김은영

제4장 회귀진단

4.3 지렛값 (leverage value)

■ 모자행렬 H의 i번째 대각원소 h_{ii}

- 디자인 행렬 X의 i번째 지렛값 → $h_{ii} = x_i^t (X^t X)^{-1} x_i$
- 중심에서 멀리 떨어져 있는 값 → 영향력 있는 관측치

$$H = X(X^t X)^{-1} X^t = (h_{ij}) \qquad (i, j = 1, \cdots, n)$$

여기서, $X_{(n \times p)} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1,p-1} \\ & & \vdots & \\ 1 & x_{n1} & \cdots & x_{n,p-1} \end{bmatrix} = \begin{bmatrix} x_1^t \\ \vdots \\ x_n^t \end{bmatrix} \quad \rightarrow \; x_i = (1 \quad x_{i1} \quad \cdots \quad x_{i,p-1})^t$

$$\Rightarrow \; h_{ij} = x_i^t (X^t X)^{-1} x_j$$

■ 특수한 경우로 단순선형회귀모형을 생각해 보자.

$y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ 여기서, $X = \begin{bmatrix} 1 & x_i \\ 1 & x_i \end{bmatrix}$

$$h_{ii} = x_i^t (X^t X)^{-1} x_i = [1 \quad x_i] \begin{bmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ x_i \end{bmatrix}$$

$$= [1 \quad x_i] \frac{1}{n\sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2} \begin{bmatrix} \sum_{i=1}^n X_i^2 & -\sum_{i=1}^n X_i \\ -\sum_{i=1}^n X_i & n \end{bmatrix} \begin{bmatrix} 1 \\ x_i \end{bmatrix} = [1 \quad x_i] \frac{1}{S_{xx}} \begin{bmatrix} \frac{\sum_{i=1}^n X_i^2}{n} & -\bar{X} \\ -\bar{X} & 1 \end{bmatrix} \begin{bmatrix} 1 \\ x_i \end{bmatrix}$$

$$= \frac{1}{S_{xx}} \left[\frac{\sum_{i=1}^n X_i^2}{n} - X_i \bar{X}, \quad -\bar{X} + X_i \right] \begin{bmatrix} 1 \\ x_i \end{bmatrix} = \frac{1}{S_{xx}} \left[\frac{\sum_{i=1}^n X_i^2}{n} - X_i \bar{X} - X_i \bar{X} + X_i^2 \right]$$

$$= \frac{1}{S_{xx}} \left[\frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2 + \bar{X}^2 - 2X_i \bar{X} + X_i^2 \right] = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{S_{xx}}$$

< 설명변수가 $p-1$개인 중선형회귀모형에서 h_{ii}의 여러 가지 성질 >

(1) $\sum_{i=1}^{n} h_{ii} = p$ → 모든 지렛값의 합은 설명변수의 개수에 1을 더한 것이다.

[pf] $\sum_{i=1}^{n} h_{ii} = tr(H) = tr\left[X(X^tX)^{-1}X^t\right] = tr\left[(X^tX)^{-1}X^tX\right] = tr(I_p) = p$

---

(2) $\frac{1}{n} \le h_{ii} \le 1$

[pf_1] 행렬 $\tilde{X}$: 디자인 행렬 X에서 각 관측값들의 평균을 뺀 행렬

→ 중심화된 행렬

→ $h_{ii} = \frac{1}{n} + (x_i - \bar{x})^t(\tilde{X}^t\tilde{X})^{-1}(x_i - \bar{x}) \ \Rightarrow\ h_{ii} \ge \frac{1}{n}$

[pf_2] H : 멱등행렬 → $H = H^2 = H \bullet H$

$h_{ii} = \sum_{j=1}^{n} h_{ij}^2 = h_{ii}^2 + \sum_{j \ne i} h_{ij}^2$ 여기서, 우변의 두 번째 항은 항상 비음

→ $h_{ii} \ge h_{ii}^2$ → $h_{ii}(1-h_{ii}) \ge 0$ 여기서, $h_{ii} > 0$

⇒ $(1-h_{ii}) \ge 0$ → $h_{ii} \le 1$

---

(3) $h_{ij}^2 \le h_{ii}h_{jj}$ → 비대각원소의 제곱은 언제나 해당되는 대각원소들의 곱보다 작거나 같다.

X^tX : 대칭행렬, 양정치행렬 → $X^tX = R^tR$이 되는 상삼각행렬 R이 존재한다.

(촐레스키 분할법)

$h_{ij} = x_i^t(X^tX)^{-1}x_j = x_i^t(R^tR)^{-1}x_j = a^tb$ 여기서, $a = (R^{-1})^t x_i$, $b = (R^{-1})^t x_j$

코오시-슈바르쯔 부등식[1]에 의해서

$h_{ij}^2 \le \left[\left\{(R^{-1})^t x_i\right\}^t (R^{-1})^t x_i\right]\left[\left\{(R^{-1})^t x_j\right\}^t (R^{-1})^t x_j\right]$

$= x_i^t(R^tR)^{-1}x_i x_j^t(R^tR)^{-1}x_j = h_{ii}h_{jj}$

1) 두 벡터 a와 b에 대해서, $(a^tb)^2 \le (a^ta)(b^tb)$가 성립한다.