Множественная регрессия

Пусть имеются случайные величины $\xi_1, \dots, \xi_n$ с совместной плотностью $f(x_1, x_2, \dots, x_n)$. Регрессией или условным математическим ожиданием случайной величины ξ_1 при фиксированных значениях случайных величин $\xi_i,\ i = 2, \dots, n$, называется функция, зависящая от переменных $x_2, \dots, x_n$, вида

$$E(\xi_1 | \xi_2 = x_2, \dots, \xi_n = x_n) = \frac{\int\limits_{-\infty}^{\infty} x_1 f(x_1, \dots, x_n) dx_1}{\int\limits_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1}$$

Поверхность регрессии описывается уравнением $x_1 = m_1(x_2, \dots, x_n)$.

Линейная среднеквадратичная регрессия

Для упрощения выводов будем считать $\mathrm{E}\xi_i = 0$. Формулы для произвольного центра тяжести будут получаться простой заменой ξ_i на $\xi_i - m_i$. Среднеквадратическая регрессия величины ξ_1 относительно $\xi_2, \ldots, \xi_n$ определяется как гиперплоскость

$$\xi_1 = \beta_{12}\xi_2 + \ldots + \beta_{1n}\xi_n, \qquad (1)$$

дающая наилучшую аппроксимацию математического ожидания квадрата разности

$$\mathrm{E}(\xi_1 - \beta_{12}\xi_2 - \ldots - \beta_{1n}\xi_n)^2 \to \min,$$

т.е. правая часть (1) является наилучшей линейной оценкой ξ_1 величинами $\xi_2, \ldots, \xi_n$ в смысле минимума (??). Коэффициенты β_{1i} называются частными коэффициентами регрессии.

Вычисление частных коэффициентов регрессии

Для их вычисления продифференцируем $\mathrm{E}(\xi_1 - \beta_{12}\xi_2 - \ldots - \beta_{1n}\xi_n)^2$ по β_{1i}.

$$\begin{cases} \frac{d}{d\beta_{12}}: \quad -2\mathrm{E}\xi_2(\xi_1 - \beta_{12}\xi_2 - \ldots - \beta_{1n}\xi_n) = 0 \\ \qquad\qquad \ldots \\ \frac{d}{d\beta_{1n}}: \quad -2\mathrm{E}\xi_n(\xi_1 - \beta_{12}\xi_2 - \ldots - \beta_{1n}\xi_n) = 0 \end{cases}$$

Это равносильно при $\lambda_{ij} = \mathrm{E}\xi_i\xi_j$ системе уравнений

$$\begin{cases} \lambda_{22}\beta_{12} + \ldots + \lambda_{2n}\beta_{1n} = \lambda_{21} \\ \qquad \ldots \\ \lambda_{n2}\beta_{12} + \ldots + \lambda_{nn}\beta_{1n} = \lambda_{n1} \end{cases} \tag{2}$$

Определитель этой системы равен алгебраическому дополнению Λ_{11} матрицы вторых моментов

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \ldots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & \ldots & \lambda_{2n} \\ & & \ldots & \\ \lambda_{n1} & \lambda_{n2} & \ldots & \lambda_{nn} \end{bmatrix}$$

По теореме Крамера, β_{12} равно отношению определителя

$$\begin{vmatrix} \lambda_{21} & \lambda_{23} & \dots & \lambda_{2\mathrm{n}} \\ & & \dots & \\ \lambda_{\mathrm{n}1} & \lambda_{\mathrm{n}3} & \dots & \lambda_{\mathrm{nn}} \end{vmatrix},$$

который равен $-\Lambda_{12}$, к определителю Λ_{11}, т.е.
$\beta_{12} = -\Lambda_{12}/\Lambda_{11}$.

Коэффициент β_{13} равен отношению определителя

$$\begin{vmatrix} \lambda_{22} & \lambda_{21} & \lambda_{24} & \dots & \lambda_{2\mathrm{n}} \\ & & & \dots & \\ \lambda_{\mathrm{n}2} & \lambda_{\mathrm{n}1} & \lambda_{\mathrm{n}4} & \dots & \lambda_{\mathrm{nn}} \end{vmatrix}, \tag{3}$$

полученного в результате замены в матрице Λ_{11} второго столбца на столбец свободных членов системы (2), к определителю Λ_{11}. Поменяем местами первые два столбца. Определитель (3) равен

$$-\begin{vmatrix} \lambda_{21} & \lambda_{22} & \lambda_{24} & \dots & \lambda_{2\mathrm{n}} \\ & & & \dots & \\ \lambda_{\mathrm{n}1} & \lambda_{\mathrm{n}2} & \lambda_{\mathrm{n}4} & \dots & \lambda_{\mathrm{nn}} \end{vmatrix} = -\Lambda_{13}, \tag{4}$$

т.е.

$$\beta_{13} = -\frac{\Lambda_{13}}{\Lambda_{11}}.$$

Аналогично получаем, что

$$\beta_{1k} = -\frac{\Lambda_{1k}}{\Lambda_{11}}. \tag{5}$$

Простой перестановкой индексов получаем

$$\beta_{ik} = -\frac{\Lambda_{ik}}{\Lambda_{ii}}. \tag{6}$$

Остатки и остаточная дисперсия

Ограничимся рассмотрением случая, когда $|\Lambda_{11}| \neq 0$ и все $\mathrm{E}\xi_{\mathrm{i}} = 0$. Случайная величина вида

$$\eta_{1\cdot 23\ldots \mathrm{n}} = \xi_1 - \beta_{12}\xi_2 - \ldots - \beta_{1\mathrm{n}}\xi_{\mathrm{n}} \tag{7}$$

называется остатком величины ξ_1 относительно $\xi_2, \ldots, \xi_{\mathrm{n}}$. Покажем, что остаток не коррелирован ни с одной из "вычитаемых" случайных величин. Из $\beta_{1\mathrm{k}} = -\frac{\Lambda_{1\mathrm{k}}}{\Lambda_{11}}$ и из того, что $1 = \Lambda_{11}/\Lambda_{11}$ имеем выражение остатка через линейную комбинацию случайных величин

$$\eta_{1\cdot 23\ldots \mathrm{n}} = \frac{1}{\Lambda_{11}} \sum_{\mathrm{k}=1}^{\mathrm{n}} \Lambda_{1\mathrm{k}}\xi_{\mathrm{k}}. \tag{8}$$

Отсюда, с учетом того, что $\mathrm{E}\xi_{\mathrm{i}} = 0$, получаем

$$\mathrm{E}\eta_{1\cdot 23\ldots \mathrm{n}} = 0. \tag{9}$$

Остаточная дисперсия

$$
\begin{aligned}
\mathrm{E}\xi_i\eta_{1\cdot 23\ldots n} &= \frac{1}{\Lambda_{11}}\sum_{k=1}^{n}\mathrm{E}\xi_i\xi_k\Lambda_{1k} = \\
&= \frac{1}{\Lambda_{11}}\sum_{k=1}^{n}\lambda_{ik}\Lambda_{1k} = \begin{cases} \frac{\Lambda}{\Lambda_{11}}, & i = 1, \\ 0, & i \neq 1, \end{cases}
\end{aligned} \qquad (10)
$$

Так как $\sum_{k=1}^{n}\lambda_{1k}\Lambda_{1k} = \Lambda$, а $\sum_{k=1}^{n}\lambda_{ik}\Lambda_{1k} = 0$. Дисперсия остатка

$$\sigma^2_{1\cdot 23\ldots n} = \mathrm{E}(\eta^2_{1\cdot 23\ldots n}).$$

Заменим одну из η на $\eta_{1\cdot 23\ldots n} = \frac{1}{\Lambda_{11}}\sum_{k=1}^{n}\Lambda_{1k}\xi_k.$. В силу (10)

$$\sigma^2_{1\cdot 23\ldots n} = \mathrm{E}\left(\frac{1}{\Lambda_{11}}\sum_{i=1}^{n}\Lambda_{1i}\xi_i\eta_{1\cdot 23\ldots n}\right) = \mathrm{E}(\xi_1\eta_{1\cdot 23\ldots n}) = \frac{\Lambda}{\Lambda_{11}}. \qquad (11)$$

Коэффициент корреляции между остатками

$$\varrho_{12\cdot 34\ldots n} = \mathrm{cor}(\eta_{1\cdot 34\ldots n}\eta_{2\cdot 34\ldots n}) = \frac{\mathrm{E}(\eta_{1\cdot 34\ldots n}\eta_{2\cdot 34\ldots n})}{\sqrt{\mathrm{E}(\eta_{1\cdot 34\ldots n})^2\mathrm{E}(\eta_{2\cdot 34\ldots n})^2}} \quad (12)$$

$$\mathrm{E}\eta^2_{1\cdot 34\ldots n} = \mathrm{E}(\xi_1\eta_{1\cdot 34\ldots n}) = \frac{\Lambda_{22}}{\Lambda_{22\cdot 11}},\ \mathrm{E}\eta^2_{2\cdot 34\ldots n} = \frac{\Lambda_{11}}{\Lambda_{11\cdot 22}},$$

$$\mathrm{E}(\eta_{1\cdot 34\ldots n}\eta_{2\cdot 34\ldots n}) = \mathrm{E}(\xi_1\eta_{2\cdot 34\ldots n}), \quad (13)$$

$$\text{так как } \eta_{1\cdot 34\ldots n} = \frac{1}{\Lambda_{22\cdot 11}}\sum_{k=1,k\neq 2}^{n}\Lambda_{22\cdot 1k}\xi_k,$$

$$\mathrm{E}\xi_k\eta_{2\cdot 34\ldots n}\begin{cases} = 0, & k = 3, 4, \ldots, n \\ \neq 0, & k = 2 \\ \neq 0, & k = 1 \end{cases}$$

Так как $\eta_{2\cdot 34\ldots n} = \frac{1}{\Lambda_{11\cdot 22}} \sum\limits_{k=2}^{n} \Lambda_{11\cdot 2k}\xi_k$, то (13) имеет вид

$$\mathrm{E}(\xi_1\eta_{2\cdot 34\ldots n}) = \frac{1}{\Lambda_{11\cdot 22}} \sum_{k=2}^{n} \Lambda_{11\cdot 2k}\mathrm{E}\xi_1\xi_k =$$

$$= \frac{1}{\Lambda_{11\cdot 22}} \sum_{k=2}^{n} \Lambda_{11\cdot 2k}\lambda_{1k} = -\frac{\Lambda_{12}}{\Lambda_{11\cdot 22}}.$$

Таким образом,

$$\varrho_{12\cdot 34\ldots n} = -\frac{\Lambda_{12}}{\sqrt{\Lambda_{11}\Lambda_{22}}} \tag{14}$$

Между частной корреляцией и соответствующим частным коэффициентом регрессии имеет место линейная зависимость. Согласно (14) и (5) имеем

$$\varrho_{12\cdot 34\ldots \mathrm{n}} = -\frac{\Lambda_{12}}{\sqrt{\Lambda_{11}\Lambda_{22}}} = -\frac{\Lambda_{12}}{\Lambda_{11}}\sqrt{\frac{\Lambda_{11}}{\Lambda_{22}}} = \beta_{12\cdot 34\ldots \mathrm{n}}\sqrt{\frac{\Lambda_{11}}{\Lambda_{22}}}. \qquad (15)$$

Для случая $\mathrm{n} = 3$ матрица вторых моментов Λ имеет вид

$$\Lambda = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\varrho_{12} & \sigma_1\sigma_3\varrho_{13} \\ \sigma_1\sigma_2\varrho_{12} & \sigma_2^2 & \sigma_2\sigma_3\varrho_{23} \\ \sigma_1\sigma_3\varrho_{13} & \sigma_2\sigma_3\varrho_{23} & \sigma_3^2 \end{bmatrix}$$

$$\varrho_{12\cdot 3} = -\frac{\Lambda_{12}}{\sqrt{\Lambda_{11}\Lambda_{22}}} = \frac{\sigma_1\sigma_2\varrho_{12}\sigma_3^2 - \sigma_1\sigma_3\varrho_{13}\sigma_2\sigma_3\varrho_{23}}{\sigma_2\sigma_3\sqrt{1-\varrho_{23}^2}\,\sigma_1\sigma_3\sqrt{1-\varrho_{13}^2}} =$$

$$= \frac{\varrho_{12} - \varrho_{13}\varrho_{23}}{\sqrt{1-\varrho_{23}^2}\sqrt{1-\varrho_{13}^2}}$$

ПРИМЕР

j	Годы	Продолж-ть жизни ξ_1	Нац.богатсво (млрд.дол.) ξ_2	Водка (бут.в год) ξ_3
1	1970	68.9	2372*	25.3
2	1975	68.1	2372*	28
3	1980	67.6	2489	30
4	1985	69.2	3379	23.5
5	1990	69.2	4130	18
6	1995	64.6	1171	38.4
7	1998	67	689	29.6

$$\varrho_{12} = 0.78,\ \varrho_{13} = -0.93,\ \varrho_{23} = -0.82.\ \sigma^2_{1\cdot 23} = \frac{\Lambda}{\Lambda_{11}} = 0.29.$$

$$\beta_{12} = -\frac{\Lambda_{12}}{\Lambda_{11}} = -\frac{-368}{13214148} = 2.8\cdot 10^{-5},\quad \varrho_{12\cdot 3} = -\frac{\Lambda_{12}}{\sqrt{\Lambda_{11}\Lambda_{22}}} = 0.032,$$

$$\beta_{13} = -\frac{\Lambda_{13}}{\Lambda_{11}} = -\frac{138823}{13214148} = -0.24,\quad \varrho_{13\cdot 2} = -\frac{\Lambda_{13}}{\sqrt{\Lambda_{11}\Lambda_{33}}} = -0.82,$$

Корреляция $\varrho_{12} = 0.78$ объясняется тем, что при снижении национального богатства увеличивается продажа водки, что влечет снижение продолжительности жизни.

Множественный коэффициент корреляции

Множественным коэффициентом корреляции $\varrho_{1(23\ldots\mathrm{n})}$ называется коэффициент корреляции между ξ_1 и $\xi_1^* = \beta_{12}\xi_2 + \ldots + \beta_{1\mathrm{n}}\xi_\mathrm{n} = \xi_1 - \eta_{1\cdot 23\ldots\mathrm{n}}$. При $\mathrm{E}\xi_\mathrm{i} = 0$

$$\varrho_{1(23\ldots\mathrm{n})} = \frac{\mathrm{E}(\xi_1\xi_1^*)}{\sqrt{\mathrm{E}\xi_1^2\mathrm{E}(\xi_1^*)^2}}, \tag{16}$$

$$\mathrm{E}(\xi_1\xi_1^*) = \mathrm{E}(\xi_1(\xi_1 - \eta_{1\cdot 23\ldots\mathrm{n}})) = \mathrm{E}\xi_1^2 - \mathrm{E}\xi_1\eta_{1\cdot 23\ldots\mathrm{n}} = \lambda_{11} - \frac{\Lambda}{\Lambda_{11}},$$

$$\mathrm{E}(\xi_1^*)^2 = \mathrm{E}(\xi_1^2 - 2\xi_1\eta_{1\cdot 23\ldots\mathrm{n}} + \eta_{1\cdot 23\ldots\mathrm{n}}^2) = \lambda_{11} - \frac{\Lambda}{\Lambda_{11}}.$$

Следовательно,

$$\varrho_{1(23\ldots\mathrm{n})} = \frac{\sqrt{\lambda_{11} - \frac{\Lambda}{\Lambda_{11}}}}{\sqrt{\lambda_{11}}} = \sqrt{1 - \frac{\Lambda}{\Lambda_{11}\lambda_{11}}}. \tag{17}$$

Для численных вычислений оказывается полезной формула

$$\varrho^2_{1(23\ldots \mathrm{n})} = 1 - \frac{\sigma^2_{1\cdot 23\ldots \mathrm{n}}}{\sigma^2_1} = \frac{\sigma^2_1 - \sigma^2_{1\cdot 23\ldots \mathrm{n}}}{\sigma^2_1}, \tag{18}$$

так как по (11) $\sigma^2_{1\cdot 23\ldots \mathrm{n}} = \frac{\Lambda}{\Lambda_{11}}$, а $\lambda_{11} = \sigma^2_1$.
Квадрат множественного коэффициента корреляции или коэффициент детерминации равен доле дисперсии, объясняемой регрессией, то есть влиянием случайных величин $\xi_2, \ldots, \xi_\mathrm{n}$ на ξ_1.
ПРИМЕР (продолжение). Коэффициент детерминации равен

$$\mathrm{R}^2 = \varrho^2_{1(23)} = \frac{\sigma^2_1 - \sigma^2_{1\cdot 23}}{\sigma^2_1} = \frac{2.31 - 0.29}{2.31} = 0.87.$$

Отношение дисперсий и частный коэффициент корреляции

$$\frac{\beta_{12}^2\sigma_{2\cdot 34\ldots \mathrm{n}}^2}{\sigma_{1\cdot 234\ldots \mathrm{n}}^2} = \frac{\varrho_{12\cdot 34\ldots \mathrm{n}}^2}{1-\varrho_{12\cdot 34\ldots \mathrm{n}}^2} \tag{19}$$

Равенство Якоби: $\Lambda\cdot\Lambda_{11\cdot 22} = \Lambda_{11}\cdot\Lambda_{22} - \Lambda_{12}\cdot\Lambda_{21}$

$$\beta_{12} = \frac{\Lambda_{12}}{\Lambda_{11}}, \quad \sigma_{1\cdot 23\ldots \mathrm{n}}^2 = \frac{\Lambda}{\Lambda_{11}}, \quad \sigma_{2\cdot 34\ldots \mathrm{n}}^2 = \frac{\Lambda_{11}}{\Lambda_{11\cdot 22}}, \quad \varrho_{12\cdot 34\ldots \mathrm{n}}^2 = \frac{\Lambda_{12}^2}{\Lambda_{11}\Lambda_{22}}.$$

получаем левую часть в виде:

$$\frac{\beta_{12}^2\sigma_{2\cdot 34\ldots \mathrm{n}}^2}{\sigma_{1\cdot 234\ldots \mathrm{n}}^2} = \frac{\Lambda_{12}^2}{\Lambda_{11}^2}\cdot\frac{\Lambda_{11}}{\Lambda_{11\cdot 22}}\cdot\frac{\Lambda_{11}}{\Lambda} = \frac{\Lambda_{12}^2}{\Lambda\cdot\Lambda_{11\cdot 22}} = \frac{\Lambda_{12}^2}{\Lambda_{11}\cdot\Lambda_{22}-\Lambda_{12}\cdot\Lambda_{21}}.$$

Правая часть совпадает с левой, так как

$$\frac{\varrho_{12\cdot 34\ldots \mathrm{n}}^2}{1-\varrho_{12\cdot 34\ldots \mathrm{n}}^2} = \frac{\Lambda_{12}^2}{\Lambda_{11}\Lambda_{22}\left(1-\frac{\Lambda_{12}^2}{\Lambda_{11}\Lambda_{22}}\right)}.$$

Равенство Якоби получается из произведения определителей $(\Lambda_{11}\Lambda_{22} - \Lambda_{12}\Lambda_{21})|\Lambda| = |\Lambda|^2\Lambda_{11\cdot 22}$ соответствующих матриц

$$\left[\begin{array}{cc|ccc} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} & \dots & \Lambda_{1n} \\ \Lambda_{21} & \Lambda_{22} & \Lambda_{23} & \dots & \Lambda_{2n} \\ \hline \multicolumn{2}{c|}{\mathbb{O}} & \multicolumn{3}{c}{\mathbb{I}} \end{array}\right] \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2n} \\ \dots & \dots & \dots & \dots \\ \lambda_{n1} & \lambda_{n2} & \dots & \lambda_{nn} \end{bmatrix} =$$

$$= \left[\begin{array}{cc|ccc} |\Lambda| & 0 & 0 & \dots & 0 \\ 0 & |\Lambda| & 0 & \dots & 0 \\ \hline \lambda_{31} & \lambda_{32} & \lambda_{33} & \dots & \lambda_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ \lambda_{n1} & \lambda_{n2} & \lambda_{3n} & \dots & \lambda_{nn} \end{array}\right]$$

Функции в R: $LM <- lm(suicide \sim vodka + auto + manager, life)$

	suicide	vodka	auto	manager
1	58027	25.30	5.50	1060
2	58027	28.00	15.30	1101
3	48195	30.00	30.20	1147
4	44562	23.50	44.50	1204
5	39150	18.00	58.60	1602
6	60953	38.40	93.30	1893
7	97276	29.60	122.00	2777

summary(LM)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-43103.03	20327.72	-2.12	0.12
vodka	1195.65	516.44	2.32	0.10
auto	-787.56	246.07	-3.20	0.05
manager	71.25	16.14	4.41	0.02

Multiple R-squared: 0.9269, Adjusted R-squared: 0.8539 F-statistic: 12.69 on 3 and 3 DF, p-value: 0.03278

Задача предсказания переменной зависимой y с помощью p независимых переменных $x_1, \ldots, x_p$

y	x_1	$\ldots$	x_p
y_1	x_{11}	$\ldots$	x_{p1}
y_2	x_{12}	$\ldots$	x_{p2}
$\vdots$	$\vdots$	$\ldots$	$\vdots$
y_n	x_{1n}	$\ldots$	x_{pn}

$$y = f(x_1, \ldots, x_p; \beta_1, \ldots, \beta_m) + e,$$

где $\beta_1, \ldots, \beta_m$ – неизвестные параметры, e – ошибка аппроксимации. В частности, если $m = p + 1$

$$f(x_1, \ldots, x_p; \beta_1, \ldots, \beta_m) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p, \qquad (20)$$

то мы имеем модель множественной линейной регрессии.

Нелинейные модели

К линейной модели сводятся многие другие варианты моделей, если независимые переменные являются функциями, например, $y = \beta_0 + \beta_1 \sin z_1 + \beta_2 \sin z_2 + e$ является моделью множественной линейной регрессии с $x_i = \sin z_i$, или $x_i = x^i$, в этом случае к модели множественной линейной регрессии сводится полиномиальная модель.

Модель линейной регрессии

Переменные $x_1, \ldots, x_p$ считаются детерминированными.

$$y_i = \alpha + \beta_1(x_{1i} - \bar{x}_1) + \ldots + \beta_p(x_{pi} - \bar{x}_p) + \delta_i, \quad i = 1, \ldots, n, (21)$$

независимые $\delta_i \sim \mathcal{N}(0, \sigma)$. Минимизирующие

$$S = \sum_{i=1}^{n} (y_i - \alpha - \beta_1(x_{1i} - \bar{x}_1) - \ldots - \beta_p(x_{pi} - \bar{x}_p))^2, \qquad (22)$$

$\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_p$, называются частными коэффициентами регрессии. Модель (21) сводится к (20) выражением $\beta_0 = \alpha - \beta_1 \bar{x}_1 - \ldots - \beta_p \bar{x}_p$.

Обозначим через

$$l_{ij} = \frac{1}{n}\sum_{\nu=1}^{n}(x_{i\nu} - \bar{x}_i)(x_{j\nu} - \bar{x}_j), \tag{23}$$

$$l_{0j} = \frac{1}{n}\sum_{\nu=1}^{n}(y_{\nu} - \bar{y})(x_{j\nu} - \bar{x}_j) \tag{24}$$

оценки вторых центральных моментов, а через L и L_{ij} будем обозначать определитель матрицы вторых выборочных центральных моментов

$$\begin{bmatrix} l_{11} & \dots & l_{1p} \\ \dots & \dots & \dots \\ l_{p1} & \dots & l_{pp} \end{bmatrix}$$

и его алгебраические дополнения. Будем предполагать, что $L \neq 0$.

Оценки по методу наименьших квадратов

По методу наименьших квадратов оценки параметров имеют вид

$$\hat{\alpha} = \bar{y}, \qquad \hat{\beta}_i = \frac{1}{L} \sum_{j=1}^{p} l_{0j} L_{ij}, \tag{25}$$

а оценкой максимального правдоподобия дисперсии является выражение:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{\nu=1}^{n} (y_\nu - \hat{y}_\nu)^2, \tag{26}$$

где через $\hat{y}_\nu = \hat{\alpha} - \hat{\beta}_1(x_{1\nu} - \bar{x}_1) - \ldots - \hat{\beta}_p(x_{p\nu} - \bar{x}_p)$ обозначено наилучшее линейное предсказание переменной y.

1. $E\hat{\alpha} = \alpha$. Покажем, что
$$\hat{\alpha} = \alpha + \bar{\delta}, \tag{27}$$
В самом деле,
$$\hat{\alpha} = \bar{y} = \frac{1}{n}\sum_{\nu=1}^{n}\left(\alpha + \sum_{i=1}^{p}\beta_i(x_{i\nu} - \bar{x}_i) + \delta_\nu\right) = \alpha + \bar{\delta},$$
$$E\hat{\alpha} = E\left(\alpha + \bar{\delta}\right) = \alpha$$
2. $E(\hat{\alpha} - \alpha)^2 = \frac{\sigma^2}{n}$ (очевидно).

$\mathrm{E}\hat{\beta}_{\mathrm{i}} = \beta_{\mathrm{i}}$

Покажем, что из $y_\nu = \alpha + \sum_{k=1}^{p} \beta_k (x_{k\nu} - \bar{x}_k) + \delta_\nu$ и $\bar{y} = \alpha + \bar{\delta}$

$$y_\nu - \bar{y} = \sum_{k=1}^{p} \beta_k (x_{k\nu} - \bar{x}_k) + \delta_\nu - \bar{\delta}.$$

$$\mathrm{E}\hat{\beta}_i = \mathrm{E}\frac{1}{L}\sum_{j=1}^{p} l_{0j} L_{ij} = \mathrm{E}\frac{1}{L}\sum_{j=1}^{p} L_{ij}\frac{1}{n}\sum_{\nu=1}^{n}(y_\nu - \bar{y})(x_{j\nu} - \bar{x}_j) =$$

$$= \mathrm{E}\frac{1}{L}\sum_{j=1}^{p} L_{ij}\frac{1}{n}\sum_{\nu=1}^{n}(x_{j\nu} - \bar{x}_j)\left(\sum_{k=1}^{p} \beta_k (x_{k\nu} - \bar{x}_k) + \delta_\nu - \bar{\delta}\right) =$$

$$= \frac{1}{L}\sum_{j=1}^{p} L_{ij}\sum_{k=1}^{p}\beta_k\frac{1}{n}\sum_{\nu=1}^{n}(x_{j\nu} - \bar{x}_j)(x_{k\nu} - \bar{x}_k) =$$

$$= \frac{1}{L}\sum_{j=1}^{p} L_{ij}\sum_{k=1}^{p}\beta_k l_{kj} = \frac{1}{L}\sum_{k=1}^{p}\beta_k\sum_{j=1}^{p} L_{ij} l_{kj} = \beta_i,$$

$$\text{так как} \quad \sum_{j=1}^{p} L_{ij} l_{kj} = \begin{cases} 0, & k \neq i, \\ L, & k = i \end{cases}$$

$E(\hat\alpha-\alpha)(\hat\beta_i-\beta_i)=0.$

С одной стороны, $\hat\alpha-\alpha=\bar\delta$, с другой, $y_\nu-\bar y-E(y_\nu-\bar y)=\delta_\nu-\bar\delta$,

$$l_{0j}-El_{0j}=\frac{1}{n}\sum_{\nu=1}^{n}(x_{j\nu}-\bar x_j)(y_\nu-\bar y-E(y_\nu-\bar y))=$$

$$=\frac{1}{n}\sum_{\nu=1}^{n}(x_{j\nu}-\bar x_j)(\delta_\nu-\bar\delta)=\frac{1}{n}\sum_{\nu=1}^{n}(x_{j\nu}-\bar x_j)\delta_\nu.$$

Таким образом,

$$\hat\beta_i-\beta_i=\frac{1}{L}\sum_{j=1}^{p}L_{ij}(l_{0j}-El_{0j})=\frac{1}{L}\sum_{j=1}^{p}L_{ij}\frac{1}{n}\sum_{\nu=1}^{n}(x_{j\nu}-\bar x_j)\delta_\nu, \tag{28}$$

следовательно,

$$E(\hat\alpha-\alpha)(\hat\beta_i-\beta_i)=\frac{1}{Ln}\sum_{j=1}^{p}L_{ij}\sum_{\nu=1}^{n}(x_{j\nu}-\bar x_j)E(\bar\delta\delta_\nu)=$$

$$=\frac{\sigma^2}{Ln^2}\sum_{j=1}^{p}L_{ij}\sum_{\nu=1}^{n}(x_{j\nu}-\bar x_j)=0.$$

Действительно,

$$E\frac{1}{L}\sum_{j=1}^{p} L_{1j}\frac{1}{n}\sum_{\mu=1}^{n}(x_{j\mu} - \bar{x}_j)\delta_\mu \cdot \frac{1}{L}\sum_{i=1}^{p} L_{2i}\frac{1}{n}\sum_{\nu=1}^{n}(x_{i\nu} - \bar{x}_i)\delta_\nu =$$

$$= \frac{\sigma^2}{n^2L^2}\sum_{\mu=1}^{n}\ \sum_{j=1}^{p} L_{1j}(x_{j\mu} - \bar{x}_j)\ \sum_{i=1}^{p} L_{2i}(x_{i\mu} - \bar{x}_i) =$$

$$= \frac{\sigma^2}{n^2L^2}\sum_{j=1}^{p} L_{1j}\sum_{i=1}^{p} L_{2i}\sum_{\mu=1}^{n}(x_{j\mu} - \bar{x}_j)(x_{i\mu} - \bar{x}_i) =$$

$$= \frac{\sigma^2}{n^2L^2}\sum_{j=1}^{p} L_{1j}\sum_{i=1}^{p} L_{2i} n l_{ij} = \frac{\sigma^2 L_{12}}{nL}.$$

В частности,

$$D(\hat{\beta}_1) = \frac{\sigma^2 L_{11}}{nL} = \frac{\sigma^2}{n s^2_{1\cdot 234\ldots p}}. \tag{29}$$

Разложение суммы квадратов ошибок

$$\sum_{\mu=1}^{n} \delta_\mu^2 = \sum_{\mu=1}^{n} \left(y_\mu - \alpha - \sum_{k=1}^{p} \beta_k (x_{k\mu} - \bar{x}_k) \right)^2 =$$

$$= \sum_{\mu=1}^{n} \left(y_\mu - \alpha - \sum_{k=1}^{p} \beta_k (x_{k\mu} - \bar{x}_k) + \right.$$

$$\left. + \hat{\alpha} - \hat{\alpha} + \sum_{k=1}^{p} \hat{\beta}_k (x_{k\mu} - \bar{x}_k) - \sum_{k=1}^{p} \hat{\beta}_k (x_{k\mu} - \bar{x}_k) \right)^2 =$$

$$= \sum_{\mu=1}^{n} \left((y_\mu - \hat{\alpha} - \sum_{k=1}^{p} \hat{\beta}_k (x_{k\mu} - \bar{x}_k)) + \right.$$

$$\left. + (\hat{\alpha} - \alpha) + \sum_{k=1}^{p} (\hat{\beta}_k - \beta_k)(x_{k\mu} - \bar{x}_k) \right)^2 =$$

$$= n\hat{\sigma}^2 + n(\hat{\alpha} - \alpha)^2 + n \sum_{k=1}^{p} \sum_{t=1}^{p} l_{kt} (\hat{\beta}_k - \beta_k)(\hat{\beta}_t - \beta_t). \tag{30}$$

Вывод последнего слагаемого

$$\sum_{\mu=1}^{n}\left(\sum_{k=1}^{p}(\hat{\beta}_k-\beta_k)(x_{k\mu}-\bar{x}_k)\right)^2 =$$

$$=\sum_{\mu=1}^{n}\left(\sum_{k=1}^{p}(\hat{\beta}_k-\beta_k)(x_{k\mu}-\bar{x}_k)\right)\left(\sum_{t=1}^{p}(\hat{\beta}_t-\beta_t)(x_{t\mu}-\bar{x}_t)\right) =$$

$$=\sum_{k=1}^{p}\sum_{t=1}^{p}(\hat{\beta}_k-\beta_k)(\hat{\beta}_t-\beta_t)\sum_{\mu=1}^{n}(x_{k\mu}-\bar{x}_k)(x_{t\mu}-\bar{x}_t) =$$

$$=\sum_{k=1}^{p}\sum_{t=1}^{p}(\hat{\beta}_k-\beta_k)(\hat{\beta}_t-\beta_t)nl_{kt}$$

Перекрестные суммы равны нулю

$$\sum_{\mu=1}^{n}\left(y_\mu - \hat{\alpha} - \sum_{k=1}^{p}\hat{\beta}_k(x_{k\mu} - \bar{x}_k)\right)\left(\sum_{t=1}^{p}(\hat{\beta}_t - \beta_t)(x_{t\mu} - \bar{x}_t)\right) =$$

$$= \sum_{t=1}^{p}(\hat{\beta}_t - \beta_t)\left(\sum_{\mu=1}^{n}(y_\mu - \bar{y})(x_{t\mu} - \bar{x}_t) - \sum_{k=1}^{p}\hat{\beta}_k\sum_{\mu=1}^{n}(x_{k\mu} - \bar{x}_k)(x_{t\mu} - \bar{x}_t)\right)$$

$$\sum_{t=1}^{p}(\hat{\beta}_t - \beta_t)n\left(l_{0t} - \sum_{k=1}^{p}\hat{\beta}_k l_{kt}\right) = 0, \text{ так как}$$

$$\sum_{k=1}^{p}\hat{\beta}_k l_{kt} = \sum_{k=1}^{p} l_{kt}\frac{1}{L}\sum_{j=1}^{p} L_{kj} l_{0j} = \frac{1}{L}\sum_{j=1}^{p} l_{0j}\sum_{k=1}^{p} l_{kt} L_{kj} = l_{0t}.$$

Перекрестные суммы равны нулю

$$\sum_{\mu=1}^{n}\left(\sum_{t=1}^{p}(\hat{\beta}_t-\beta_t)(x_{t\mu}-\bar{x}_t)\right)(\hat{\alpha}-\alpha)=0,$$

$$\sum_{\mu=1}^{n}\left(y_\mu-\hat{\alpha}-\sum_{k=1}^{p}\hat{\beta}_k(x_{k\mu}-\bar{x}_k)\right)(\hat{\alpha}-\alpha)=$$

$$=(\hat{\alpha}-\alpha)\left(n\bar{y}-n\hat{\alpha}-\sum_{k=1}^{p}\hat{\beta}_k\sum_{\mu=1}^{n}(x_{k\mu}-\bar{x}_k)\right)=0,$$

Распределения частных коэффициентов регрессии: Соотношения ортогональности для статистик

Оценкой σ^2 является выборочное значение остаточной дисперсии величины y относительно $x_1, \ldots, x_p$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{\nu=1}^{n} \left(y_\nu - \hat{\alpha} - \sum_{k=1}^{p} \hat{\beta}_k (x_{k\mu} - \bar{x}_k) \right)^2 = s^2_{0 \cdot 12 \ldots p}. \qquad (31)$$

Покажем, что $\hat{\sigma}^2$ независима от $\hat{\alpha}$ и $\hat{\beta}_i$, и что $\frac{n\hat{\sigma}^2}{\sigma^2}$ имеет распределение хи-квадрат с $n - p - 1$ степенями свободы.

Случай диагональной матрицы вторых моментов L.

$$\mathrm{l_{ij}} = 0, \quad \mathrm{L_{ij}} = 0 \quad \text{при } \mathrm{i} \neq \mathrm{j}. \tag{32}$$

т.е. $\mathrm{s_i^2} = \mathrm{l_{ii}} = \frac{\mathrm{L}}{\mathrm{L_{ii}}}$, $\mathrm{D}(\beta_\mathrm{i}) = \frac{\sigma^2}{\mathrm{ns_i^2}}$. Выражения $\sqrt{\mathrm{n}}(\hat{\alpha} - \alpha)$ и $\mathrm{s_i}\sqrt{\mathrm{n}}(\hat{\beta}_\mathrm{i} - \beta_\mathrm{i})$ есть линейные комбинации от δ_ν.
Действительно, согласно (27) и (28),

$$\sqrt{\mathrm{n}}(\hat{\alpha} - \alpha) = \frac{1}{\sqrt{\mathrm{n}}} \sum_{\nu=1}^{\mathrm{n}} \delta_\nu \text{ и}$$

$$\mathrm{s_i}\sqrt{\mathrm{n}}(\hat{\beta}_\mathrm{i} - \beta_\mathrm{i}) = \mathrm{s_i}\sqrt{\mathrm{n}} \left(\frac{1}{\mathrm{L}} \sum_{\mathrm{j}=1}^{\mathrm{p}} \mathrm{L_{ij}} \frac{1}{\mathrm{n}} \sum_{\mu=1}^{\mathrm{n}} (\mathrm{x_{j\mu}} - \bar{\mathrm{x}}_\mathrm{j}) \delta_\mu \right) =$$

$$= \mathrm{s_i}\sqrt{\mathrm{n}} \left(\frac{\mathrm{L_{ii}}}{\mathrm{Ln}} \sum_{\mu=1}^{\mathrm{n}} (\mathrm{x_{i\mu}} - \bar{\mathrm{x}}_\mathrm{i}) \delta_\mu \right) = \frac{1}{\mathrm{s_i}\sqrt{\mathrm{n}}} \sum_{\mu=1}^{\mathrm{n}} (\mathrm{x_{i\mu}} - \bar{\mathrm{x}}_\mathrm{i}) \delta_\mu$$

Эти линейные комбинации от δ_ν удовлетворяют условиям ортогональности:

$$\sum_{\mu=1}^{\mathrm{n}} \left(\frac{1}{\sqrt{\mathrm{n}}}\right)^2 = 1,$$

$$\sum_{\mu=1}^{\mathrm{n}} \left(\frac{\mathrm{x}_{\mathrm{i}\mu} - \bar{\mathrm{x}}_{\mathrm{i}}}{\mathrm{s}_{\mathrm{i}}\sqrt{\mathrm{n}}}\right)^2 = 1,$$

$$\sum_{\mu=1}^{\mathrm{n}} \left(\frac{\mathrm{x}_{\mathrm{i}\mu} - \bar{\mathrm{x}}_{\mathrm{i}}}{\mathrm{s}_{\mathrm{i}}\sqrt{\mathrm{n}}\sqrt{\mathrm{n}}}\right) = 0.$$

$$\sum_{\mu=1}^{n} \delta_{\mu}^{2} = n\hat{\sigma}^{2} + n(\hat{\alpha} - \alpha)^{2} + n\sum_{k=1}^{p}\sum_{t=1}^{p} l_{kt}(\hat{\beta}_{k} - \beta_{k})(\hat{\beta}_{t} - \beta_{t})$$

В случае диагональной матрицы вторых моментов (32)

$$n\sum_{k=1}^{p}\sum_{t=1}^{p} l_{kt}(\hat{\beta}_{k} - \beta_{k})(\hat{\beta}_{t} - \beta_{t}) = \sum_{k=1}^{p} ns_{k}^{2}(\hat{\beta}_{k} - \beta_{k})^{2},\ \text{поэтому}$$

$$\sum_{\mu=1}^{n} \delta_{\mu}^{2} = n\hat{\sigma}^{2} + n(\hat{\alpha} - \alpha)^{2} + \sum_{k=1}^{p} ns_{k}^{2}(\hat{\beta}_{k} - \beta_{k})^{2}$$

C ортогональная матрица со специальными $\mathrm{p}+1$ строками

$$\mathrm{C}=\begin{bmatrix} \frac{1}{\sqrt{\mathrm{n}}} & \dots & \frac{1}{\sqrt{\mathrm{n}}} \\ \frac{\mathrm{x}_{11}-\bar{\mathrm{x}}_1}{\mathrm{s}_1\sqrt{\mathrm{n}}} & \dots & \frac{\mathrm{x}_{1\mathrm{n}}-\bar{\mathrm{x}}_1}{\mathrm{s}_1\sqrt{\mathrm{n}}} \\ \vdots & \dots & \vdots \\ \frac{\mathrm{x}_{\mathrm{p}1}-\bar{\mathrm{x}}_\mathrm{p}}{\mathrm{s}_\mathrm{p}\sqrt{\mathrm{n}}} & \dots & \frac{\mathrm{x}_{\mathrm{pn}}-\bar{\mathrm{x}}_\mathrm{p}}{\mathrm{s}_\mathrm{p}\sqrt{\mathrm{n}}} \\ \vdots & \dots & \vdots \end{bmatrix}$$

Рассмотрим вектор $\eta=(\eta_1,\dots\eta_\mathrm{n})^\mathrm{T}$, который получен из $\eta=\mathrm{C}\delta$, где $\delta=(\delta_1,\dots,\delta_\mathrm{n})^\mathrm{T}$, очевидно,
$\eta_1=\sqrt{\mathrm{n}}(\hat{\alpha}-\alpha)$, $\eta_\mathrm{k}=\sqrt{\mathrm{n}}\mathrm{s}_\mathrm{k}(\hat{\beta}_\mathrm{k}-\beta_\mathrm{k})$, $\quad \mathrm{k}=2,\dots,\mathrm{p}+1.$

$$\sum_{\mathrm{j}=1}^{\mathrm{n}}\eta_\mathrm{j}^2=\eta^\mathrm{T}\eta=(\mathrm{C}\delta)^\mathrm{T}\mathrm{C}\delta=\delta^\mathrm{T}\mathrm{C}^\mathrm{T}\mathrm{C}\delta=\delta^\mathrm{T}\delta=\sum_{\mathrm{j}=1}^{\mathrm{n}}\delta_\mathrm{j}^2,$$

$$\frac{\mathrm{n}\hat{\sigma}^2}{\sigma^2}=\frac{1}{\sigma^2}\left(\sum_{\mathrm{j}=1}^{\mathrm{n}}\delta_\mathrm{j}^2-\eta_1^2-\sum_{\mathrm{k}=2}^{\mathrm{p}+1}\eta_\mathrm{k}^2\right)\ \sim\ \chi^2(\mathrm{n}-\mathrm{p}-1).$$

Проверка значимости коэффициентов регрессии

В общем случае, когда матрица вторых моментов не является диагональной, величины x_i следует при помощи ортогонального преобразования заменить на величины x_i', для которых матрица вторых моментов будет диагональной. К параметрам β_i следует применить контрградиентное преобразование. Т.е. если $Y = A + B^T(X - \bar{X})$ и $(X - \bar{X}) = CZ$, где Z – вектор с некоррелированными компонентами, то $Y = A + B^TCZ = A + (C^TB)^TZ$ и $B_1 = C^TB$, $B = CB_1$.

Отсюда следует, что величины

$$t_0 = \frac{(\hat{\alpha} - \alpha)\sqrt{n}}{\sigma\sqrt{\frac{1}{n-p-1}\frac{n\hat{\sigma}^2}{\sigma^2}}} = \frac{\hat{\alpha} - \alpha}{s_{0\cdot 12\ldots p}}\sqrt{n - p - 1},$$

$$t_1 = \frac{(\hat{\beta}_1 - \beta_1)\sqrt{n}s_{1\cdot 23\ldots p}}{\sigma\sqrt{\frac{1}{n-p-1}\frac{n\hat{\sigma}^2}{\sigma^2}}} = \frac{(\hat{\beta}_1 - \beta_1)s_{1\cdot 23\ldots p}}{s_{0\cdot 12\ldots p}}\sqrt{n - p - 1}, \quad (33)$$

и аналогичные величины с $\beta_2, \ldots, \beta_p$ имеют распределение Стьюдента с $n - p - 1$ степенями свободы.

Значимость частных коэффициентов корреляции

Гипотеза $H_0 : \beta_1 = 0$ может быть использована для проверки равенства нулю частного коэффициента корреляции, но вместо статистики (33), согласно (19) используется эквивалентная ей статистика вида

$$\frac{\hat{\varrho}_{01\cdot 234\ldots p}}{\sqrt{1 - \varrho^2_{01\cdot 234\ldots p}}}\sqrt{n - p - 1} \sim T(n - p - 1). \qquad (34)$$

Обозначим через $Q_2 = \sum_{\nu=1}^{n}\left(y_\nu - \hat{\alpha} - \sum_{k=1}^{p}\hat{\beta}_k(x_{k\mu} - \bar{x}_k)\right)^2$. Если использовать несмещенную оценку дисперсии $S^2 = \frac{1}{n-p-1}Q_2$, то

$$t_0 = \frac{\hat{\alpha} - \alpha}{S}\sqrt{n},$$

$$t_1 = \frac{(\hat{\beta}_1 - \beta_1)s_{1\cdot 23\ldots p}}{S}\sqrt{n} = \frac{(\hat{\beta}_1 - \beta_1)}{\sigma_{\hat{\beta}_1}}.$$

Выражение $\frac{S}{\sqrt{n}s_{1\cdot 2,3,\ldots,p}}$ соответствует оценке стандартного отклонения σ_{β_1} оценки $\hat{\beta}_1$ из (29). Поэтому при вычислениях используются дисперсии оценок частных коэффициентов регрессии, на корни из которых делятся оценки параметров для получения нужных статистик Стьюдента.

Проверка значимости прогноза $H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$

Убедимся в том, что $\frac{n}{\sigma^2}\sum_{k=1}^{p}\sum_{t=1}^{p} l_{kt}(\hat{\beta}_k - \beta_k)(\hat{\beta}_t - \beta_t)$ имеет распределение хи-квадрат с p степенями свободы.
При диагональной матрице вторых моментов (32) это очевидно,

$$n\sum_{k=1}^{p}\sum_{t=1}^{p} l_{kt}(\hat{\beta}_k - \beta_k)(\hat{\beta}_t - \beta_t) = \sum_{k=1}^{p} n s_k^2(\hat{\beta}_k - \beta_k)^2,$$

$$\text{где} \quad \sqrt{n} s_k(\hat{\beta}_k - \beta_k) \sim \mathcal{N}(0, \sigma).$$

Случай недиагональной L

В общем случае используем утверждение о том, что для любого нормального вектора Z с нулевым средним и ковариационной матрицей $\boldsymbol{\Sigma} = \mathrm{EZZ}^{\mathrm{T}}$ справедливо $\mathrm{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathrm{Z} \ \sim \ \chi^2(\mathrm{p})$.
Пусть $\mathrm{Y} = \mathrm{BZ}$ вектор с независимыми компонентами, так что $\mathrm{Y}^{\mathrm{T}}\mathrm{Y} \ \sim \ \chi^2(\mathrm{p})$. С другой стороны,

$$\mathrm{Y}^{\mathrm{T}}\mathrm{Y} = \mathrm{Z}^{\mathrm{T}}(\mathrm{B}^{\mathrm{T}}\mathrm{B})\mathrm{Z} \ \sim \ \chi^2(\mathrm{p}),$$

а ковариационная матрица вектора Z с коррелируемыми компонентами имеет тот же вид

$$\boldsymbol{\Sigma} = \mathrm{EZZ}^{\mathrm{T}} = \mathrm{EB}^{-1}\mathrm{YY}^{\mathrm{T}}(\mathrm{B}^{-1})^{\mathrm{T}} = \mathrm{B}^{-1}(\mathrm{B}^{-1})^{\mathrm{T}} = (\mathrm{B}^{\mathrm{T}}\mathrm{B})^{-1}.$$

Статистика Фишера

Поскольку ковариационная матрица вектора параметров имеет вид:

$$\mathrm{cov}(\beta) = \left\{ \frac{\sigma^2 \mathrm{L_{ij}}}{\mathrm{nL}} \right\}_{\mathrm{i,j=1}}^{\mathrm{p}} = \frac{\sigma^2}{\mathrm{n}} \mathrm{L}^{-1}, \quad \text{то}$$

$$(\hat{\beta} - \beta)^{\mathrm{T}} \left(\frac{\sigma^2}{\mathrm{n}} \mathrm{L}^{-1} \right)^{-1} (\hat{\beta} - \beta) = \frac{\mathrm{n}}{\sigma^2} (\hat{\beta} - \beta)^{\mathrm{T}} \mathrm{L} (\hat{\beta} - \beta) =$$

$$= \frac{\mathrm{n}}{\sigma^2} \sum_{\mathrm{k=1}}^{\mathrm{p}} \sum_{\mathrm{t=1}}^{\mathrm{p}} \mathrm{l_{kt}} (\hat{\beta}_{\mathrm{k}} - \beta_{\mathrm{k}})(\hat{\beta}_{\mathrm{t}} - \beta_{\mathrm{t}}) \ \sim \ \chi^2(\mathrm{p}).$$

Следовательно, статистика

$$\mathrm{F} = \frac{\mathrm{n - p - 1}}{\mathrm{p}} \cdot \frac{\sum\limits_{\mathrm{k=1}}^{\mathrm{p}} \sum\limits_{\mathrm{t=1}}^{\mathrm{p}} \mathrm{l_{kt}} (\hat{\beta}_{\mathrm{k}} - \beta_{\mathrm{k}})(\hat{\beta}_{\mathrm{t}} - \beta_{\mathrm{t}})}{\mathrm{s}^2_{0 \cdot 12 \ldots \mathrm{p}}} \ \sim \ \mathrm{F(p, n - p - 1)}. \quad (35)$$

Статистику F можно выразить через множественный коэффициент корреляции. Для этого, используя то, что при $\beta_1 = \ldots = \beta_p = 0$ выражение $l_{00} = \frac{1}{n}\sum\limits_{\nu=1}^{n}(y_\nu - \bar{y})^2$ за счет $y_\nu - \bar{y} = \alpha + \sum\limits_{k=1}^{p}\beta_k(x_{k\nu} - \bar{x}_k) + \delta_\nu - \alpha - \bar{\delta}$ превращается в

$$l_{00} = \frac{1}{n}\sum_{\nu=1}^{n}(\delta_\nu - \bar{\delta})^2 = \frac{1}{n}\sum_{\nu=1}^{n}\delta_\nu^2 - \bar{\delta}^2, \text{ из (30) получаем}$$

$$\sum_{k=1}^{p}\sum_{t=1}^{p} l_{kt}\hat{\beta}_k\hat{\beta}_t = \frac{1}{n}\sum_{\nu=1}^{n}\delta_\nu^2 - \hat{\sigma}^2 - (\hat{\alpha} - \alpha)^2 =$$

$$= l_{00} - \hat{\sigma}^2 = l_{00}\left(1 - \frac{s^2_{0\cdot 12\ldots p}}{l_{00}}\right) = l_{00}r^2_{0(12\ldots p)}.$$

А так как $s^2_{0\cdot 12\ldots p} = l_{00}(1 - r^2_{0(12\ldots p)})$, то статистику (35) представим в виде:

$$F = \frac{n-p-1}{p}\cdot\frac{r^2_{0(12\ldots p)}}{1 - r^2_{0(12\ldots p)}} \sim F(p, n-p-1). \qquad (36)$$

Таблица анализа дисперсий

Пусть $\hat{Y} = Xb$ вектор наилучшего в среднеквадратичном смысле линейного предсказания переменной Y по переменным $X_1, \ldots, X_p$.

Источник дисперсии	Сумма квадратов	степени свободы	средний квадрат
Регрессия	$SS_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	$\nu_R = p$	$MS_R = SS_R/\nu_R$
Откл.от регр.	$SS_E = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	$\nu_E = n - p - 1$	$MS_E = SS_E/\nu_E$
полная	$SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2$	$\nu_T = n - 1$	$MS_T = SS_T/\nu_T$

Оценка коэффициента детерминации $R^2 = \frac{SS_R}{SS_T}$

При справедливости нулевой гипотезы о равенстве нулю всех частных коэффициентов регрессии статистика $F = \frac{MS_R}{MS_E}$ имеет распределение Фишера с p и $n - p - 1$ степенями свободы.

Задача регрессии в матричном виде

Уравнения линейной регрессии представим в виде $Y = X\beta + e$, где зависимая переменная задается вектором $Y = (y_1, \ldots, y_n)^T$, матрица

$$X = \begin{bmatrix} 1 & x_{11} & \ldots & x_{p1} \\ 1 & x_{12} & \ldots & x_{p2} \\ & & \ldots & \\ 1 & x_{1n} & \ldots & x_{pn} \end{bmatrix} \tag{37}$$

называется матрицей плана, $\beta = (\beta_0, \ldots, \beta_p)^T$ – вектор параметров, e – нормально распределенный вектор с нулевым средним и ковариационной матрицей вида $\sigma^2 I$.

Дифференцирование по вектору параметров

$$A\beta = \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nm} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \dots \\ \beta_m \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} a_{1i}\beta_i \\ \dots \\ \sum_{i=1}^{m} a_{ni}\beta_i \end{bmatrix}$$

$$\frac{\partial A\beta}{\partial \beta} = \begin{bmatrix} \frac{\partial A\beta}{\partial \beta_1} & \dots & \frac{\partial A\beta}{\partial \beta_m} \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nm} \end{bmatrix} = A$$

Оценки МНК

$$\mathcal{L}_1 = (A\beta)'(A\beta) = \left(\sum_{i=1}^{m} a_{1i}\beta_i\right)^2 + \ldots + \left(\sum_{i=1}^{m} a_{ni}\beta_i\right)^2 .$$

Система нормальных уравнений имеет вид

$$\begin{cases} \frac{\partial \mathcal{L}_1}{\partial \beta_1} = 2\left(a_{11}\left(\sum\limits_{i=1}^{m} a_{1i}\beta_i\right) + \ldots + a_{n1}\left(\sum\limits_{i=1}^{m} a_{ni}\beta_i\right)\right) = 0, \\ \ldots \\ \frac{\partial \mathcal{L}_1}{\partial \beta_m} = 2\left(a_{1m}\left(\sum\limits_{i=1}^{m} a_{1i}\beta_i\right) + \ldots + a_{nm}\left(\sum\limits_{i=1}^{m} a_{ni}\beta_i\right)\right) = 0, \end{cases}$$

$$\frac{\partial (A\beta)'(A\beta)}{\partial \beta} = 0 \iff \quad 2\frac{\partial (A'\beta)}{\partial \beta} A\beta = 2A'A\beta = 0\,.$$

Оценки МНК с ограничением на параметры

$$\mathcal{L}_2 = \lambda'(1,\mathrm{s})\mathrm{H}'(\mathrm{s},\mathrm{m})\beta(\mathrm{m},1) =$$

$$= (\lambda_1,\ldots,\lambda_\mathrm{s})\begin{bmatrix} \mathrm{h}_{11} & \ldots & \mathrm{h}_{\mathrm{m}1} \\ \ldots & \ldots & \ldots \\ \mathrm{h}_{1\mathrm{s}} & \ldots & \mathrm{h}_{\mathrm{ms}} \end{bmatrix}\begin{bmatrix} \beta_1 \\ \ldots \\ \beta_\mathrm{m} \end{bmatrix} =$$

$$= (\lambda_1,\ldots,\lambda_\mathrm{s})\begin{bmatrix} \sum\limits_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{h}_{\mathrm{i}1}\beta_\mathrm{i} \\ \ldots \\ \sum\limits_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{h}_{\mathrm{is}}\beta_\mathrm{i} \end{bmatrix} = \sum_{\mathrm{j}=1}^{\mathrm{s}} \lambda_\mathrm{j}\left(\sum_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{h}_{\mathrm{ij}}\beta_\mathrm{i}\right).$$

Система нормальных уравнений

$$\begin{cases} \frac{\partial\mathcal{L}_2}{\partial\beta_1} = \sum\limits_{\mathrm{j}=1}^{\mathrm{s}} \lambda_\mathrm{j}\mathrm{h}_{1\mathrm{j}} = 0, \\ \qquad\ldots \\ \frac{\partial\mathcal{L}_2}{\partial\beta_\mathrm{m}} = \sum\limits_{\mathrm{j}=1}^{\mathrm{s}} \lambda_\mathrm{j}\mathrm{h}_{\mathrm{mj}} = 0, \end{cases} \qquad \Longleftrightarrow$$

$$\Longleftrightarrow \quad \begin{bmatrix} \mathrm{h}_{11} & \ldots & \mathrm{h}_{1\mathrm{s}} \\ \ldots & \ldots & \ldots \\ \mathrm{h}_{\mathrm{m}1} & \ldots & \mathrm{h}_{\mathrm{ms}} \end{bmatrix}\begin{bmatrix} \lambda_1 \\ \ldots \\ \lambda_\mathrm{s} \end{bmatrix} = \mathrm{H}\lambda = 0.$$

Оценки МНК в задаче регрессии

При минимизации квадратичной формы $S = (Y - X\beta)^T(Y - X\beta)$ вектор оценок $b = (b_0, \ldots, b_p)^T$ является решением системы нормальных уравнений

$$(X^T X)\beta = X^T Y \tag{38}$$

Действительно, дифференцирование по вектору параметров приводит к

$$2X^T(Y - X\beta) = 0 \Longleftrightarrow X^T Y = X^T X\beta.$$

Следовательно,

$$b = \hat{\beta} = (X^T X)^{-1} X^T Y. \tag{39}$$

Ковариация оценок

$$\mathrm{b} = (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{Y}, \quad \mathbb{E}\mathrm{b} = (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\mathbb{E}\mathrm{Y}$$

Ковариационная матрица оценок частных коэффициентов регрессии можно представить в виде:

$$\begin{aligned}\mathrm{Cov}(\mathrm{b}) = \mathbb{E}(\mathrm{b}-\mathbb{E}\mathrm{b})(\mathrm{b}-\mathbb{E}\mathrm{b})^{\mathrm{T}} =\\ = \mathbb{E}((\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}(\mathrm{Y}-\mathbb{E}\mathrm{Y}))((\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}(\mathrm{Y}-\mathbb{E}\mathrm{Y}))^{\mathrm{T}} =\\ = (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{Cov}(\mathrm{Y})((\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}})^{\mathrm{T}} =\\ = \sigma^2(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{X}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1} = \sigma^2(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}.\end{aligned}$$

$$\hat{\sigma}^2 = \frac{1}{\mathrm{n}}(\mathrm{Y}-\mathrm{X}\hat{\beta})^{\mathrm{T}}(\mathrm{Y}-\hat{\beta}\mathrm{X})$$

Остаточная сумма квадратов

Разложение определителя матрицы блочной структуры

$$|U| = |D|(a - b^T D^{-1} c) = a|D| - \sum_i b_i \sum_k D_{ki} c_k,$$

$$\text{где} \quad U = \begin{bmatrix} a & b^T \\ c & D \end{bmatrix}.$$

Остаточная дисперсия через матрицу вторых моментов $\sigma^2_{1\cdot 23\ldots n} = \frac{\Lambda}{\Lambda_{11}}$ в предположении центрированности.

$$\begin{aligned}(Y - X\hat{\beta})^T(Y - X\hat{\beta}) = Y^T Y - Y^T X\hat{\beta} = \\ = Y^T Y - Y^T X(X^T X)^{-1} X^T Y = \\ = \begin{vmatrix} Y^T Y & Y^T X \\ X^T Y & X^T X \end{vmatrix} / |X^T X| = \frac{1}{G_{11}},\end{aligned} \qquad (40)$$

где $G = S^{-1}$, $S = Z^T Z$ матрица вторых выборочных моментов с точностью до множителя n, $Z = [Y, X]$ – матрица размерности n на $p + 1$, X – матрица центрированного плана размерности n на p.

График соотношения эмпирических и теоретических квантилей

$x < -\mathrm{rnorm}(n, \mathrm{mean} = 1, \mathrm{sd} = 2)$
plot(ecdf(x)) эмпирическая функция распределения
curve(pnorm, −3, 3) график функции распределения F(x)
Qe эмпирические и Qt теоретические квантили для всех значений
$a = F^{-1}\left(\frac{i-0.5}{n}\right)$

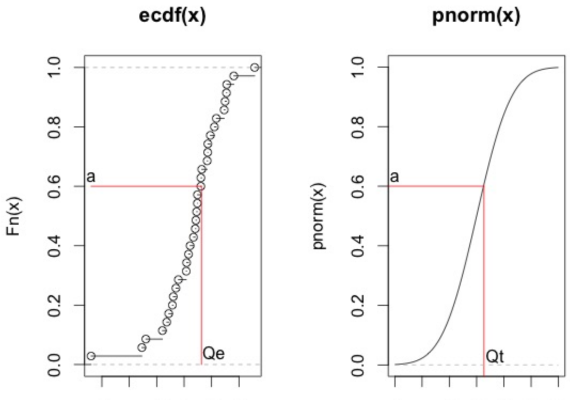
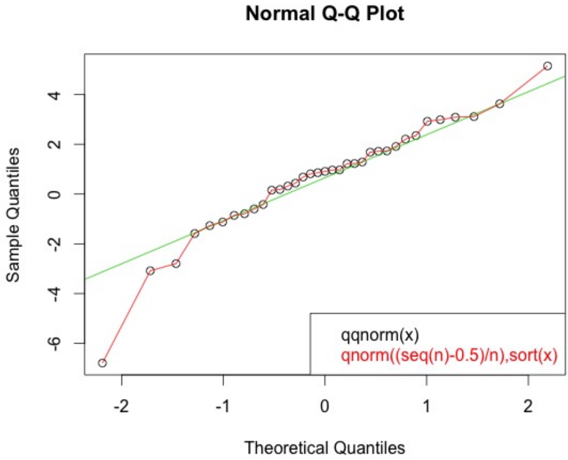
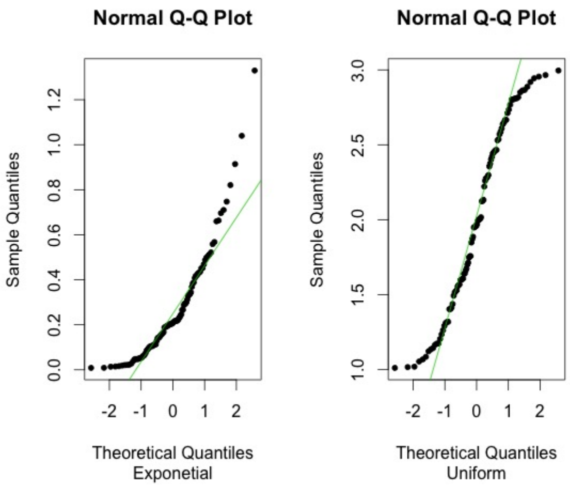


График квантиль-квантиль qqnorm(x)

По вертикальной оси откладываются упорядоченные значения $x_1, \ldots, x_n$, по горизонтальной $F^{-1}\left(\frac{i-0.5}{n}\right)$, в случае нормального распределения $\mathrm{qnorm}((\mathrm{seq}(n) - 0.5)/n)$.



QQ – plot при несогласованных распределениях



Примеры графиков квантиль-квантиль с теоретическими нормальными квантилями и эмпирическими квантилями равномерного и экспоненциального распределений