

Обозначим

$$\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N dy_1 dy_2 \dots dy_n = K$$

– величина, не зависящая ни от номера u опыта, ни от характера зависимости $\psi(\mathbf{x}_u)$ отклика от входных переменных. Тогда вероятность получить значения, близкие к наблюдаемым на опыте:

$$P = K \exp\left(-\frac{1}{2\sigma^2} \sum_{u=1}^N [y_u - \psi(\mathbf{x}_u)]^2\right). \quad (5.2)$$

Эта вероятность возрастает вместе с увеличением показателя степени и максимальна, если сумма в показателе достигает минимума:

$$\sum_{u=1}^N [y_u - \psi(\mathbf{x}_u)]^2 = \min. \quad (5.3)$$

Но функция $\psi(\mathbf{x})$ неизвестна, а задача как раз и состоит в нахождении модели $f(\mathbf{x})$, которая была бы близка к ней. Поэтому при сделанных предположениях из принципа максимального правдоподобия следует, что *наилучшей моделью будет такая, для которой сумма квадратов отклонений эмпирических значений y_u от значений, предсказанных моделью, обращается в минимум:*

$$S(y_u, f, \mathbf{x}_u) = \sum_{u=1}^N [y_u - f(\mathbf{x}_u)]^2 = \min. \quad (5.4)$$

Методом наименьших квадратов называют метод отыскания параметров модели, который обеспечивает обращение в минимум суммы квадратов отклонений (невязок) наблюдаемых и предсказанных значений.

Метод наименьших квадратов является важнейшим методом обработки эмпирической информации. Он привлекает минимум дополнительных предположений о природе опытных данных и в силу этого имеет высокую общность. Обычно этот метод принято рассматривать в курсе анализа, в разделе теории функций нескольких переменных. При этом метод излагается в терминах анализа применительно к узкому кругу приложений (как правило – применительно к отысканию параметров линейной регрессии). Такое изложение является недостаточным для практики, так как причины выбора критерия близости функции $f(\mathbf{x})$ к множеству экспериментальных точек $\{(\mathbf{x}, y)\}$ остаются нераскрытыми.

5.3 Однофакторная линейная регрессия

Пусть в процессе исследования варьировалась одна независимая переменная x и после проведения N экспериментов получены значения y_u , $u = \overline{1, N}$. Требуется методом наименьших квадратов подобрать параметры линейной экспериментально-статистической модели

$$y = ax + b.$$

Для модели указанного вида сумма квадратов отклонений будет равна:

$$S = \sum_{u=1}^N (y_u - (ax_u + b))^2 .$$

Считая эту сумму функцией неизвестных параметров $S = S(a, b)$, потребуем выполнения необходимого условия локального экстремума:

$$\begin{cases} \frac{\partial S}{\partial a} = 0 \\ \frac{\partial S}{\partial b} = 0 \end{cases} .$$

Дифференцируя и приравнявая частные производные к нулю, получим:

$$\begin{cases} \sum_{u=1}^N (y_u - ax_u - b)x_u = 0 \\ \sum_{u=1}^N (y_u - ax_u - b) = 0 \end{cases} .$$

Изменим порядок суммирования:

$$\begin{cases} a \sum_{u=1}^N x_u^2 + b \sum_{u=1}^N x_u = \sum_{u=1}^N y_u x_u \\ a \sum_{u=1}^N x_u + bN = \sum_{u=1}^N y_u \end{cases} .$$

Если все значения x_u различны, то полученная система двух линейных уравнений (которую называют *нормальной системой*) имеет единственное решение. Можно доказать, что это решение действительно соответствует точке локального минимума функции $S = S(a, b)$.

5.4 Регрессионный анализ: линейная по параметрам модель

Нормальная система, возникающая в процессе применения метода наименьших квадратов, определяется видом регрессионной модели $y = f(\mathbf{x})$. В большинстве случаев нормальная система является нелинейной и решается численно.

Существует достаточно широкий класс практически важных моделей, для которых нормальная система является линейной и допускает простое и компактное представление. Этот класс представлен *моделями, линейными по параметрам*:

$$f(\mathbf{x}) = b_1 \varphi_1(\mathbf{x}) + b_2 \varphi_2(\mathbf{x}) + \dots + b_L \varphi_L(\mathbf{x}) .$$

В этих моделях функции $\varphi_j(\mathbf{x})$, $j = \overline{1, L}$ называются *базисными функциями*. Для линейной по параметрам модели

$$f(\mathbf{x}) = \sum_{j=1}^L b_j \varphi_j(\mathbf{x})$$

сумма квадратов отклонений предсказанных $f(\mathbf{x}_u)$ и экспериментальных значений y_u имеет вид

$$S(b_1, b_2, \dots, b_L) = \sum_{u=1}^N (y_u - f(\mathbf{x}_u))^2 = \sum_{u=1}^N \left(y_u - \sum_{j=1}^L b_j \varphi_j(\mathbf{x}_u) \right)^2.$$

Считая эту сумму функцией от L неизвестных параметров, потребуем выполнения необходимого условия локального экстремума

$$\sum_{u=1}^N \left[\left(y_u - \sum_{j=1}^L b_j \varphi_j(\mathbf{x}_u) \right) \left(\frac{\partial}{\partial b_i} \sum_{j=1}^L b_j \varphi_j(\mathbf{x}_u) \right) \right] = 0, \quad i = \overline{1, L}.$$

Базисные функции не зависят от параметров, поэтому при всех $i \neq j$ производные

$$\frac{\partial}{\partial b_i} b_j \varphi_j(\mathbf{x}_u)$$

равны нулю. Следовательно, в каждой из сумм $\frac{\partial}{\partial b_i} \sum_{j=1}^L b_j \varphi_j(\mathbf{x}_u)$ имеется только одно ненулевое слагаемое:

$$\frac{\partial}{\partial b_i} \sum_{j=1}^L b_j \varphi_j(\mathbf{x}_u) = \varphi_i(\mathbf{x}_u).$$

Нормальная система принимает вид

$$\sum_{u=1}^N \left[\left(y_u - \sum_{j=1}^L b_j \varphi_j(\mathbf{x}_u) \right) \varphi_i(\mathbf{x}_u) \right] = 0, \quad i = \overline{1, L}.$$

Изменим порядок суммирования и перенесем в правую часть слагаемые, в которые входят эмпирические значения отклика:

$$\sum_{u=1}^N \left(\varphi_i(\mathbf{x}_u) \sum_{j=1}^L b_j \varphi_j(\mathbf{x}_u) \right) = \sum_{u=1}^N y_u \varphi_i(\mathbf{x}_u), \quad i = \overline{1, L}.$$

Вновь меняя в левой части порядок суммирования, запишем нормальную систему в виде

$$\sum_{j=1}^L \left(b_j \sum_{u=1}^N \varphi_i(\mathbf{x}_u) \varphi_j(\mathbf{x}_u) \right) = \sum_{u=1}^N y_u \varphi_i(\mathbf{x}_u), \quad i = \overline{1, L}. \quad (5.5)$$

Введем обозначения:

$$\mathbf{X} = \begin{pmatrix} \varphi_1(\mathbf{x}_1) & \varphi_2(\mathbf{x}_1) & \dots & \varphi_L(\mathbf{x}_1) \\ \varphi_1(\mathbf{x}_2) & \varphi_2(\mathbf{x}_2) & \dots & \varphi_L(\mathbf{x}_2) \\ \dots & \dots & \dots & \dots \\ \varphi_1(\mathbf{x}_N) & \varphi_2(\mathbf{x}_N) & \dots & \varphi_L(\mathbf{x}_N) \end{pmatrix} -$$

матрица размера $N \times L$, называемая *матрицей базисных функций*;

$$\mathbf{V} = (b_1, b_2, \dots, b_L)^T -$$

вектор-столбец высоты L , называемый *вектором искомых параметров*;

$$\mathbf{Y} = (y_1, y_2, \dots, y_N)^T -$$

вектор-столбец высоты N , называемый *вектором откликов*.

Тогда нормальную систему (5.5) можно записать в матричной форме:

$$(\mathbf{X}^T \mathbf{X})\mathbf{B} = \mathbf{X}^T \mathbf{Y}. \quad (5.6)$$

Входящая в эту систему симметрическая квадратная матрица

$$\mathbf{M} = \mathbf{X}^T \mathbf{X} = \begin{pmatrix} \varphi_1(\mathbf{x}_1) & \varphi_1(\mathbf{x}_2) & \dots & \varphi_1(\mathbf{x}_N) \\ \varphi_2(\mathbf{x}_1) & \varphi_2(\mathbf{x}_2) & \dots & \varphi_2(\mathbf{x}_N) \\ \dots & \dots & \dots & \dots \\ \varphi_L(\mathbf{x}_1) & \varphi_L(\mathbf{x}_2) & \dots & \varphi_L(\mathbf{x}_N) \end{pmatrix} \begin{pmatrix} \varphi_1(\mathbf{x}_1) & \varphi_2(\mathbf{x}_1) & \dots & \varphi_L(\mathbf{x}_1) \\ \varphi_1(\mathbf{x}_2) & \varphi_2(\mathbf{x}_2) & \dots & \varphi_L(\mathbf{x}_2) \\ \dots & \dots & \dots & \dots \\ \varphi_1(\mathbf{x}_N) & \varphi_2(\mathbf{x}_N) & \dots & \varphi_L(\mathbf{x}_N) \end{pmatrix} =$$

$$= \begin{pmatrix} \sum_{u=1}^N \varphi_1^2(\mathbf{x}_u) & \sum_{u=1}^N \varphi_1(\mathbf{x}_u)\varphi_2(\mathbf{x}_u) & \dots & \sum_{u=1}^N \varphi_1(\mathbf{x}_u)\varphi_L(\mathbf{x}_u) \\ \sum_{u=1}^N \varphi_2(\mathbf{x}_u)\varphi_1(\mathbf{x}_u) & \sum_{u=1}^N \varphi_2^2(\mathbf{x}_u) & \dots & \sum_{u=1}^N \varphi_2(\mathbf{x}_u)\varphi_L(\mathbf{x}_u) \\ \dots & \dots & \dots & \dots \\ \sum_{u=1}^N \varphi_L(\mathbf{x}_u)\varphi_1(\mathbf{x}_u) & \sum_{u=1}^N \varphi_L(\mathbf{x}_u)\varphi_2(\mathbf{x}_u) & \dots & \sum_{u=1}^N \varphi_L^2(\mathbf{x}_u) \end{pmatrix},$$

порядок которой совпадает с числом базисных функций (и с числом слагаемых регрессионной модели), называется *матрицей моментов*.

Рассмотренная ранее одномерная линейная регрессия

$$y = b + ax$$

является частным случаем модели, линейной по параметрам. Для этой модели базисные функции

$$\varphi_1(x) = 1, \quad \varphi_2(x) = x,$$

поэтому ее матрица базисных функций (имеющая N строк и два столбца):

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_N \end{pmatrix}.$$

Матрица моментов одномерной линейной регрессии:

$$\mathbf{M} = \mathbf{X}^T \mathbf{X} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_N \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_N \end{pmatrix} = \begin{pmatrix} N & \sum_{u=1}^N x_u \\ \sum_{u=1}^N x_u & \sum_{u=1}^N x_u^2 \end{pmatrix};$$

нетрудно видеть, что элементами этой матрицы являются именно те суммы, которые входят в левые части записанной ранее нормальной системы.

Матрица, обратная к матрице моментов,

$$\mathbf{D} = \mathbf{M}^{-1} = (\mathbf{X}^T \mathbf{X})^{-1}$$

называется *ковариационной матрицей*, или *матрицей ошибок*. Искомый столбец параметров равен

$$\mathbf{B} = \mathbf{DX}^T \mathbf{Y}. \quad (5.7)$$

Применение соотношения (5.7) приводит к L -кратному увеличению вычислительных затрат по сравнению с (5.6). Несмотря на это для нахождения параметров целесообразно использовать именно соотношение (5.7). Это связано с тем, что диагональные элементы ковариационной матрицы характеризуют дисперсии параметров модели, а внедиагональные – их взаимное влияние. Ковариационная матрица требуется на этапе статистического анализа построенной ЭС-модели.

Для рассмотренной ранее одномерной линейной регрессии ковариационная матрица равна

$$\mathbf{D} = \begin{pmatrix} N & \sum_{u=1}^N x_u \\ \sum_{u=1}^N x_u & \sum_{u=1}^N x_u^2 \end{pmatrix}^{-1} = \frac{1}{N \sum_{u=1}^N x_u^2 - \sum_{u=1}^N x_u} \begin{pmatrix} \sum_{u=1}^N x_u^2 & -\sum_{u=1}^N x_u \\ -\sum_{u=1}^N x_u & N \end{pmatrix}.$$

Взаимное влияние параметров будет минимальным, если внедиагональные элементы обратятся в ноль; получаем условие

$$\sum_{u=1}^N x_u = 0. \quad (5.8)$$

При проведении *активного эксперимента* исследователь почти всегда имеет возможность выбрать значения входных переменных так, чтобы обеспечить выполнение условий, подобных (5.8). В этом случае говорят, что *план эксперимента* в том или ином смысле *оптимален*.

Если для линейной регрессионной модели условие (5.8) выполнено, то ее ковариационная матрица

$$\mathbf{D} = \frac{1}{N \sum_{u=1}^N x_u^2} \begin{pmatrix} \sum_{u=1}^N x_u^2 & 0 \\ 0 & N \end{pmatrix} = \begin{pmatrix} \frac{1}{N} & 0 \\ 0 & \left(\sum_{u=1}^N x_u^2 \right)^{-1} \end{pmatrix},$$

поэтому дисперсии параметров модели пропорциональны величинам

$$D_a = \left(\sum_{u=1}^N x_u^2 \right)^{-1}, \quad D_b = \frac{1}{N}.$$

Как и следовало ожидать, дисперсия свободного члена b обратно пропорциональна числу опытов. Дисперсия коэффициента a при входной переменной уменьшается вместе с возрастанием числа опытов и увеличением абсолютных величин тех значений входной переменной, для которых измеряются значения отклика.