김은영

제3장 중선형회귀모형

3.6 중회귀분석에서의 추론 Ⅱ

두 개 이상의 회귀계수에 대한 개별적인 t−검정을 동시에 하는 것은 가설검정의 개념에서 보면 정확하지 않을 수 있다. 이 경우 F−분포를 이용한 검정들은 정확한 결과를 제공해 준다.

3.6.1 부분 F-검정

분산분석표의 F−검정에서 $H_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$ 을 기각하는 경우 이것은 모든 β_j가 0이 아니라는 것을 의미하지 않는다. 여전히 β_j들의 일부는 0이 되어 해당 설명변수는 반응변수와 선형관계가 없음을 나타낼 수 있다.

< β_j들의 일부가 0이라는 가설에 대한 검정방법 >

$$H_0 : \beta_r = \beta_{r+1} = \cdots = \beta_{p-1} = 0$$

→ $p-r$개의 설명변수$(X_r, X_{r+1}, \cdots, X_{p-1})$는 반응변수와 선형관계가 없다.

※ 유의성이 의심되는 설명변수의 위치는 일정하지 않지만 이들을 회귀모형식에서 전부 뒤쪽으로 옮기면 일반성을 잃지 않고 위의 가설 형태를 생각할 수 있다.

[Idea]
만약 상기 귀무가설이 옳다면 해당 설명변수들$(X_r, X_{r+1}, \cdots, X_{p-1})$은 반응변수와 선형관계가 없으므로 현재 $p-1$개의 설명변수들이 포함된 모형에서 이들을 제거한다 하더라도 회귀모형에 의하여 설명되는 변동의 크기는 변하지 않아 SSR의 크기는 변하지 않는다.
역으로 이들 설명변수들을 회귀모형에서 제거하였을 때 SSR의 크기가 변하지 않으면 귀무가설은 옳은 것이고, SSR이 많이 감소하면 귀무가설이 옳지 않다고 할 수 있다. 따라서 SSR의 값의 변화 정도는 상기 가설에 대한 검정에서 중요한 역할을 한다.
→ $SSR(F) - SSR(R)$의 크기

- 완전모형(Full Model): 모든 설명변수들을 포함하는 모형
  완전모형 하에서의 SSE와 SSR : $SSE(F)$, $SSR(F)$
  → $SST = SSE(F) + SSR(F)$

■ 축소모형(Reduced Model): $H_0 : \beta_r = \beta_{r+1} = \cdots = \beta_{p-1} = 0$ 하에서 성립하는 모형

축소모형 하에서의 SSE와 SSR : $SSE(R)$, $SSR(R)$

$\rightarrow$ $SST = SSE(R) + SSR(R)$

■ 완전모형(Full Model) : $y = X\beta + \epsilon$ 여기서, $X_{(n \times p)}$, $\beta_{(p \times 1)}$

$y = X_1\beta_1 + X_2\beta_2 + \epsilon$ 여기서, $X_{2\,[n \times (p-r)]}$, $\beta_{2\,[(p-r) \times 1]}$

■ 축소모형(Reduced Model): $y = X_1\beta_1 + \epsilon$ 여기서, $X_{1(n \times r)}$, $\beta_{1(r \times 1)}$

$$y = X_1\beta_1 + X_2\beta_2 + \epsilon$$

$$X = \begin{bmatrix} X_1 & X_2 \end{bmatrix},\ \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

$H_0 : \beta_2 = 0$

$$\begin{aligned} SST &= SSR(F) + SSE(F) \\ &= SSR(R) + SSE(R) \end{aligned}$$

$$SSR(F) + SSE(F) = SSR(R) + SSE(R)$$
$$SSR(F) - SSR(R) = SSE(R) - SSE(F)$$

$$R(\beta_2|\beta_1) = SSR(F) - SSR(R) = SSR(\beta_1, \beta_2) - SSR(\beta_1)$$

여기서, $R(\beta_2|\beta_1)$: β_1에 대해 수정된 β_2의 추가제곱합

$R(\beta_2|\beta_1) = SSE(R) - SSE(F) = e_R^t e_R - e^t e$ 여기서, e_R: $H_0 : \beta_2 = 0$ 하의 잔차벡터

$= y^t(I - H_1)y - y^t(I - H)y$ 여기서, $H_{1(n \times n)} = X_1\left(X_1^t X_1\right)^{-1} X_1^t$

$H_{(n \times n)} = X\left(X^t X\right)^{-1} X^t$

$= y^t(H - H_1)y$

※ 추가제곱합 $R(\beta_2|\beta_1) = SSE(R) - SSE(F)$은 두 제곱합들의 자유도의 차이인 $(p-r)$의 자유도를 갖는다.

$H_0 : \beta_2 = 0$를 검정하기 위한 검정통계량은

$$F = \frac{\{SSE(R) - SSE(F)\}/\{df_R - df_F\}}{SSE(F)/df_F}$$

여기서, $df_F = n - p$: 완전모형 하의 자유도

$df_R = n - r$: 축소모형 하의 자유도

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

※ $\{SSE(R) - SSE(F)\}$와 $SSE(F)$ 간의 독립성에 대한 증명은 (3.6.2) 참조

$H_0 : \beta_2 = 0$ 하에서 검정통계량 $F = \dfrac{\{SSE(R) - SSE(F)\}/(p-r)}{s^2} \sim F(p-r,\ n-p)$