10주차 계량경제학 과제 201547130 윤요섭

1. 다중회귀모형에 관한 기본 가정을 열거하고 설명하시오.

MR1. $y_i = \beta_1 + \beta_2 x_{i2} + ... + \beta_k x_{ik} + e_i = 1, ..., N$

MR2. $E(y_i) = \beta_1 + \beta_2 x_{i2} + ... + \beta_k x_{ik} \Leftrightarrow E(e_i) = 0$

MR3. $var(y_i) = var(e_i) = \sigma^2$

확률변수 y 및 e는 동일한 분산을 갖는다. 왜냐하면 이들은 단지 일정한 상수만큼 차이가 나기 때문이다.

MR4. $cov(y_i, y_j) = cov(e_i, e_j) = 0$

무작위 오차 e가 통계적으로 독립적인 경우 종속변수 y의 값도 통계적으로 독립적이라고 할 경우 더욱 강한 가정이 된다.

MR5. x_{ik}의 값은 확률적이 아니며 다른 설명변수와 정확히 선형함수 관계에 있지 않다.

MR6. $y_i \sim N[(\beta_1 + \beta_2 x_{i2} + ... + \beta_k x_{ik}), \sigma^2)] \Leftrightarrow e_i \sim N(0, \sigma^2)$

다중회귀모형은 종속변수에 영향을 미치는 독립변수가 늘어났다는 점 외에 단순회귀모형의 기본 가정과 차이가 없다.

2. 다중회귀모형에서 신뢰구간 추정 절차에 대해 설명하시오.

1. 95%의 구간 추정 값을 구하고 싶을 때, $P(-t_c < t_{(n-k)} < t_c) = 0.95$를 구해야한다.

2. $P(-t_c \le \frac{b_k - \beta_k}{se(b_k)} \le t_c) = 1 - \alpha = 0.95$ 이다.

3. 신뢰구간의 경우 $1-\alpha/2$이며, 분포의 좌우 각 꼬리부분에 0.025가 필요하므로 $100(1-\alpha)\%$ 신뢰구간을 일반적으로 나타내면 다음과 같다.

$[b_k - t_{(1-\alpha/2, N-K)} \times se(b_2), b_k + t_{(1-\alpha/2, N-K)} \times se(b_2)]$

3. 다중회귀모형에서 단일계수 유의성 검정 절차를 설명하시오.

1. $H_0 : \beta_k = 0$, $H_1 : \beta_k \neq 0$으로 귀무가설과 대립가설을 설정한다.

2. $\beta_k = 0$이므로 $t = \frac{b_k}{se(b_k)}$에서 임계값 $-t_c$와 t_c를 비교하여 임계값보다 작거나 크면 귀무가설을 기각하고, 그렇지 않으면 귀무가설을 채택한다.

4. 다중회귀모형에서 적합도 측정 방법을 설명하시오.

$SST = SSR + SSE$이고, $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$인데, 독립변수가 증가하면서 SSE값이 작아지면 R^2값이 증가하게 된다. 따라서 식을 변형하여 $\overline{R^2} = 1 - \frac{SSE/(n-k)}{SST/(n-1)}$ 으로 조절된 R^2을 이용한다.

5. 다중회귀모형에서 독립변수들 사이에 상관관계가 크다는 뜻은 무엇일까? 예를 들어 설명하시오. 만일 상관관계가 높은 독립변수들을 다중회귀모형에 투입할 경우 추정결과에 어떤 문제가 발생하는가? 이 경우 어떻게 효과적으로 대처해야 할까?에 대해 설명하시오.

다중회귀모형에서 독립변수 사이에 상관관계가 커버리면 공산성문제가 생기며, 또한 $var(b_2)$이 무한대로 커진다. 이 경우 t값이 매우 낮아져서, 통계적으로 유의하지 않다고(종속변수와 독립변수간에 인과관계가 없다고) 추론할 우려가 있다. 또한 β_k의 추정값을 구할 수 없으며 최소제곱법을 시행하는데 필요한 가정 중 하나인 MR5를 위반한다.

$$var(b_2) = \frac{\widehat{\sigma^2}}{(1-\gamma_{2,3}^2)*\sum_{i=1}^{n}(x_{i,2}-\overline{x_{i2}})^2}\ ,\ \gamma_{2,3} = \frac{\sum_{i=1}^{n}(x_{i,2}-\overline{x_{i,2}})(x_{i,3}-\overline{x_{i,3}})}{\sqrt{\sum_{i=1}^{n}(x_{i,2}-\overline{x_{i,2}})^2\sum_{i=1}^{n}(x_{i,3}-\overline{x_{i,3}})^2}}$$

이를 해결하기 위해서는 설명변수들 짝 사이의 표본 상관계수를 이용하여 공선관계를 찾아낸다. 이런 방법으로 찾아낼 수 없을 시에는 보조적인 회귀를 추정하여 해결할 수 있다.