김은영

제3장 중선형회귀모형

3.5 중회귀분석에서의 추론 I

3.5.2 공동신뢰영역과 가설검정

1. 두 개 이상의 모수에 대해 t−검정을 동시에 수행하는 경우 유의수준의 해석에 주의
→ 두 개 이상의 회귀계수에 대한 95% 신뢰구간을 구할 때 이 신뢰구간들에 대한 신뢰도가 각각 모두 95% 된다고 해석하는 것은 옳지 않다. 실제로 이 신뢰도는 회귀계수들의 개수가 많아질수록 95%보다 훨씬 작아질 수 있다.
⇒ 정확하게 신뢰도 95%를 만족하는 영역을 구하는 것은 중요한 문제

2. 확률벡터$(\hat{\beta}-\beta)$의 이차형식을 생각해 보자.

$$Q=(\hat{\beta}-\beta)^t X^t X(\hat{\beta}-\beta)$$

$(\hat{\beta}-\beta)\sim N_p\left(0,\ \sigma^2(X^tX)^{-1}\right)$

$\Rightarrow\ \dfrac{Q}{\sigma^2}=(\hat{\beta}-\beta)^t\dfrac{X^tX}{\sigma^2}(\hat{\beta}-\beta)\sim\chi^2\left[rank(AV),\ \dfrac{1}{2}\mu^tA\mu\right]$ ← [정리 3.1]

$AV=\dfrac{X^tX}{\sigma^2}\cdot\sigma^2(X^tX)^{-1}=I_p \quad\rightarrow\quad rank(AV)=rank(I_p)=p$

$\dfrac{1}{2}\mu^tA\mu=\dfrac{1}{2}0^t\dfrac{X^tX}{\sigma^2}0=0$

$\Rightarrow\ \dfrac{Q}{\sigma^2}=(\hat{\beta}-\beta)^t\dfrac{X^tX}{\sigma^2}(\hat{\beta}-\beta)\sim\chi^2(p)$

[정리 3.3]을 이용하여 Q와 SSE는 서로 독립임을 보일 수 있다.

$F=\dfrac{\dfrac{Q}{\sigma^2}/p}{\dfrac{SSE}{\sigma^2}/(n-p)}$ 여기서, $\dfrac{SSE}{\sigma^2}\sim\chi^2(n-p)$

Q와 SSE 가 서로 독립임을 확인하기 위해서

$Q=(\hat{\beta}-\beta)^tX^tX(\hat{\beta}-\beta)$ → $(\hat{\beta}-\beta)$: p-차원 벡터

$SSE=e^te=y^t(I-H)y$ → y : n-차원 벡터

⇒ [정리 3.3]을 이용하여 Q와 SSE는 서로 독립임을 보일 수 없다.

$X,\ Z$: 독립 $\rightarrow$ $g(X),\ h(Z)$: 독립

$Q=(\hat{\beta}-\beta)^t X^t X(\hat{\beta}-\beta) \rightarrow Q$: $\hat{\beta}=(X^tX)^{-1}X^t y$의 함수
SSE: $e=(I-H)y$의 함수

$\hat{\beta},\ e$: y에 대해 선형, 정규분포를 따름
$\rightarrow$ $Cov(\hat{\beta},\ e)=0$임을 증명 $\Rightarrow$ Q와 SSE 가 서로 독립

$$Cov(\hat{\beta},\ e)=Cov\left[(X^tX)^{-1}X^t y,\ (I-H)y\right]=(X^tX)^{-1}X^t Cov(y)(I-H)$$
$$=\sigma^2(X^tX)^{-1}X^t\left(I-X(X^tX)^{-1}X^t\right)=\sigma^2\left[(X^tX)^{-1}X^t-(X^tX)^{-1}X^tX(X^tX)^{-1}X^t\right]=0$$

$\rightarrow$ Q와 SSE 가 서로 독립

$$F=\frac{\frac{Q}{\sigma^2}/p}{\frac{SSE}{\sigma^2}/(n-p)}=\frac{Q}{ps^2}=\frac{(\hat{\beta}-\beta)^t X^t X(\hat{\beta}-\beta)}{ps^2}\sim F(p,n-p) \quad \text{여기서, } s^2=\frac{SSE}{n-p}$$

< β에 대한 $100(1-\alpha)\%$ 공동신뢰영역 >

$$\frac{(\hat{\beta}-\beta)^t X^t X(\hat{\beta}-\beta)}{ps^2}\le F_\alpha(p,n-p)$$

여기서, $Q=(\hat{\beta}-\beta)^t X^t X(\hat{\beta}-\beta)$: 이차형식 $\rightarrow$ 비음

p와 s^2: 비음

$\rightarrow$ $\left\{\beta : (\hat{\beta}-\beta)^t X^t X(\hat{\beta}-\beta)\le ps^2 \cdot F_\alpha(p,n-p)\right\}$

단순회귀에서도 회귀계수 β_0와 β_1에 대한 공동신뢰영역을 구할 수 있다.

$$\beta=\begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix},\ \hat{\beta}=\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix},\ X^tX=\begin{bmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2\end{bmatrix}$$

"β에 대한 $100(1-\alpha)\%$ 공통신뢰영역" 식에 대입하면

$$n(\beta_0-\hat{\beta}_0)^2+2(\beta_0-\hat{\beta}_0)(\beta_1-\hat{\beta}_1)\sum_{i=1}^n X_i+(\beta_1-\hat{\beta}_1)^2\sum_{i=1}^n X_i^2\le 2s^2F_\alpha(2,\ n-2)$$

[예제 3.6] 2장의 책의 가격에 대한 예제에서 β_0와 β_1에 대한 95% 공동신뢰영역을 구하고, 이를 각 계수에 대한 95% 신뢰구간과 함께 나타내어라.

《 풀이 》

$y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

< 2장에서 구한 결과들 >

$\sum_{i=1}^{20} X_i = 8299$, $\sum_{i=1}^{20} X_i^2 = 3999411$, $\hat{\beta}_0 = 2.2018$, $\hat{\beta}_1 = 0.0350$, $s^2 = 2.053$

$F_{0.05}(2, 18) = 3.55$

β에 대한 95% 공동신뢰영역은

$$n(\beta_0 - \hat{\beta}_0)^2 + 2(\beta_0 - \hat{\beta}_0)(\beta_1 - \hat{\beta}_1)\sum_{i=1}^{n} X_i + (\beta_1 - \hat{\beta}_1)^2 \sum_{i=1}^{n} X_i^2 \le 2s^2 F_\alpha(2, n-2)$$

$$20(\beta_0 - 2.2018)^2 + 2(\beta_0 - 2.2018)(\beta_1 - 0.0350)8299 + (\beta_1 - 0.0350)^2 3999411 \le 2(2.053)(3.55)$$

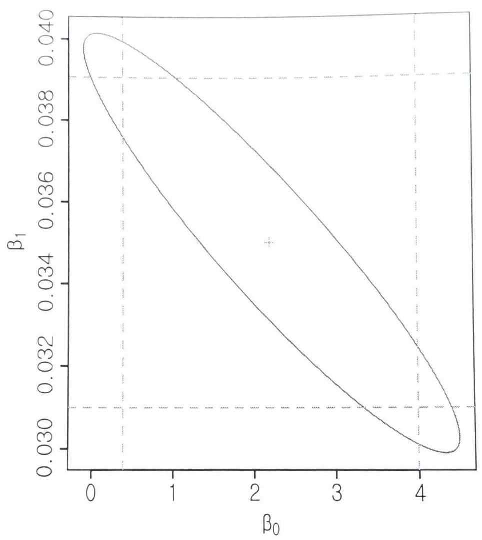
① β_0와 β_1에 대한 $100(1-\alpha)\%$ 신뢰구간

β_0: $\hat{\beta}_0 \pm t_{\alpha/2} \cdot SE(\hat{\beta}_0)$

β_1: $\hat{\beta}_1 \pm t_{\alpha/2} \cdot SE(\hat{\beta}_1)$

② β_0와 β_1에 대한 $100(1-\alpha)\%$ 공동신뢰영역

$$(\hat{\beta} - \beta)^t X^t X(\hat{\beta} - \beta) = \begin{bmatrix} \hat{\beta}_0 - \beta_0 & \hat{\beta}_1 - \beta_1 \end{bmatrix} \begin{bmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \end{bmatrix}$$



(그림 3.2) 책의 가격 예제에서 두 회귀계수에 대한 공동신뢰영역(타원 내부)과 각각의 회귀계수에 대한 신뢰구간(직사각형 내부). 그림에서 + 기호는 점추정치의 값을나타낸다.