4. Hypothesis Testing

Why we want to test?

Linear hypothesis

The examples of hypotheses can fall into a general class.

Let $p-1=5$. $\quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_5 \end{pmatrix} \in \mathbb{R}^6$

① $H_0: \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_5 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}}_{5\times 1} \Leftrightarrow \underbrace{\big(\underbrace{0}_{5\times1} \quad \underbrace{I}_{5\times5}\big)}_{5\times 6} \beta = 0$

$\Leftrightarrow \underbrace{A}_{5\times6} \underbrace{\beta}_{6\times1} = \underbrace{0}_{5\times1}$

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}_{5\times 6}$$

$$A\beta = \big(\underbrace{0}_{5\times1} \quad \underbrace{I}_{5\times5}\big) \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_5 \end{pmatrix} = I_{5\times5} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_5 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_5 \end{pmatrix}$$

② $H_0: \beta_1 = 0 \Leftrightarrow \underbrace{(0, 1, 0, \cdots 0)}_{\beta_1} \beta = 0$

$A\beta = 0$

$H_0: \beta_i = 0 \Leftrightarrow \underbrace{(0, 0, \cdots 1, 0 \cdots 0)}_{1\times 6} \beta = 0$

$i \geq 1$

③ $H_0: \beta_1 = \beta_2 = 0 \Leftrightarrow (0, 1, 0, 0, 0, 0)\, \beta = \beta_1 = 0$

$\begin{cases} \beta_1 = 0 \\ \beta_2 = 0 \end{cases} \qquad (0, 0, 1, 0, 0, 0)\, \beta = \beta_2 = 0$

$$\Leftrightarrow \quad A\beta = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}_{2\times 6}$$

④ $H_0: \beta_1 = c \quad \Leftrightarrow \quad (0,1,0,0,0,0)\beta = \beta_1 = c$

$\downarrow$ fixed value $\quad \Leftrightarrow \quad A\beta = c$

$A = (0,1,0,0,0,0)$

⑤ $H_0: \beta_1 = \beta_2 \quad \Leftrightarrow \quad (0,1,-1,0,0,0)\underset{6\times 1}{\beta}$

$\Leftrightarrow \beta_1 - \beta_2 = 0 \qquad = \beta_1 - \beta_2 = 0$

$\Leftrightarrow \quad A\beta = 0$

$A = (0,1,-1,0,0,0)$

Common thing: $A\beta = c$

Difference: Specify A and c differently

General theory for testing $H_0: A\beta = c$

Under full-rank model $\qquad$ Matrix

$$Y_i = \beta_0 + X_{i1}\beta_1 + \cdots + X_{i,p-1}\beta_{p-1} + \epsilon_i \qquad (Y = X\beta + \epsilon)$$

If $\underset{q\times p}{A} = \begin{pmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_q^T \end{pmatrix}$, $A\beta = c \quad \Leftrightarrow$ q constraints $\begin{cases} a_1^T\beta = c_1 \\ a_2^T\beta = c_2 \\ \vdots \\ a_q^T\beta = c_q \end{cases}$

Assume w.l.g. that rows of A are linearly independent

(otherwise, we can keep a subset of a_i^T's that are linearly independent.)

$\Rightarrow$ $\text{rank}(A) = q$ $(q \le p)$

H_0: $A\beta = c$ specifies a constrained nest model

H_A: $Y = X\beta + \epsilon$ is a full model without constraints.

To compare 2 models $\Rightarrow$ Likelihood Ratio Test

(F - test)

Suppose $L(\theta \mid Y)$ is the likelihood of parameters $\theta = (\beta, \sigma^2)$

coefficient error var

Likelihood ratio statistic (LR)

$$\frac{\max_{\theta \in \Omega} L(\theta \mid Y)}{\max_{\theta \in \omega} L(\theta \mid Y)}$$

Ω: full parameter space Ω : $(\beta, \sigma^2) \in \mathbb{R}^{p+1}$

ω: $\omega \subseteq \Omega$ (ω is a subset of Ω under H_0)

Principle: Reject H_0 if LR statistic is too large

(How large is too large?) We need to charactize distributions.

[Test 1] Likelihood Ratio Test & F-test

Given the full linear model

$$Y = X\beta + \epsilon \qquad \epsilon \sim N(0, \sigma^2 I_n) \qquad rank(X) = p$$

(1) Likelihood function of (β, σ^2)

$$L(\beta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \|Y - X\beta\|^2\right\}$$

(Notes Oct 17)

(2) MLE under the full model

$$\hat{\beta} = (X^T X)^{-1} X^T Y \qquad \text{(OLS)}$$

$$\hat{\sigma}^2 = \frac{1}{n} \|Y - X\hat{\beta}\|^2 \qquad \left\{ \begin{array}{l} \text{Derivative} = 0 \\ \text{MLE} \end{array} \right.$$

Maximum value of the likelihood

$$L(\hat{\beta}, \hat{\sigma}^2) = (2\pi\hat{\sigma}^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\hat{\sigma}^2} \|Y - X\hat{\beta}\|^2\right\}$$

$$= (2\pi\hat{\sigma}^2)^{-\frac{n}{2}} \exp\left(-\frac{n}{2}\right)$$

(3) MLE subjects to the constraints $H_0: A\beta = c$

(HW 1) Lagrange multiplier

$L(\beta, \sigma^2)$ linear constraints: $A\beta - c = 0$

The Lagrange function is given by

(log likelihood)

$$\mathcal{L}(\beta, \sigma^2, \lambda) = \log L(\beta, \sigma^2) - \lambda^T (A\beta - c)$$

$$= -\frac{n}{2}\log \sigma^2 - \frac{1}{2\sigma^2}\|Y - X\beta\|^2 - \lambda^T(A\beta - c) + \text{constant}$$

① ② ③

$\text{rank}(A) = q$ $\qquad \lambda^T(A\beta - c) = \sum_{i=1}^{q} \lambda_i (a_i^T \beta - c_i)$

$$A = \begin{pmatrix} a_1^T \\ \vdots \\ a_q^T \end{pmatrix}$$

$$\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_q \end{pmatrix} \in \mathbb{R}^q$$

③ $= -\lambda^T A\beta + \lambda^T c$

$$\frac{\partial ③}{\partial \beta} = -(\lambda^T A)^T$$

(by Quiz 0)

Find stationary point by taking derivatives

$$(*)\begin{cases} \dfrac{\partial \mathcal{L}(\beta, \sigma^2, \lambda)}{\partial \beta} = \dfrac{1}{\sigma^2} X^T (Y - X\beta) - A^T\lambda & (*1) \\[2ex] \dfrac{\partial \mathcal{L}}{\partial \sigma^2} = -\dfrac{n}{2} \times \dfrac{1}{\sigma^2} + \dfrac{1}{2(\sigma^2)^2}\|Y - X\beta\|^2 & (*2) \\[2ex] \dfrac{\partial \mathcal{L}}{\partial \lambda} = -(A\beta - c) & (*3) \end{cases}$$

Similar to Notes Oct 17, we consider 2 steps

[Step 1] Given σ^2 (1) find $\hat{\beta}_H$ and $\hat{\lambda}_H$ solve (*) $= 0$

(2) Prove

$$L(\hat{\beta}_H, \sigma^2) \geq L(\beta, \sigma^2)$$

for any $\beta \in \mathbb{R}^p$ subject to $A\beta = c$

[1.(1)] (*1) $= 0$

$$\Rightarrow X^T X \hat{\beta}_H = X^T Y - A^T \hat{\lambda}_H \sigma^2$$

$$\Rightarrow \hat{\beta}_H = (X^T X)^{-1} X^T Y - (X^T X)^{-1} A^T \hat{\lambda}_H \sigma^2 \quad (*1.1)$$