Sep 25

3. Properties of estimators.

① Dist. of $\hat\beta$: $\hat\beta \sim N(\beta, \sigma^2(X^TX)^{-1})$

Properties: – Unbiased: $E\hat\beta = \beta$.

– Best Linear Unbiased estimator (BLUE). See exercise.

② Dist. of $\hat\sigma^2 = \frac{RSS}{n-p'}$: $\frac{(n-p')\hat\sigma^2}{\sigma^2} = \frac{RSS}{\sigma^2} \sim \chi^2_{n-p'}$

③ $\hat\beta \perp\!\!\!\perp \hat\sigma^2$

All of above are derived based on the assumption $e_i \overset{iid}{\sim} N(0, \sigma^2)$

4. CI & HT

Idea: Derive the pivots in CI or the test stats in HT from above properties.

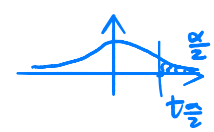
$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + e_i$, $i = 1, \cdots, n$

4.1 T test

HT $H_0: \beta_j = \beta_{j0}$, $j = 0, \cdots, p$

Test stat: $t = \frac{\hat\beta_j - \beta_{j0}}{\hat\sigma\sqrt{c_{jj}}} \overset{H_0}{\sim} t_{n-p'}$ where $C = (X^TX)^{-1}$

Reject H_0 if $|t| > t_{\frac{\alpha}{2}, n-p'}$. $P(|t| > t_{\frac{\alpha}{2}, n-p'}) = \alpha$



CI Pivot: $t = \frac{\hat\beta_j - \beta_j}{\hat\sigma\sqrt{c_{jj}}} \sim t_{n-p'}$

$$\Rightarrow P(|t| \le t_{\frac{\alpha}{2}, n-p'}) = 1-\alpha$$

$\Rightarrow$ $1-\alpha$ c.I. of β_j: $\hat\beta_j \mp t_{\frac{\alpha}{2}, n-p'} \hat\sigma\sqrt{c_{jj}}$.

* See P20 of lecture notes for the case of $p=1$.

Simultaneous C.I. / Mutiple H.T.:

$$\hat\beta_i \mp t_{\frac{\alpha}{2m}, n-p'} \cdot \hat\sigma\sqrt{\sigma_{ii}}, \quad i = 1, \cdots, m.$$

4.2 F test.

4.2.1 All coefficients = 0 .

① Partitioning total variability

$$TSS = RegSS + ResSS$$

$$\Sigma(y_i - \bar{y})^2 = \Sigma(\hat{y}_i - \bar{y})^2 + \Sigma(y_i - \hat{y}_i)^2$$

② Distributions.

$$\frac{ResSS}{\sigma^2} \sim \chi^2_{n-p'} \perp\!\!\!\perp \frac{RegSS}{\sigma^2} \sim \chi^2_{p,\lambda} \Rightarrow \frac{TSS}{\sigma^2} \sim \chi^2_{n-1,\lambda} \,, \quad \lambda = \frac{1}{\sigma^2} \sum_{i,j=1}^{p} \beta_i \beta_j S_{x_i x_j}$$

Derive the properties of estimators based on matrix form

Assumption $e_i \overset{iid}{\sim} N(0, \sigma^2)$

1. $\hat{\beta} = (X^TX)^{-1}X^Ty \sim N(\beta, \sigma^2(X^TX)^{-1})$

Pf. Recall $y = X\beta + e$, $e \sim N(0, \sigma^2 I) \Rightarrow y \sim N(X\beta, \sigma^2 I)$

$\hat{\beta}$ is a linear transformation of y.

$\Rightarrow \hat{\beta} \sim$ Normal.

Derive mean. $\mathbb{E}\hat{\beta} = (X^TX)^{-1}X^T\,\mathbb{E}y = (X^TX)^{-1}X^TX\beta = \beta$.

Derive covariance: $\operatorname{Cov}\hat{\beta} = (X^TX)^{-1}X^T\operatorname{Cov}(y)\cdot X(X^TX)^{-1}$

$$= \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1} = \sigma^2(X^TX)^{-1}$$

2. BLUE

Def. Linear Unbiased Estimator (LUE) of β:

$$\{Ay : E(Ay) = \beta,\ A \in \mathbb{R}^{p' \times n}\}.$$

Def. Best Linear Unbiased Estimator (BLUE): the LUE with the minimum variance.

Thm. Under the assumption $e \sim N(0, \sigma^2 I)$, $\hat{\beta}_{LSE}$ is the BLUE.

Pf. (partial): Let $\tilde{\beta} = Cy$ be any LUE of β with $C = (X^TX)^{-1}X^T + D$ where $D \in \mathbb{R}^{p' \times n} \neq 0$.

"UE" $\Rightarrow \mathbb{E}\tilde{\beta} = \beta$

$$= C\,\mathbb{E}y = [(X^TX)^{-1}X^T + D]X\beta = (X^TX)^{-1}X^TX\beta + DX\beta = \beta + DX\beta$$

$\Rightarrow DX = 0$

$$\operatorname{Cov}(\tilde{\beta}) = C\operatorname{Cov}(y)C^T = \sigma^2 CC^T = \sigma^2[(X^TX)^{-1}X^T + D][X(X^TX)^{-1} + D^T]$$

$$= \sigma^2[(X^TX)^{-1} + (X^TX)^{-1}(DX)^T + DX(X^TX)^{-1} + DD^T]$$

$$= \sigma^2(X^TX)^{-1} + \sigma^2 DD^T$$

$$= \operatorname{Cov}(\hat{\beta}) + \sigma^2 DD^T \succeq \operatorname{Cov}(\hat{\beta}),$$ since DD^T is p.s.d.

3. $\hat{\beta} \perp\!\!\!\perp$ RSS

Lemma: Let $X \sim N(0, I_n)$, $A_{n\times n}$ symmetric. $B_{m\times n}$. $BA = 0$.

Then $BX \perp\!\!\!\perp X'AX$.

Pf. $\hat{\beta} = (X^TX)^{-1}X^Ty = (X^TX)^{-1}X^T(X\beta+e) = \beta + (X^TX)^{-1}X^Te$.

$RSS = \frac{1}{n}(y-\hat{y})^T(y-\hat{y}) = \frac{1}{n} y^T(I - X(X^TX)^{-1}X^T)y$

$\hat{y} = X\hat{\beta} = X(X^TX)^{-1}X\beta \Rightarrow y-\hat{y} = (I - X(X^TX)^{-1}X^T)y$.

$(X^TX)^{-1}X^T(I - X(X^TX)^{-1}X^T) = (X^TX)^{-1}X^T - (X^TX)^{-1}X^T = 0$

Applying lemma, we have $\hat{\beta} \perp\!\!\!\perp$ RSS.