Метод наименьших квадратов (МНК)

МНК есть тем классическим методом, с которого собственно, и надо было начинать рассылку о методах прогнозирования. Его краткое представление ниже.

Пусть в качестве исходных данных имеем таблицу

X	x_1	x_2	…	x_n
Y	y_1	y_2	…	y_n

содержащую статистические данные, или данные экспериментов. Если в качестве X выступает время, то имеем динамический ряд (тогда x_i размещены в возрастающем порядке). Необходимо получить аналитическую зависимость

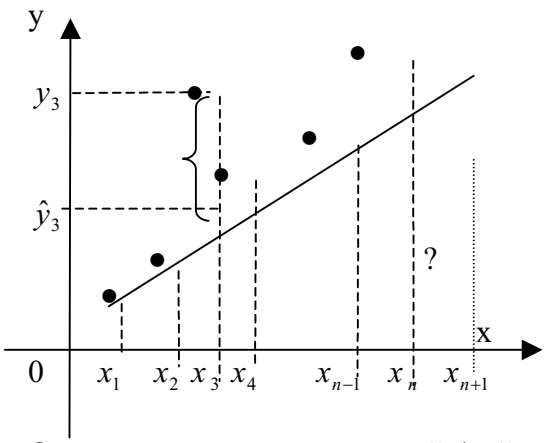
$$\hat{Y} = f(x), \qquad (*)$$

которая наилучшим образом описывает начальные данные. Словосочетание «наилучшим образом», будем понимать в смысле минимума суммы квадратов отклонений значений y_i, данных в таблице от $\hat{y}_i$, рассчитанных по (*):

$$E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \ . \qquad (**)$$

Определение зависимости (*) необходимо, в т.ч., и для нахождения $\hat{Y}_{n+1} = f(x_{n+1})$, что уже представляет собой задачу прогнозирования.

Нанесём точки из таблицы на координатную плоскость и сделаем предположение, что зависимость (*) есть линейной $\hat{Y} = a + bx$, а отклонения от прямой вызваны случайными факторами.



Определим уравнение прямой (найдем значения коэффициентов a и b), так, чтобы получить решение задачи $E \to \min$, т.е. необходимо найти минимум функции

$$E = \sum_{i=1}^{n} (y_i - (a + bx_i))^2 \ .$$

Функция $E = E(a,b)$. Продифференцируем E по a и по b. Получим:

$$\frac{\partial E}{\partial a} = -2\sum_{i=1}^{n} (y_i - (a + bx_i)) \ ,$$

$$\frac{\partial E}{\partial b} = -2\sum_{i=1}^{n} (y_i - (a + bx_i))x_i \ .$$

Для того, чтобы найти минимум функции E(a,b), приравняем нулю производные и упростим систему:

$$\begin{cases} -2\sum_{i=1}^{n}(y_i-(a+bx_i))=0, \\ -2\sum_{i=1}^{n}(y_i-(a+bx_i))x_i=0, \end{cases} \Rightarrow \begin{cases} \sum_{i=1}^{n}y_i-na-b\sum_{i=1}^{n}x_i=0 \\ \sum_{i=1}^{n}x_iy_i-a\sum_{i=1}^{n}x_i-b\sum_{i=1}^{n}x^2{}_i=0, \end{cases} \Rightarrow \begin{cases} na+b\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}y_i, \\ a\sum_{i=1}^{n}x_i+b\sum_{i=1}^{n}x^2{}_i=\sum_{i=1}^{n}x_iy_i. \end{cases}$$

Последнюю систему можно представить в матричном виде:

$$\begin{pmatrix} n & \sum_{i=1}^{n}x_i \\ \sum_{i=1}^{n}x_i & \sum_{i=1}^{n}x^2{}_i \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=\begin{pmatrix} \sum_{i=1}^{n}y_i \\ \sum_{i=1}^{n}x_iy_i \end{pmatrix}$$

Решая её получим:

$$b=\frac{n\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{(\sum_{i=1}^{n}x_i)^2+n\sum_{i=1}^{n}x^2{}_i}, \qquad a=\bar{y}-b\bar{x}\ .$$

Вычислив a и b, получим функцию $\hat{y}=a+bx$, которая в классе линейных функций наилучшим образом описывает табличную зависимость в смысле минимума суммы квадратов отклонений. Теперь можно рассчитать и прогноз

$$\hat{y}_{n+1}=a+bx_{n+1}\ .$$