Гаусовская линейная модель

X, Y — выборка $\qquad y = X\theta + \varepsilon$

Опр. лин. модель — гаусовская, если

$$\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$$

Хар-ки коэф-ов (θ) из statsmodels

← крит $H_0: \theta = 0$ vs $H_1: \theta \neq 0$

name	$\hat\theta$	$\hat\sigma^2$	$T_j^0(X)$	p-value	$\hat\theta$ - conf_inf	
var 1	$\hat\theta_1$	$\hat\sigma^2 (X^T X)^{-1}_{jj}$	$T_j^0(X)$	ур. крит.	$\theta_j - \hat\sigma \cdot \sqrt{(X^T X)^{-1}_{jj}} \cdot T_{n-d, 1-\frac{\alpha}{2}}$	$-\|\|-$ $T_{n-d, 1+\frac{\alpha}{2}}$
var 2	$\hat\theta_2$					
$\vdots$						

① $\hat{\theta} = (X^T X)^{-1} X^T Y \sim N\left(\theta, \sigma^2 (X^T X)^{-1}\right)$

② $\hat{\sigma} = \dfrac{RSS(\hat{\theta})}{n-d}$

③ $T_j = \dfrac{\hat{\theta}_j - \theta_j}{\hat{\sigma}\sqrt{(X^T X)^{-1}_{jj}}} \sim T_{n-d}$

$H_0 : \theta_j = 0$

④ $S = \{ |T_j^0| > T_{n-d,\, 1-\alpha/2} \}$

$\text{p-value}(t) = P(|x| > |t|)$

⑤ Дов. инт. $\theta_j \in \left(\hat{\theta}_j \pm \hat{\sigma}\sqrt{(X^T X)^{-1}_{jj}}\; T_{n-d,\, 1-\alpha/2}\right)$

Сведение к гауссовской модели

Пример Наблюдения (X, Y) такие что

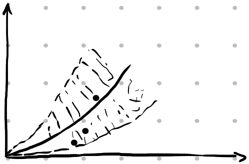
для $Y_1 - Y_{n/2}$ $\quad \varepsilon \sim N(0, \sigma^2)$

для $Y_{\frac{n}{2}+1} - Y_n$ $\quad \varepsilon \sim N(0, 4\sigma^2)$

$(X_{\frac{n}{2}+1} - X_n, \ Y_{\frac{n}{2}+1} - Y_n) \ /= 2$

$$\frac{Y}{2} = \frac{X}{2}\theta + \frac{\varepsilon}{2}$$

Пример Стандартные преобразования



Бокса-Кокса ?

Просто перебираем стандартное преобразование

и смотрим что получится

Задача $X_1 \ldots X_n \sim \mathcal{N}(\theta_1, \sigma^2)$

$Y_1 \ldots Y_m \sim \mathcal{N}(\theta_2, \sigma^2)$ $\quad X, Y$ - нез.

Проверить $H_0: \theta_1 = \theta_2$ vs. $H_1: \theta_1 \neq \theta_2$

Построим модель
$$W = \begin{pmatrix} X_1 \\ \vdots \\ X_n \\ Y_1 \\ \vdots \\ Y_m \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}1\\1\\1\end{matrix}}\right\} n \\ \left.\vphantom{\begin{matrix}1\\1\\1\end{matrix}}\right\} m \end{matrix} \quad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$

$$W = Z\theta + \varepsilon, \quad \text{где} \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Критерий с лекции $T \in \mathbb{R}^{k \times d}$

$$H_0: T\theta = \tau$$

$\hat{\theta}$ – МНК оценка $\Rightarrow$ $\hat{t} = T\hat{\theta}$ – оценка τ

$$B = T(X^T X)^{-1} T^T$$

Крит Фишера: $F(X,Y) = \dfrac{(\hat{t}-\tau)^T B^{-1} (\hat{t}-\tau)}{\|y - X\hat{\theta}\|^2} \cdot \dfrac{n-d}{k} \sim F_{k,n-d}$

В нашем случае: $T = (-1, 1)$ $\tau = 0$

$F(Z, W) = $ [ниже]

$$Z^T Z = \begin{pmatrix} 1 \dots 1 & 0 \dots 0 \\ 0 \dots 0 & 1 \dots 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} n & 0 \\ 0 & m \end{pmatrix}$$

$$\hat{\theta} = (Z^T Z)^{-1} Z^T W = \begin{pmatrix} \overline{X} \\ \overline{Y} \end{pmatrix}$$

$$\hat{t} = T(Z^T Z)^{-1} Z^T W$$

$$\hat{t} = T\hat{\theta} = \bar{Y} - \bar{X}$$

$$B = T \begin{pmatrix} 1/n & 0 \\ 0 & 1/m \end{pmatrix} T^T = \frac{n+m}{nm}$$

$$\| W - Z\hat{\theta} \|^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{m} (Y_i - \bar{Y})^2 = n\hat{\sigma}_x^2 + m\hat{\sigma}_y^2$$

$$W - Z\hat{\theta} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \\ y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \bar{X} \\ \bar{Y} \end{pmatrix} = \begin{pmatrix} X_1 - \bar{X} \\ \vdots \\ x_n - \bar{x} \\ Y_1 - \bar{Y} \\ \vdots \\ X_n - \bar{Y} \end{pmatrix}$$

$$F(Z, W) = \frac{(\bar{Y} - \bar{X})^T (\bar{Y} - \bar{X})}{n\hat{\sigma}_x^2 + m\hat{\sigma}_y^2} \frac{nm}{n+m} \cdot \frac{n+m-2}{1} \sim F_{1,\, n+m-2}$$

$\blacksquare$

$\theta = (\hat{\theta}_1, \theta_2)$, $\hat{\theta}_1 \in \mathbb{R}^{d-k}$, $\theta_2 \in \mathbb{R}^k$ $H_0: \theta_2 = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ // $Y = X\theta$, $Y_1 = X_1\theta_1$

$\hat{\theta}$ – МНК для θ $\hat{\theta}_1$ – МНК для θ_1

Утв. $\dfrac{RSS(\hat{\theta}) - RSS(\hat{\theta}_1)}{RSS(\hat{\theta})} \cdot \dfrac{n-d}{k} \overset{H_0}{\sim} F_{k, n-d}$

---- иза будет мясо

▲ т. о разл. гаус. вектора $L = L_1 \oplus \ldots \oplus L_k$

ξ – гаус вектор $\sim N(a, \sigma^2 I_n)$

$\eta_j = \mathrm{proj}_{L_j} \xi$ $E\eta_j = \mathrm{proj}_{L_j} a$

то $\dfrac{1}{\sigma^2} \| \eta_j - E\eta_j \|^2 \sim \chi^2_{\dim(L_j)}$

Возьмем $L = \{X\theta\}$ $\qquad L \oplus L^{\perp}$

$$\text{proj}_L Y = X\hat{\theta} \quad \Rightarrow \quad \text{proj}_{L^{\perp}} Y = Y - X\hat{\theta}$$

$\Rightarrow$ по т. о р. г. в. $\quad \frac{1}{\sigma^2} \| y - X\hat{\theta} \|^2 \sim \chi^2_{n-d}$

$$\text{proj}_{L^{\perp}} \theta = 0$$

Теперь рассмотрим разбиение

$$L_1(X) \oplus L_2(X) \oplus L^{\perp}(X)$$

где $L_1 = \{X\theta_1 \mid \theta \in \mathbb{R}^{d-k}\}$

$L_2 = \{X\theta_2 \mid \theta \in \mathbb{R}^{k}\}$

$$\frac{1}{\sigma^2} \| \text{proj}_{L_2} y - E\,\text{proj}_{L_2} y \|^2 \; - \; ?$$

Найдём $E\,\mathrm{proj}_{L_2} Y$

$$\mathrm{proj}_{L_2} Y = \mathrm{proj}_L Y - \mathrm{proj}_{L_1} Y$$

$$E\,\mathrm{proj}_{L_2} Y = \theta - \theta = 0$$

Найдём $\mathrm{proj}_{L_2} Y$

$$\underbrace{L_1 \oplus L_2}_{L} \oplus L^{\perp}, \qquad L_1 \oplus \underbrace{L_2 \oplus L^{\perp}}_{L_1^{\perp}}$$

$c^2 = a^2 + b^2$

$$\| \mathrm{proj}_{L_1^{\perp}} Y \|^2 = \| \mathrm{proj}_{L_2} Y \|^2 + \| \mathrm{proj}_{L^{\perp}} Y \|^2$$

$$\| \mathrm{proj}_{L_2} Y \|^2 = RSS(\hat{\theta}_1) - RSS(\hat{\theta})$$

$$\frac{1}{\sigma^2} \| \mathrm{proj}_{L_2} Y - E\,\mathrm{proj}_{L_2} Y \|^2 = \frac{1}{\sigma^2}\left[RSS(\hat{\theta}_1) - RSS(\hat{\theta}) \right] \sim \chi^2_k$$

Т.к. $\mathrm{proj}_{L_2} Y \perp\!\!\!\perp \mathrm{proj}_{L^{\perp}} Y$, то $\dfrac{RSS(\hat{\theta}_1) - RSS(\hat{\theta})}{RSS(\hat{\theta})} \cdot \dfrac{n-d}{k} \sim F_{k,n-d}$ $\blacksquare$

Значимость регрессии

Рассматриваем $R^2 = 1 - \frac{RSS(\hat{\theta})}{\| y - \bar{y} \|^2} = 1 - \frac{RSS(\hat{\theta})}{RSS(\hat{\theta}_1)} = \frac{RSS(\hat{\theta}_1) - RSS(\hat{\theta})}{RSS(\hat{\theta}_1)}$

$$\frac{R^2}{1-R^2} = \frac{RSS(\hat{\theta}_1) - RSS(\hat{\theta})}{RSS(\hat{\theta})} \Rightarrow \frac{R^2}{1-R^2} \quad \frac{n-d}{d-1} \sim F_{d-1,\, n-d}$$