Классическая линейная модель множественной регрессии

лекция 2

8 февраля 2022 года

Что изучим

- Постановку задачи
- Методы оценивания коэффициентов
- Проверку гипотез и прогнозирование в КЛММР

Постановка задачи и исходные данные

Ставится задача построения и исследования зависимости результирующего признака y от объясняющих переменных $x_1, x_2, \ldots, x_k$ на основе результатов наблюдений признаков на “n” объектах $O_1, O_2, \ldots O_n$, $n >> k$.

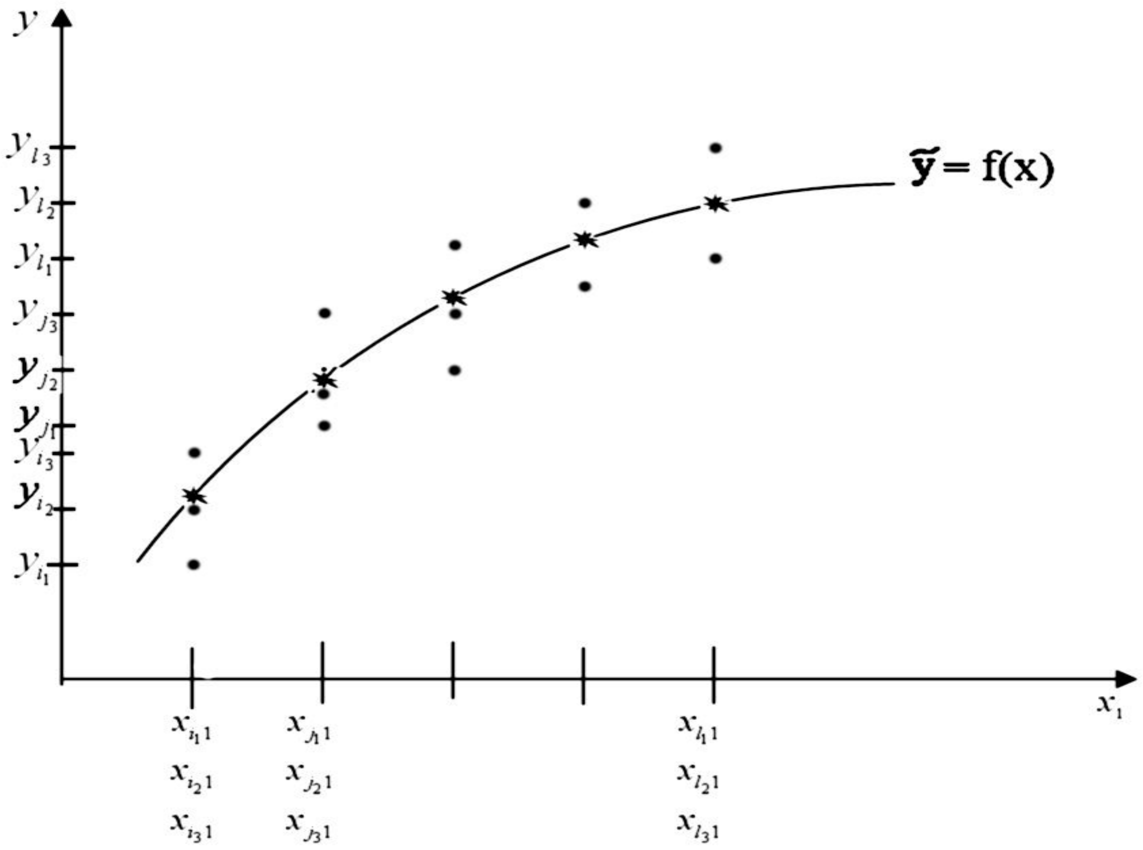
Результаты наблюдений результирующего признака и объясняющих переменных представлены вектором $Y_{n\times 1} = (y_1 \quad y_2 \quad \cdots \quad y_n)^T$ и матрицей X типа «объект-свойство»:

$$X_{n\times k} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

где y_i – наблюденное значение результативного признака для i-го объекта;

x_{ij} – значение j-го признака на i-м объекте наблюдения $i = \overline{1,n}$, $j = \overline{1,k}$.

Графическая иллюстрация для k = 1



Постановка задачи – переход к модели

Будем строить линейную относительно коэффициентов, в общем случае, но нелинейную по объясняющим переменным регрессионную зависимость:

$$\tilde{y} = \beta_0\psi_0(x_1, x_2, \dots, x_k) + \beta_1\psi_1(x_1, x_2, \dots, x_k) + \cdots + \beta_k\psi_k(x_1, x_2, \dots, x_k)$$
$$= \beta_0\psi_0(x) + \beta_1\psi_1(x) + \ldots + \beta_k\psi_k(x),$$

где $\psi_i(x)$, $i = \overline{0,k}$ - линейно независимые базисные функции; обычно полагают

$\psi_0(x) \equiv 1$;

$\beta = (\beta_0\beta_1 \ldots \beta_k)^T$ - вектор коэффициентов функции регрессии.

Для каждого i-объекта наблюдения:

Запишем с учетом ненаблюдаемого $\tilde{y}_i$ и наблюдённого y_i и вводя величину расхождения $z_i = y_i - \tilde{y}_i$

Регрессионная модель в векторно-матричном виде (апостериорная)

Введем матрицу:

$$\Psi_{\square} = \begin{pmatrix} \psi_0(x_{11}x_{12}\ldots x_{1k}) & \psi_1(x_{11}x_{12}\ldots x_{1k}) & \ldots. & \psi_k(x_{11}x_{12}\ldots x_{1k}) \\ \psi_0(x_{21}x_{22}\ldots x_{2k}) & \psi_1(x_{21}x_{22}\ldots x_{2k}) & \ldots. & \psi_k(x_{21}x_{22}\ldots x_{2k}) \\ \ldots. & \ldots. & \ldots & \ldots. \\ \psi_0(x_{n1}x_{n2}\ldots x_{nk}) & \psi_1(x_{n1}x_{n2}\ldots x_{nk}) & \ldots & \psi_k(x_{n1}x_{n2}\ldots x_{nk}) \end{pmatrix}$$

$$= \begin{pmatrix} \psi_0(x^{(1)}) & \psi_1(x^{(1)}) & \ldots & \psi_k(x^{(1)}) \\ \psi_0(x^{(2)}) & \psi_1(x^{(2)}) & \ldots & \psi_k(x^{(2)}) \\ \ldots & \ldots. & \ldots & \ldots. \\ \psi_0(x^{(n)}) & \psi_1(x^{(n)}) & \ldots & \psi_k(x^{(n)}) \end{pmatrix}$$

Тогда апостериорная регрессионная модель в векторно-матричном виде:

Регрессионная модель в векторно-матричном виде (априорная)

Обозначим

$\eta_{1,n} = (\eta_1, \eta_2, \ldots, \eta_n)^T$- случайный априорный вектор, возможными значениями которого является вектор $Y = (y_1, y_2 \ldots y_n)^T$;

$\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$- случайный априорный вектор регрессионных остатков, возможными значениями которого является вектор $Z = (z_1, \ldots, z_n)^T$.

Тогда априорная регрессионная модель:

Условия Гаусса-Маркова

1) $x_1,\ldots,x_к$ – детерминированные (неслучайные) переменные;

т.к. $x_1,\ldots,x_к$ – детерминированные, то и базисные функции от детерминированных величин тоже неслучайные, следовательно, матрица Ψ - детерминированная матрица.

2) rang Ψ = к+1- среди базисных функций нет линейно зависимых;

3) $M\varepsilon_i = 0,\ i = \overline{1,n}$ - нет систематических ошибок в измерении y;

4) $D\varepsilon_i = M\varepsilon_i^2 = \sigma^2, \quad i = \overline{1,n}$ - условие гомоскедастичности регрессионных остатков (равноточности измерений);

5) $cov(\varepsilon_i, \varepsilon_j) = M(\varepsilon_i \cdot \varepsilon_j) = 0, \quad i \neq j, \quad i = \overline{1,n}\ j = \overline{1,n}$ - условие некоррелированности регрессионных остатков.

Условия 4 – 5 можно записать в векторной форме:

4') $\Sigma_\varepsilon = M\varepsilon\varepsilon^T = \sigma^2 E_n$,

где Σ_ε- ковариационная матрица вектора регрессионных остатков ε;

E_n - единичная матрица.

Методы оценивания коэффициентов КЛММР

1) минимума суммы модулей отклонений наблюдаемых значений y_i от «значений» функции регрессии $\tilde{y}_i$:

$$\sum_{i=1}^{n} |y_i - \tilde{y}_i| = \sum_{i=1}^{n} |z_i| \to \min_{\beta};$$

2) минимума максимального модуля отклонения наблюдаемых значений y_i от «значений» функции регрессии $\tilde{y}_i$:

$$\max_{i=1,n} |y_i - \tilde{y}_i| = \max_{i=1,n} |z_i| \to \min_{\beta};$$

3) минимума суммы квадратов отклонений наблюдаемых значений y_i от «значений» функции регрессии (метод наименьших квадратов):

$$\sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 = \sum_{i=1}^{n} z_i^2 \to \min_{\beta};$$

4) максимума функции правдоподобия.

МНК-оценка коэффициентов КЛММР

$$\phi(\beta_0, \beta_1, \dots \beta_\kappa) = \sum_{i=1}^{n} z_i^2 = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 = (Y - \tilde{Y})^T (Y - \tilde{Y})$$

$$= (Y - \Psi\beta)^T (Y - \Psi\beta) = Y^T Y - \beta^T \Psi^T Y - Y^T \Psi\beta + \beta^T \Psi^T \Psi\beta$$

$$= Y^T Y - \beta^T \Psi^T Y - \underbrace{Y^T \Psi\beta}_{\beta^T \Psi^T Y} + \beta^T \Psi^T \Psi\beta$$

$$= Y^T Y - 2\beta^T \Psi^T Y + \beta^T \Psi^T \Psi\beta \to \min_{\beta}$$

$$\mathrm{H} = \frac{\partial^2 \phi}{\partial \beta^2} = \Psi^T \Psi$$

Тогда используя

$$\frac{\partial \phi}{\partial \beta} = 2\Psi^T \Psi\beta - 2\Psi^T Y = 0$$

Получаем

И в итоге МНК-оценка

ММП-оценка коэффициентов КЛММР

$$\varepsilon_i \in N(0,\sigma^2), \quad i=\overline{1,n}, \ \ \varepsilon \in N(0,\sigma^2 E_n)$$

$$L(z_1,..,z_n/\beta;\sigma^2) = \qquad = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{1}{2\sigma^2}(z_i - M\varepsilon_i)^2}$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{1}{2\sigma^2}(y_i - \tilde{y}_i)^2} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{1}{2\sigma^2}(y_i - \psi(x^{(i)})\beta)^2}$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} exp(-\frac{1}{2\sigma^2}(Y-\Psi\beta)^T(Y-\Psi\beta))$$

$$l = ln\, L$$

Анализ вариации результативного признака.

$$Q_{общ} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}((y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}))^2$$

$$= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) =$$

$$= Q_{ост} + Q_{факт} + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$$

Анализ вариации результативного признака.

$$Q_{общ} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}((y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}))^2$$

$$= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) =$$

$$= Q_{ост} + Q_{факт} + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$$

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n} z_i(\hat{y}_i - \bar{y})$$

$$= \sum_{i=1}^{n} \hat{z}_i\big(\hat{\beta}_0\psi_0(x_{i1}x_{i2\dots}x_{ik}) + \hat{\beta}_1\psi_1(x_{i1}x_{i2\dots}x_{ik}) + \dots \hat{\beta}_k\psi_k(x_{i1}x_{i2\dots}x_{ik}) - \bar{y}\big) =$$

$$= \hat{\beta}_0\sum_{i=1}^{n} \hat{z}_i\,\psi_0(x^{(i)}) + \hat{\beta}_1\sum_{i=1}^{n} \hat{z}_i\psi_1(x^{(i)}) + \dots + \hat{\beta}_k\sum_{i=1}^{n} \hat{z}_i\psi_k(x^{(i)}) - \bar{y}\sum_{i=1}^{n} \hat{z}_i = 0$$

Анализ вариации результативного признака.

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n} z_i(\hat{y}_i - \bar{y})$$

$$= \sum_{i=1}^{n} \hat{z}_i\left(\hat{\beta}_0\psi_0(x_{i1}x_{i2\dots}x_{ik}) + \hat{\beta}_1\psi_1(x_{i1}x_{i2\dots}x_{ik}) + \dots \hat{\beta}_k\psi_k(x_{i1}x_{i2\dots}x_{ik}) - \bar{y}\right) =$$

$$= \hat{\beta}_0 \sum_{i=1}^{n} \hat{z}_i\, \psi_0(x^{(i)}) + \hat{\beta}_1 \sum_{i=1}^{n} \hat{z}_i \psi_1(x^{(i)}) + \dots + \hat{\beta}_k \sum_{i=1}^{n} \hat{z}_i \psi_k(x^{(i)}) - \bar{y} \sum_{i=1}^{n} \hat{z}_i = 0$$

$$\begin{pmatrix} \sum_{i=1}^{n} \psi_0(x^{(i)}) \cdot \hat{z}_i \\ \sum_{i=1}^{n} \psi_1(x^{(i)}) \cdot \hat{z}_i \\ \dots \\ \sum_{i=1}^{n} \psi_k(x^{(i)}) \cdot \hat{z}_i \end{pmatrix} = \begin{pmatrix} \psi_0(x^{(1)}) & \psi_0(x^{(2)}) & \dots & \psi_0(x^{(n)}) \\ \psi_1(x^{(1)}) & \psi_1(x^{(2)}) & \dots & \psi_1(x^{(n)}) \\ \dots & \dots & \dots & \dots \\ \psi_k(x^{(1)}) & \psi_k(x^{(2)}) & \dots & \psi_k(x^{(n)}) \end{pmatrix} \cdot \begin{pmatrix} \hat{z}_1 \\ \hat{z}_2 \\ \dots \\ \hat{z}_n \end{pmatrix} = \Psi^T Z = \Psi^T(Y - \Psi\hat{\beta})$$

$$= \Psi^T Y - \Psi^T \Psi \hat{\beta} =$$

$$= \Psi^T Y - \Psi^T \Psi(\Psi^T \Psi)^{-1} \Psi^T Y = \Psi^T Y - \Psi^T Y = 0$$

то есть $\sum_{i=1}^{n} \psi_0(x^{(i)})\hat{z}_i = 0$, $\sum_{i=1}^{n} \psi_l(x^{(i)})\hat{z}_i = 0$, $l = \overline{0, k}$.

Если дополнительно потребовать, что $\psi_0(x^{(i)}) \equiv 1$, то $\sum_{i=1}^{n} \hat{z}_i = 0$.

Коэффициент детерминации

КЛММР: проверка гипотез

и доверительные интервалы

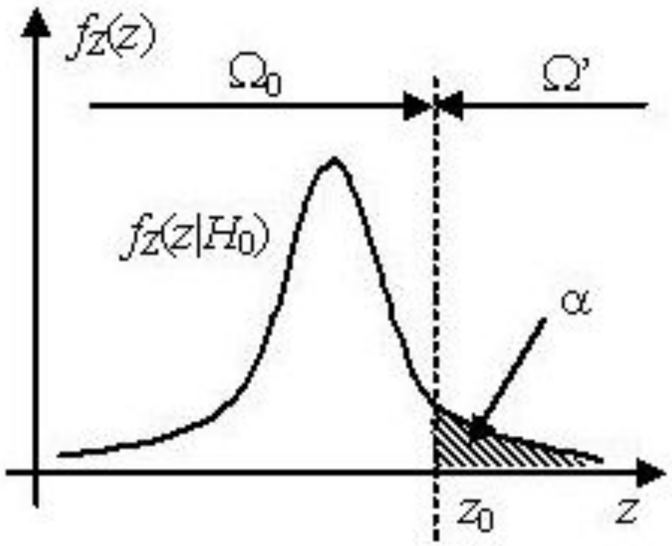
Многомерное обобщение теоремы Фишера

Проверка значимости модели в целом (адекватности модели выборочным данным)

- H0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$

H1: $\sum_{j=1}^{k} \beta_j^2 > 0$

$$F(\eta_{1,n}) = \frac{\frac{Q_{\text{факт}}(\eta_{1,n})}{k}}{\frac{Q_{\text{ост}}(\eta_{1,n})}{n-k-1}} = \frac{\frac{\hat{R}^2(\eta_{1,n})}{k}}{\frac{1-\hat{R}^2(\eta_{1,n})}{n-k-1}} \sim F(k; n-k-1)$$



$F_{\text{набл}} = F(y_{1,n})$

$F_{\text{крит}} = z_0 = F_{1-\alpha}(k; n-k-1)$ - квантиль уровня $1-\alpha$ распределения Фишера-Снедекора с числом степеней свободы k в числителе и n-k-1 в знаменателе

$F_{\text{набл}} > F_{\text{крит}} \Rightarrow$ H0 отвергается

$P\left(F(\eta_{1,n}) > F(y_{1,n})\right) < \alpha \Rightarrow$ H0 отвергается

Проверка гипотезы о значении отдельного коэффициента

$\hat{\beta}_{\text{МНК}}(\eta_{1,n}) \in N\ \ (\beta; \sigma^2(\Psi^T\Psi)^{-1})$

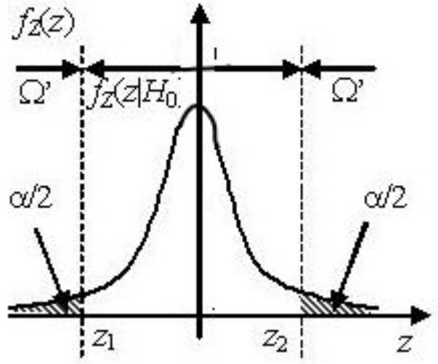
- H0: $\beta_j = a$
  H1: $\beta_j \neq a$

H0: $\beta_j = a$
H1: $\beta_j > a$

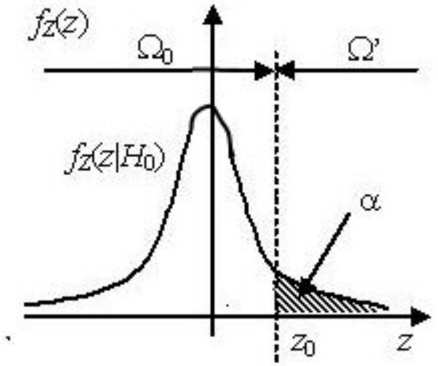
H0: $\beta_j = a$
H1: $\beta_j < a$

$$t_j(\eta_{1,n}) = \frac{\hat{\beta}_{j,\text{мнк}}(\eta_{1,n}) - a}{S_{\hat{\beta}_j}(\eta_{1,n})} \sim t(n-k-1)$$

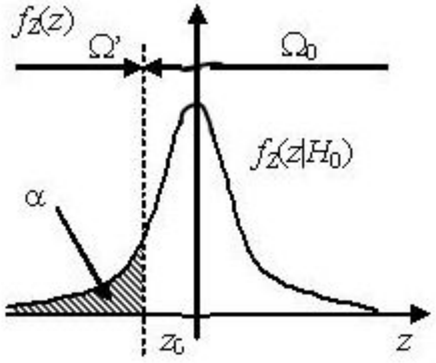
$$j = 0,1,\ldots,k, \quad S_{\hat{\beta}_j}(\eta_{1n}) = S_{\text{ост}}(\eta_{1n})\sqrt{[(\Psi^T\Psi)^{-1}]_{j+1\,j+1}}.$$



$t_{\text{крит}} = z_2 = t_{1-\alpha/2}(n-k-1)$ - квантиль уровня $1-\alpha/2$ распределения Стьюдента с n-k-1 ст.св.



$t_{\text{крит}} = z_0 = t_{1-\alpha}(n-k-1)$ - квантиль уровня $1-\alpha$ распределения Стьюдента с n-k-1 ст.св.

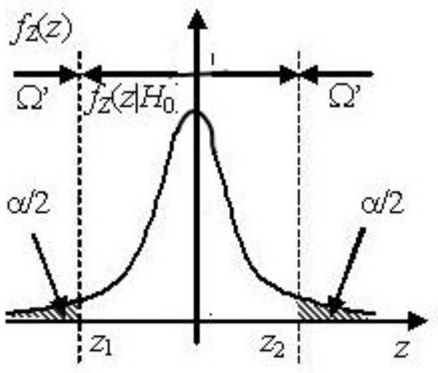


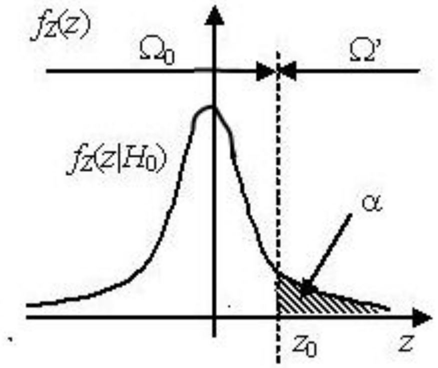
$t_{\text{крит}} = z_0 = t_{\alpha}(n-k-1)$ - квантиль уровня α распределения Стьюдента с n-k-1 ст.св.

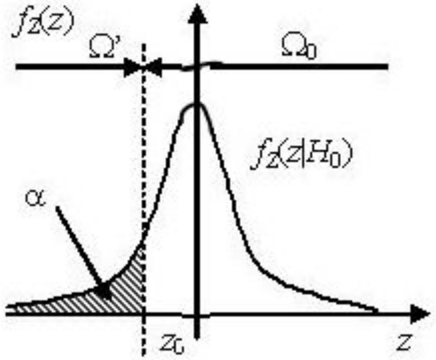
Проверка значимости отдельного коэффициента

H0: $\beta_j = 0$	H0: $\beta_j = 0$	H0: $\beta_j = 0$
H1: $\beta_j \neq 0$	H1: $\beta_j > 0$	H1: $\beta_j < 0$

$$t_j(\eta_{1,n}) = \frac{\hat{\beta}_{j,\text{мнк}}(\eta_{1,n})}{S_{\hat{\beta}_j}(\eta_{1,n})} \sim t(n-k-1), \quad j = 0,1,\ldots,k, \quad S_{\hat{\beta}_j}(\eta_{1n}) = S_{\text{ост}}(\eta_{1n})\sqrt{[(\Psi^T\Psi)^{-1}]_{j+1j+1}}.$$







Построение доверительных интервалов для коэффициентов

- $$\hat{\beta}_{\text{МНК}}(\eta_{1,n}) \in N \quad (\beta; \sigma^2(\Psi^T\Psi)^{-1})$$

$$t_j(\eta_{1,n}) = \frac{\hat{\beta}_{j,\text{мнк}}(\eta_{1,n}) - \beta_j}{\hat{S}_{\hat{\beta}_j}(\eta_{1,n})} \sim t(n-k-1)$$

$$j = 0,1,\ldots,k, \quad \hat{S}_{\hat{\beta}_j}(\eta_{1,n}) = \hat{S}_{\text{ост}}\sqrt{[(\Psi^T\Psi)^{-1}]_{j+1\,j+1}}$$

$$\left|\frac{\hat{\beta}_{j,\text{мнк}} - \beta_j}{\hat{S}_{\hat{\beta}_j}}\right| < \delta_\gamma$$

$$\hat{\beta}_j - \hat{S}_{\hat{\beta}_j} t_{1-\alpha/2} \le \beta_j \le \hat{\beta}_j + \hat{S}_{\hat{\beta}_j} t_{1-\alpha/2}$$

γ – доверительная вероятность, $1 - \gamma = \alpha$

$\delta_\gamma = t_{1-\alpha/2}$ – квантиль уровня $1 - \alpha/2$ распределения Стьюдента с n-k-1 степенями свободы

Проверка гипотезы о незначимости группы коэффициентов

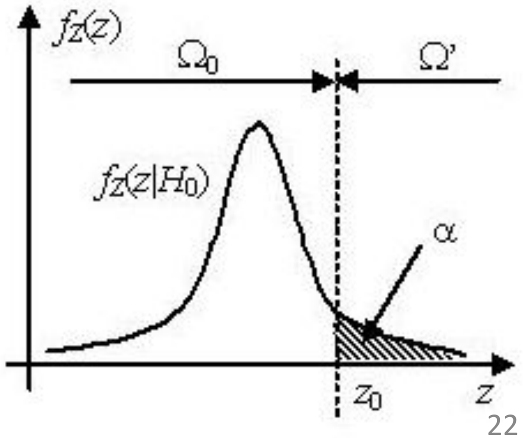
- Длинная регрессия (unrestricted, UR) : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + z$
- Короткая регрессия (restricted, R) : $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{k-q} x_{k-q} + z$

(наложили ограничение на q коэффициентов)

H0: $\beta_{k-q+1} = \ldots = \beta_k = 0$ (выбираем короткую регрессию)

H1: $\sum_{j=k-q+1}^{k} \beta_j^2 > 0$ (выбираем длинную регрессию)

$$F(\eta_{1,n}) = \frac{\frac{Q_{\text{ост}}^{R}(\eta_{1,n}) - Q_{\text{ост}}^{UR}(\eta_{1,n})}{q}}{\frac{Q_{\text{ост}}^{UR}(\eta_{1,n})}{n-k-1}} = \frac{\frac{R_{UR}^2(\eta_{1,n}) - R_R^2(\eta_{1,n})}{q}}{\frac{(1 - R_{UR}^2)(\eta_{1,n})}{n-k-1}} \sim F(q; n-k-1)$$



Проверка линейных ограничений общего вида

- Пусть $\mathrm{H}_{q\times(k+1)}$ и $\mathrm{r}_{\mathrm{q}\times 1}$ –матрица и вектор, задающие ограничения на коэффициенты $\mathrm{H}\beta = \mathrm{r}$.

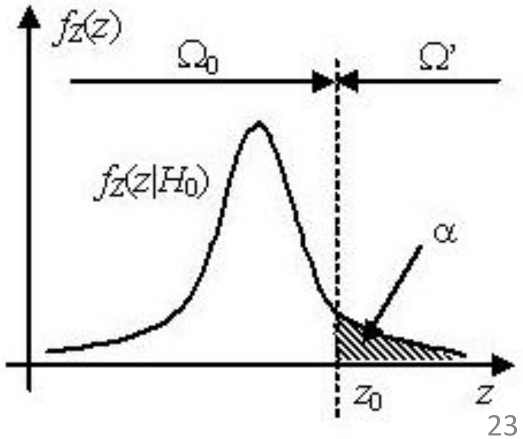
Ограничений меньше, чем параметров $q \le k + 1,$ ограничения линейно независимы $rank(\mathrm{H}) = q$.

Пример:$\mathrm{H}\beta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} = r \sim \begin{cases} \beta_0 = 2 \\ \beta_1 - \beta_2 = 0 \end{cases} \sim \begin{cases} \beta_0 = 2 \\ \beta_1 = \beta_2 \end{cases}$

H0: $\mathrm{H}\beta = \mathrm{r}$ (выбираем короткую регрессию, с ограничениями, restricted, R)

H1: $\mathrm{H}\beta \ne \mathrm{r}$ (выбираем длинную регрессию, без ограничений, unrestricted, UR)

$$F(\eta_{1,n}) = \frac{\frac{Q^R_{\text{ост}}(\eta_{1,n}) - Q^{UR}_{\text{ост}}(\eta_{1,n})}{q}}{\frac{Q^{UR}_{\text{ост}}(\eta_{1,n})}{n-k-1}} = \frac{\frac{R^2_{UR}(\eta_{1,n}) - R^2_R(\eta_{1,n})}{q}}{\frac{(1 - R^2_{UR})(\eta_{1,n})}{n-k-1}} \sim F(q; n-k-1)$$



Проверка линейных ограничений общего вида – пример преобразования

- Пример: $H\beta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} = r \sim \begin{cases} \beta_0 = 2 \\ \beta_1 - \beta_2 = 0 \end{cases} \sim \begin{cases} \beta_0 = 2 \\ \beta_1 = \beta_2 \end{cases}$

Длинная регрессия - регрессия без ограничений (unrestricted, UR) :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + z$$

Короткая регрессия - регрессия с ограничениями (restricted, R) :

$$y = 2 + \beta_1 x_1 + \beta_1 x_2 + z$$

$$y - 2 = \beta_1 (x_1 + x_2) + z$$

Прогнозирование в ЛММР

- Функция регрессии $\tilde{y} = \beta_0\psi_0(x) + \beta_1\psi_1(x) + \ldots + \beta_k\psi_k(x)$

  Оценка функции регрессии $\hat{y} = \hat{\beta}_0\psi_0(x) + \hat{\beta}_1\psi_1(x) + \ldots + \hat{\beta}_k\psi_k(x)$

  Пример: y – стоимость квартиры, млн. руб., x1 – жилая площадь квартиры, кв.м, x2 – площадь кухни, кв.м.

$$\hat{y} = 0{,}5 + 0{,}05 \cdot x1 + 0{,}06 \cdot x2$$

  Вопрос 1. Сколько в среднем стоит квартира жилой площадью 30 кв.м. и площадью кухни 10 кв. м.? (прогноз условного математического ожидания, то есть среднего значения результативного показателя – точечный прогноз $\tilde{y}_{\psi(x^{(n+1)})}$)

  Вопрос 2. Сколько будет стоить конкретная квартира жилой площадью 30 кв.м. и площадью кухни 10 кв. м.? (прогноз индивидуального значения результативного показателя – точечный прогноз $y_{\psi(x^{(n+1)})} = y_{n+1}$)

  Решение: $x^{(n+1)} = (1 \quad 30 \quad 10)$, $\hat{y} = 0{,}5 + 0{,}05 \cdot 30 + 0{,}06 \cdot 10 = 2{,}6$ млн. руб.

Построение доверительного интервала для значения функции регрессии (<u>среднего</u> значения результативного признака в точке)

- Вопрос 1. Сколько в среднем стоит квартира жилой площадью 30 кв.м. и площадью кухни 10 кв. м.? (прогноз условного математического ожидания, то есть среднего значения результативного показателя – точечный прогноз $\tilde{y}_{\psi(x^{(n+1)})}$). А если нужна интервальная оценка?

Тогда:

$$M\hat{y}_{\psi(x^{(n+1)})}(\eta_{1n}) = M(\psi^T(x^{(n+1)}) \cdot \hat{\beta}(\eta_{1n})) = \psi^T(x^{(n+1)})M\hat{\beta}(\eta_{1n}) = \psi^T(x^{(n+1)})\beta \ = \tilde{y}_{\psi(x^{(n+1)})}$$

$$\hat{y}_{\psi(x^{(n+1)})}(\eta_{1n}) = D(\psi^T(x^{(n+1)})\hat{\beta}(\eta_{1n})) = \sigma^2\psi^T(x^{(n+1)})(\Psi^T\Psi)^{-1}\psi(x^{(n+1)}),$$

$$t(\eta_{1n}) = \frac{\hat{y}_{\psi\left(x^{(n+1)}\right)} - \tilde{y}_{\psi\left(x^{(n+1)}\right)}}{S_{\text{ост}}\ (\eta_{1n})\sqrt{\psi^T(x^{(n+1)})(\Psi^T\Psi)^{-1}\psi(x^{(n+1)})}} \sim \mathrm{t(n-k-1)}$$

$$\hat{y}_{\psi(x^{(n+1)})} - \delta_\gamma S_{\text{ост}}\ (\eta_{1n})\sqrt{\psi^T(x^{(n+1)})(\Psi^T\Psi)^{-1}\psi(x^{(n+1)})} \le \tilde{y}_{\psi(x^{(n+1)})} \le \hat{y}_{\psi(x^{(n+1)})} + \delta_\gamma S_{\text{ост}}\ (\eta_{1n})\sqrt{\psi^T(x^{(n+1)})(\Psi^T\Psi)^{-1}\psi(x^{(n+1)})}$$

γ – доверительная вероятность, $1 - \gamma = \alpha$

$\delta_\gamma = t_{1-\alpha/2}$ – квантиль уровня $1 - \alpha/2$ распределения Стьюдента с n-k-1 степенями свободы

Построение доверительного интервала для <u>индивидуального</u> значения результативного признака в точке

- Вопрос 2. Сколько будет стоить конкретная квартира жилой площадью 30 кв.м. и площадью кухни 10 кв. м.? (прогноз индивидуального значения результативного показателя – точечный прогноз $y_{\psi(x^{(n+1)})} = y_{n+1}$). А если нужна интервальная оценка?

  Тогда:

$$t(\eta_{1n}) = \frac{\hat{y}_{\psi\left(x^{(n+1)}\right)} - y_{n+1}}{S_{\text{ост}}\,(\eta_{1n})\sqrt{\psi^T(x^{(n+1)})(\Psi^T\Psi)^{-1}\psi(x^{(n+1)})}} \sim \mathrm{t(n-k-1)}$$

$$\hat{y}_{\psi(x^{(n+1)})} - \delta_\gamma S_{\text{ост}}\,(\eta_{1n})\sqrt{\psi^T(x^{(n+1)})(\Psi^T\Psi)^{-1}\psi(x^{(n+1)}) + 1} \le y_{n+1} \le \hat{y}_{\psi(x^{(n+1)})} + \delta_\gamma S_{\text{ост}}\,(\eta_{1n})\sqrt{\psi^T(x^{(n+1)})(\Psi^T\Psi)^{-1}\psi(x^{(n+1)}) + 1}.$$

γ – доверительная вероятность, $1 - \gamma = \alpha$

$\delta_\gamma = t_{1-\alpha/2}$ – квантиль уровня $1 - \alpha/2$ распределения Стьюдента с n-k-1 степенями свободы

Спасибо за внимание!

- Идем в Moodle
- Делаем тест после лекции 2 (5 вопросов) – время на выполнение 5 минут.