김은영

제4장 회귀진단

4.5 영향력 측도(influence measure)

- 소거법(deletion method)
- 무한소 교란법(infinitesimal perturbation method)
- 국소 영향력(local influence)
- 대치법(replacement method)

4.5.1 소거법에 의한 측도

(1) 쿡 통계량(Cook's distance, Cook 1977)

i-번째 관측치의 쿡 통계량을 정의해 보자.

① i-번째 관측치의 $\hat{\beta}$에 대한 영향력: $\hat{\beta}-\hat{\beta}_{(i)}$

$$\hat{\beta}-\hat{\beta}_{(i)} = \frac{(X^tX)^{-1}x_ie_i}{(1-h_{ii})}$$

여기서, 잔차: $e_i = y_i - \hat{y}_i$

레버리지: $h_{ii} = x_i^t(X^tX)^{-1}x_i$

→ $\hat{\beta}-\hat{\beta}_{(i)}$은 e_i와 h_{ii}의 증가함수

[pf]

$A_{(p\times p)}$, u, v: p-벡터

$$(A-uv^t)^{-1} = A^{-1} + \frac{A^{-1}uv^tA^{-1}}{1-v^tA^{-1}u}\text{: updating formula}$$

$X^tX = X_{(i)}^t X_{(i)} + x_ix_i^t \ \rightarrow\ X_{(i)}^t X_{(i)} = X^tX - x_ix_i^t$

$X^ty = X_{(i)}^t y_{(i)} + x_iy_i \ \rightarrow\ X_{(i)}^t y_{(i)} = X^ty - x_iy_i$

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

→ $\hat{\beta}_{(i)} = \left(X^tX - x_ix_i^t\right)^{-1}\left(X^ty - x_iy_i\right)$

여기서, $X^tX = A,\ x_i = u,\ x_i^t = v^t \ \rightarrow\ (A-uv^t)^{-1} = A^{-1} + \dfrac{A^{-1}uv^tA^{-1}}{1-u^tA^{-1}v}$

$$\hat{\beta}_{(i)} = \left(X^tX - x_ix_i^t\right)^{-1}\left(X^ty - x_iy_i\right)$$

$$= \left\{\left(X^tX\right)^{-1} + \frac{\left(X^tX\right)^{-1}x_ix_i^t\left(X^tX\right)^{-1}}{1 - x_i^t\left(X^tX\right)^{-1}x_i}\right\}\left(X^ty - x_iy_i\right)$$

$$= \hat{\beta} + \frac{1}{1-h_{ii}}\left(X^tX\right)^{-1}x_ix_i^t\left(X^tX\right)^{-1}\left(X^ty - x_iy_i\right) - \left(X^tX\right)^{-1}x_iy_i$$

$$= \hat{\beta} + \frac{\left(X^tX\right)^{-1}x_ix_i^t\hat{\beta}}{1-h_{ii}} - \left(X^tX\right)^{-1}x_iy_i - \frac{\left(X^tX\right)^{-1}x_ih_{ii}y_i}{1-h_{ii}}$$

$$= \hat{\beta} + \frac{\left(X^tX\right)^{-1}}{1-h_{ii}}\left\{x_i\hat{y}_i - (1-h_{ii})x_iy_i - x_ih_{ii}y_i\right\}$$

$$= \hat{\beta} - \frac{\left(X^tX\right)^{-1}}{1-h_{ii}}\left\{x_i\left(y_i - \hat{y}_i\right)\right\} = \hat{\beta} - \frac{\left(X^tX\right)^{-1}x_ie_i}{1-h_{ii}}$$

$$\Rightarrow \hat{\beta} - \hat{\beta}_{(i)} = \frac{\left(X^tX\right)^{-1}x_ie_i}{\left(1-h_{ii}\right)}$$

② 쿡 통계량(Cook's distance, Cook 1977)

$$C_i = \frac{1}{p}\left(\hat{\beta} - \hat{\beta}_{(i)}\right)^t Cov\left(\hat{\beta}\right)^{-1}\left(\hat{\beta} - \hat{\beta}_{(i)}\right)$$

i-번째 쿡 통계량

$$C_i = \frac{1}{p}\left[\frac{\left(X^tX\right)^{-1}x_ie_i}{1-h_{ii}}\right]^t\left\{\left(X^tX\right)^{-1}\sigma^2\right\}^{-1}\left[\frac{\left(X^tX\right)^{-1}x_ie_i}{1-h_{ii}}\right]$$

$$= \frac{1}{p\sigma^2} \cdot \frac{1}{(1-h_{ii})^2} \cdot e_ix_i^t\left(X^tX\right)^{-1}\left(X^tX\right)\left(X^tX\right)^{-1}x_ie_i$$

$$= \frac{1}{p\sigma^2} \cdot \frac{1}{(1-h_{ii})^2} \cdot e_ix_i^t\left(X^tX\right)^{-1}x_ie_i$$

$$= \frac{1}{p\sigma^2} \cdot \frac{h_{ii}e_i^2}{(1-h_{ii})^2}$$

σ^2: 미지의 모수 → s^2으로 대체

$C_i = \dfrac{1}{ps^2} \cdot \dfrac{h_{ii}e_i^2}{(1-h_{ii})^2}$: e_i와 h_{ii}의 증가함수

③ 영향력 관측치군(influential set, set of influential observations)

- 하나의 영향력 관측치를 제거하는 것만으로 충분하지 않을 때가 있다.
- 이때는 영향력 관측치군을 제거해야 한다.

k개의 영향력 관측치군을 제거하는 경우를 생각해 보자.

$K=\{i_1,\ i_2,\ \cdots,\ i_k\}$: 크기가 k인 index set

$X_{(K)}$: $X_{i1},\ \cdots,\ X_{ik}$를 제거한 후의 $(n-k)\times p$ 행렬

X_K : $X_{i1},\ \cdots,\ X_{ik}$로 구성된 $(k\times p)$ 행렬

H_K : $(k\times k)$ 행렬. $H_K=\begin{bmatrix} h_{i1,i1} & h_{i1,i2} & \cdots & h_{i1,ik} \\ & h_{i2,i2} & \cdots & h_{i2,ik} \\ & & \ddots & \\ & & & h_{ik,ik} \end{bmatrix}$

$e_K=\{e_{i1},\ \cdots,\ e_{ik}\}$: $(k\times 1)$ 벡터

$\rightarrow\ H_K=X_K(X^tX)^{-1}X_K^t$: $(k\times k)$ 행렬

$h_{k,k}=x_k(X^tX)^{-1}x_k$

집합 K에 속하는 k개 관측치들의 $\hat{\beta}$에 대한 쿡 통계량

$$C_K\equiv\frac{1}{p}\left(\hat{\beta}-\hat{\beta}_{(K)}\right)^t Cov(\hat{\beta})^{-1}\left(\hat{\beta}-\hat{\beta}_{(K)}\right)$$

여기서, $\hat{\beta}_{(K)}=\left(X_{(K)}^t X_{(K)}\right)^{-1}X_{(K)}^t y_{(K)}=\hat{\beta}-(X^tX)^{-1}X_K^t(I-H_K)^{-1}e_K$

$$C_K=\frac{1}{p}\left[(X^tX)^{-1}X_K^t(I-H_K)^{-1}e_K\right]^t\left[(X^tX)^{-1}\sigma^2\right]^{-1}\left[(X^tX)^{-1}X_K^t(I-H_K)^{-1}e_K\right]$$

$$=\frac{1}{p\sigma^2}e_K^t(I-H_K)^{-1}X_K(X^tX)^{-1}X^tX(X^tX)^{-1}X_K^t(I-H_K)^{-1}e_K$$

$$=\frac{1}{p\sigma^2}e_K^t(I-H_K)^{-1}X_K(X^tX)^{-1}X_K^t(I-H_K)^{-1}e_K$$

$$=\frac{1}{p\sigma^2}e_K^t(I-H_K)^{-1}H_K(I-H_K)^{-1}e_K$$

$$C_i=\frac{1}{p\sigma^2}e_i(1-h_{ii})^{-1}h_{ii}(1-h_{ii})^{-1}e_i$$

④ $s_{(i)}^2 = s^2 \cdot \dfrac{n-p-r_i^2}{n-p-1}$

여기서, $r_i = \dfrac{e_i - E(e_i)}{S.E.(e_i)} = \dfrac{e_i}{s\sqrt{1-h_{ii}}}$

$$r_i^* = \frac{e_i}{s_{(i)}\sqrt{1-h_{ii}}}$$

[pf]

■ $s_{(i)}^2$의 정의: i-번째 관측치 제거 후 $(n-1)$개 관측치 기반 σ^2의 추정량

$$(n-p-1)s_{(i)}^2 = \sum_{j \neq i}\left(y_j - x_j^t\hat{\beta}_{(i)}\right)^2 = \sum_{j=1}^{n}\left(y_j - x_j^t\hat{\beta}_{(i)}\right)^2 - \left(y_i - x_i^t\hat{\beta}_{(i)}\right)^2$$

$$= \sum_{j=1}^{n}\left(y_j - x_j^t\hat{\beta} + x_j^t\hat{\beta} - x_j^t\hat{\beta}_{(i)}\right)^2 - \left(y_i - x_i^t\hat{\beta} + x_i^t\hat{\beta} - x_i^t\hat{\beta}_{(i)}\right)^2$$

$$= \sum_{j=1}^{n}\left(y_j - x_j^t\hat{\beta} + x_j^t(\hat{\beta} - \hat{\beta}_{(i)})\right)^2 - \left(y_i - x_i^t\hat{\beta} + x_i^t(\hat{\beta} - \hat{\beta}_{(i)})\right)^2 \quad \text{여기서, } \hat{\beta} - \hat{\beta}_{(i)} = \frac{(X^tX)^{-1}x_ie_i}{(1-h_{ii})}$$

$$= \sum_{j=1}^{n}\left[e_j + x_j^t\left\{\frac{(X^tX)^{-1}x_ie_i}{1-h_{ii}}\right\}\right]^2 - \left[e_i + x_i^t\left\{\frac{(X^tX)^{-1}x_ie_i}{1-h_{ii}}\right\}\right]^2$$

$$= \sum_{j=1}^{n}\left[e_j + \frac{h_{ji}e_i}{1-h_{ii}}\right]^2 - \left[e_i + \frac{h_{ii}e_i}{1-h_{ii}}\right]^2 \qquad \text{여기서, } h_{ji} = h_{ij}$$

$$= \sum_{j=1}^{n}\left[e_j^2 + 2\frac{h_{ji}e_ie_j}{1-h_{ii}} + \frac{h_{ji}^2e_i^2}{(1-h_{ii})^2}\right] - \frac{e_i^2}{(1-h_{ii})^2}$$

$$= \sum_{j=1}^{n}e_j^2 + 2\frac{e_i}{1-h_{ii}}\sum_{j=1}^{n}h_{ji}e_j + \frac{e_i^2}{(1-h_{ii})^2}\sum_{j=1}^{n}h_{ji}^2 - \frac{e_i^2}{(1-h_{ii})^2}$$

여기서, $\sum_{j=1}^{n}h_{ji}e_j$: $He[=0]$의 i-번째 원소로 그 값은 0

H: 멱등행렬 $\rightarrow$ $H^2 = H \rightarrow \sum_{j=1}^{n}h_{ji}^2 = h_{ii}$

$$= \sum_{j=1}^{n}e_j^2 + \frac{e_i^2h_{ii}}{(1-h_{ii})^2} - \frac{e_i^2}{(1-h_{ii})^2} = (n-p)s^2 + \frac{e_i^2}{(1-h_{ii})^2}(h_{ii}-1) = (n-p)s^2 - \frac{e_i^2}{(1-h_{ii})}$$

$$= (n-p)s^2 - \frac{s^2e_i^2}{s^2(1-h_{ii})} = s^2\left\{(n-p) - \frac{e_i^2}{s^2(1-h_{ii})}\right\} = s^2\left(n-p-r_i^2\right)$$

$\Rightarrow s_{(i)}^2 = s^2 \cdot \dfrac{n-p-r_i^2}{n-p-1}$