김은영

제3장 중선형회귀모형

3.5 중회귀분석에서의 추론 I

3.5.4 Y의 평균에 대한 추론

설명변수들의 값이 주어져 있는 경우

1. y의 평균값에 대한 추정
2. 새로운 y값에 대한 예측구간

은 중회귀분석에서도 중요한 문제가 된다.

(1) y의 평균값에 대한 추정

$x_0 = (1, x_{01}, x_{02}, \cdots, x_{0,p-1})$에서 $E(y)$를 추정해보자.

- 점추정량: $x_0^t\hat{\beta} \quad \leftarrow \; E(y) = x_0^t\beta$
- $Var(x_0^t\hat{\beta}) = x_0^t Cov(\hat{\beta})x_0 = x_0^t\left[(X^tX)^{-1}\sigma^2\right]x_0 = x_0^t(X^tX)^{-1}x_0\sigma^2$

$\hat{\beta}$의 정규성에 의해서 $x_0^t\hat{\beta}$ 역시 정규분포를 따른다. → σ^2 대신 s^2을 사용할 수 있다.

$$\frac{\dfrac{x_0^t\hat{\beta} - x_0^t\beta}{\sigma\sqrt{x_0^t(X^tX)^{-1}x_0}}}{\sqrt{\dfrac{SSE}{\sigma^2}/(n-p)}} = \frac{x_0^t\hat{\beta} - x_0^t\beta}{s\sqrt{x_0^t(X^tX)^{-1}x_0}} \sim t(n-p)$$

< $x_0^t\beta$에 대한 $100(1-\alpha)\%$ 신뢰구간 >

$$x_0^t\beta = x_0^t\hat{\beta} \pm t_{\alpha/2}(n-p) \cdot s\sqrt{x_0^t(X^tX)^{-1}x_0}$$

[응용] y의 평균값에 대한 신뢰구간 공식은 회귀계수들의 임의의 선형함수에 대한 신뢰구간을 구하는 것에도 사용될 수 있다.

$$q^t\beta = q_0\beta_0 + q_1\beta_1 + \cdots + q_{p-1}\beta_{p-1}$$

'$x_0^t\beta$에 대한 $100(1-\alpha)\%$ 신뢰구간'에서 x_0 대신에 q를 대입하면 '$q^t\beta$에 대한 $100(1-\alpha)\%$ 신뢰구간'을 다음과 같이 구할 수 있다.

$$q^t\beta = q^t\hat{\beta} \pm t_{\alpha/2}(n-p) \cdot s\sqrt{q^t(X^tX)^{-1}q}$$

■ $p=5$인 경우, $\beta_2-\beta_3$에 대한 신뢰구간을 계산해 보자.

$q^t\beta = q^t\hat{\beta} \pm t_{\alpha/2}(n-p) \cdot s\sqrt{q^t(X^tX)^{-1}q}$

$\beta = (\beta_0,\ \beta_1,\ \beta_2,\ \beta_3,\ \beta_4)^t$

$\beta_2-\beta_3 \ \rightarrow \ q=(0,\ 0,\ 1,\ -1,\ 0) \ \Rightarrow \ q^t\beta = \beta_2-\beta_3$

(2) 새로운 y값에 대한 예측구간

설명변수들의 값이 주어져 있을 경우, 새로운 y값에 대한 예측구간도 단순회귀에서와 같은 논리에 의해서 구할 수 있다. 즉, $x=x_0$에서 y의 예측값 $y(x_0)$를 포함할 확률이 $1-\alpha$인 예측구간은 다음과 같다.

$$x_0^t\hat{\beta} \pm t_{\alpha/2}(n-p) \cdot s\sqrt{1+x_0^t(X^tX)^{-1}x_0}$$

[예제 3.9] 풀어보기