⊛ Test 1 (ANOVA Test)

i) $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$

$H_1 : \beta_j \neq 0$ for at least one j

ii) Model

FM : $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$

RM : $y = \beta_0 + \varepsilon$ ← under H_0

iii) 검정통계량 F

$$F = \frac{SSR / k}{SSE / (n-p)} \sim F(k, n-p) \quad \text{under } H_0$$

iv) 기각역

$R : F > F_\alpha(k, n-p)$ F분포는 단측검정만 가능

※ idea

H_0가 옳다면 SSE가 조금만 줄어들것이다.

$$\frac{\Delta SSE}{SSE(FM)} = \frac{\text{변동된 것}}{\text{남은 것}}$$

$\Delta SSE + SSE(FM) = SSE(RM)$

/ 남은것보다 변동된 것이 많으면

그 설명변수의 설명력이 좋은 것.

∴ F가 클수록 H_0 rejected!

※ k : 변수 갯수

p : 모수 갯수 $(= k+1)$

⊛ Test 2 (Single regression coefficient)

i) $H_0 : \beta_j = 0$ or $\beta_j = \beta_{j0}$ ← constant

$H_1 : \beta_j \neq 0$ $\quad \beta_j \neq \beta_{j0}$

ii) $\hat{\beta}_j \sim N(\beta_j, C_{jj}\sigma^2)$ $\quad C = (X'X)^{-1}$

iii) 검정통계량 t

$$t = \frac{\hat{\beta}_j - \beta_{j0}}{S\sqrt{C_{jj}}} \sim t(n-p)$$

iv) 기각역

$R : |t| > t_{\alpha/2}(n-p)$ t분포는 양측, 단측 검정 모두가능

Partial F-test 가능

i) 좌측과 동일

ii) Model

FM : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$

RM : $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{j-1} x_{j-1} + \beta_{j+1} x_{j+1} + \cdots + \beta_k x_k + \varepsilon$

iii) 검정통계량

$$F = \frac{SSE(RM) - SSE(FM) / 1}{SSE(FM) / (n-p)} \sim F(1, n-p)$$

iv) 기각역

$R : F > F_\alpha(1, n-p)$

⊛ Test 3 (Sequential F-test)

i) $H_0 : \beta_{j1} = \beta_{j2} = \cdots = \beta_{jr} = 0 \quad 1 \le r \le k$

$H_1 : \beta_{jl} \neq 0$ for at least one l

$\Leftrightarrow$ $H_0 : \underset{\sim}{\beta_1} = 0$

$H_1 :$ not H_0

ii) Model

FM : $\underset{\sim}{y} = X_2 \underset{\sim}{\beta_2} + X_1 \underset{\sim}{\beta_1} + \underset{\sim}{\varepsilon}$

RM : $\underset{\sim}{y} = X_2 \underset{\sim}{\beta_2} + \underset{\sim}{\varepsilon}$

$$\underset{\sim}{\beta_1} = \begin{pmatrix} \beta_{j1} \\ \beta_{j2} \\ \vdots \\ \beta_{jr} \end{pmatrix}$$

$\underset{\sim}{\beta_2}$: β_1에 들어가지 않은 회귀계수

& β_0 항은 포함 (test의 대상 X)

: 보정을 위한 term

iii) 검정통계량 F

$$F = \frac{\{SSE(RM) - SSE(FM)\}/r}{SSE(FM)/(n-p)} \sim F(r, n-p)$$

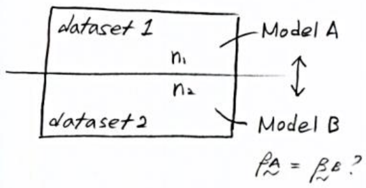
iv) 기각역

$$R : F > F_{\alpha}(r, n-p)$$

(*) Test 4 (the general linear hypothesis)

: 독립적인 dataset 에서의 독립적인 회귀방정식을 비교 (모수 비교)

But ε의 분포는 동일 *



i) Model과 가설

혹은 변수여러개 가능

dataset 1 : $y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i \quad i = 1, 2, \cdots, n_1$

dataset 2 : $y_i = \gamma_0 + \gamma_1 x_{i1} + \varepsilon_i \quad i = n_1+1, n_1+2, \cdots, n_1+n_2$

$H_0 : C\underset{\sim}{\beta} = \underset{\sim}{d}$

$H_1 : C\underset{\sim}{\beta} \neq \underset{\sim}{d}$

$\underset{\sim}{\beta} : p \times 1 \quad C : r \times p \quad \underset{\sim}{d} : r \times 1$

if they are the same line, (under H_0)

$H_0 : \beta_0 = \gamma_0$ & $\beta_1 = \gamma_1$ $\longrightarrow$ $\underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \gamma_0 \\ \gamma_1 \end{pmatrix} \quad C = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \quad \underset{\sim}{d} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

☆ 하나의 dataset 에서 Sequential F-test 가능!

(eg) ① $H_0 : \beta_2 = \beta_3 = 0$

$k=6$

Model

FM : $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_6 x_6 + \varepsilon$

RM : $y = \beta_0 + \beta_1 x_1 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \varepsilon$

$$\underset{C}{\begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}} \underset{\underset{\sim}{\beta}}{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_6 \end{pmatrix}} = \underset{\underset{\sim}{d}}{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}$$

② $H_0 : \beta_2 + \beta_3 = 3$

$$\underset{C}{\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}} \underset{\underset{\sim}{\beta}}{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_6 \end{pmatrix}} = \underset{\underset{\sim}{d}}{3}$$

ii) 검정 통계량 F

$$F = \frac{(C\hat{\beta} - d)'(C(X'X)^{-1}C')^{-1}(C\hat{\beta} - d)/r}{SSE(FM)/(n-p)} \sim F_{(r,\, n-p)}$$

iii) 기각역

$$R : F > F_{\alpha}(r,\, n-p)$$

※ 지시변수를 이용한 형태

dataset 1 : $y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i \quad i = 1, 2, \cdots, n_1$

dataset 2 : $y_i = \gamma_0 + \gamma_1 x_{i1} + \varepsilon_i \quad i = n_1+1, n_1+2, \cdots, n_1+n_2$

$\Rightarrow y_i = \beta_0 Z_{i1} + \beta_1 x_{i1} Z_{i1} + \gamma_0 Z_{i2} + \gamma_1 x_{i1} Z_{i2} + \varepsilon_i$, $\quad Z_1 = \begin{cases} 1 & i = 1, 2, \cdots, n_1 \\ 0 & \text{otherwise} \end{cases}$, $\quad Z_2 = \begin{cases} 1 & i = n_1+1, n_1+2, \cdots, n_1+n_2 \\ 0 & \text{otherwise} \end{cases}$

$$\underset{\sim}{y} = X \begin{pmatrix} \beta_0 \\ \beta_1 \\ \gamma_0 \\ \gamma_1 \end{pmatrix} + \underset{\sim}{\varepsilon}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_1} \\ \hline y_{n_1+1} \\ y_{n_1+2} \\ \vdots \\ y_{n_1+n_2} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & 0 & 0 \\ 1 & x_{21} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n_1 1} & 0 & 0 \\ \hline 0 & 0 & 1 & x_{n_1+1} \\ 0 & 0 & 1 & x_{n_1+2} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & x_{n_1+n_2} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \gamma_0 \\ \gamma_1 \end{pmatrix} + \underset{\sim}{\varepsilon}$$

$\Rightarrow$ 독립적 dataset 이므로
β_0, β_1 추정과 γ_0, γ_1 추정은
서로 영향을 주지않음.
(orthogonal)

* Example 3.9