CHAPTER 3 Hypothesis test and confidence interval.

Maximum likelihood estimator

$$y = X\beta + \varepsilon. \qquad \varepsilon \sim N(0, \sigma^2 I_n)$$

maximum likelihood estimator of β

$$\hat{\beta} = (X'X)^{-1} X'y$$

maximum likelihood estimator of σ^2

$$\hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta}) = \frac{1}{n} e'e$$

proof: $y_i \sim N\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}, \ \sigma^2\right)$

$$\underset{n\times 1}{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}} = \underset{n\times(k+1)}{\begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & & & \\ 1 & x_{i1} & x_{i2} & \cdots & x_{ik} \\ \vdots & \vdots & & & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}} \cdot \underset{(k+1)\times 1}{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_i \\ \vdots \\ \beta_k \end{pmatrix}} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$f(y_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(y_i - \mu_i)^2\right]. \qquad \mu_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}$$

$$L = \prod_{i=1}^{n} f(y_i) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu_i)^2\right]$$

$$= \sigma^{-n}(2\pi)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right]$$

$$\rightarrow \max\ l = \log L = -\frac{n}{2}\ln(\sigma^2) - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)$$

$$\begin{cases} \frac{\partial l}{\partial \beta} = 0 \\ \frac{\partial l}{\partial \sigma^2} = 0 \end{cases} \Rightarrow \quad \hat{\beta} = (X'X)^{-1}X'y$$

$$\hat{\sigma}^2 = \frac{(y-X\beta)'(y-X\beta)}{n}.$$

$$\left.\begin{aligned} \hat{\sigma}^2 &= \frac{(n-k-1)s^2}{n} \\ E(\hat{\sigma}^2) &= \sigma^2 - \frac{k+1}{n}\cdot\sigma^2 \neq \sigma^2. \end{aligned}\right\} \ \hat{\sigma}^2 \text{ is biased}$$

Cochran's theorem.

$y \sim N(\mu, \sigma^2 I_n)$, M space $n\times(k+1)$, P is the orthogonal projection matrix on M.

(1) $Py \sim N(Py, \sigma^2 P)$ (2) Py and $y-Py$ independant.

(3) $\frac{(Py-P\mu)'(Py-P\mu)}{\sigma^2} = \frac{(y-\mu)'P(y-\mu)}{\sigma^2} \sim \chi^2_{k+1}$

(4). in particular, $\frac{y'Ay}{\sigma^2} \sim \chi^2_r$ where $y \sim N(0, \sigma^2 I_n)$

A is a symmetric and idempotent matrix. $tr(A) = r$.

Distribution of estimators: known variance.

1. $\hat{\beta} \sim N(\beta, \underbrace{\sigma^2 (X'X)^{-1}}_{cov(\hat{\beta})})$

2. $\frac{(n-k-1)S^2}{\sigma^2} \sim \chi^2_{n-k-1}$

3. $\hat{\beta}$ and s^2 are independant.

Distribution of estimators: unknown variance.

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

1. $T_j = \frac{\hat{\beta}_j - \beta_j}{S\sqrt{[(X'X)^{-1}]_{jj}}} \sim t_{n-k-1}$. $\forall j = 0, 1, \dots k$.

2. $\frac{1}{(k+1)S^2}(\hat{\beta}-\beta)'(X'X)(\hat{\beta}-\beta) \sim F_{k+1,\, n-k-1}$.

proof: (1) $$z_j = \frac{\hat{\beta}_j - \beta_j}{\sigma\sqrt{[(X'X)^{-1}]_{jj}}} \sim N(0,1)$$

$$K_1 = (n-k-1)\frac{s^2}{\sigma^2} \sim \chi^2_{n-k-1}$$

z_j and K_1 independant,

$$T_j = \frac{z_j}{\sqrt{\frac{K_1}{n-k-1}}} = \frac{\hat{\beta}_j - \beta_j}{s\sqrt{[(X'X)^{-1}]_{jj}}} \sim t_{n-k-1}.$$

(2) $$Z = \frac{\hat{\beta} - \beta}{\sigma\sqrt{(X'X)^{-1}}} = \frac{(X'X)^{\frac{1}{2}}(\hat{\beta}-\beta)}{\sigma} \sim N_{k+1}(0, I).$$

$(A^{-1})' = (A')^{-1}$

$$K_2 = Z'Z = \frac{(\hat{\beta}-\beta)'\left[(X'X)^{\frac{1}{2}}\right]' \cdot (X'X)^{\frac{1}{2}}(\hat{\beta}-\beta)}{\sigma^2}$$

$$= \frac{(\hat{\beta}-\beta)'(X'X)^{\frac{1}{2}}(X'X)^{\frac{1}{2}}(\hat{\beta}-\beta)}{\sigma^2}$$

$$= \frac{(\hat{\beta}-\beta)'(X'X)(\hat{\beta}-\beta)}{\sigma^2} \sim \chi^2_{k+1}$$

$\because K_1 \sim \frac{(n-k-1)s^2}{\sigma^2} \sim \chi^2_{n-k-1}$ and $K_2 = \frac{(\hat{\beta}-\beta)'(X'X)(\hat{\beta}-\beta)}{\sigma^2} \sim \chi^2_{k+1}$

ind.

$$\Rightarrow F = \frac{K_2/(k+1)}{K_1/(n-k-1)} = \frac{(\hat{\beta}-\beta)'(X'X)(\hat{\beta}-\beta)}{(k+1)s^2} \sim F_{k+1,\, n-k-1}.$$

Linear hypothesis.

$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & \vdots & & & \\ \vdots & & & & \\ 1 & x_{i1} & x_{i2} & \cdots & x_{ik} \\ & \vdots & & & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$H_0: C\beta = d.$ vs $H_1: C\beta \neq d.$ C : matrix$_{m \times (k+1)}$

$$F = \frac{1}{ms^2}(C\hat{\beta}-d)'\left[C(X'X)^{-1}C'\right]^{-1}(C\hat{\beta}-d) \sim F_{m,\, n-k-1}.$$

rejete H_0 when $F = \frac{1}{ms^2}(c\hat{\beta}-d)'[c(X'X)^{-1}c']^{-1}$

$(c\hat{\beta}-d) \geq F_{m,\, n-k-1,\, \alpha}$.

Maximum likelihood Test

1. without constraints.

$$y = X\beta + \varepsilon. \qquad \varepsilon \sim N(0, \sigma^2 I_n)$$

$$L(\beta, \sigma^2 | y) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right]$$

maximum likelihood estimator of β.

$$\hat{\beta} = (X'X)^{-1}X'y.$$

maximum likelihood estimator of σ^2

$$\hat{\sigma}_{H_1} = \hat{\sigma}^2 = \frac{1}{n}(y-X\hat{\beta})'(y-X\hat{\beta}) = \frac{1}{n}e'e.$$

2. with constraints.

H_0: $c\beta = d$. $\qquad H_1$: $c\beta \neq d$.

$$\min\ Q = (y-X\beta)'(y-X\beta)$$

$$\text{s.t.}\quad c\beta - d = 0.$$

$$\hat{\beta}_{H_0} = \hat{\beta} + (X'X)^{-1} C' \left[C (X'X)^{-1} C' \right]^{-1} (d - C\hat{\beta})$$

$$\hat{\sigma}^2_{H_0} = \frac{1}{n} (y - X\hat{\beta}_{H_0})' (y - X\hat{\beta}_{H_0})$$

$$\frac{\hat{\sigma}^2_{H_0}}{\hat{\sigma}^2_{H_1}} = 1 + \frac{(C\hat{\beta} - d)' \left[C (X'X)^{-1} C' \right]^{-1} (C\hat{\beta} - d)}{(y - X\hat{\beta})' (y - X\hat{\beta})}$$

Likelihood ratio test

$H_0 : C\beta - d = 0$. vs $H_1 : C\beta - d \neq 0$

$$\Lambda = \frac{\max_{H_1} L(\beta, \sigma^2 | y)}{\max_{H_0} L(\beta, \sigma^2 | y)} = \left(\frac{\hat{\sigma}^2_{H_0}}{\hat{\sigma}^2_{H_1}} \right)^{n/2}$$

rejete H_0 if Λ is large.

Theorem: Distribution of maximum likelihood test.

$\varepsilon_i \sim N(0, \sigma^2)$ $H_0 : C\beta - d = 0$.

$$F = \frac{(C\hat{\beta} - d)' \left[C (X'X)^{-1} C' \right]^{-1} (C\hat{\beta} - d) / m}{S^2} \sim F_{m, n-k-1}.$$

Comparaison of regression equation

model 1: $y_1 = X_1\beta_1 + \varepsilon_1$. X_1: $n_1 \times p$. $\beta_1 : p\times 1$

model 2: $y_2 = X_2\beta_2 + \varepsilon_2$ X_2: $n_2\times p$. $\beta_2 = p \times 1$.

Suppose that the first elements of β_1 and β_2 are equal. $\beta_1 = \begin{pmatrix} \beta^{(1)} \\ \beta_1^{(2)} \end{pmatrix}$ $\beta_2 = \begin{pmatrix} \beta^{(1)} \\ \beta_2^{(2)} \end{pmatrix}$

$\beta^{(1)}$: $r\times 1$ $\beta_1^{(2)}$ and $\beta_2^{(2)}$: $(p-r)\times 1$.

H_0: $\beta_1^{(2)} = \beta_2^{(2)}$ vs H_1: $\beta_1^{(2)} \neq \beta_2^{(2)}$

$X_1 : [\, {X_1^{(1)}}_{n_1\times r} \quad {X_1^{(2)}}_{n_1\times(p-r)} \,]$ $X_2 : [\, {X_2^{(1)}}_{n_2\times r} \quad {X_2^{(2)}}_{n_2\times(p-r)} \,]$

$$y_1 = X_1^{(1)}\beta^{(1)} + X_1^{(2)} \cdot \beta_1^{(2)} + \varepsilon_1$$

$$y_2 = X_2^{(1)}\beta^{(1)} + X_2^{(2)}\beta_2^{(2)} + \varepsilon_2$$

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \quad X = \begin{bmatrix} X_1^{(1)} & X_1^{(2)} & 0 \\ X_2^{(1)} & 0 & X_2^{(2)} \end{bmatrix} \quad \beta = \begin{pmatrix} \beta^{1} \\ \beta_1^{(1)} \\ \beta_2^{(2)} \end{pmatrix}$$

The model becomes:

$$y = X\beta + \varepsilon. \qquad \text{s.t. } H_0: C\beta = 0.$$

$$C = \left[\, 0_{(p-r)\times r} \quad I_{p-r} \quad -I_{(p-r)} \,\right]$$

$$C = {\scriptstyle p-r}\left\{ \left[\underbrace{\begin{matrix} 0 & 0 & \cdots & 0 \\ \vdots & & & \\ \vdots & & & \\ 0 & 0 & \cdots & 0 \end{matrix}}_{r} \quad \underbrace{\begin{matrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & & 1 \end{matrix}}_{p-r} \quad \underbrace{\begin{matrix} -1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & \cdots & & -1 \end{matrix}}_{p-r} \right] \right._{(p-r)\times(2p-r)}$$

$$F = \frac{(C\hat{\beta} - d)'\left[C(X'X)^{-1}C'\right]^{-1}(C\hat{\beta} - d)}{(p-r)\, s^2} \sim F_{p-r,\,(n_1+n_2-2p+r)}$$

Confidence interval of parameters

(1) two-sided confidence interval level $1-\alpha$ of β_j

$$IC = \left[\hat{\beta}_j - t_{n-k-1,\frac{\alpha}{2}} \cdot se(\hat{\beta}_j), \right.$$
$$\left. \hat{\beta}_j + t_{n-k-1,\frac{\alpha}{2}} \cdot se(\hat{\beta}_j) \right]$$

$$se(\hat{\beta}_j) = S \cdot \sqrt{[(X'X)^{-1}]_{jj}}$$

(2) two-sided confidence interval level $1-\alpha$ of $c\beta$

$$c\hat{\beta} \pm t_{n-k-1,\frac{\alpha}{2}} \cdot S \cdot \sqrt{c(X'X)^{-1}c'}$$

$$c: \quad 1 \times (k+1)$$

(3) Confidence region level $1-\alpha$ of β.

$$R = \Big\{ \beta : \underbrace{(\hat{\beta}-\beta)' \underbrace{(X'X)}_{(k+1)\cdot(k+1)} (\hat{\beta}-\beta)}_{\text{Constante.}} \leq (k+1) \cdot S^2 F_{k+1,\, n-k-1;\, \alpha} \Big\}.$$

(4) Confidence region of level $1-\alpha$ for $c\beta$.

$$R = \{ c\beta : (c\hat{\beta} - c\beta)' \left[c(X'X)^{-1} c' \right]^{-1} (c\hat{\beta} - c\beta) \leq m s^2 F_{m, n-k-1, \alpha} \}.$$

$\dim C$: $m \times (k+1)$

Confidence interval for the mean of predicted value.

$x_0' = [1 \; x_{01} \; x_{02} \; \cdots \; x_{0k}]$ new vector of explicative variables.

$$\hat{y}_0 = x_0' \cdot \hat{\beta}$$

(1) $$T = \frac{\hat{y}_0 - x_0'\beta}{s \cdot \sqrt{x_0'(X'X)^{-1} x_0}} \sim t_{n-k-1}$$

(2) two-sided confidence interval $1-\alpha$ level for

$$x_0'\beta : \quad \hat{y}_0 \pm t_{n-k-1, \frac{\alpha}{2}} \, s \cdot \sqrt{x_0' (X'X)^{-1} x_0}$$

x_0': $1 \times (k+1)$ $\quad X'$: $(k+1) \times n$ $\quad (X'X)^{-1} = (k+1) \times (k+1)$

Confidence interval for a future value of the dependant variable.

$$x_0' = [1 \quad x_{01} \ \cdots \ x_{0K}].$$

$$y_0 = x_0' \cdot \beta + \varepsilon_0 \qquad \varepsilon_0 \sim N(0, \sigma^2)$$

$$\hat{y}_0 = x_0' \cdot \hat{\beta} \qquad E(y_0 - \hat{y}_0) = 0.$$

$$\operatorname{var}(\hat{y}_0 - y_0) = \operatorname{var}(x_0'(\hat{\beta} - \beta) - \varepsilon_0)$$
$$= \sigma^2 [x_0' (X'X)^{-1} x_0 + 1]$$

Confidence interval of y_0.

(1) $$T = \frac{\hat{y}_0 - y_0}{s\sqrt{1 + x_0'(X'X)^{-1} x_0}} \sim t_{n-K-1}$$

(2) two-sided confidence interval level $1-\alpha$.

$$\hat{y}_0 \pm t_{n-K-1, \frac{\alpha}{2}} \cdot S \cdot \sqrt{1 + x_0'(X'X)^{-1} x_0}$$