Тема 6. Введение в регрессионный анализ.

1. Теоретическая и практическая задачи регрессионного анализа.

Основная задача регрессионного анализа заключается в исследовании зависимости между случайной величиной $\eta(\omega)$ и совокупностью случайных величин $(\xi_1(\omega),..., \xi_k(\omega))$.

Теоретическая задача регрессионного анализа.

В процессе анализа одним из методов строится некоторая случайная величина $\hat{\eta}(\omega)$, зависящая только от $(\xi_1(\omega),..., \xi_k(\omega))$, которая сравнивается с исходной величиной $\eta(\omega)$.

Условие «зависящая только от $(\xi_1(\omega),..., \xi_k(\omega))$» означает, что случайная величина $\hat{\eta}(\omega)$ не является самостоятельной, а является детерминированной функцией от случайных величин $(\xi_1(\omega),..., \xi_k(\omega))$:

$$\hat{\eta}(\omega) = \hat{\eta}(\xi_1(\omega),..., \xi_k(\omega)) .$$

Пусть на некотором множестве элементарных событий A вектор случайных величин $(\xi_1(\omega),..., \xi_k(\omega))$ принимает одно и тоже значение $(x_1,..., x_k)$:

$$\forall \omega \in A : (\xi_1(\omega),..., \xi_k(\omega)) = (x_1,..., x_k) ,$$

тогда на множестве A случайная величина $\hat{\eta}(\omega) = \hat{\eta}(\xi_1(\omega),..., \xi_k(\omega))$ может принимать только одно постоянное значение $\hat{\eta}(\omega) = \hat{\eta}(x_1,..., x_k) = y$, а случайная величина $\eta(\omega)$ может принимать совершенно произвольные значения. Если диапазон значений $\eta(\omega)$ на множестве A окажется широким, то каким бы образом ни выбиралась постоянная y «расхождение» между величиной $\eta(\omega)$ и величиной $\hat{\eta}(\omega) = y$ на множестве A окажется значительным. Напротив, если диапазон значений $\eta(\omega)$ на множестве A окажется узким, то, подобрав подходящее значение y, можно сделать «расхождение» между $\eta(\omega)$ и $\hat{\eta}(\omega) = y$ незначительным.

Если между величиной $\eta(\omega)$ и величинами $(\xi_1(\omega),..., \xi_k(\omega))$ имеется «сильная вероятностная зависимость», тогда на множествах, где вектор $(\xi_1(\omega),..., \xi_k(\omega))$ принимает постоянные значения, величина $\eta(\omega)$ изменяется незначительно (по заданным значениям $(\xi_1(\omega),..., \xi_k(\omega))$ можно с хорошим приближением указать значение $\eta(\omega)$), тогда, выбирая подходящим образом постоянные значения для величины $\hat{\eta}(\omega)$ на этих множествах, можно сделать «суммарное расхождение» между $\eta(\omega)$ и $\hat{\eta}(\omega)$ малым.

Если между величиной $\eta(\omega)$ и величинами $(\xi_1(\omega),..., \xi_k(\omega))$ имеется лишь «слабая вероятностная зависимость», тогда на множествах, где вектор $(\xi_1(\omega),..., \xi_k(\omega))$ принимает постоянные значения, величина $\eta(\omega)$ изменяется значительно (по заданным значениям $(\xi_1(\omega),..., \xi_k(\omega))$ невозможно предугадать значение $\eta(\omega)$), тогда каким бы образом ни выбирались значения для $\hat{\eta}(\omega)$ «суммарное расхождение» между $\eta(\omega)$ и $\hat{\eta}(\omega)$ нельзя сделать малым.

Пусть «суммарное расхождение» между величинами $\eta(\omega)$ и $\hat{\eta}(\omega)$ измеряется среднеквадратичным отклонением $\delta(\eta, \hat{\eta})$:

$$\delta(\eta, \hat{\eta}) = M[|\eta - \hat{\eta}|^2] ,$$

и требуется найти такую величину $\hat{\eta}(\omega) = \hat{\eta}(\xi_1(\omega),..., \xi_k(\omega))$, которая доставляет минимальное значение $\delta(\eta, \hat{\eta})$. Известно, что такой величиной является условное математическое ожидание $M[\eta \mid \xi_1,..., \xi_k]$ случайной величиной η относительно случайных величин ξ_1, …, ξ_k.

Определение 6.1.

Регрессией случайной величины $\eta(\omega)$ *по* $\xi_1(\omega), \ldots, \xi_k(\omega)$ называется условное математическое ожидание $M[\eta \mid \xi_1,\ldots, \xi_k](\omega)$.

Условное математическое ожидание является случайной величиной особого вида, для построения которой требуется совместное вероятностное распределение случайного вектора $(\eta(\omega), \xi_1(\omega),\ldots, \xi_k(\omega))$. В общем случае, нахождение условного математического ожидания может представлять сложную задачу, тем не менее, в некоторых частных простых случаях можно получить явное выражение для условного математического ожидания. Например, если $k = 1$ и вектор $(\eta(\omega), \xi_1(\omega))$ имеет нормальное распределение, тогда:

$$M[\eta \mid \xi_1](\omega) = M[\eta] + \frac{\operatorname{cov}(\eta, \xi_1)}{D[\xi_1]}(\xi_1(\omega) - M[\xi_1]) .$$

Практическая задача регрессионного анализа.

На практике совестное вероятностное распределение $(\eta(\omega), \xi_1(\omega),\ldots, \xi_k(\omega))$ редко бывает известно, тем не менее, в некоторых случаях можно вполне обоснованно предположить, что условное математическое ожидание является функцией определенного вида:

$$\hat{\eta}(\omega) = M[\eta \mid \xi_1,\ldots, \xi_k](\omega) = f(\xi_1(\omega),\ldots, \xi_k(\omega); \tilde{\theta}_1,\ldots, \tilde{\theta}_m),$$

где $\tilde{\theta} = (\tilde{\theta}_1,\ldots, \tilde{\theta}_m)$ – некоторый неизвестный числовой вектор.

Пусть случайная величина $\varepsilon(\omega) = \eta(\omega) - \hat{\eta}(\omega)$, тогда:

$$\eta(\omega) = \hat{\eta}(\omega) + \varepsilon(\omega),$$

$$\eta(\omega) = f(\xi_1(\omega),\ldots, \xi_k(\omega); \tilde{\theta}_1,\ldots, \tilde{\theta}_m) + \varepsilon(\omega) .$$

В качестве исходных данных для задачи построения регрессии выступает совокупность наблюдаемых величин $\eta = (\eta_1,\ldots, \eta_n)$, в котором каждая случайная величина η_i получена при условии, что $\xi_1 = x_{i1}$, …, $\xi_k = x_{ik}$:

$$\eta_1 = f(x_{11} \ldots, x_{1k}; \tilde{\theta}_1,\ldots, \tilde{\theta}_m) + \varepsilon_1,$$

$$\ldots$$

$$\eta_n = f(x_{n1} \ldots, x_{nk}; \tilde{\theta}_1,\ldots, \tilde{\theta}_m) + \varepsilon_n,$$

где $\eta = (\eta_1,\ldots, \eta_n)$ – векторная случайная величина, x_{ij} – числовые значения ($i = \overline{1, n}$, $j = \overline{1, k}$), $\varepsilon = (\varepsilon_1,\ldots, \varepsilon_n)$ – векторная случайная величина. Введем обозначение X для матрицы числовых значений x_{ij}: $\|X\|_{ij} = x_{ij}$.

Задача построения регрессии заключается в нахождении оценки неизвестного вектора параметров $\tilde{\theta} = (\tilde{\theta}_1,\ldots, \tilde{\theta}_m)$ на основе наблюдаемого вектора $\eta = (\eta_1,\ldots, \eta_n)$, например, в нахождении такого вектора параметров $\hat{\theta} = (\hat{\theta}_1,\ldots, \hat{\theta}_m)$, при котором среднеквадратичное отклонение $\delta(\eta, f(X, \theta))$:

$$\delta(\eta, f(X; \theta)) = \|\eta - f(X; \theta)\|_2^2$$

$$f(X; \theta) = \begin{pmatrix} f(x_{11},\ldots, x_{1k}; \theta_1,\ldots, \theta_m) \\ \ldots \\ f(x_{n1},\ldots, x_{nk}; \theta_1,\ldots, \theta_m) \end{pmatrix}.$$

принимает наименьшее возможное значение:

$$\delta(\eta, f(X; \hat{\theta})) = \min_{\theta} \delta(\eta, f(X; \theta)) .$$

Вектор $\hat{\theta}$ в общем случае зависит от вектора случайных величин η, поэтому $\hat{\theta}$, вообще говоря, является некоторой векторной случайной величиной.

2. Задача линейной регрессии и оценка по методу наименьших квадратов.

Задача линейной регрессии является частным случаем практической задачи, изложенной в предыдущем пункте, в котором априорно известно (или предполагается), что условное математическое ожидание $M[\eta \mid \xi_1,..., \xi_k](\omega)$ является функцией линейно зависящей от параметра $\tilde{\theta} = (\tilde{\theta}_1,..., \tilde{\theta}_m)$:

$$M[\eta \mid \xi_1,..., \xi_k] = f(\xi_1,..., \xi_k; \tilde{\theta}_1,..., \tilde{\theta}_m) =$$
$$= \varphi_1(\xi_1,..., \xi_k)\tilde{\theta}_1 + ... + \varphi_m(\xi_1,..., \xi_k)\tilde{\theta}_m = \sum_{j=1}^{m} \varphi_j(\xi_1,..., \xi_k)\tilde{\theta}_j \qquad (6.1)$$

где $\tilde{\theta} = (\tilde{\theta}_1,..., \tilde{\theta}_m)$ – фиксированный, но неизвестный числовой вектор, и $\varphi_i(x_1,..., x_k)$ – известные детерминированные функции ($i = \overline{1, m}$). В этом важном частном случае может быть получено уравнение для нахождения вектора параметров $\hat{\theta} = (\hat{\theta}_1,..., \hat{\theta}_m)$, при котором среднеквадратичное отклонение $\delta(\eta, f(X;\theta))$ принимает наименьшее значение.

Из (6.1) следует, что:

$$f(X;\theta) = \begin{pmatrix} f(x_{11},..., x_{1k};\theta) \\ ... \\ f(x_{n1},..., x_{nk};\theta) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{m} \varphi_i(x_{11},..., x_{1k})\theta_i \\ ... \\ \sum_{i=1}^{m} \varphi_i(x_{n1},..., x_{nk})\theta_i \end{pmatrix} = Z\theta, \qquad (6.2)$$

где $Z = \begin{pmatrix} \varphi_1(x_{11},..., x_{1k}) & ... & \varphi_m(x_{11},..., x_{1k}) \\ ... & ... & ... \\ \varphi_1(x_{n1},..., x_{nk}) & ... & \varphi_m(x_{n1},..., x_{nk}) \end{pmatrix} = \begin{pmatrix} \varphi_1(x^{(1)}) & ... & \varphi_m(x^{(1)}) \\ ... & ... & ... \\ \varphi_1(x^{(n)}) & ... & \varphi_m(x^{(n)}) \end{pmatrix}$,

$x^{(i)} = (x_{i1},..., x_{ik})$ – i-ая строка матрицы X, и среднеквадратичное отклонение принимает вид:

$$\delta(\eta, f(X;\theta)) = \|\eta - f(X;\theta)\|_2^2 = \|\eta - Z\theta\|_2^2 = \sum_{i=1}^{n} \left(\eta_i - \sum_{j=1}^{m} \phi_j(x^{(i)})\theta_j\right)^2.$$

Решение задачи нахождения вектора $\hat{\theta} = (\hat{\theta}_1,..., \hat{\theta}_m)$, при котором среднеквадратичное отклонение принимает наименьшее значение, дается методом наименьших квадратов, а сам вектор $\hat{\theta} = (\hat{\theta}_1,..., \hat{\theta}_m)$ называют *оценкой по методу наименьших квадратов*.

Необходимое условие экстремума $\delta(\eta, f(X;\theta))$ как функции параметра $\theta = (\theta_1,..., \theta_m)$ заключается в равенстве частных производных $\delta(\eta, f(X;\theta))$ нулю в точке $\hat{\theta} = (\hat{\theta}_1,..., \hat{\theta}_m)$:

$$\left.\frac{\partial \delta(\eta, f(X;\theta))}{\partial \theta_l}\right|_{\theta=\hat{\theta}} = 0, \; l = \overline{1, m}.$$

Вычислим частные производные:

$$\left.\frac{\partial \delta(\eta, f(X;\theta))}{\partial \theta_l}\right|_{\theta=\hat{\theta}} = \sum_{i=1}^{k} 2\left(\eta_i - \sum_{j=1}^{m} \varphi_j(x^{(i)})\hat{\theta}_j\right)\varphi_l(x^{(i)}) = 0, \; l = \overline{1, m},$$

$$\sum_{i=1}^{k} \varphi_l(x^{(i)})\left(\eta_i - \sum_{j=1}^{m} \varphi_j(x^{(i)})\hat{\theta}_j\right) = 0, \; l = \overline{1, m},$$

$$Z^T(\eta - Z\hat{\theta}) = 0,$$

$$Z^T Z\hat{\theta} = Z^T \eta \qquad (6.3)$$

Полученное уравнение называется *нормальным уравнением* метода наименьших квадратов. Нормальное уравнение играет роль необходимого условия: «если $\delta(\eta, f(X;\theta))$

принимает наименьшее значение в точке $\hat{\theta}$, тогда $\hat{\theta}$ удовлетворяет нормальному уравнению $Z^T Z\hat{\theta} = Z^T \eta$ ». Отсюда следует, что среди всех решений нормального уравнения есть решения, доставляющие минимальное значение среднеквадратичному отклонению. Однако, отсюда еще не следует, что любое решение нормального уравнения будет доставлять минимальное значение среднеквадратичному отклонению. Тем не менее, оказывается, что это действительно так, необходимое условие в некоторых случаях является достаточным: «если $\hat{\theta}$ удовлетворяет нормальному уравнению $Z^T Z\hat{\theta} = Z^T \eta$, тогда $\delta(\eta, f(X;\theta))$ принимает наименьшее значение в точке $\hat{\theta}$», доказательству этого факта посвящено утверждение 6.3, основанное на утверждении 6.2.

Утверждение 6.2.

Для матрицы Z порядка $n \times m$:

1) матрица $Z^T Z$ неотрицательно определена, $Z^T Z \geq 0$;

2) если ранг матрицы Z равен m, тогда матрица $Z^T Z$ положительно определена, $Z^T Z > 0$;

3) если ранг матрицы Z равен m, тогда матрица $| Z^T Z | \neq 0$.

Доказательство:

1) Пусть x – произвольный вектор столбец порядка m, тогда:

$$x^T Z^T Zx = (Zx)^T Zx = \|Zx\|_2^2 \geq 0 .$$

2) Согласно первому пункту имеет место неотрицательная определенность $Z^T Z \geq 0$, покажем, что если ранг матрицы Z равен m, то имеет место положительная определенность $Z^T Z > 0$, для этого покажем, что для всех векторов столбцов $x \neq \overline{0}$ порядка m справедливо $x^T Z^T Zx \neq 0$.

Допустим противное, пусть существует вектор $x \neq \overline{0}$ такой, что $x^T Z^T Zx = 0$, тогда:

$$0 = x^T Z^T Zx = \|Zx\|_2^2 .$$

По определению нормы равенство $\|Zx\|_2^2 = 0$ возможно тогда и только тогда, когда $Zx = \overline{0}$, но тогда столбцы матрицы Z линейно зависимы и следовательно ранг матрицы Z не может быть равен m.

3) Пусть ранг матрицы Z равен m, тогда из уже доказанного пункта 2 матрица $Z^T Z$ является положительно определенной. Покажем, что отсюда следует $| Z^T Z | \neq 0$ от противного: предположим, что $| Z^T Z | = 0$, тогда существует нетривиальное решение $x \neq \overline{0}$ однородной системы:

$$Z^T Zx = 0 .$$

Отсюда следует, что:

$$x^T Z^T Zx = x^T (Z^T Zx) = x^T 0 = 0 ,$$

$$x^T Z^T Zx = 0 ,$$

что противоречит положительной определенности матрицы $Z^T Z$.

Утверждение доказано.

Утверждение 6.3.

Пусть η – вектор порядка n и Z – матрица порядка $n \times m$, где $n \geq m$, если $\hat{\theta}$ – решение нормального уравнения:

$$Z^T Z\hat{\theta} = Z^T \eta ,$$

тогда,

$$\delta(\eta, f(X;\hat{\theta})) = \min_{\theta} \delta(\eta, f(X;\theta)) .$$

Если дополнительно ранг матрицы Z равен m, тогда решение нормального уравнения $\hat{\theta}$ единственно, дается равенством:

$$\hat{\theta} = (Z^T Z)^{-1} Z^T \eta$$

и является единственным вектором, при котором $\delta(\eta, f(X;\theta))$ принимает наименьшее значение.

Доказательство:

1) Преобразуем выражение для $\delta(\eta, f(X;\theta))$:

$$\delta(\eta, f(X;\theta)) = \left\|\eta - Z\hat{\theta} + Z\hat{\theta} - Z\theta\right\|_2^2 = (\eta - Z\hat{\theta} + Z\hat{\theta} - Z\theta)^T (\eta - Z\hat{\theta} + Z\hat{\theta} - Z\theta) =$$

$$= (\eta - Z\hat{\theta})^T (\eta - Z\hat{\theta}) + (Z\hat{\theta} - Z\theta)^T (\eta - Z\hat{\theta}) + (\eta - Z\hat{\theta})^T (Z\hat{\theta} - Z\theta) + (Z\hat{\theta} - Z\theta)^T (Z\hat{\theta} - Z\theta) =$$

$$= \left\|(\eta - Z\hat{\theta})\right\|_2^2 + (\hat{\theta} - \theta)^T Z^T (\eta - Z\hat{\theta}) + (\eta - Z\hat{\theta})^T Z(\hat{\theta} - \theta) + (\hat{\theta} - \theta)^T Z^T Z(\hat{\theta} - \theta) =$$

$$= \left\|(\eta - Z\hat{\theta})\right\|_2^2 + (\hat{\theta} - \theta)^T (Z^T\eta - Z^T Z\hat{\theta}) + (Z^T\eta - Z^T Z\hat{\theta})^T (\hat{\theta} - \theta) + (\hat{\theta} - \theta)^T Z^T Z(\hat{\theta} - \theta).$$

Поскольку $\hat{\theta}$ является решением нормального уравнения, то $Z^T Z\hat{\theta} - Z^T\eta = 0$, тогда:

$$\delta(\eta, f(X;\theta)) = \left\|(\eta - Z\hat{\theta})\right\|_2^2 + (\hat{\theta} - \theta)^T Z^T Z(\hat{\theta} - \theta) = \delta(\eta, f(X,\hat{\theta})) + (\hat{\theta} - \theta)^T Z^T Z(\hat{\theta} - \theta).$$

Согласно утверждению 6.2 матрица $Z^T Z$ всегда неотрицательно определена, поэтому для любого вектора, в том числе и для $\hat{\theta} - \theta$, выражение $(\hat{\theta} - \theta)^T Z^T Z(\hat{\theta} - \theta) \geq 0$, отсюда:

$$\delta(\eta, f(X;\theta)) = \delta(\eta, f(X;\hat{\theta})) + (\hat{\theta} - \theta)^T Z^T Z(\hat{\theta} - \theta) \geq \delta(\eta, f(X;\hat{\theta})).$$

Таким образом, для произвольного θ:

$$\delta(\eta, f(X;\theta)) \geq \delta(\eta, f(X;\hat{\theta})),$$

Тогда,

$$\delta(\eta, f(X;\hat{\theta})) = \min_{\theta} \delta(\eta, f(X;\theta)).$$

2) Пусть ранг матрицы Z равен m, тогда согласно утверждению 6.2 пункт 3 $\left|Z^T Z\right| \neq 0$. Поскольку $\left|Z^T Z\right| \neq 0$, то, как известно, решение системы $Z^T Z\hat{\theta} = Z^T\eta$ с невырожденной матрицей $Z^T Z$ существует и единственно, и существует обратная матрица $(Z^T Z)^{-1}$, тогда, умножая равенство $Z^T Z\hat{\theta} = Z^T\eta$ слева на $(Z^T Z)^{-1}$, получим:

$$\hat{\theta} = (Z^T Z)^{-1} Z^T \eta.$$

Пусть ранг матрицы Z равен m, тогда согласно утверждению 6.2 пункт 2 $Z^T Z > 0$, тогда на основании пункта 1 для любого вектора $\theta \neq \hat{\theta}$ значение $(\hat{\theta} - \theta)^T Z^T Z(\hat{\theta} - \theta) > 0$, тогда:

$$\delta(\eta, f(X;\theta)) > \delta(\eta, f(X;\hat{\theta})),$$

то есть $\hat{\theta}$ является единственным вектором, при котором $\delta(\eta, f(X;\theta))$ принимает наименьшее значение.

Утверждение доказано.

3. Свойства оценки по методу наименьших квадратов в задаче линейной регрессии.

Пусть известно, что выполняется (6.1), тогда вектор $\eta = (\eta_1, \ldots, \eta_n)$ имеет вид:

$$\begin{aligned} \eta_1 &= \varphi_1(x_{11}, \ldots, x_{1k})\tilde{\theta}_1 + \ldots + \varphi_m(x_{11}, \ldots, x_{1k})\tilde{\theta}_m + \varepsilon_1, \\ &\ldots, \\ \eta_n &= \varphi_1(x_{n1}, \ldots, x_{nk})\tilde{\theta}_1 + \ldots + \varphi_m(x_{n1}, \ldots, x_{nk})\tilde{\theta}_m + \varepsilon_n, \end{aligned} \tag{6.4}$$

Будем дополнительно предполагать, что случайные величины ε_1, …, ε_n имеют следующие свойства:

$$
\begin{aligned}
&1)\ M[\varepsilon_i] = 0,\ i = \overline{1, n};\\
&2)\ D[\varepsilon_i] = \sigma^2 > 0,\ i = \overline{1, n};\\
&3)\ \mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0,\ i \neq j,\ i, j = \overline{1, n}.
\end{aligned}
\tag{6.5}
$$

Предположим, что на основании результатов предыдущего пункта получена оценка по методу наименьших квадратов $\hat{\theta} = (\hat{\theta}_1,..., \hat{\theta}_m)$. Оценка $\hat{\theta}$, зависящая от наблюдаемого вектора $\eta = (\eta_1,..., \eta_n)$, является векторной случайной величиной, поэтому возникает вопрос о свойствах оценки $\hat{\theta}$ и о том, каким образом связаны между собой оценка $\hat{\theta}$ и неизвестный вектор параметров $\tilde{\theta}$.

Теорема 6.4.

Если в задаче линейной регрессии (6.1)-(6.5) ранг матрицы Z равен m, тогда:

1) оценка $\hat{\theta}$ является несмещенной оценкой $\tilde{\theta}$: $M[\hat{\theta}] = \tilde{\theta}$;

2) дисперсионная матрица $D[\hat{\theta}]$ оценки $\hat{\theta}$ имеет вид:

$$D[\hat{\theta}] = \sigma^2 (Z^T Z)^{-1};$$

3) если $\theta^* = (\theta_1^*,..., \theta_m^*)$ несмещенная и линейная по η оценка $\tilde{\theta}$, тогда:

$$D[\hat{\theta}_i] \leq D[\theta_i^*],$$
$$i = \overline{1, m}.$$

то есть компоненты $\hat{\theta}_i$ в оценке $\hat{\theta} = (\hat{\theta}_1,..., \hat{\theta}_m)$ имеют наименьшую дисперсию среди всех несмещенных и линейных по η оценок $\tilde{\theta}_i$.

Доказательство:

1) Поскольку ранг матрицы Z равен m, то из утверждения 6.2 пункт 3 $\left|Z^T Z\right| \neq 0$, тогда существует обратная матрица $(Z^T Z)^{-1}$, следовательно, из (6.3) при умножении левой и правой частей на $(Z^T Z)^{-1}$:

$$\hat{\theta} = (Z^T Z)^{-1} Z^T \eta .$$

Из (6.2) и (6.4) следует, что $\eta = Z\tilde{\theta} + \varepsilon$, тогда:

$$\hat{\theta} = (Z^T Z)^{-1} Z^T (Z\tilde{\theta} + \varepsilon) = (Z^T Z)^{-1} Z^T Z\tilde{\theta} + (Z^T Z)^{-1} Z^T \varepsilon = \tilde{\theta} + (Z^T Z)^{-1} Z^T \varepsilon .$$

Легко видеть, что математическое ожидание $M[\hat{\theta}]$:

$$M[\hat{\theta}] = M[\tilde{\theta} + (Z^T Z)^{-1} Z^T \varepsilon] = \tilde{\theta} + (Z^T Z)^{-1} Z^T M[\varepsilon] = \tilde{\theta} + (Z^T Z)^{-1} Z^T \overline{0}_n = \tilde{\theta},$$

поскольку из (6.5) пункт 1 следует $M[\varepsilon] = \overline{0}_n$.

2) Дисперсионная матрица $D[\hat{\theta}]$:

$$D[\hat{\theta}] = D[\tilde{\theta} + (Z^T Z)^{-1} Z^T \varepsilon] = D[(Z^T Z)^{-1} Z^T \varepsilon].$$

Согласно свойству дисперсионной матрицы $D[C\varepsilon] = CD[\varepsilon]C^T$, тогда:

$$D[\hat{\theta}] = (Z^T Z)^{-1} Z^T D[\varepsilon]((Z^T Z)^{-1} Z^T)^T = (Z^T Z)^{-1} Z^T \sigma^2 E_n Z((Z^T Z)^{-1})^T =$$
$$= \sigma^2 (Z^T Z)^{-1} Z^T Z ((Z^T Z)^{-1})^T = \sigma^2 ((Z^T Z)^{-1})^T,$$

где $D[\varepsilon] = \sigma^2 E_n$, поскольку в силу условия (6.5) пункты 2 и 3 диагональные элементы $\|D[\varepsilon]\|_{ii} = D[\varepsilon_i] = \sigma^2$ ($i = \overline{1, n}$), а остальные элементы $\|D[\varepsilon]\|_{ij} = \mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$ ($i \neq j,\ i, j = \overline{1, n}$).

Поскольку $Z^T Z$ – симметричная матрица, то обратная матрица $(Z^T Z)^{-1}$ тоже симметричная, поэтому $((Z^T Z)^{-1})^T = (Z^T Z)^{-1}$, тогда:

$$D[\hat{\theta}] = \sigma^2 (Z^T Z)^{-1}$$

3) В пункте 1 было показано, что в условиях теоремы $\hat{\theta} = (Z^T Z)^{-1} Z^T \eta$ и, как легко видеть, оценка $\hat{\theta}$ линейно зависит от η. Покажем, что среди всех несмещенных и линейных по η оценок $\tilde{\theta}$ компоненты $\hat{\theta}_i$ в оценке $\hat{\theta} = (\hat{\theta}_1,..., \hat{\theta}_m)$ имеют наименьшую дисперсию, для этого рассмотрим произвольную несмещенную и линейную по η оценку $\theta^* = (\theta_1^*,..., \theta_m^*)$ и покажем, что дисперсии $D[\theta_i^*]$ компонент θ_i^* оказываются минимальными только тогда, когда оценка θ^* совпадает с оценкой по методу наименьших квадратов $\hat{\theta}$.

Пусть θ^* – произвольная несмещенная и линейная по η оценка $\tilde{\theta}$, тогда:

$$\theta^* = L\eta ,$$

где L – некоторая матрица порядка $m \times n$. Из условия несмещенности оценки θ^* следует:

$$M[\theta^*] = M[L\eta] = LM[\eta] = LZ\tilde{\theta} = \tilde{\theta} ,$$

причем равенство должно быть справедливо для произвольного $\tilde{\theta}$, это возможно тогда и только тогда, когда:

$$LZ = E_m \tag{6.6}$$

где E_m – единичная матрица порядка $m \times m$.

Вычислим дисперсионную матрицу $D[\theta^*]$ оценки θ^*:

$$D[\theta^*] = M[(\theta^* - M[\theta^*])(\theta^* - M[\theta^*])^T] = M[(L\eta - M[L\eta])(L\eta - M[L\eta])^T] =$$

$$= M[(L\eta - LM[\eta])(L\eta - LM[\eta])^T] = M[L(\eta - M[\eta])(L(\eta - M[\eta]))^T] =$$

$$= M[L(\eta - M[\eta])(\eta - M[\eta])^T L^T] = LM[(\eta - M[\eta])(\eta - M[\eta])^T]L^T =$$

$$= LD[\eta]L^T = L\sigma^2 E_n L^T = \sigma^2 LL^T . \tag{6.7}$$

Диагональные элементы дисперсионной матрицы $D[\theta^*]$ являются дисперсиями компонент вектора $\theta^* = (\theta_1^*,..., \theta_m^*)$:

$$\left\| D[\theta^*] \right\|_{i,i} = D[\theta_i^*] ,$$

$$i = \overline{1, m} .$$

Какова должна быть матрица L, чтобы с одной стороны выполнялось равенство (6.6) (которое следует из требования несмещенности) и с другой стороны диагональные элементы $D[\theta^*] = \sigma^2 LL^T$ были как можно меньше? Покажем, что матрица L должна быть равна матрице $(Z^T Z)^{-1} Z^T$, при которой получается оценка по методу наименьших квадратов $\hat{\theta}$.

Пусть для матрицы L выполняется (6.6), рассмотрим квадрат разности $L - (Z^T Z)^{-1} Z^T$:

$$(L - (Z^T Z)^{-1} Z^T)(L - (Z^T Z)^{-1} Z^T)^T = (L - (Z^T Z)^{-1} Z^T)(L^T - Z((Z^T Z)^{-1})^T) =$$

$$= LL^T - (Z^T Z)^{-1} Z^T L^T - LZ((Z^T Z)^{-1})^T + (Z^T Z)^{-1} Z^T Z((Z^T Z)^{-1})^T =$$

$$= LL^T - (Z^T Z)^{-1}(LZ)^T - E_m((Z^T Z)^{-1})^T + E_m((Z^T Z)^{-1})^T = LL^T - (Z^T Z)^{-1} E_m = LL^T - (Z^T Z)^{-1} .$$

Таким образом, если для L справедливо (6.6), тогда:

$$LL^T = (Z^T Z)^{-1} + (L - (Z^T Z)^{-1} Z^T)(L - (Z^T Z)^{-1} Z^T)^T$$

В правой части от матрицы L зависит только второе слагаемое, поэтому диагональные элементы LL^T принимают наименьшее значение, когда наименьшее значение принимают диагональные элементы матрицы $(L - (Z^T Z)^{-1} Z^T)(L - (Z^T Z)^{-1} Z^T)^T$.

Пусть $C = L - (Z^T Z)^{-1} Z^T$, заметим, что диагональный элемент в i-ой строке матрицы CC^T есть сумма квадратов элементов расположенных в i-ой строке матрицы C:

$$\left\| CC^T \right\|_{ii} = \sum_{j=1}^{n} \left(\left\| C \right\|_{ij} \right)^2 .$$

Сумма квадратов будет минимальна, если все элементы равны нулю, отсюда во всех строках все элементы матрицы C должны быть равны нулю ($0_{m\times n}$ – матрица порядка $m\times n$ состоящая из нулей):

$$C = 0_{m\times n}$$

$$L-(Z^TZ)^{-1}Z^T = 0_{m\times n}$$

$$L=(Z^TZ)^{-1}Z^T .$$

Легко видеть, что $(Z^TZ)^{-1}Z^TZ=E_m$, то есть матрица $(Z^TZ)^{-1}Z^T$ удовлетворяет (6.6). Таким образом, среди всех матриц L, удовлетворяющих (6.6), только при матрице $L=(Z^TZ)^{-1}Z^T$ диагональные элементы матрицы $D[\theta^*]$, то есть дисперсии $D[\theta_i^*]$, принимают наименьшие значения, но при $L=(Z^TZ)^{-1}Z^T$ оценка $\theta^*=L\eta$ становится оценкой по методу наименьших квадратов $\hat{\theta}=(Z^TZ)^{-1}Z^T\eta$. Таким образом, каждая компонента $\hat{\theta}_i$ в оценке $\hat{\theta}=(\hat{\theta}_1,\dots,\hat{\theta}_m)$ имеет наименьшую возможную дисперсию среди всех несмещенных и линейных по η оценок $\tilde{\theta}_i$.

Теорема доказана.

4. Оценка остаточной дисперсии.

В условиях (6.1)-(6.5) рассмотрим вопрос оценки остаточной дисперсии σ^2. Из (6.4) следует, что $\eta_i-(\varphi_1(x_{i1},\dots,x_{ik})\tilde{\theta}_1+\dots+\varphi_m(x_{i1},\dots,x_{ik})\tilde{\theta}_m)=\varepsilon_i$ и величины ε_i играют роль «остатков», поэтому дисперсия $D[\varepsilon_i]=\sigma^2$ называется остаточной.

Утверждение 6.5.

Если в задаче линейной регрессии (6.1)-(6.5) ранг матрицы Z равен m и статистика:

$$\hat{\sigma}^2(\eta_1,\dots,\eta_n)=\frac{\delta(\eta,\, f(X;\hat{\theta}))}{n-m},$$

тогда:

1) $\hat{\sigma}^2(\eta_1,\dots,\eta_n)$ является несмещенной оценкой σ^2;

2) $\hat{\sigma}^2(\eta_1,\dots,\eta_n)=\dfrac{\eta^T(E_n-Z(Z^TZ)^{-1}Z^T)\eta}{n-m}$.

Доказательство:

1) Пусть $\hat{\theta}$ – оценка по методу наименьших квадратов в задаче (6.1)-(6.5). При доказательстве утверждения 6.3 в пункте 1 было установлено, что для любого вектора параметров $\theta=(\theta_1,\dots,\theta_m)$:

$$\delta(\eta,\, f(X;\theta))=\delta(\eta,\, f(X;\hat{\theta}))+(\hat{\theta}-\theta)^TZ^TZ(\hat{\theta}-\theta),$$

где X – матрица значений x_{ij} из (6.4), $\|X\|_{ij}=x_{ij}$. Отсюда следует, что для вектора параметров $\tilde{\theta}=(\tilde{\theta}_1,\dots,\tilde{\theta}_m)$:

$$\delta(\eta,\, f(X;\tilde{\theta}))=\delta(\eta,\, f(X;\hat{\theta}))+(\hat{\theta}-\tilde{\theta})^TZ^TZ(\hat{\theta}-\tilde{\theta}),$$

$$\delta(\eta,\, f(X;\hat{\theta}))=\delta(\eta,\, f(X;\tilde{\theta}))-(\hat{\theta}-\tilde{\theta})^TZ^TZ(\hat{\theta}-\tilde{\theta}).$$

Вычислим математическое ожидание левой и правой частей:

$$M[\delta(\eta,\, f(X;\hat{\theta}))]=M[\delta(\eta,\, f(X;\tilde{\theta}))]-M[(\hat{\theta}-\tilde{\theta})^TZ^TZ(\hat{\theta}-\tilde{\theta})].$$

Поскольку $\delta(\eta,\, f(X;\tilde{\theta}))=\|\eta-Z\tilde{\theta}\|_2^2$ и из (6.2) и (6.4) следует $\eta=Z\tilde{\theta}+\varepsilon$ (где вектор столбец $\varepsilon=(\varepsilon_1,\dots,\varepsilon_n)$), то:

$$\delta(\eta,\, f(X;\tilde{\theta}))=\|\eta-Z\tilde{\theta}\|_2^2=\|\varepsilon\|_2^2=\sum_{i=1}^{n}\varepsilon_i^2 .$$

В силу условия (6.5) пункты 1 и 2:

$$M[\delta(\eta, f(X;\tilde{\theta}))] = M\left[\sum_{i=1}^{n} \varepsilon_i^2\right] = \sum_{i=1}^{n} M[\varepsilon_i^2] = \sum_{i=1}^{n} D[\varepsilon_i^2] = \sum_{i=1}^{n} \sigma^2 = n\sigma^2 .$$

Легко видеть, что:

$$M[(\hat{\theta} - \tilde{\theta})^T Z^T Z(\hat{\theta} - \tilde{\theta})] = M\left[\sum_{i=1}^{m}\sum_{j=1}^{m} (\hat{\theta}_i - \tilde{\theta}_i)\left\|Z^T Z\right\|_{ij} (\hat{\theta}_j - \tilde{\theta}_j)\right] =$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{m} M[(\hat{\theta}_i - \tilde{\theta}_i)\left\|Z^T Z\right\|_{ij} (\hat{\theta}_j - \tilde{\theta}_j)] = \sum_{i=1}^{m}\sum_{j=1}^{m} \left\|Z^T Z\right\|_{ij} M[(\hat{\theta}_i - \tilde{\theta}_i)(\hat{\theta}_j - \tilde{\theta}_j)] .$$

Поскольку по условию ранг матрицы Z равен m, то в силу теоремы 6.4 оценка $\hat{\theta}$ является несмещенной оценкой $\tilde{\theta}$, то есть:

$$M[\hat{\theta}] = \tilde{\theta} ,$$

$$M\left[\begin{pmatrix} \hat{\theta}_1 \\ \ldots \\ \hat{\theta}_m \end{pmatrix}\right] = \begin{pmatrix} M[\hat{\theta}_1] \\ \ldots \\ M[\hat{\theta}_m] \end{pmatrix} = \begin{pmatrix} \tilde{\theta}_1 \\ \ldots \\ \tilde{\theta}_m \end{pmatrix},$$

$$M[\hat{\theta}_i] = \tilde{\theta}_i , \ i = \overline{1, m} .$$

Таким образом,

$$M[(\hat{\theta} - \tilde{\theta})^T Z^T Z(\hat{\theta} - \tilde{\theta})] = \sum_{i=1}^{m}\sum_{j=1}^{m} \left\|Z^T Z\right\|_{ij} M[(\hat{\theta}_i - M[\hat{\theta}_i])(\hat{\theta}_j - M[\hat{\theta}_j])] =$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{m} \left\|Z^T Z\right\|_{ij} \operatorname{cov}(\hat{\theta}_i, \hat{\theta}_j) = \sum_{i=1}^{m}\sum_{j=1}^{m} \left\|Z^T Z\right\|_{ij} \left\|D[\hat{\theta}]\right\|_{ij} ,$$

где $D[\hat{\theta}]$ – дисперсионная матрица оценки $\hat{\theta}$, причем согласно теореме 6.4 пункт 2 (по условию утверждения ранг матрицы Z равен m):

$$D[\hat{\theta}] = \sigma^2 (Z^T Z)^{-1},$$

тогда,

$$M[(\hat{\theta} - \tilde{\theta})^T Z^T Z(\hat{\theta} - \tilde{\theta})] = \sum_{i=1}^{m}\sum_{j=1}^{m} \left\|Z^T Z\right\|_{ij} \left\|\sigma^2 (Z^T Z)^{-1}\right\|_{ij} = \sigma^2 \sum_{i=1}^{m}\sum_{j=1}^{m} \left\|Z^T Z\right\|_{ij} \left\|(Z^T Z)^{-1}\right\|_{ij} .$$

Поскольку $Z^T Z$ симметричная матрица, то $\left\|Z^T Z\right\|_{ij} = \left\|Z^T Z\right\|_{ji}$, тогда:

$$M[(\hat{\theta} - \tilde{\theta})^T Z^T Z(\hat{\theta} - \tilde{\theta})] = \sigma^2 \sum_{j=1}^{m}\sum_{i=1}^{m} \left\|Z^T Z\right\|_{ji} \left\|(Z^T Z)^{-1}\right\|_{ij} = \sigma^2 \sum_{j=1}^{m} \left\|Z^T Z (Z^T Z)^{-1}\right\|_{jj}$$

$$= \sigma^2 \sum_{j=1}^{m} \left\|E_m\right\|_{jj} = \sigma^2 m .$$

Таким образом,

$$M[\delta(\eta, f(X;\hat{\theta}))] = \sigma^2 n - \sigma^2 m = (n - m)\sigma^2 .$$

Отсюда следует, что статистика $\hat{\sigma}^2(\eta_1,\ldots,\eta_n)$:

$$\hat{\sigma}^2(\eta_1,\ldots,\eta_n) = \frac{\delta(\eta, f(X;\hat{\theta}))}{n - m} \qquad (6.8)$$

является несмещенной оценкой для остаточной дисперсии σ^2, действительно:

$$M[\hat{\sigma}^2(\eta_1,\ldots,\eta_n)] = M\left[\frac{\delta(\eta, f(X;\hat{\theta}))}{n - m}\right] = \frac{M[\delta(\eta, f(X;\hat{\theta}))]}{n - m} = \frac{n - m}{n - m}\sigma^2 = \sigma^2 .$$

2) Заметим, что если ранг матрицы Z равен m, тогда $\left|Z^T Z\right| \neq 0$ (утверждение 6.2) и из (6.3) оценка $\hat{\theta} = (Z^T Z)^{-1} Z^T \eta$. В этом случае для статистики $\hat{\sigma}^2(\eta_1,..., \eta_n)$ может быть получено выражение только через вектор величин $\eta = (\eta_1,..., \eta_n)$:

$$\hat{\sigma}^2(\eta_1,...,\eta_n) = \frac{\delta(\eta, f(X;\hat{\theta}))}{n-m} = \frac{\left\|\eta - Z\hat{\theta}\right\|_2^2}{n-m} = \frac{(\eta - Z\hat{\theta})^T(\eta - Z\hat{\theta})}{n-m} =$$

$$= \frac{(\eta - Z(Z^TZ)^{-1}Z^T\eta)^T(\eta - Z(Z^TZ)^{-1}Z^T\eta)}{n-m} = \frac{(\eta^T - \eta^T Z(Z^TZ)^{-1}Z^T)(\eta - Z(Z^TZ)^{-1}Z^T\eta)}{n-m} =$$

$$= \frac{\eta^T(E_n - Z(Z^TZ)^{-1}Z^T)(E_n - Z(Z^TZ)^{-1}Z^T)\eta}{n-m} =$$

$$= \frac{\eta^T(E_n - Z(Z^TZ)^{-1}Z^TE_n - E_nZ(Z^TZ)^{-1}Z^T + Z(Z^TZ)^{-1}Z^TZ(Z^TZ)^{-1}Z^T)\eta}{n-m} =$$

$$= \frac{\eta^T(E_n - Z(Z^TZ)^{-1}Z^T - Z(Z^TZ)^{-1}Z^T + Z(Z^TZ)^{-1}Z^T)\eta}{n-m} =$$

$$= \frac{\eta^T(E_n - Z(Z^TZ)^{-1}Z^T)\eta}{n-m}. \tag{6.9}$$

где E_n – единичная матрица порядка $n \times n$.

Утверждение доказано.

5. Качество регрессии и коэффициент детерминации.

Предположим, что в задаче линейной регрессии (6.1)-(6.5) построена оценка по методу наименьших квадратов $\hat{\theta}$, вслед за построением оценки, в ряде случаев на практике возникает вопрос о «качестве» построенной регрессии. Вопрос о «качестве» возникает также и при сравнении различных вариантов регрессий, когда в отношении условного математического ожидания (6.1) выдвигаются различные предположения с различным числом параметров $\tilde{\theta}_i$ и различным набором функций φ_j, а также когда увеличивается количество k случайных величин $(\xi_1(\omega),..., \xi_k(\omega))$. Для определения «качества» регрессии используется числовая характеристика, которая называется коэффициент детерминации.

Пусть в задаче линейной регрессии (6.1)-(6.5) ранг матрицы Z равен m, $\eta = (\eta_1,..., \eta_n)$ – вектор наблюдаемых величин, $\hat{\theta}$ – оценка по методу наименьших квадратов и $\hat{\eta} = f(X, \hat{\theta})$ – вектор регрессионных значений. Можно показать, что:

$$\sum_{i=1}^{n}(\eta_i - \hat{m}_1(\eta))^2 = \sum_{i=1}^{n}(\eta_i - \hat{\eta}_i)^2 + \sum_{i=1}^{n}(\hat{\eta}_i - \hat{m}_1(\eta))^2,$$

где $\hat{m}_1(\eta) = \frac{1}{n}\sum_{i=1}^{n}\eta_i$, или в векторной форме:

$$\left\|\eta - \overline{\eta}\right\|_2^2 = \left\|\eta - \hat{\eta}\right\|_2^2 + \left\|\hat{\eta} - \overline{\eta}\right\|_2^2,$$

где $\overline{\eta} = \left(\frac{1}{n}\sum_{i=1}^{n}\eta_i,..., \frac{1}{n}\sum_{i=1}^{n}\eta_i\right)$ – вектор, составленный из средних значений. Полученное соотношение показывает, что вариацию величин вектора η относительно среднего значения $\frac{1}{n}\sum_{i=1}^{n}\eta_i$ можно разложить на отклонение регрессионных значений $\hat{\eta}$ от величин вектора η и вариацию регрессионных значений $\hat{\eta}$ относительно среднего $\frac{1}{n}\sum_{i=1}^{n}\eta_i$. Разделив левую и правую части на $\|\eta - \overline{\eta}\|_2^2$ получим:

$$1 = \frac{\|\eta - \hat{\eta}\|_2^2}{\|\eta - \overline{\eta}\|_2^2} + \frac{\|\hat{\eta} - \overline{\eta}\|_2^2}{\|\eta - \overline{\eta}\|_2^2},$$

$$\frac{\|\hat{\eta} - \overline{\eta}\|_2^2}{\|\eta - \overline{\eta}\|_2^2} = 1 - \frac{\|\eta - \hat{\eta}\|_2^2}{\|\eta - \overline{\eta}\|_2^2}.$$

Определение 6.6.

Коэффициентом детерминации R^2 называется число:

$$R^2 = 1 - \frac{\|\eta - \hat{\eta}\|_2^2}{\|\eta - \overline{\eta}\|_2^2}.$$

В координатной форме коэффициент детерминации R^2 имеет вид:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\eta_i - \hat{\eta}_i)^2}{\sum_{i=1}^{n}(\eta_i - \hat{m}_1(\eta))^2}.$$

Коэффициент детерминации отражает меру расхождения между наблюдаемыми величинами η и регрессионными значениями $\hat{\eta}$, отнесенную к вариации величин вектора η.

Если R^2 – величина близкая к 0, то дробь – величина близкая к 1, и, следовательно, регрессионные значения $\hat{\eta}$ сосредоточены вблизи постоянного значения $\hat{m}_1(\eta)$, что является показателем регрессии «плохого» качества – регрессия не повторяет (не улавливает) вариацию наблюдаемых величин η, что, скорее всего, обусловлено неудачным выбором набора функций φ_i или указывает на «слабую вероятностную зависимость» между случайной величиной $\eta(\omega)$ и случайными величинами $(\xi_1(\omega),..., \xi_k(\omega))$.

Если R^2 – величина близка к 1, то дробь – величина близка к 0, тогда регрессионные значения $\hat{\eta}$ оказываются близкими к величинам вектора η, что свидетельствует о регрессии «хорошего» качества – регрессия весьма точно воспроизводит изменение величин вектора η, что означает «сильную вероятностную зависимость» между случайной величиной $\eta(\omega)$ и случайными величинами $(\xi_1(\omega),..., \xi_k(\omega))$.

Коэффициент детерминации R^2 может быть использован и для сравнения двух регрессий с одинаковым числом k случайных величин $\xi_i(\omega)$. Из двух вариантов регрессии лучшим следует признать тот вариант, для которого коэффициент детерминации R^2 больше.

Использование коэффициента детерминации R^2 для сравнения регрессий с различным количеством k случайных величин $\xi_i(\omega)$ может оказаться некорректным, поскольку исходно варианты регрессии оказываются в различных условиях: регрессия с большим количеством k теоретически оказывается в более выгодном положении, хотя при неудачном выборе функций φ_i (в (6.1)) в итоге может оказаться хуже. Тем не менее, для корректного сравнения регрессий в данном случае следует использовать скорректированный коэффициент детерминации, учитывающий количество случайных величин k.

Определение 6.7.

Скорректированным коэффициентом детерминации R^2_{adj} называется величина:

$$R_{adj}^2 = 1 - \frac{\frac{1}{n-k}\left\|\eta - \hat{\eta}\right\|_2^2}{\frac{1}{n}\left\|\eta - \overline{\eta}\right\|_2^2}.$$

По-прежнему из двух вариантов регрессии лучшим следует считать тот вариант, для которого величина R_{adj}^2 больше.

6. Задача нормальной линейной регрессии.

Рассмотрим задачу линейной регрессии (6.1)-(6.4), в которой предположение (6.5) заменено более сильным предположением в отношении вектора $\varepsilon = (\varepsilon_1,..., \varepsilon_n)$:

вектор $(\varepsilon_1,..., \varepsilon_n)$ имеет нормальное распределение $N(\overline{0}_n, \sigma^2 E_n)$, $\sigma^2 > 0$, (6.10)

где $\overline{0}_n$ – нулевой вектор порядка n и E_n – единичная матрица порядка $n \times n$.

Задача линейной регрессии (6.1)-(6.4), (6.10) называется *задачей нормальной линейной регрессии* (иногда кратко задачей *нормальной регрессии*).

Из (6.2) и (6.4) следует, что вектор величин $\eta = Z\tilde{\theta} + \varepsilon$, тогда из (6.10) следует, что вектор $\eta = (\eta_1,..., \eta_n)$ имеет нормальное распределение (как линейное преобразование вектора ε, имеющего нормальное распределение). Из пункта 1 теоремы 6.4 (в пункте 1 предположение $\left|Z^T Z\right| \neq 0$ не используется), следует, что $M[\eta] = Z\tilde{\theta}$ и $D[\eta] = D[\varepsilon] = \sigma^2 E_n$. Таким образом, вектор $\eta = (\eta_1,..., \eta_n)$ имеет нормальное распределение $N(Z\tilde{\theta}, \sigma^2 E_n)$ с плотностью вероятности:

$$p_\eta(y_1,..., y_n \mid \tilde{\theta}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\left|D[\eta]\right|}} e^{-\frac{1}{2}(y - M[\eta])^T (D[\eta])^{-1}(y - M[\eta])} = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\left|\sigma^2 E_n\right|}} e^{-\frac{1}{2}(y - Z\tilde{\theta})^T(\sigma^2 E_n)^{-1}(y - Z\tilde{\theta})} =$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\sigma^{2n}}} e^{-\frac{1}{2}(y - Z\tilde{\theta})^T \frac{1}{\sigma^2}E_n (y - Z\tilde{\theta})} = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} e^{-\frac{1}{2\sigma^2}(y - Z\tilde{\theta})^T(y - Z\tilde{\theta})} =$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} e^{-\frac{1}{2\sigma^2}\left\|y - Z\tilde{\theta}\right\|_2^2},$$

где $y = (y_1,..., y_n)$.

Рассмотрим задачу построения оценки θ^* неизвестного вектора параметров $\tilde{\theta}$ по методу максимального правдоподобия. Легко видеть, что функция правдоподобия,

$$L(y_1,..., y_n \mid \theta) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} e^{-\frac{1}{2\sigma^2}\left\|y - Z\theta\right\|_2^2},$$

принимает наибольшее значение при таком значении вектора параметров θ, при котором наименьшее значение принимает квадрат нормы $\left\|y - Z\theta\right\|_2^2$. Отсюда следует, что МП-оценка $\theta^*(\eta_1,..., \eta_n)$ определяется условием:

$$\left\|\eta - Z\theta^*\right\|_2^2 = \min_{\theta} \left\|\eta - Z\theta\right\|_2^2,$$

но из (6.2) квадрат нормы $\left\|\eta - Z\theta\right\|_2^2$ совпадает со среднеквадратичным отклонением $\delta(\eta, f(X;\theta))$ и наименьшее значение $\delta(\eta, f(X;\theta))$ достигается при оценке по методу наименьших квадратов $\hat{\theta}$. Отсюда следует, что в задаче нормальной линейной регрессии

(6.1)-(6.4), (6.10) оценка по методу наименьших квадратов одновременно является оценкой максимального правдоподобия вектора параметров $\tilde{\theta}$.

Заметим, что из условия (6.10) следует условие (6.5), поэтому теорема 6.4 о свойствах оценки по методу наименьших квадратов остается в силе и при замене (6.5) на (6.10). Условие (6.10) более сильное, поэтому позволяет установить ряд фактов о распределениях как самой оценки по методу наименьших квадратов, так и сопутствующих статистик в частности статистики (6.8), оценивающей остаточную дисперсию σ^2 .

Прежде всего рассмотрим вспомогательное утверждение.

Утверждение 6.8.

Пусть $\left|Z^T Z\right| \neq 0$, тогда существуют матрицы Q и D :

1) $Z^T Z = Q^T D D^T Q$, где $Q^{-1} = Q^T$ и $\left|D\right| \neq 0$.

2) $Z^T Z = Q^T \tilde{D}^T \tilde{D} Q$, где $\tilde{D} = D^{-1}$.

Доказательство:

1) Поскольку $\left|Z^T Z\right| \neq 0$, то у матрицы $Z^T Z$ существует базис из собственных векторов, который можно сделать ортономированным, тогда:

$$Z^T Z = Q^T \hat{D} Q ,$$

где столбцы матрицы Q являются ортономированными векторами базиса (отсюда следует, что $Q^T Q = E$ и $\left|Q\right| \neq 0$, поскольку векторы базиса линейно независимы, тогда существует Q^{-1} и следовательно $Q^T Q Q^{-1} = E Q^{-1}$ и $Q^T = Q^{-1}$), и матрица $\hat{D}$ является диагональной матрицей, на главной диагонали которой располагаются собственные числа матрицы $Z^T Z$.

Заметим, что все собственные числа матрицы $Z^T Z$ положительны, действительно, допустим противное:

а) пусть число $\lambda = 0$ является собственным числом матрицы $Z^T Z$ и λ соответствует собственный вектор $x \neq 0$, тогда:

$$Z^T Z x = \lambda x = 0 x = \overline{0} .$$

Поскольку $x \neq 0$, то отсюда следует, что столбцы матрицы $Z^T Z$ линейно зависимы, и тогда $\left|Z^T Z\right| = 0$, что противоречит условию утверждения $\left|Z^T Z\right| \neq 0$.

б) пусть число $\lambda < 0$ является собственным число матрицы $Z^T Z$ и λ соответствует собственный вектор $x \neq 0$, тогда с одной стороны:

$$x^T Z^T Z x = (Zx)^T (Zx) = \left\|Zx\right\|_2^2 \geq 0$$

а с другой стороны:

$$x^T Z^T Z x = x^T \lambda x = \lambda \left\|x\right\|_2^2 < 0 ,$$

поскольку $\lambda < 0$ и $\left\|x\right\|_2^2 \neq 0$ так как $x \neq 0$. Одно неравенство противоречит другому.

Таким образом, все элементы на главной диагонали $\hat{D}$ являются положительными числами, и следовательно корректным является следующее определение матрицы D :

$$\left\|D\right\|_{i,j} = \begin{cases} \sqrt{\left\|\hat{D}\right\|_{i,j}} & ,i = j \\ 0 & i \neq j \end{cases} .$$

Поскольку D – диагональная матрица, то $D^T = D$, и $D D^T = \hat{D}$, тогда:

$$Z^T Z = Q^T D D^T Q .$$

2) В пункте 1 было показано, что все диагональные элементы матрицы $\hat{D}$ положительны, поэтому в матрице D все диагональные элементы также положительны, тогда корректным является определение матрицы $\tilde{D}$:

$$\left\|\tilde{D}\right\|_{i,j} = \begin{cases} \dfrac{1}{\left\|D\right\|_{i,j}} & , i = j \\ 0 & i \neq j \end{cases}.$$

Легко видеть, что $\left|D\right| \neq 0$, $\tilde{D}D = E$ и $D\tilde{D} = E$, поэтому $\tilde{D} = D^{-1}$.

Непосредственной проверкой легко убедиться, что $(Z^T Z)^{-1} = Q^T \tilde{D}^T \tilde{D} Q$, используя разложение $Z^T Z = Q^T D D^T Q$, доказанное в пункте 1, действительно:

$$Z^T Z Q^T \tilde{D}^T \tilde{D} Q = Q^T D D^T Q Q^T \tilde{D}^T \tilde{D} Q = Q^T D D^T \tilde{D}^T \tilde{D} Q = Q^T D(\tilde{D} D)^T \tilde{D} Q = Q^T D \tilde{D} Q = Q^T Q = E,$$

$$Q^T \tilde{D}^T \tilde{D} Q Z^T Z = Q^T \tilde{D}^T \tilde{D} Q Q^T D D^T Q = Q^T \tilde{D}^T \tilde{D} D D^T Q = Q^T \tilde{D}^T D^T Q = Q^T (D\tilde{D})^T Q = Q^T Q = E.$$

Поскольку по условию $\left|Z^T Z\right| \neq 0$, то $(Z^T Z)^{-1}$ существует и единственна, тогда:

$$(Z^T Z)^{-1} = Q^T \tilde{D}^T \tilde{D} Q.$$

Утверждение доказано.

Везде далее для краткости использовано обозначение $\delta(\theta) = \delta(\eta, f(X;\theta))$.

Теорема 6.9.

Если в задаче нормальной линейной регрессии (6.1)-(6.4), (6.10) ранг матрицы Z равен m, тогда:

1) оценка $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_m)$ имеет нормальное распределение $N(\tilde{\theta}, \sigma^2 (Z^T Z)^{-1})$;

2) случайная величина $\dfrac{\delta(\tilde{\theta}) - \delta(\hat{\theta})}{\sigma^2}$ имеет распределение $\chi^2(m)$;

3) случайная величина $\dfrac{\delta(\hat{\theta})}{\sigma^2}$ имеет распределение $\chi^2(n-m)$;

4) случайный вектор $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_m)$ и случайная величина $\delta(\hat{\theta})$ независимы;

5) случайные величины $\delta(\hat{\theta})$ и $\delta(\tilde{\theta}) - \delta(\hat{\theta})$ независимы.

Доказательство:

1) Поскольку ранг матрицы Z равен m, то согласно утверждению 6.2 пункт 3 $\left|Z^T Z\right| \neq 0$, тогда существует обратная матрица $(Z^T Z)^{-1}$ и из (6.3) при умножении левой и правой частей на $(Z^T Z)^{-1}$:

$$\hat{\theta} = (Z^T Z)^{-1} Z^T \eta.$$

Из (6.2) и (6.4) следует, что $\eta = Z\tilde{\theta} + \varepsilon$, тогда:

$$\hat{\theta} = (Z^T Z)^{-1} Z^T (Z\tilde{\theta} + \varepsilon) = (Z^T Z)^{-1} Z^T Z \tilde{\theta} + (Z^T Z)^{-1} Z^T \varepsilon = \tilde{\theta} + (Z^T Z)^{-1} Z^T \varepsilon. \quad (6.11)$$

Поскольку ε имеет нормальное распределение и $\hat{\theta}$ является линейным преобразованием ε, то $\hat{\theta}$ имеет нормальное распределение. Причем из теоремы 6.4 $M[\hat{\theta}] = \tilde{\theta}$ и $D[\hat{\theta}] = \sigma^2 (Z^T Z)^{-1}$. Таким образом, $\hat{\theta}$ имеет нормальное распределение $N(\tilde{\theta}, \sigma^2 (Z^T Z)^{-1})$.

2) В пункте 1 утверждения 6.3 было показано, что из (6.3) следует:

$$\delta(\eta, f(X;\theta)) = \delta(\eta, f(X, \hat{\theta})) + (\hat{\theta} - \theta)^T Z^T Z (\hat{\theta} - \theta),$$

тогда для $\delta(\eta, f(X;\tilde{\theta}))$, используя сокращенные обозначения, получим:

$$\delta(\tilde{\theta}) = \delta(\hat{\theta}) + (\hat{\theta} - \tilde{\theta})^T Z^T Z (\hat{\theta} - \tilde{\theta}),$$

$$\delta(\tilde{\theta}) - \delta(\hat{\theta}) = (\hat{\theta} - \tilde{\theta})^T Z^T Z (\hat{\theta} - \tilde{\theta}).$$

Откуда с учетом (6.11) следует:

$$\delta(\tilde{\theta}) - \delta(\hat{\theta}) = (\tilde{\theta} + (Z^T Z)^{-1} Z^T \varepsilon - \tilde{\theta})^T Z^T Z (\tilde{\theta} + (Z^T Z)^{-1} Z^T \varepsilon - \tilde{\theta}) =$$

$$= ((Z^T Z)^{-1} Z^T \varepsilon)^T Z^T Z (Z^T Z)^{-1} Z^T \varepsilon = \varepsilon^T Z ((Z^T Z)^{-1})^T Z^T \varepsilon =$$

$$= \varepsilon^T Z (Z^T Z)^{-1} Z^T \varepsilon , \tag{6.12}$$

поскольку $(Z^T Z)^{-1}$ – симметричная матрица (как всякая обратная матрица симметричной матрицы).

Поскольку $\left| Z^T Z \right| \neq 0$, то согласно утверждению 6.8 существует разложение:

$$\begin{gathered} Z^T Z = Q^T D D^T Q , \\ (Z^T Z)^{-1} = Q^T \tilde{D}^T \tilde{D} Q , \\ Q^{-1} = Q^T , \ \tilde{D} = D^{-1} . \end{gathered} \tag{6.13}$$

Подставляя разложение для $(Z^T Z)^{-1}$ из (6.13) в (6.12), получим:

$$\delta(\tilde{\theta}) - \delta(\hat{\theta}) = \varepsilon Z Q^T \tilde{D}^T \tilde{D} Q Z^T \varepsilon = (\tilde{D} Q Z^T \varepsilon)^T \tilde{D} Q Z^T \varepsilon .$$

тогда,

$$\frac{\delta(\tilde{\theta}) - \delta(\hat{\theta})}{\sigma^2} = \left(\frac{1}{\sigma} \tilde{D} Q Z^T \varepsilon \right)^T \frac{1}{\sigma} \tilde{D} Q Z^T \varepsilon$$

Пусть вектор столбец $\tilde{\varepsilon} = \frac{1}{\sigma} \tilde{D} Q Z^T \varepsilon$, тогда:

$$\frac{\delta(\tilde{\theta}) - \delta(\hat{\theta})}{\sigma^2} = \tilde{\varepsilon}^T \tilde{\varepsilon} = \left\| \tilde{\varepsilon} \right\|_2^2 = \sum_{i=1}^{m} \tilde{\varepsilon}_i^2 \tag{6.14}$$

Вектор $\tilde{\varepsilon}$ является линейным преобразованием вектора ε и согласно (6.10) вектор ε имеет нормальное распределение, тогда вектор $\tilde{\varepsilon}$ также имеет нормальное распределение. Математическое ожидание $M[\tilde{\varepsilon}]$:

$$M[\tilde{\varepsilon}] = M\left[\frac{1}{\sigma} \tilde{D} Q Z^T \varepsilon \right] = \frac{1}{\sigma} \tilde{D} Q Z^T M[\varepsilon] = \frac{1}{\sigma} \tilde{D} Q Z^T \overline{0}_n = \overline{0}_m ,$$

где $M[\varepsilon] = \overline{0}_n$ согласно (6.10) и $\overline{0}_n$ – нулевой вектор порядка n. Дисперсионная матрица $D[\tilde{\varepsilon}]$:

$$D[\tilde{\varepsilon}] = D\left[\frac{1}{\sigma} \tilde{D} Q Z^T \tilde{\varepsilon} \right] = \frac{1}{\sigma} \tilde{D} Q Z^T D[\tilde{\varepsilon}] \left(\frac{1}{\sigma} \tilde{D} Q Z^T \right)^T = \frac{1}{\sigma^2} \tilde{D} Q Z^T \sigma^2 E_n Z Q^T \tilde{D}^T = \tilde{D} Q Z^T Z Q^T \tilde{D}^T ,$$

используя разложение для $Z^T Z$ из (6.13), получим:

$$D[\tilde{\varepsilon}] = \tilde{D} Q Q^T D D^T Q Q^T \tilde{D}^T = \tilde{D} D D^T \tilde{D}^T = \tilde{D} D D^T \tilde{D}^T = (\tilde{D} D)^T = E_m ,$$

где E_m – единичная матрица порядка $m \times m$.

Таким образом, вектор $\tilde{\varepsilon}$ имеет нормальное распределение $N(\overline{0}_m, E_m)$, откуда следует, что в сумме квадратов (6.14) все случайные величины имеют нормальное распределение $N(0,1)$ и некоррелированы, но для совместно нормальных случайных величин из некоррелированности следует независимость, поэтому сумма квадратов $\sum_{i=1}^{m} \tilde{\varepsilon}_i^2$ имеет распределение $\chi^2(m)$ и, следовательно, случайная величина $\frac{\delta(\tilde{\theta}) - \delta(\hat{\theta})}{\sigma^2}$ имеет распределение $\chi^2(m)$.

3) При доказательстве пункта 2 в утверждения 6.5 условие (6.5) не использовалось, но использовалось условие $\left| Z^T Z \right| \neq 0$, справедливое в условиях теоремы, поэтому из (6.9) следует:

$$\frac{\delta(\eta, f(X; \hat{\theta}))}{n - m} = \hat{\sigma}^2(\eta_1, \ldots, \eta_n) = \frac{\eta^T (E_n - Z (Z^T Z)^{-1} Z^T) \eta}{n - m} ,$$

$$\delta(\hat{\theta}) = \eta^T (E_n - Z (Z^T Z)^{-1} Z^T) \eta .$$

Из (6.2) и (6.4) следует, что $\eta = Z\tilde{\theta} + \varepsilon$, тогда:

$$\eta^T (E_n - Z(Z^T Z)^{-1} Z^T)\eta = (Z\tilde{\theta} + \varepsilon)^T (E_n - Z(Z^T Z)^{-1} Z^T)(Z\tilde{\theta} + \varepsilon) =$$
$$= (\tilde{\theta}^T Z^T + \varepsilon^T)(E_n - Z(Z^T Z)^{-1} Z^T)(Z\tilde{\theta} + \varepsilon) =$$
$$= \tilde{\theta}^T Z^T (E_n - Z(Z^T Z)^{-1} Z^T)Z\tilde{\theta} + \tilde{\theta}^T Z^T (E_n - Z(Z^T Z)^{-1} Z^T)\varepsilon +$$
$$+ \varepsilon^T (E_n - Z(Z^T Z)^{-1} Z^T)Z\tilde{\theta} + \varepsilon^T (E_n - Z(Z^T Z)^{-1} Z^T)\varepsilon =$$
$$= \tilde{\theta}^T (Z^T Z - Z^T Z(Z^T Z)^{-1} Z^T Z)\tilde{\theta} + \tilde{\theta}^T (Z^T - Z^T Z(Z^T Z)^{-1} Z^T)\varepsilon +$$
$$+ \varepsilon^T (Z - Z(Z^T Z)^{-1} Z^T Z)\tilde{\theta} + \varepsilon^T (E_n - Z(Z^T Z)^{-1} Z^T)\varepsilon =$$
$$= \tilde{\theta}^T (Z^T Z - Z^T Z)\tilde{\theta} + \tilde{\theta}^T (Z^T - Z^T)\varepsilon + \varepsilon^T (Z - Z)\tilde{\theta} + \varepsilon^T (E_n - Z(Z^T Z)^{-1} Z^T)\varepsilon =$$
$$= \varepsilon^T (E_n - Z(Z^T Z)^{-1} Z^T)\varepsilon .$$

Таким образом,

$$\delta(\hat{\theta}) = \varepsilon^T (E_n - Z(Z^T Z)^{-1} Z^T)\varepsilon \tag{6.15}$$

Из (6.15) непосредственно следует:

$$\frac{\delta(\hat{\theta})}{\sigma^2} = \frac{\varepsilon^T\varepsilon - \varepsilon^T Z(Z^T Z)^{-1} Z^T\varepsilon}{\sigma^2} = \frac{\varepsilon^T\varepsilon}{\sigma^2} - \frac{\varepsilon^T Z(Z^T Z)^{-1} Z^T\varepsilon}{\sigma^2} = \left(\frac{\varepsilon}{\sigma}\right)^T \left(\frac{\varepsilon}{\sigma}\right) - \frac{\varepsilon^T Z(Z^T Z)^{-1} Z^T\varepsilon}{\sigma^2},$$

$$\frac{\delta(\hat{\theta})}{\sigma^2} + \frac{\varepsilon^T Z(Z^T Z)^{-1} Z^T\varepsilon}{\sigma^2} = \left(\frac{\varepsilon}{\sigma}\right)^T \left(\frac{\varepsilon}{\sigma}\right),$$

$$\frac{\delta(\hat{\theta})}{\sigma^2} + \frac{\varepsilon^T Z(Z^T Z)^{-1} Z^T\varepsilon}{\sigma^2} = (\varepsilon^*)^T \varepsilon^*, \tag{6.16}$$

где $\varepsilon^* = \dfrac{\varepsilon}{\sigma}$.

Покажем, что в (6.16) правая часть $(\varepsilon^*)^T \varepsilon^*$ имеет распределение $\chi^2(n)$, в левой части второе слагаемое $\dfrac{\varepsilon^T Z(Z^T Z)^{-1} Z^T\varepsilon}{\sigma^2}$ имеет распределение $\chi^2(m)$ и не зависит от первого слагаемого $\dfrac{\delta(\hat{\theta})}{\sigma^2}$. Отсюда будет следовать, что первое слагаемое $\dfrac{\delta(\hat{\theta})}{\sigma^2}$ имеет распределение $\chi^2(n-m)$ (свойство характеристических функций распределения χ^2).

Легко видеть, что правая часть (6.16) есть сумма квадратов:

$$(\varepsilon^*)^T \varepsilon^* = \left\|\varepsilon^*\right\|_2^2 = \sum_{i=1}^{n} (\varepsilon_i^*)^2 \tag{6.17}$$

Кроме того, вектор ε^* имеет нормальное распределение как линейное преобразование вектора ε, имеющего согласно (6.10) нормальное распределение. Математическое ожидание $M[\varepsilon^*]$:

$$M[\varepsilon^*] = M\left[\frac{\varepsilon}{\sigma}\right] = \frac{1}{\sigma} M[\varepsilon] = \frac{1}{\sigma}\bar{0}_n = \bar{0}_n,$$

дисперсионная матрица $D[\varepsilon^*]$:

$$D[\varepsilon^*] = D\left[\frac{\varepsilon}{\sigma}\right] = \frac{1}{\sigma^2} D[\varepsilon] = \frac{1}{\sigma^2}\sigma^2 E_n = E_n .$$

Таким образом, вектор ε^* имеет нормальное распределение $N(\bar{0}_n, E_n)$, поэтому все случайные величины ε_i^* в сумме в (6.17) имеют нормальное распределение $N(0,1)$ и некоррелированы, следовательно независимы (поскольку для совместно нормальных случайных величин из некоррелированности следует независимость), поэтому сумма $\sum_{i=1}^{n}(\varepsilon_i^*)^2$ и величина $(\varepsilon^*)^T\varepsilon^*$ имеют распределение $\chi^2(n)$.

Величина $\dfrac{\varepsilon^T Z(Z^TZ)^{-1}Z^T\varepsilon}{\sigma^2}$ рассматривалась в пункте 2 доказательства, из (6.12):

$$\frac{\varepsilon^T Z(Z^TZ)^{-1}Z^T\varepsilon}{\sigma^2}=\frac{\delta(\tilde{\theta})-\delta(\hat{\theta})}{\sigma^2},$$

и было показано, что $\dfrac{\delta(\tilde{\theta})-\delta(\hat{\theta})}{\sigma^2}$ имеет распределение $\chi^2(m)$, тогда случайная величина $\dfrac{\varepsilon^T Z(Z^TZ)^{-1}Z^T\varepsilon}{\sigma^2}$ имеет такое же распределение $\chi^2(m)$.

Из условия $\left|Z^TZ\right|\neq 0$ следует разложение для $(Z^TZ)^{-1}$ (6.13), тогда:

$$\begin{gathered}\varepsilon^T Z(Z^TZ)^{-1}Z^T\varepsilon=\left(\tilde{D}QZ^T\varepsilon\right)^T\left(\tilde{D}QZ^T\varepsilon\right)=\dot{\varepsilon}^T\dot{\varepsilon},\\ \dot{\varepsilon}=\tilde{D}QZ^T\varepsilon\end{gathered}\qquad(6.18)$$

Заметим, что в правой части (6.15) матрица $B=E_n-Z(Z^TZ)^{-1}Z^T$ является идемпотентной ($B^2=B$), действительно:

$$\begin{aligned}B^2&=(E_n-Z(Z^TZ)^{-1}Z^T)(E_n-Z(Z^TZ)^{-1}Z^T)=\\&=E_nE_n-E_nZ(Z^TZ)^{-1}Z^T-Z(Z^TZ)^{-1}Z^TE_n+Z(Z^TZ)^{-1}Z^TZ(Z^TZ)^{-1}Z^T)=\\&=E_n-Z(Z^TZ)^{-1}Z^T-Z(Z^TZ)^{-1}Z^T+Z(Z^TZ)^{-1}Z^T)=E_n-Z(Z^TZ)^{-1}Z^T=B.\end{aligned}$$

Кроме того, матрица B является симметричной ($B^T=B$), действительно:

$$B^T=(E_n-Z(Z^TZ)^{-1}Z^T)^T=E_n-(Z^T)^T((Z^TZ)^{-1})^TZ^T=E_n-Z(Z^TZ)^{-1}Z^T=B,$$

где $((Z^TZ)^{-1})^T=(Z^TZ)^{-1}$, поскольку Z^TZ – симметричная матрица и обратная матрица всякой симметричной матрицы тоже симметрична.

Таким образом, из (6.15):

$$\begin{gathered}\delta(\hat{\theta})=\varepsilon^T(E_n-Z(Z^TZ)^{-1}Z^T)\varepsilon=\varepsilon^TB\varepsilon=\varepsilon^TB^2\varepsilon=\varepsilon^TBB\varepsilon=\varepsilon^TB^TB\varepsilon=(B\varepsilon)^T(B\varepsilon)=\ddot{\varepsilon}^T\ddot{\varepsilon},\\ \ddot{\varepsilon}=B\varepsilon.\end{gathered}\qquad(6.19)$$

Покажем, что векторы $\ddot{\varepsilon}=B\varepsilon$ и $\dot{\varepsilon}=\tilde{D}QZ^T\varepsilon$ из (6.18) независимы. Рассмотрим расширенный вектор $(\ddot{\varepsilon},\dot{\varepsilon})=\left(B\varepsilon,\tilde{D}QZ^T\varepsilon\right)$, легко видеть, что расширенный вектор $(B\varepsilon,\tilde{\varepsilon})$ является линейным преобразованием вектора ε:

$$\begin{pmatrix}\ddot{\varepsilon}\\ \dot{\varepsilon}\end{pmatrix}=\begin{pmatrix}B\\ \tilde{D}QZ^T\end{pmatrix}\varepsilon.$$

Согласно (6.10) вектор ε имеет нормальное распределение, тогда вектор $(\ddot{\varepsilon},\dot{\varepsilon})$ также имеет нормальное распределение. Для совместно нормальных величин независимость следует из некоррелированности, поэтому для доказательства независимости $\ddot{\varepsilon}$ и $\dot{\varepsilon}$ достаточно показать, что $\ddot{\varepsilon}$ и $\dot{\varepsilon}$ некоррелированы, то есть ковариационная матрица $K(\ddot{\varepsilon},\dot{\varepsilon})$ является нулевой. Ковариационная матрица $K(\ddot{\varepsilon},\dot{\varepsilon})$, образованная ковариациями $\mathrm{cov}(\ddot{\varepsilon}_i,\dot{\varepsilon}_j)$, имеет вид:

$$K(\ddot{\varepsilon},\dot{\varepsilon})=M[(\ddot{\varepsilon}-M[\ddot{\varepsilon}])(\dot{\varepsilon}-M[\dot{\varepsilon}])^T].$$

Заметим, что $M[\ddot{\varepsilon}]=M[B\varepsilon]=BM[\varepsilon]=\overline{0}_n$ и $M[\dot{\varepsilon}]=M[\tilde{D}QZ^T\dot{\varepsilon}]=\tilde{D}QZ^TM[\dot{\varepsilon}]=\overline{0}_m$, тогда:

$$\begin{aligned}K(\ddot{\varepsilon},\dot{\varepsilon})&=M[\ddot{\varepsilon}\dot{\varepsilon}^T]=M[B\varepsilon(\tilde{D}QZ^T\varepsilon)^T]=M[B\varepsilon\varepsilon^TZQ^T\tilde{D}^T]=BM[\varepsilon\varepsilon^T]ZQ^T\tilde{D}^T=BD[\varepsilon]ZQ^T\tilde{D}^T=\\&=BD[\varepsilon]ZQ^T\tilde{D}^T=B\sigma^2E_nZQ^T\tilde{D}^T=\sigma^2BZQ^T\tilde{D}^T=\sigma^2(E_n-Z(Z^TZ)^{-1}Z^T)ZQ^T\tilde{D}^T=\end{aligned}$$

$$= \sigma^2 (ZQ^T \tilde{D}^T - Z(Z^T Z)^{-1} Z^T ZQ^T \tilde{D}^T) = \sigma^2 (ZQ^T \tilde{D}^T - ZQ^T \tilde{D}^T) = 0_{n\times m},$$

где $0_{n\times m}$ – нулевая матрица порядка $n \times m$, поскольку из (6.10) $M[\varepsilon] = \overline{0}_n$ и $D[\varepsilon] = \sigma^2 E_n$.

Таким образом, векторы $\ddot{\varepsilon}$ и $\dot{\varepsilon}$ независимы, тогда из (6.18) и (6.19) независимы величины $\delta(\hat{\theta})$ и $\varepsilon^T Z(Z^T Z)^{-1} Z^T \varepsilon$, тогда независимы величины $\dfrac{\delta(\hat{\theta})}{\sigma^2}$ и $\dfrac{\varepsilon^T Z(Z^T Z)^{-1} Z^T \varepsilon}{\sigma^2}$ в (6.16). Откуда с учетом доказанных ранее распределений для $\dfrac{\varepsilon^T Z(Z^T Z)^{-1} Z^T \varepsilon}{\sigma^2}$ и $(\varepsilon^*)^T \varepsilon^*$ величина $\dfrac{\delta(\hat{\theta})}{\sigma^2}$ имеет распределение $\chi^2(n-m)$.

4) Для доказательства независимости вектора $\hat{\theta}$ и величины $\delta(\hat{\theta})$ воспользуемся представлениями (6.11) и (6.19). Рассмотрим расширенный вектор $(\hat{\theta}, \ddot{\varepsilon})$, легко видеть, что:

$$\begin{pmatrix} \hat{\theta} \\ \ddot{\varepsilon} \end{pmatrix} = \begin{pmatrix} \tilde{\theta} \\ \overline{0}_n \end{pmatrix} + \begin{pmatrix} (Z^T Z)^{-1} Z^T \\ B \end{pmatrix} \varepsilon,$$

поэтому $(\hat{\theta}, \ddot{\varepsilon})$ имеет нормальное распределение, как линейное преобразование вектора ε, имеющего согласно (6.10) нормальное распределение. Для совместно нормальных случайных величины независимость следует из некоррелированности, поэтому для доказательства независимости $\hat{\theta}$ и $\ddot{\varepsilon}$ достаточно показать, что матрица коваріаций $K(\hat{\theta}, \ddot{\varepsilon})$ является нулевой матрицей:

$$K(\hat{\theta}, \ddot{\varepsilon}) = M[(\hat{\theta} - M[\hat{\theta}])(\ddot{\varepsilon} - M[\ddot{\varepsilon}])^T] =$$

$$= M[(\tilde{\theta} - (Z^T Z)^{-1} Z^T \varepsilon - M[\tilde{\theta} - (Z^T Z)^{-1} Z^T \varepsilon])(B\varepsilon - M[B\varepsilon])^T] =$$

$$= M[(\tilde{\theta} - (Z^T Z)^{-1} Z^T \varepsilon - \tilde{\theta} - (Z^T Z)^{-1} Z^T M[\varepsilon])(B\varepsilon - BM[\varepsilon])^T] =$$

$$= M[((Z^T Z)^{-1} Z^T \varepsilon)(B\varepsilon)^T] = M[(Z^T Z)^{-1} Z^T \varepsilon\varepsilon^T B^T] = (Z^T Z)^{-1} Z^T M[\varepsilon\varepsilon^T] B^T = (Z^T Z)^{-1} Z^T D[\varepsilon] B^T =$$

$$= (Z^T Z)^{-1} Z^T \sigma^2 E_n B^T = \sigma^2 (Z^T Z)^{-1} Z^T B^T = \sigma^2 (Z^T Z)^{-1} Z^T B = \sigma^2 (Z^T Z)^{-1} Z^T (E_n - Z(Z^T Z)^{-1} Z^T) =$$

$$= \sigma^2 ((Z^T Z)^{-1} Z^T - (Z^T Z)^{-1} Z^T Z(Z^T Z)^{-1} Z^T) = \sigma^2 ((Z^T Z)^{-1} Z^T - (Z^T Z)^{-1} Z^T) = 0_{m\times n},$$

где $0_{m\times n}$ – нулевая матрица порядка $m \times n$, поскольку из (6.10) $M[\varepsilon] = \overline{0}_n$ и $D[\varepsilon] = \sigma^2 E_n$, и при доказательстве пункта 3 было показано, что матрица B является симметричной ($B^T = B$).

Таким образом, векторы $\hat{\theta}$ и $\ddot{\varepsilon}$ независимы, тогда независимы вектор $\hat{\theta}$ и величина $\delta(\hat{\theta}) = \ddot{\varepsilon}^T \ddot{\varepsilon}$.

5) В конце пункта 3 доказательства было доказано, что величины $\delta(\hat{\theta})$ и $\varepsilon^T Z(Z^T Z)^{-1} Z^T \varepsilon$ независимы. Согласно (6.12) $\delta(\tilde{\theta}) - \delta(\hat{\theta}) = \varepsilon^T Z(Z^T Z)^{-1} Z^T \varepsilon$, тогда величины $\delta(\hat{\theta})$ и $\delta(\tilde{\theta}) - \delta(\hat{\theta})$ независимы.

Теорема доказана.

7. Интервальное оценивание в задаче нормальной линейной регрессии.

В задаче нормальной линейной регрессии (6.1)-(6.4), (6.10), в случае если ранг матрицы Z равен m теорема 6.9 позволяет строить: а) доверительные интервалы для компонент $\tilde{\theta}_i$ неизвестного вектора параметров $\tilde{\theta}$, б) доверительный интервал для остаточной дисперсии σ^2, в) доверительную область для вектора $\tilde{\theta}$, г) критерий проверки гипотезы об отсутствии зависимости.

Доверительный интервал для компонент $\tilde{\theta}_i$.

Согласно пункту 1 теоремы 6.9 векторная случайная величина $\hat{\theta} = (\hat{\theta}_1,..., \hat{\theta}_m)$ имеет нормальное распределение $N(\tilde{\theta}, \sigma^2 (Z^T Z)^{-1})$, откуда каждая случайная величина $\hat{\theta}_i$ имеет нормальное распределение $N\left(\tilde{\theta}_i, \sigma^2 \left\|(Z^T Z)^{-1}\right\|_{i,i}\right)$, тогда случайная величина $\dfrac{\hat{\theta}_i - \tilde{\theta}_i}{\sigma\sqrt{\left\|(Z^T Z)^{-1}\right\|_{i,i}}}$ имеет стандартное нормальное распределение $N(0,1)$. Согласно пункту 3 теоремы 6.9 случайная величина $\dfrac{\delta(\hat{\theta})}{\sigma^2}$ имеет распределение $\chi^2(n-m)$ и согласно пункту 4 случайные величины $\hat{\theta}_i$ и $\delta(\hat{\theta})$ независимы, тогда статистика:

$$\frac{\dfrac{\hat{\theta}_i - \tilde{\theta}_i}{\sigma\sqrt{\left\|(Z^T Z)^{-1}\right\|_{i,i}}}}{\sqrt{\dfrac{\dfrac{\delta(\hat{\theta})}{\sigma^2}}{n-m}}} = \frac{\hat{\theta}_i - \tilde{\theta}_i}{\sqrt{\left\|(Z^T Z)^{-1}\right\|_{i,i} \dfrac{\delta(\hat{\theta})}{n-m}}}$$

по определению имеет распределение Стьюдента $T(n-m)$. Отсюда следует, что доверительным интервалом для неизвестного параметра $\tilde{\theta}_i$ с уровнем доверия $P_д$ будет интервал:

$$\left(\hat{\theta}_i - y\sqrt{\left\|(Z^T Z)^{-1}\right\|_{i,i} \frac{\delta(\hat{\theta})}{n-m}}; \hat{\theta}_i + y\sqrt{\left\|(Z^T Z)^{-1}\right\|_{i,i} \frac{\delta(\hat{\theta})}{n-m}}\right),$$

где y – квантиль уровня $\dfrac{1+P_д}{2}$ распределения Стьюдента $T(n-m)$.

Доверительный интервал для остаточной дисперсии σ^2.

Построение доверительного интервала для неизвестной остаточной дисперсии σ^2 основывается на пункте 3 теоремы 6.9, согласно которому статистика $\dfrac{\delta(\hat{\theta})}{\sigma^2}$ имеет распределение $\chi^2(n-m)$. Легко видеть, что доверительным интервалом для σ^2 с уровнем доверия $P_д$ является интервал:

$$\left(\frac{\delta(\hat{\theta})}{y_2}; \frac{\delta(\hat{\theta})}{y_1}\right),$$

где y_1 и y_2 – квантили уровней $\dfrac{1-P_д}{2}$ и $\dfrac{1+P_д}{2}$ распределения $\chi^2(n-m)$.

Доверительная область для вектора $\tilde{\theta}$.

Поскольку $\hat{\theta}$ является решением нормального уравнения, то, как было показано при доказательстве утверждения 6.3 пункт 1, для любого вектора параметров θ:

$$\delta(\theta) = \delta(\hat{\theta}) + (\hat{\theta} - \theta)^T Z^T Z(\hat{\theta} - \theta),$$

тогда для вектора $\tilde{\theta}$,

$$\delta(\tilde{\theta}) - \delta(\hat{\theta}) = (\hat{\theta} - \tilde{\theta})^T Z^T Z(\hat{\theta} - \tilde{\theta}),$$

отсюда,

$$\frac{\delta(\tilde{\theta}) - \delta(\hat{\theta})}{\delta(\hat{\theta})} = \frac{(\hat{\theta} - \tilde{\theta})^T Z^T Z(\hat{\theta} - \tilde{\theta})}{\delta(\hat{\theta})},$$

$$\frac{\frac{1}{m}\frac{\delta(\tilde{\theta})-\delta(\hat{\theta})}{\sigma^2}}{\frac{1}{n-m}\frac{\delta(\hat{\theta})}{\sigma^2}}=\frac{n-m}{m}\frac{(\hat{\theta}-\tilde{\theta})^T Z^T Z(\hat{\theta}-\tilde{\theta})}{\delta(\hat{\theta})} \qquad (6.20)$$

Заметим, что по теореме 6.9 из пункта 3 следует, что случайная величина $\frac{\delta(\hat{\theta})}{\sigma^2}$ имеет распределение $\chi^2(n-m)$, из пункта 2 следует, что случайная величина $\frac{\delta(\tilde{\theta})-\delta(\hat{\theta})}{\sigma^2}$ имеет распределение $\chi^2(m)$ и согласно пункту 6 случайные величины $\delta(\hat{\theta})$ и $\delta(\tilde{\theta})-\delta(\hat{\theta})$ независимы, тогда статистика в левой части равенства (6.20) по определению имеет распределение Фишера $F(m,n-m)$, отсюда следует, что такое же распределение имеет и правая часть (6.20):

$$\frac{n-m}{m}\frac{(\hat{\theta}-\tilde{\theta})^T Z^T Z(\hat{\theta}-\tilde{\theta})}{\delta(\hat{\theta})}\sim F(m,n-m)$$

Пусть $P_д>0$ есть некоторый уровень доверия и $y(P_д)$ обозначает квантиль уровня $P_д$ распределения Фишера $F(m,n-m)$, тогда:

$$P\left\{\omega:\frac{n-m}{m}\frac{(\hat{\theta}(\omega)-\tilde{\theta})^T Z^T Z(\hat{\theta}(\omega)-\tilde{\theta})}{\delta(\hat{\theta}(\omega))}<y(P_д)\right\}=F_{m,n-m}(y(P_д))=P_д,$$

где $F_{m,n-m}(x)$ – функция распределения для распределения Фишера $F(m,n-m)$. Отсюда следует, что:

$$P\left\{(\hat{\theta}-\tilde{\theta})^T Z^T Z(\hat{\theta}-\tilde{\theta})<\frac{m}{n-m}\delta(\hat{\theta})y(P_д)\right\}=P_д \qquad (6.21)$$

Рассмотрим область $G(\hat{\theta},C)$:

$$G(\hat{\theta},C)=\{\theta:(\hat{\theta}-\theta)^T Z^T Z(\hat{\theta}-\theta)<C\}.$$

Заметим, что область $G(\hat{\theta},C)$ является внутренностью эллипсоида с центром в точке $\hat{\theta}$: действительно, если ранг матрицы Z равен m, то согласно утверждению 6.2 матрица $Z^T Z$ является положительно определенной, откуда следует, что геометрическое место точек θ удовлетворяющих уравнению $(\hat{\theta}-\theta)^T Z^T Z(\hat{\theta}-\theta)=C$, где $C>0$, является эллипсоидом с центром в точке $\hat{\theta}$.

Таким образом, (6.21) означает, что при всяком векторе $\tilde{\theta}$:

$$P\left\{\tilde{\theta}\in G\left(\hat{\theta},\frac{m}{n-m}\delta(\hat{\theta})y(P_д)\right)\right\}=P_д.$$

Отсюда следует, что область $G\left(\hat{\theta},\frac{m}{n-m}\delta(\hat{\theta})y(P_д)\right)$ является доверительной областью для $\tilde{\theta}$ с уровнем доверия $P_д$.

Критерий проверки гипотезы об отсутствии зависимости.

Если в (6.1) все величины $\tilde{\theta}_i$ равны нулю, то это означает отсутствие зависимости между случайной величиной $\eta(\omega)$ и случайными величинами $\varphi_i(\xi_1(\omega),\ldots,\ \xi_k(\omega))$. Пусть рассматривается гипотеза H_0 об отсутствии зависимости:

$$H_0:\tilde{\theta}_1=0,\ \ldots,\ \tilde{\theta}_m=0,$$

и требуется построить критерий проверки гипотезы H_0. Из теоремы 6.8 следует, что статистика:

$$\frac{\frac{1}{m}\frac{\delta(\tilde{\theta})-\delta(\hat{\theta})}{\sigma^2}}{\frac{1}{n-m}\frac{\delta(\hat{\theta})}{\sigma^2}}=\frac{n-m}{m}\frac{(\hat{\theta}-\tilde{\theta})^T Z^T Z(\hat{\theta}-\tilde{\theta})}{\delta(\hat{\theta})}$$

имеет распределение Фишера $F(m,n-m)$ при всяком параметре $\tilde{\theta}$. Отсюда, если гипотеза H_0 верна, тогда статистика:

$$T(\eta_1,..,\eta_n)=\frac{n-m}{m}\frac{\hat{\theta}^T Z^T Z\hat{\theta}}{\delta(\hat{\theta})}=\frac{n-m}{m}\frac{(Z\hat{\theta})^T Z\hat{\theta}}{\delta(\hat{\theta})}=\frac{n-m}{m}\frac{\left\|Z\hat{\theta}\right\|_2^2}{\delta(\hat{\theta})},$$

(где $\hat{\theta}=(Z^T Z)^{-1}Z^T\eta$, $\eta=(\eta_1,..,\eta_n)$)

имеет распределение Фишера $F(m,n-m)$.

Большие значения статистики $T(\eta_1,..,\eta_n)$ свидетельствуют против гипотезы H_0, действительно, если гипотеза H_0 не верна, то есть $\tilde{\theta}\neq\overline{0}_m$, тогда вектор наблюдаемых величин $\eta=Z\tilde{\theta}+\varepsilon$ имеет смещение $M[\eta]=Z\tilde{\theta}$. Если ранг матрицы Z равен m, то для всякого ненулевого вектора $\tilde{\theta}$ вектор $Z\tilde{\theta}$ тоже ненулевой (в противном случае, если $Z\tilde{\theta}=\overline{0}_n$, то столбцы матрицы Z линейно зависимы и ранг матрицы Z не может быть равен m). Таким образом, вектор η имеет ненулевое смещение $M[\eta]=Z\tilde{\theta}$ и с большой вероятностью попадает в окрестность ненулевого вектора $Z\tilde{\theta}$. Оценка по методу наименьших квадратов $\hat{\theta}$ минимизирует величину среднеквадратичного отклонения $\delta(\hat{\theta})=\left\|\eta-Z\hat{\theta}\right\|_2^2$, поэтому вектор регрессионных значений $Z\hat{\theta}$ оказывается близким к вектору η и тоже с большой вероятностью оказывается в окрестности ненулевого вектора $Z\tilde{\theta}$, отсюда величина $\left\|Z\hat{\theta}\right\|_2^2$ не является малой, а величина $\delta(\hat{\theta})=\left\|\eta-Z\hat{\theta}\right\|_2^2$ является малой, поэтому отношение $\frac{\left\|Z\hat{\theta}\right\|_2^2}{\delta(\hat{\theta})}$ не является величиной близкой к нулю и умножается на возрастающий при $n\to\infty$ коэффициент $\frac{n-m}{m}$. Таким образом, если гипотеза H_0 не верна, то статистика $T(\eta_1,..,\eta_n)$ с большой вероятностью принимает значения не из окрестности нуля.

Отсюда следует, что в качестве критической области Γ_α гипотезы H_0 следует выбирать область вида:

$$\Gamma_\alpha=\{t: t=T(\eta_1,..,\eta_n)\geq h_\alpha\},$$

где h_α квантиль распределения $F(m,n-m)$ уровня $1-\alpha$ и α – заданный уровень значимости. Действительно, вероятность отклонения гипотезы H_0 при условии, что H_0 верна:

$$P\{T(\eta_1,..,\eta_n)\in\Gamma_\alpha\}=P\left\{\frac{n-m}{m}\frac{\left\|Z\hat{\theta}\right\|_2^2}{\delta(\hat{\theta})}\geq h_\alpha\right\}=1-P\left\{\frac{n-m}{m}\frac{\left\|Z\hat{\theta}\right\|_2^2}{\delta(\hat{\theta})}<h_\alpha\right\}=$$

$$=1-F_{m,n-m}(h_\alpha)=1-(1-\alpha)=\alpha,$$

где $F_{m,n-m}(x)$ – функция распределения для распределения Фишера $F(m, n-m)$.