2021.05.05.
김은영

제3장 중선형회귀모형

3.5 중회귀분석에서의 추론 I

3.5.3 동시신뢰구간

1. 특히 $p \geq 4$일 때
   동시신뢰영역(joint confidence region)은 계산하기도 어렵고 해석하기도 어렵다.

2. 동시신뢰영역에 대한 대안으로,
   동시신뢰구간(simultaneous confidence interval)을 사용할 수 있다.

(1) Bonferroni t 동시신뢰구간

$E_1, E_2, \cdots, E_g$: 사건

$P\left(\bigcup_{i=1}^{g} E_i\right) \leq \sum_{i=1}^{g} P(E_i)$: 본페로니 부등식 (Bonferroni Inequality)

구간추정의 문제에서 g개의 모수 $\theta_1, \theta_2, \cdots, \theta_g$에 대한 신뢰구간을 구해보자.

$E_i =$ { C.I. for parameter θ_i is correct }

↳ E_i : '모수 θ_i에 대한 신뢰구간이 옳다'라는 사건

각 신뢰구간의 계산에서 신뢰도 $1-\alpha^*$가 사용되었다면

$$P(E_i) \geq 1-\alpha^* \quad \text{또는} \quad P(E_i^C) < \alpha^*$$

모든 모수들에 대한 신뢰구간들이 옳다는 사건은 사건 E_i들의 공통집합으로 주어진다.

$Q =$ { all the C.I. for g parameters $\theta_1, \theta_2, \cdots, \theta_g$ is correct }

↳ Q : '모든 모수들에 대한 신뢰구간들이 옳다'는 사건

$$P(Q) = P\left(\bigcap_{i=1}^{n} E_i\right) = 1 - P\left(\bigcup_{i=1}^{g} E_i^C\right) \geq 1 - \sum_{i=1}^{g} P(E_i^C) \quad \leftarrow \text{본페로니 부등식}$$

$$\geq 1 - g\alpha^*$$

→ 본페로니: $1 - g\alpha^* \equiv 1-\alpha$ for some $\alpha > 0 \quad \Rightarrow \quad \alpha^* = \alpha/g$

[예] $g=10$개의 신뢰구간에서 각각 $1-\alpha^{*}=0.95$의 신뢰도를 사용하여($\alpha^{*}=0.05$) 모든 신뢰구간이 옳을 확률이 50% 밖에 넘지 않는다는 것을 보여준다. 그러므로 모든 θ_i에 대한 신뢰구간이 옳을 확률이 적어도 $1-\alpha$가 되기 위해서는 개개의 신뢰구간에서의 신뢰도가 $1-\alpha/g$가 되어야 한다.

본페로니 동시신뢰구간은 위의 사실을 t분포에 기초한 신뢰구간에 적용시킨 것이다.

[예] g개의 회귀계수들에 대한 신뢰구간은 동시에 구하고자 하면 임의의 β_j에 대한 $100(1-\alpha)\%$ 본페로니 동시신뢰구간의 한계는

$$\beta_j=\hat{\beta}_j \pm t_{\alpha/2g} SE\left(\hat{\beta}_j\right)$$

$$cf)\ \text{개별 C.I. : } \beta_j=\hat{\beta}_j \pm t_{\alpha/2} SE\left(\hat{\beta}_j\right)$$

$cf)$ β에 대한 $100(1-\alpha)\%$ 공동신뢰영역 $\left\{\beta : (\hat{\beta}-\beta)^t X^t X(\hat{\beta}-\beta) \le ps^2 \cdot F_\alpha(p, n-p)\right\}$

각 신뢰구간에 $1-\alpha/g$의 신뢰도가 주어지므로 구간의 계산에 사용되는 t분포의 백분위수의 값이 커진다는 차이가 있다.

< 본페로니 동시신뢰구간의 장단점 >

- 장점: 계산이 쉽고 해석이 단순하다.
- 단점: 신뢰구간의 개수 g가 큰 경우는 해당되는 t분포의 상위백분위수의 값이 매우 크게 되어 신뢰구간의 폭이 너무 넓어진다. 신뢰구간의 폭이 넓으면 정확도가 상대적으로 떨어진다.

(2) Scheffe F 동시신뢰구간

$$\beta_j=\hat{\beta}_j \pm \sqrt{gF_\alpha(g, n-p)}\, SE\left(\hat{\beta}_j\right)$$

(3) 최대계수(Maximum Modulus) t 동시신뢰구간

$$\beta_j=\hat{\beta}_j \pm u_\alpha(g, n-p) SE\left(\hat{\beta}_j\right)$$

여기서, $u_\alpha(g, n-p)$: 자유도가 $n-p$인 t분포를 갖는 g개의 서로 독립인 확률변수들 중에서 절대값이 제일 큰 확률변수가 갖는 분포의 상위 $100\alpha\%$ 백분위수. 지정교재 부록B<표6>

[예제 3.8] 풀어보기