Ch7. Hypothesis testing and Confidence Intervals

7.1 Hypothesis testing: General Hypothesis

Let $\boldsymbol{y} \sim N(\boldsymbol{X}\beta, \sigma^2 \boldsymbol{I})$, Then

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \qquad \hat{\beta} \sim N(\beta, (\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma^2)$$

any given matrix

.

- We are now interested in testing the following hypothesis:

$$H_0 : \boldsymbol{K}'\beta = \boldsymbol{\mu} \qquad (7.1)$$

constant

where $\boldsymbol{K}'$ is $q \times (p+1)$, and $\boldsymbol{K}'$ is assumed to be full row rank.

7.1 Hypothesis testing: General Hypothesis

- The test statistics

$$\begin{aligned} F(H) &= \frac{Q/q}{SSE/[n-r(\boldsymbol{X})]} = \frac{Q}{q\hat{\sigma}^2} \\ &\sim F_{[q, N-r(\boldsymbol{X}), \frac{1}{2\sigma^2}(\boldsymbol{K}'\beta-\boldsymbol{\mu})'[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{K}'\beta-\boldsymbol{\mu})]} \end{aligned}$$

  Here $\hat{\sigma}^2 = SSE/(n-r(\boldsymbol{X}))$ which is the unbiased estimator of σ^2.

- Under $H_0 : \boldsymbol{K}'\beta = \boldsymbol{\mu}$, we have

$$F(\boldsymbol{H}) \overset{H_0}{\sim} F_{\{q, n-r(\boldsymbol{X})\}}.$$

p-value = $Pr(F_{q, n-r(X)} \geq F(H))$

calculated based on the set of observations!

7.1 Hypothesis testing: General Hypothesis

Note that

- $$\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{\mu} \sim N[\boldsymbol{K}'\boldsymbol{\beta} - \boldsymbol{\mu}, \boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}\sigma^2]$$
- $(\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K})^{-1}$ is symmetric.
- Let
$$Q = (\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{\mu})'[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{\mu})$$
Thm 5.1
(Q is sometimes denoted by SSH) then
$$\frac{Q}{\sigma^2} \sim \chi^2_{\{q, \frac{1}{2\sigma^2}(\boldsymbol{K}'\boldsymbol{\beta} - \boldsymbol{\mu})'[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{K}'\boldsymbol{\beta} - \boldsymbol{\mu})\}}$$
- Q and SSE are independent

$SSE = (\underset{\sim}{y} - \hat{\underset{\sim}{y}})'(\underset{\sim}{y} - \hat{\underset{\sim}{y}}) = \underset{\sim}{y}'(\underset{\sim}{I} - \underset{\sim}{H})\underset{\sim}{y}$, $H = \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'$

SSE and $\hat{\beta}$ are independent $\Rightarrow$ SSE and Q are independent

7.1 Hypothesis testing: General Hypothesis

Consider the following model

$$\boldsymbol{y} \sim N(\boldsymbol{X}\boldsymbol{a}, \sigma^2 \boldsymbol{I})$$

under the constraint (the null hypothesis) H_0 : $\boldsymbol{K}'\boldsymbol{a} = \boldsymbol{\mu}$.

- Denote the LS estimator of $\boldsymbol{a}$ by $\tilde{\boldsymbol{a}}$.
- To obtain $\tilde{\boldsymbol{a}}$ by minimizing

$$(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a}) + 2\theta'(\boldsymbol{K}'\boldsymbol{a} - \boldsymbol{\mu})$$

with respect to $\boldsymbol{a}$ and θ. Note that 2θ is a vector of Lagrange multipliers. After the minimization,

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

$$\begin{aligned} \tilde{\boldsymbol{a}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{K}(\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K})^{-1}(\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{\mu})) \\ &= \hat{\beta} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}(\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K})^{-1}(\boldsymbol{K}'\hat{\beta} - \boldsymbol{\mu}) \end{aligned} \tag{7.2}$$

- $\hat{\beta} - \tilde{\boldsymbol{a}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}(\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K})^{-1}(\boldsymbol{K}'\hat{\beta} - \boldsymbol{\mu})$;
- $\tilde{\boldsymbol{a}}$ is the BLUE for the model with the constraint.

7.1 Proof of that $\tilde{a}$ is the BLUE for the model with the constraint

Proof.

Let $\mathbf{L} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a})$. For $\forall \boldsymbol{a}_0$ with constraint $\boldsymbol{K}'\boldsymbol{a}_0 = \boldsymbol{\mu}$, we have

$$
\begin{aligned}
\mathbf{L}_0 &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a}_0)'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a}_0) \\
&= (\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{a}} + \boldsymbol{X}\tilde{\boldsymbol{a}} - \boldsymbol{X}\boldsymbol{a}_0)'(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{a}} + \boldsymbol{X}\tilde{\boldsymbol{a}} - \boldsymbol{X}\boldsymbol{a}_0) \\
&= (\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{a}})'(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{a}}) + (\boldsymbol{X}\tilde{\boldsymbol{a}} - \boldsymbol{X}\boldsymbol{a}_0)'(\boldsymbol{X}\tilde{\boldsymbol{a}} - \boldsymbol{X}\boldsymbol{a}_0) \\
&\quad + 2(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{a}})'(\boldsymbol{X}\tilde{\boldsymbol{a}} - \boldsymbol{X}\boldsymbol{a}_0)
\end{aligned}
$$

7.1 Proof (continued)

But

$$\begin{aligned}
& (\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{a}})'(\boldsymbol{X}\tilde{\boldsymbol{a}} - \boldsymbol{X}\boldsymbol{a}_0) \\
= & (\boldsymbol{y}'\boldsymbol{X} - \tilde{\boldsymbol{a}}'\boldsymbol{X}'\boldsymbol{X})(\tilde{\boldsymbol{a}} - \boldsymbol{a}_0) \\
= & [\boldsymbol{y}'\boldsymbol{X} - \hat{\beta}'(\boldsymbol{X}'\boldsymbol{X}) + (\boldsymbol{K}'\hat{\beta} \\
& -\mu)'(\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K})^{-1}\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{X})](\tilde{\boldsymbol{a}} - \boldsymbol{a}_0) \\
= & (\boldsymbol{K}'\hat{\beta} - \mu)'(\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K})^{-1}\boldsymbol{K}'(\tilde{\boldsymbol{a}} - \boldsymbol{a}_0) = 0
\end{aligned}$$

Because $(\boldsymbol{K}'\tilde{\boldsymbol{a}} = \boldsymbol{K}'\boldsymbol{a}_0 = \mu)$. Thus,

$$\mathbf{L}_0 = (\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{a}})'(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{a}}) + (\tilde{\boldsymbol{a}} - \boldsymbol{a}_0)'\boldsymbol{X}'\boldsymbol{X}(\tilde{\boldsymbol{a}} - \boldsymbol{a}_0).$$

Hence,

$$\boldsymbol{a}_0 = \tilde{\boldsymbol{a}} \text{ minimize } \mathbf{L}_0.$$

7.1 Hypothesis testing: General Hypothesis

- Without the null hypothesis, $SSE = (\boldsymbol{y} - \boldsymbol{X}\hat{\beta})'(\boldsymbol{y} - \boldsymbol{X}\hat{\beta})$.
- Under the null hypothesis (reduced Model)

$$\begin{aligned} SSE_{H_0} &= (\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{a}})'(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{a}}) \\ &= [\boldsymbol{y} - \boldsymbol{X}\hat{\beta} + \boldsymbol{X}\hat{\beta} - \boldsymbol{X}\tilde{\boldsymbol{a}}]'[\boldsymbol{y} - \boldsymbol{X}\hat{\beta} + \boldsymbol{X}\hat{\beta} - \boldsymbol{X}\tilde{\boldsymbol{a}}] \\ &= [\boldsymbol{y} - \boldsymbol{X}\hat{\beta} + \boldsymbol{X}(\hat{\beta} - \tilde{\boldsymbol{a}})]'[\boldsymbol{y} - \boldsymbol{X}\hat{\beta} + \boldsymbol{X}(\hat{\beta} - \tilde{\boldsymbol{a}})] \\ &= (\boldsymbol{y} - \boldsymbol{X}\hat{\beta})'(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}) + (\hat{\beta} - \tilde{\boldsymbol{a}})'\boldsymbol{X}'\boldsymbol{X}(\hat{\beta} - \tilde{\boldsymbol{a}}) \end{aligned}$$

Since $((\hat{\beta} - \tilde{\boldsymbol{a}})'\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}) = 0)$

- From (7.2),

$$\begin{aligned} SSE_{H_0} &= SSE + (\boldsymbol{K}'\hat{\beta} - \mu)'[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X} \\ &\quad \boldsymbol{K}[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{K}\hat{\beta} - \mu) \\ &= SSE + (\boldsymbol{K}'\hat{\beta} - \mu)'(\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K})^{-1}(\boldsymbol{K}'\hat{\beta} - \mu) \\ &= SSE + Q \\ &\geq SSE \end{aligned}$$

"Q

7.1 Special cases. (1) $H_0 : \beta = \beta_0$

$\beta_0 = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

In this case, $\boldsymbol{K}' = \boldsymbol{I}, q = p + 1, \boldsymbol{\mu} = \boldsymbol{\beta_0}$

1.

$$Q = (\hat{\beta} - \beta_0)'\boldsymbol{X}'\boldsymbol{X}(\hat{\beta} - \beta_0)$$

2. $F(H) = \frac{Q}{(p+1)\hat{\sigma}^2}$
3. Under the null hypothesis,

$$F(H) \sim F_{\{p+1, n-(p+1)\}}$$

under H_0

4. $\tilde{\boldsymbol{a}} = \hat{\beta} - (\hat{\beta} - \boldsymbol{a}_0) = \boldsymbol{\beta_0}$

7.1 Special cases. (2) $H_0 : \lambda'\beta = m$

special case $\lambda = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \to j\text{-th}$

In this case, $\boldsymbol{K}' = \lambda'$, $q = 1$, $\boldsymbol{\mu} = m$

$H_0: \lambda'\beta = 0 \Leftrightarrow \beta_j = 0$

1
$$Q = (\lambda'\hat{\beta} - m)'[\lambda'(\boldsymbol{X}'\boldsymbol{X})^{-1}\lambda]^{-1}(\lambda'\hat{\beta} - m)$$
$$= (\lambda'\hat{\beta} - m)^2/\lambda'(\boldsymbol{X}'\boldsymbol{X})^{-1}\lambda;$$

$Q = \frac{\hat{\beta}_j^2}{g_{jj}}$, g_{jj} – j-th diagonal element of $(X'X)^{-1}$

2 $F(H) = \frac{Q}{\hat{\sigma}^2}$;

3 Under the null hypothesis,
$$F(H) \sim F_{(1, n-r(\boldsymbol{X}))}$$

Note: $\sqrt{F(H)} = \frac{\sqrt{Q}}{\hat{\sigma}} \sim t_{n-r(\boldsymbol{X})}$;

$\Rightarrow \frac{\hat{\beta}_j^2}{g_{jj}^{1/2} \cdot \hat{\sigma}} \sim t_{n-p-1}$

t-statistic

4
$$\tilde{\boldsymbol{a}} = \hat{\beta} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\lambda[\lambda'(\boldsymbol{X}'\boldsymbol{X})^{-1}\lambda]^{-1}(\lambda'\hat{\beta} - \boldsymbol{\mu})$$
$$= \hat{\beta} - \frac{(\lambda'\hat{\beta} - \boldsymbol{\mu})}{\lambda'(\boldsymbol{X}'\boldsymbol{X})^{-1}\lambda}(\boldsymbol{X}'\boldsymbol{X})^{-1}\lambda.$$

used in R-function "lm"

Note: $\lambda'\hat{\beta} - \boldsymbol{\mu} \sim N(\lambda'\beta - \boldsymbol{\mu}, \lambda'(\boldsymbol{X}'\boldsymbol{X})^{-1}\lambda\sigma^2)$

7.1 Special cases. (3) $H_0 : \beta_2 = 0$

Here, $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$.

$$Y = (X_1, X_2)\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \varepsilon$$

Remark 7.1 $\underbrace{\underset{h\times(p+1)}{K'} = (\underset{h\times(p+1-h)}{0} \; \vdots \; \underset{h\times h}{I_h})}$ $\cancel{\beta}$

$$\underset{p+1}{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}\begin{matrix} \}p+1-h \\ \}h \end{matrix}$$

$H_0: \beta_2 = 0 \iff H_0: K'\beta = 0$ $\cancel{\beta}$ β

$$Q = (K'\hat{\beta})' \underline{[K'(X'X)^{-1}K]^{-1}} [K'\hat{\beta}]$$

$\hat{\beta}_2$

$$X'X = \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix} \overset{d}{=} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

$$(X'X)^{-1} = \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}B^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}B^{-1} \\ -B^{-1}A_{21}A_{11}^{-1} & B^{-1} \end{pmatrix} \overset{d}{=} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

$\underline{(*)}$

$$\left[\underset{\sim}{K}'(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{K}\right]^{-1} = \underset{\sim}{B} = \underset{\sim}{X}_2'\underset{\sim}{X}_2 - \underset{\sim}{X}_2'\underbrace{\underset{\sim}{X}_1(\underset{\sim}{X}_1'\underset{\sim}{X}_1)^{-1}\underset{\sim}{X}_1'}_{H_1}\underset{\sim}{X}_2$$

$$= \underset{\sim}{X}_2'(\underset{\sim}{I} - \underset{\sim}{H}_1)\underset{\sim}{X}_2$$

$$\begin{pmatrix}\hat{\beta}_1 \\ \hat{\beta}_2\end{pmatrix} = \hat{\beta} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22}\end{pmatrix}\begin{pmatrix}\underset{\sim}{X}_1'\underset{\sim}{Y} \\ \underset{\sim}{X}_2'\underset{\sim}{Y}\end{pmatrix}$$

$$= \begin{pmatrix} B_{11}\underset{\sim}{X}_1'\underset{\sim}{Y} + B_{12}\underset{\sim}{X}_2'\underset{\sim}{Y} \\ \underline{B_{21}\underset{\sim}{X}_1'\underset{\sim}{Y} + B_{22}\underset{\sim}{X}_2'\underset{\sim}{Y}}\end{pmatrix}$$

$$\Rightarrow \underset{\sim}{\hat{\beta}_2} = \underset{\sim}{B}^{-1}\underset{\sim}{X}_2'(\underset{\sim}{I} - \underset{\sim}{H}_1)\underset{\sim}{Y}$$

$$\underline{\underline{Q}} = \hat{\beta}_2'\left[\underset{\sim}{X}_2'(\underset{\sim}{I} - \underset{\sim}{H}_1)\underset{\sim}{X}_2\right]\underset{\sim}{\hat{\beta}_2}$$

$$= \underset{\sim}{Y}'(I - \underset{\sim}{H}_1)\underset{\sim}{X}_2 \underset{\sim}{B}^{-1} \underbrace{\left[\underset{\sim}{X}_2'(I - \underset{\sim}{H}_1)\underset{\sim}{X}_2\right]}_{= \underset{\sim}{B}} \underset{\sim}{B}^{-1} \underset{\sim}{X}_2'(I - \underset{\sim}{H}_1)\underset{\sim}{Y}$$

$$= \underset{\sim}{Y}' \underbrace{(I - H_1)\underset{\sim}{X}_2 \underset{\sim}{B}^{-1} \underset{\sim}{X}_2'(I - \underset{\sim}{H}_1)}_{\underset{\sim}{A}} \underset{\sim}{Y}$$

$$\frac{Q}{\sigma^2} \sim \chi^2\left(h, \frac{1}{2\sigma^2} \beta_2 \underset{\sim}{X}_2'(I - H_1)\underset{\sim}{X}_2 \beta_2\right)$$

Thm I 1.
$\lambda = \frac{1}{2\sigma^2} \underset{\sim}{\mu}' \underset{\sim}{A} \underset{\sim}{\mu}$,
$\underset{\sim}{\mu} = \underset{\sim}{X} \underset{\sim}{\beta}$

$$F = \frac{Q}{h \hat{\sigma}^2} \overset{H_0}{\sim} F(h, n-p-1)$$

Using (*), it is easy to prove that

$$Q = \underset{\sim}{Y}'(\underset{\sim}{H} - \underset{\sim}{H}_1)\underset{\sim}{Y}$$

Thus
$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \underset{\sim}{Y}'\underset{\sim}{Y} - n\bar{y}^2$$

$$= \underbrace{\left(y'y - \hat{\beta}'X'y\right)}_{SSE} + \underbrace{\left(\hat{\beta}'X'y - \hat{\beta}_1^{*\prime}X_1'y\right)}_{SS(\beta_2|\beta_1) = Q}$$

$$+ \underbrace{\left(\hat{\beta}_1^{*\prime}X_1'y - n\bar{y}^2\right)}_{SSR \text{ (reduced)}}$$

$$\hat{\beta}_1^* = (X_1'X_1)^{-1}X_1'y \leftarrow y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

when $\beta_2 = 0$

Special case: $Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$

$$\beta = \begin{pmatrix} \beta_0 \\ \hline \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \hline \beta_2 \end{pmatrix} \quad H_0: \beta_2 = 0 \text{ v.s. } \beta_2 \neq 0$$

$\Leftrightarrow H_0: \beta_1 = \cdots = \beta_p = 0$ v.s. H_1: at least one of $\beta_1 \cdots \beta_p \neq 0$

$k = p$. $Q = Y'(H - \frac{1}{n}11')Y = SSR$.

$$F = \frac{Q}{k\hat{\sigma}^2} = \frac{SSR/p}{SSE/n-p-1} \overset{H_0}{\sim} F(p, n-p-1).$$

7.1 Likelihood Ratio Test

Theorem: If $\boldsymbol{y}$ is $N_n(\boldsymbol{x}\boldsymbol{a}, \sigma^2\mathbf{I})$, where (rank of $\boldsymbol{x}$ is $p+1$), the likelihood ratio for H_0: $\boldsymbol{a} = \mathbf{0}$ can be based on

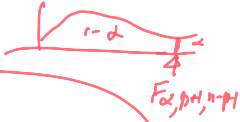
$$F = \frac{\hat{\mathbf{a}}'\mathbf{x}'\mathbf{y}/(p+1)}{(\mathbf{y}'\mathbf{y} - \hat{\mathbf{a}}'\mathbf{x}'\mathbf{y})/(n-p-1)}.$$

H_0 is rejected if $F > F_{\alpha,p+1,n-p-1}$.

$$LR = \frac{\max_{\underset{\sim}{a}:H_0} L(\underset{\sim}{a}, \sigma^2)}{\max_{H_1} L(\underset{\sim}{a}, \sigma^2)}$$

7.2 Confidence intervals and prediction intervals

- Confidence region for $\boldsymbol{\beta}$
- Confidence interval for β_j
- Confidence interval for $\boldsymbol{\lambda}'\boldsymbol{\beta}$
- Confidence interval for $E(y^*)$ given $\boldsymbol{x} = \boldsymbol{x}^*$
- Prediction interval for a future observation
- Confidence interval for σ^2
- Simultaneous intervals
    - Familywise confidence level
    - Bonferroni procedure
    - Scheffé procedure



7.2 Confidence region for β

Since

$$P\left[\frac{(\hat{\boldsymbol\beta}-\boldsymbol\beta)'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol\beta}-\boldsymbol\beta)}{(p+1)\hat\sigma^2} \le F_{\alpha,p+1,n-p-1}\right] = 1-\alpha,$$

a $100(1-\alpha)\%$ joint confidence region for $\beta_0, \beta_1, ..., \beta_p$ is defined to consist of all vectors $\boldsymbol\beta$ that satisfy

$$(\hat{\boldsymbol\beta}-\boldsymbol\beta)'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol\beta}-\boldsymbol\beta) \le (p+1)\hat\sigma^2 F_{\alpha,p+1,n-p-1}.$$

?

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

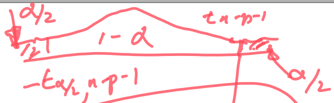
$\frac{SSE}{\sigma^2} \sim \chi^2_{n-p-1}$

$\hat\sigma^2 = SSE/(n-p-1)$

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

$$\frac{(\hat\beta-\beta)'\frac{(X'X)}{\sigma^2}(\hat\beta-\beta)/p+1}{\frac{SSE}{\sigma^2}/n-p-1} \sim F_{p+1,n-p-1}$$

7.2 Confidence interval for β_j



Since

$$P[-t_{\alpha/2,n-p-1} \leq \frac{\hat{\beta}_j - \beta_j}{S\{\hat{\beta}_j\}} \leq t_{\alpha/2,n-p-1}] = 1 - \alpha,$$

$t_{\alpha/2,\,n-p-1}$

hence, a $100(1-\alpha)\%$ confidence interval for β_j is

$$\hat{\beta}_j \pm t_{\alpha/2,n-p-1} S\{\hat{\beta}_j\}.$$

$\hat\beta_j \sim N(\beta_j,\ g_{jj}\sigma^2)$, g_{jj} – jth diagonal element of $(X'X)^{-1}$

$\dfrac{\hat\beta_j - \beta_j}{g_{jj}^{1/2}\sigma^2} \sim N(0,1)$ $\Rightarrow$ $\dfrac{\hat\beta_j - \beta_j}{g_{jj}^{1/2}\hat\sigma} = \cdots \sim t_{n-p-1}$

$\hat\sigma^2 = \dfrac{SSE}{n-p-1}$

7.2 Confidence interval for $\underset{1\times(p+1)}{\lambda'}\beta$

Note that

$$t = \frac{\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} - \boldsymbol{\lambda}'\boldsymbol{\beta}}{\hat{\sigma}\sqrt{\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\lambda}}} \sim t_{n-p-1},$$

hence, a $100(1-\alpha)\%$ confidence interval for $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is

$$\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-p-1}\hat{\sigma}\sqrt{\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\lambda}}.$$

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

7.2 Confidence interval for $E(y^*)$ given $x = x^*$

Given that $\boldsymbol{x} = \boldsymbol{x}^*$,

$$E(y^*) = \boldsymbol{x}^{*'}\boldsymbol{\beta},$$
$$\widehat{E(y^*)} = \boldsymbol{x}^{*'}\hat{\boldsymbol{\beta}},$$

$\underset{\sim}{\lambda} = \underset{\sim}{x}^*$

and

$$Var(E(y^*) - \widehat{E(y^*)}) = [\boldsymbol{x}^{*'}(\boldsymbol{X'X})^{-1}\boldsymbol{x}^*]\sigma^2,$$

hence, a $100(1-\alpha)\%$ confidence interval for $E(y^*)$ is

$$\boldsymbol{x}^{*'}\hat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-p-1}\hat{\sigma}\sqrt{\boldsymbol{x}^{*'}(\boldsymbol{X'X})^{-1}\boldsymbol{x}^*}.$$

7.2 Prediction interval for a future observation

Given that $\boldsymbol{x} = \boldsymbol{x}^*$. Let $y^* = \boldsymbol{x}'\boldsymbol{\beta} + \epsilon^*$ be a future value of y when $\boldsymbol{x} = \boldsymbol{x}^*$ that needs to be predicted.

$$\hat{y}^* = \boldsymbol{x}^{*'}\hat{\boldsymbol{\beta}},$$

$$Var(y^* - \hat{y}^*) = [1 + \boldsymbol{x}^{*'}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}^*]\sigma^2,$$

hence, a $100(1-\alpha)\%$ prediction interval for y^* is

$$\hat{y}^* \pm t_{\alpha/2,n-p-1}\hat{\sigma}\sqrt{1 + \boldsymbol{x}^{*'}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}^*}.$$

$y^* - \hat{y}^* \sim N(0, \sigma^2(x^{*'}(x'x)^{-1}x^* + 1))$

7.2 Confidence interval for σ^2

Note that $(n-p-1)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{(n-p-1)}$. Therefore,

$$P[\chi^2_{1-\alpha/2,n-p-1} \leq \frac{(n-p-1)\hat{\sigma}^2}{\sigma^2} \leq \chi^2_{\alpha/2,n-p-1}] = 1-\alpha,$$

hence, a $100(1-\alpha)\%$ confidence interval for σ^2 is

$$\left(\frac{(n-p-1)\hat{\sigma}^2}{\chi^2_{\alpha/2,n-p-1}}, \frac{(n-p-1)\hat{\sigma}^2}{\chi^2_{1-\alpha/2,n-p-1}} \right).$$

7.2 Simultaneous intervals

- Familywise confidence level: $1-\alpha_f$ implies that we are $100(1-\alpha_f)\%$ confident that every interval contains its respective parameter.
- Bonferroni confidence intervals
  - Individual confidence level $1-\alpha_c$;
  - m intervals;
  - If we choose $\alpha_c = \alpha_f/m$, familywise confidence level $\geq 1-\alpha_f$;
  - For m linear functions $\boldsymbol{\lambda}_1'\boldsymbol{\beta}, \boldsymbol{\lambda}_2'\boldsymbol{\beta}, \ldots, \boldsymbol{\lambda}_m'\boldsymbol{\beta}$, for $i = 1, \ldots, m$, the $100(1-\alpha)\%$ Bonferroni confidence intervals are

$$\boldsymbol{\lambda}_i'\hat{\beta} \pm t_{\alpha/2m, n-p-1}\hat{\sigma}\sqrt{\boldsymbol{\lambda}_i'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{\lambda}_i}.$$

  - Scheffé confidence intervals for all possible linear functions $\boldsymbol{\lambda}'\boldsymbol{\beta}$: the $100(1-\alpha)\%$ conservative confidence interval for any and all $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is

$$\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} \pm \hat{\sigma}\sqrt{(p+1)F_{\alpha,p+1,n-p-1}\boldsymbol{\lambda}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{\lambda}}$$

$A = \{\beta_1 \in CI_1\}$ $\quad P(A) = P\{\beta_1 \in CI_1\} = 1 - \alpha_1$

$B = \{\beta_2 \in CI_2\}$ $\quad P(B) = P(\beta_2 \in CI_2\} = 1 - \alpha_2$

$$
\begin{aligned}
P(A \cap B) &= 1 - P(\overline{A \cap B}) \\
&= 1 - P(\bar{A} \cup \bar{B}) \\
&\geq 1 - (P(\bar{A}) + P(\bar{B})) \\
&= 1 - (\alpha_1 + \alpha_2) \\
&= 1 - \alpha
\end{aligned}
$$

$\alpha_1 + \alpha_2 = \alpha$

Multiple testing H_{0j} v.s H_{1j}, $j = 1, \cdots J$

False discovery rate (FDR)

(Multiple comparison)