таким образом, на основании формул (5.16)–(5.18) и табл. 5.6 можно записать:

$$\bar{y} = \begin{bmatrix} 10 \\ 12 \\ 17 \\ 13 \\ 15 \\ 10 \\ 14 \\ 12 \\ 16 \\ 18 \end{bmatrix}; \qquad X = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 2 & 2 \\ 1 & 7 & 10 \\ 1 & 2 & 4 \\ 1 & 6 & 8 \\ 1 & 3 & 4 \\ 1 & 5 & 7 \\ 1 & 3 & 3 \\ 1 & 9 & 10 \\ 1 & 10 & 11 \end{bmatrix}; \qquad B = \begin{vmatrix} b_0 \\ b_1 \\ \dots \\ b_N \end{vmatrix} = ?$$

Для вычисления коэффициентов воспользуемся формулой (5.22). Тогда:

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 2 & 2 & 8 & \dots & 10 \\ 1 & 2 & 10 & \dots & 11 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 & 1 \\ 1 & 2 & 2 \\ 1 & 8 & 10 \\ \vdots & \vdots & \vdots \\ 1 & 10 & 11 \end{bmatrix} = \begin{bmatrix} 10 & 50 & 60 \\ 50 & 336 & 398 \\ 60 & 398 & 480 \end{bmatrix};$$

$$\left(X^T X\right)^{-1} = \begin{bmatrix} 0{,}402 & -0{,}017 & -0{,}036 \\ -0{,}017 & 0{,}168 & -0{,}136 \\ -0{,}036 & -0{,}137 & 0{,}120 \end{bmatrix};$$

$$X^T \overline{y} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 2 & 2 & 8 & \dots & 10 \\ 1 & 2 & 10 & \dots & 11 \end{bmatrix} \cdot \begin{bmatrix} 10 \\ 12 \\ 17 \\ \vdots \\ 18 \end{bmatrix} = \begin{bmatrix} 137 \\ 756 \\ 908 \end{bmatrix}.$$

После этого находим

$$B = \left(X^T X\right)^{-1} X^T Y = \begin{bmatrix} 0{,}402 & -0{,}017 & -0{,}036 \\ -0{,}017 & 0{,}168 & -0{,}137 \\ -0{,}036 & -0{,}137 & 0{,}120 \end{bmatrix} \cdot \begin{bmatrix} 137 \\ 756 \\ 908 \end{bmatrix} = \begin{bmatrix} 9{,}39 \\ 0{,}13 \\ 0{,}61 \end{bmatrix}.$$

Таким образом, $b_0 = 9{,}39$; $b_1 = 0{,}13$; $b_2 = 0{,}61$ и уравнение регрессии имеет вид $\overline{y} = 9{,}39 + 0{,}13x_1 + 0{,}61x_2$.