김은영

제4장 회귀진단

4.4 영향력 관측치

■ 한 개 또는 소수 개의 관측치가 회귀분석에서의 추정량 $\hat{\beta}$, s^2 등에 큰 영향을 미칠 때 이들 관측치를 '영향력 관측치(influential observation)'라 부른다.

■ 어떤 관측치가 이상치이거나 높은 지렛점이라고 해서 반드시 영향력 관측치가 될 필요는 없다.

■ '영향력 관측치'를 탐색할 때 중요하게 고려되어야 할 점

(1) 어떤 추정량에 영향을 미치는가?(influential on what?)
→ 영향력 관측치는 반드시 어떤 추정치에 관한 것인지가 명시되어야 한다.

(2) 관측치 하나 하나의 영향력과 관측치 군의 영향력이 구별되어져야 한다.
- (개인적) 영향력 관측치(singly influential observation)
- 영향력 관측치군(influential set, set of influential observations)

■ 회귀모형: $y = X\beta + \epsilon$

■ $\hat{\beta}_{(i)}$: i-번째 관측치 제거 후 $(n-1)$개 관측치로 계산된 β의 추정량 $cf)$ $\hat{\beta}$

$$y_{(n\times 1)} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},\quad X_{(n\times p)} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1,p-1} \\ & & \vdots & \\ 1 & x_{n1} & \cdots & x_{n,p-1} \end{bmatrix} = \begin{bmatrix} x_1^t \\ \vdots \\ x_n^t \end{bmatrix}$$

$$y_{(i)\,[(n-1)\times 1]} = \begin{bmatrix} y_1 \\ \vdots \\ y_{i-1} \\ y_{i+1} \\ \vdots \\ y_n \end{bmatrix},\quad X_{(i)\,[(n-1)\times p]} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1,p-1} \\ & & \vdots & \\ 1 & x_{i-1,1} & \cdots & x_{i-1,p-1} \\ 1 & x_{i+1,1} & \cdots & x_{i+1,p-1} \\ & & \vdots & \\ 1 & x_{n1} & \cdots & x_{n,p-1} \end{bmatrix} = \begin{bmatrix} x_1^t \\ \vdots \\ x_{i-1}^t \\ x_{i+1}^t \\ \vdots \\ x_n^t \end{bmatrix}$$

$\hat{\beta} = (X^t X)^{-1} X^t y$: n개의 관측치로 계산된 β의 추정량

$\hat{\beta}_{(i)} = \left(X_{(i)}^t X_{(i)}\right)^{-1} X_{(i)}^t y_{(i)}$: i-번째 관측치 제거 후 $(n-1)$개 관측치로 계산된 β의 추정량

※ 상기 두 개 β의 추정량의 차원은 모두 $(p\times 1)$

$$X^t X = [x_1 \quad \cdots \quad x_n]\begin{bmatrix} x_1^t \\ \vdots \\ x_n^t \end{bmatrix} = \sum_{i=1}^{n} x_i x_i^t$$ 여기서, $x_{i(p\times 1)}$, $x^t_{i(1\times p)}$

$$X^t_{(i)} X_{(i)} = \sum_{j\neq i}^{n} x_j x_j^t = X^t X - x_i x_i^t$$

$$X^t_{(i)} y_{(i)} = [x_1 \quad \cdots \quad x_{i-1} \quad x_{i+1} \quad \cdots \quad x_n]\begin{bmatrix} y_1 \\ \vdots \\ y_{i-1} \\ y_{i+1} \\ \vdots \\ y_n \end{bmatrix} = \sum_{y\neq i}^{n} x_j y_j = X^t y - x_i y_i$$