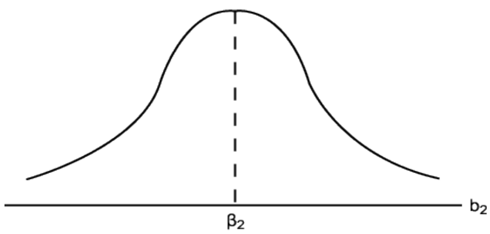
응용계량경제 5주차 과제 201547130 윤요섭

1. 최소제곱추정량으로 구한 표본회귀계수 b_2는 모수 β_2에 대해 어떤 차이가 있는가?

단순 선형회귀모형에 있어서 모수 β_2는 모회귀선의 기울기에 해당하며 변하지 않는 수 이다.

그러나 우리는 전수조사를 할 수 없기 때문에 표본을 추출하여 β_2를 추정한다. 이 때 β_2의 최소제곱 추정 값이 확률변수 b_2가 된다.

- 진회귀선에서 모수(β_2)는 오직 하나의 값을 갖는다. 변수가 아니라 모수(파라메타)이다.
- 특정 표본이 여러 개 있다면, 여러 개의 표본회귀선과 표본회귀계수(b_2: 파라메타 추정치) 가 존재한다.
- 고로 $b_2 = \dfrac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum(x_i - \overline{x})^2} = \beta_2$이 아니다.
- 표본회귀계수 b_2는 확률변수이다.
- 표본회귀계수 b_2는 평균과 분산을 갖는 표본분포를 한다.



2. 단순회귀모형에 관한 앞의 가정(SR1-SR5) 하에서 최소제곱 추정량(estimator)과 최소제곱추정치(추정값, estimate)를 명확히 구분하고 이해하자.

- 특정 표본이 여러 개 있다면, 여러 개의 표본에 대해 표본회귀선과 표본회귀계수(b_1, b_2)가 여러 개 존재한다. 이 경우 최소제곱추정량은 평균과 분산을 갖는 확률분포를 보인다.
- 그러나 특정 표본이 한 개만 있다면, 이에 대한 표본회귀선에서 표본회귀계수(b_1, b_2)는 오직 하나이다. 이 경우 b_1, b_2는 최소제곱추정치이다.
- 학생들에게 표본을 샘플링해서 각각의 표본에 대해 표본회귀선을 추정해서 한계소비성향(b_2)을 파악하려고 한다고 하자. 이때 전체 학생들이 최소제곱원칙에 의해 추정한 한계소비성향(b_2) 전체를 우리는 최소제곱 추정량(estimator)이라고 한다. 반면에 개별 학생이 최소제곱의 원칙으로 추정한 한계소비성향(b_2)은 우리는 최소제곱 추정치(estimate)라고 한다. 하지만 학생들이 개별적으로 산출한 최소제곱 추정치들의 전체

는 추정량이다.

- 그렇다면 최소제곱추정량은 어떤 분포를 하고 있는가? 평균과 분산은 얼마인가? 최소제곱추정량은 어떤 특징이 있는가?에 대한 의문이 있게 된다. → 최소제곱추정량의 특성 이해하기
- 또 다른 의문은 최소제곱추정량 중 어느 학생의 추정치를 신뢰하고, 어떤 추정치는 신뢰할 수 없는가에 대한 문제가 있게 된다. 어떤 판단 기준하에서 취사선택할 것인가?라는 의문이 있게 된다. → 신뢰구간과 가설검정 이해하기

3. 최소제곱추정량은 어떤 특징이 있는가? 또 어떤 분포를 하고 있는가? 평균과 분산은? → 최소제곱추정량의 특성 이해하기

3.1. 단순회귀모형에서 가정(SR1-SR5)을 설명하시오. 단, 모회귀모형을 기준으로 하시오.

• SR1. x의 각 값에 대해 y값은 다음과 같다.

$y = \beta_1 + \beta_2 x + e$

• SR2. 무작위 오차 e의 기댓값은 다음과 같다.

$E(e) = 0$

왜냐하면 다음과 같이 가정하였기 때문이다.

$E(y) = \beta_1 + \beta_2 x$

• SR3. 무작위 오차 e의 분산은 다음과 같다.

$var(e) = \sigma^2 = var(y)$

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

• SR4. 무작위 오차의 한 쌍인 e_i, e_j의 공분산은 다음과 같다.

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

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

• SR5. 변수 x는 확률적이지 않으며 최소한 2개의 상이한 값을 가져야 한다.

3.2 최소제곱원칙을 이용한 표본회귀선의 절편 및 기울기, 즉 통상적인 최소제곱추정량(b_1, b_2)은 아래식을 이용하면 산출된다. 단, 여기서는 최소제곱추정량 b_2에 한정해서 답하시오.

$$b_1 = \overline{y} - b_2 \overline{x} \ , \ b_2 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2}$$

– 최소제곱추정량이 어떤 특징이 있는가?

최소제곱추정량은 최소제곱추정 값들의 집합으로, 평균과 분산을 갖는 확률분포를 보인다.

이 때, 최소제곱추정량에서 b_1과 b_2는 $b_1 = \overline{y} - b_2\overline{x}$ $\quad b_2 = \dfrac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum(x_i - \overline{x})^2}$으로, 확률변수 라면 기댓값은 $E(b_1) = \beta_1, E(b_2) = \beta$가 되며, 분산은

$var(b_1) = \sigma^2[\dfrac{\Sigma x_i^2}{N\Sigma(x_i - \overline{x})^2}], var(b_2) = [\dfrac{\sigma^2}{\Sigma(x_i - \overline{x})^2}]$이 된다.

최소제곱추정량은 분산이 적을수록 해당 추정량의 표본추출 정확성은 커지게 되며, 무작위 오차항의 분산인 σ^2이 클수록 통계모형의 불확실성이 커지며 최소제곱 추정량의 분산과 공분산이 증가한다.

· 최소제곱추정량 b_2가 y_i에 대해 선형식으로 표현됨을 증명하시오.

$b_2 = \dfrac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum(x_i - \overline{x})^2}$를 가정 SR1과 약간의 대수를 이용하여 선형 추정량으로 나타낼 수 있다. $b_2 = \Sigma w_i y_i (i = 1\text{부터 } N\text{까지})$, $w_i = \dfrac{x_i - \overline{x}}{\Sigma(x_i - \overline{x})^2}$인데, 위에서 w_i항은 확률적이 아닌 x_i에만 의존하므로, w_i도 역시 확률적이지 않다. 그리고 나서 대수학을 이용하면 이론적으로 편리한 방법으로, b_2를 $b_2 = \beta_2 + \Sigma w_i e_i$으로 나타낼 수 있다.

$b_2 = \Sigma w_i y_i (i = 1\text{부터 } N\text{까지})$

$y_i = \beta_1 + \ \beta_2 x_i + e$

$b_2 = \Sigma w_i y_i = \Sigma w_i(\beta_1 + \beta_2 x_i + e_i) = \beta_1 \Sigma w_i + \beta_2 \Sigma w_i x_i + \Sigma w_i e_i = \beta_2 + \Sigma w_i e_i$

$\because \ \Sigma w_i = \Sigma[\dfrac{(x_i - \overline{x})}{\Sigma(x_i - \overline{x})}] = \dfrac{1}{\Sigma(x_i - \overline{x})^2}\Sigma(x_i - \overline{x}) = 0$

$\Sigma(x_i - \overline{x})^2 = \Sigma(x_i - \overline{x})(x_i - \overline{x}) = \Sigma(x_i - \overline{x})x_i - \overline{x}\Sigma(x_i - \overline{x}) = \Sigma(x_i - \overline{x})x_i$

$\therefore \ \Sigma w_i x_i = \dfrac{\Sigma(x_i - \overline{x})x_i}{\Sigma(x_i - \overline{x})^2} = \dfrac{\Sigma(x_i - \overline{x})x_i}{\Sigma(x_i - \overline{x})x_i} = 1$

$\therefore \ b_2 = \Sigma w_i y_i = \Sigma w_i(\beta_1 + \beta_2 x_i + e_i) = \beta_1 \Sigma w_i + \beta_2 \Sigma w_i x_i + \Sigma w_i e_i = \beta_2 + \Sigma w_i e_i$ Q.E.D

· 최소제곱추정량 b_2의 기댓값(또는 평균)이 모수 β_2가 됨을 증명하시오.

$E(b_2) = E(\beta_2 + \Sigma w_i e_i) = E(\beta_2 + w_1 e_1 + w_2 e_2 + ... + w_N e_N) = E(\beta_2) + E(w_1 e_1) + E(w_2 e_2) + ... + E(w_N e_N)$
$= E(\beta_2) + \Sigma E(w_i e_i) = \beta_2 + \Sigma w_i E(e_i) = \beta_2$

(w_i는 확률적이 아닌 상수, $E(e_i) = 0$ 가정)

· **최소제곱추정량 b_2의 분산이 산출되는 과정을 유도하시오.**

$$Var(b_2) = \frac{\delta^2}{\Sigma(x_i - \overline{x})^2}$$

첫 번째 유도방법

$b_2 = \beta_2 + \Sigma w_i e_i$
$var(b_2) = E[b_2 - E(b_2)]^2$
$var(b_2) = E(\beta_2 + \Sigma w_i e_i - \beta_2)^2 = E(\Sigma w_i e_i)^2 = E(\Sigma w_i^2 e_i^2 + 2\Sigma\Sigma w_i w_j e_i e_j)$ (단, $i \neq j$)
$= \Sigma w_i^2 E(e_i^2) + 2\Sigma\Sigma w_i w_j E(e_i e_j)$ (단, $i \neq j$)
$= \sigma^2 \Sigma w_i^2 = \dfrac{\sigma^2}{\Sigma(x_i - \overline{x})^2}$

$\because$ $w_i \neq$ 확률적

$\because var(e_i) = E(e^2)$ (첫 번째 가정), $cov(e_i, e_j) = 0$ (두 번째 가정)

$$\because \Sigma w_i^2 = \Sigma[\frac{(x_i - \overline{x})^2}{\Sigma(x_i - \overline{x})^{2^2}}] = \frac{\Sigma(x_i - \overline{x})^2}{\Sigma(x_i - \overline{x})^{2^2}} = \frac{1}{\Sigma(x_i - \overline{x})^2}$$

$\therefore$ $Var(b_2) = \dfrac{\delta^2}{\Sigma(x_i - \overline{x})^2}$ Q.E.D

두 번째 유도방법

$var(aX + bY) = a^2 var(X) + b^2 var(Y) + 2ab cov(X, Y)$

$b_2 = \beta_2 + \Sigma w_i e_i$

$var(b_2) = var(\beta_2 + \Sigma w_i e_i) = var(\Sigma w_i e_i)$
$= \Sigma w_i^2 var(e_i) + \Sigma\Sigma w_i w_j cov(e_i, e_j)$ (단, $i \neq j$)
$= \Sigma w_i^2 var(e_i)$
$= \sigma^2 \Sigma w_i^2 = \dfrac{\sigma^2}{\Sigma(x_i - \overline{x})^2}$

$\because$ $\beta_2 =$ 상수, $cov(ce_i, e_j) = 0, var(e_i) = \sigma^2$

$\therefore$ $Var(b_2) = \dfrac{\delta^2}{\Sigma(x_i - \overline{x})^2}$ Q.E.D

· **최소제곱추정량 b_2의 분산을 작게 할 수 있는 방법을 소개하시오.**

최소제곱추정량 b_2의 분산을 작게 하기 위해서는 무작위 오차항의 분산인 σ^2가 작아져야 한다. 두 번째로 $\Sigma(x_i - \overline{x})$가 커져야 한다. 마지막으로 표본크기 N이 커져야 한다.

· **최소제곱추정량 b_2의 분산을 무한히 작게 해서 $b_2 = \beta_2$가 될 수 있는가?를 설명하시오.**

개별적인 추정값인 b_2는 β_2에 근접할 수도 큰 차이가 날 수도 있다. 또한, β_2는 알 수 없기 때문에 b_2의 분산을 무한히 작게 한다면 β_2에 근접할 수도 있고 하지 않을 수도 있다.

3.3 가우스-마코프(Gauss-Markov) 정리를 설명하시오.

가우스-마코프 정리는 선형회귀 모형에 관한 과정 SR1 - SR5 하에서 추정량 b_1 및 b_2이 β_1 및 β_2의 모든 선형 및 불편 추정량 중에서 최소의 분산을 가지며 β_1 및 β_2의 최우수 선형 불편 추정량이다. 가우스-마코프의 정리는 추정량 b_1과 b_2가 선 형 및 불편한 유사한 추정량들과 비교하여 최우수하며, 추정량 b_1과 b_2가 같은 부류 내에서 분산이 최소이므로 최우수하다고 본다. 가우스-마코프 정리가 지켜지기 위해서는 가정 SR1-SR5가 준수되어야만 하고, SR6에 의존하지 않는다. 가우스-마코프 정리는 최소제곱 추정량에 적용되지만, 단 하나의 표본에 기초한 최소제곱 추정값에는 적용되지 않는다.

3.4 단순회귀모형에서 가정(SR6, 정규성 가정)이 충족될 경우 최소제곱추정량 b_2는 어떤 분포를 하는지 설명하시오.

단순회귀모형에서 정규성 가정이 충족될 경우 최소제곱 추정량의 확률도 정규분포하게 된다. 최소제곱추정량 b_2의 정규 확률변수의 합계는 정규분포한다.

$$b_1 \sim N(\beta_1, \frac{\delta^2 \Sigma x^{2_i}}{N\Sigma(x_i - \overline{x})^2}) \quad b_2 \sim N(\beta_2, \frac{\delta^2}{\Sigma(x_i - \overline{x})^2})$$