

1. 모회귀선과 표본회귀선에 대한 질문이다.

두 회귀선의 차이(모수와 계수, 오차와 잔차, 모집단과 표본, 회귀선 표기 방법 등)를 비교하시오.

모회귀선과 표본회귀선의 수식은 각각 다음과 같다.

식(1) $y_i = \beta_1 + \beta_2 x_i + e_i$ (모회귀선 추정식)

식(2) $\hat{y} = b_1 + b_2 x_i + \hat{e}$ (표본회귀선 추정식)

모회귀선의 β_1, β_2 는 모회귀선에서 사용되는 모수(parameter)이며, b_1, b_2 는 표본회귀선에서 사용되는 계수이자 모수의 추정값이다. 한편 모회귀선에서 사용되는 오차항 e_i 는 종속변수 y 값과의 $E[y|x]$ 차이에서 생기는 차이이며, 잔차항 \hat{e} 는 표본회귀선에서 이용되는 $y_i - \hat{y}_i$ 값의 차이이다. 덧붙여 모회귀선은 모집단에서 도출되는 직선의 선형함수이며 표본회귀선은 표본 추출한 집단에서 표출되는 직선의 선형함수이다.

2. 계량경제모형에서 통계적 또는 확률적 관계식을 쓰는 이유는? 이 경우 오차항 또는 잔차항에 대한 기본 가정을 설명하시오. 이때 오차항과 잔차항의 기본가정은 같은가?

- 사회과학 현상에서 나타나는 변수들 간의 관계 중엔 확정적 관계로 표현될 수 있는 경우가 거의 없기 때문이다. 예를 들어 소득수준이 같은 근로자 가구라도 소비 지출액은 가구마다 각각 다르게 나타난다.
- SR1-SR6를 적는다.
- 오차항과 잔차항의 기본가정은 같다. 다만 교과서에는 오차항에 대해서만 SR1-SR6을 제시하고 있다.

3. 다음 주장에 대해 옳고 그름을 논하시오.

3.1. 모회귀선(또는 진회귀선)은 오직 하나뿐이다.

O/ 모회귀선은 하나밖에 존재하지 않는다. 단, 표본회귀선은 표본의 개수만큼 기대할 수 있다

3.2. 표본회귀선은 무수히 존재한다.

O/ 표본추출된 집단에 비등하게 존재한다.

3.3. 표본회귀선에서 계수 b_2 는 평균과 분산을 갖는 추정량이다.

O/ 표본회귀계수 b_2 는 표본분포를 하며 평균과 분산을 갖는다.

3.4 표본회귀모형에서 종속변수 실제 값과 추정값의 차이는 오차(e_i)이다.

X/ 표본회귀모형에서는 이를 잔차 \hat{e} 라고 한다.

4. 다음과 같이 개인별 연간 소득과 소비지출액이 있다. 다음 질문에 답하시오.

- 4.1. 종속변수, 독립변수, 그리고 인과관계를 나타내는 표본회귀모형을 제시하시오.
- 4.2 한계소비성향(b_2)은 얼마인가?
- 4.3 소득이 전혀 없는 경우에 소비지출액(b_1)은 얼마인가?
- 4.4 소득이 1억이면 소비지출액은 어느 정도 될까?

개인구분(i)	소득(x_i)	소비지출액(y_i)
1.	1000만원	500만원
2.	2000만원	500만원
3	3000만원	1000만원
4	4000만원	1000만원
5	5000만원	1000만원

4-1. 개인별 연간 소득과 소비지출액은 선형관계를 가지므로,

$$\text{식(3)} \quad y_i = b_2 x_i + b_1 + \hat{e}$$

라는 단순선형회귀함수로 나타낼 수 있다.

4-2. 한계소비성향 b_2 는 다음과 같은 식으로 구할 수 있다.

$$\text{식(4)} \quad b_2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \text{이므로,}$$

$$\text{식(5)} \quad \bar{x} = \frac{1,000 + 2,000 + 3,000 + 4,000 + 5,000}{5} = 3,000$$

$$\text{식(6)} \quad \bar{y} = \frac{500 + 500 + 1,000 + 1,000 + 1,000}{5} = 800$$

$$\text{식(7)} \quad b_2 = \frac{(1,000 - 3,000)(500 - 800) + \dots + (5,000 - 3,000)(1,000 - 800)}{(1,000 - 3,000)^2 + \dots + (5,000 - 3,000)^2} = 0.15$$

$$\therefore b_2 = 0.15$$

4-3. $x_i = 0$ 일 경우, $b_1 = \bar{y} - b_2 \bar{x}$ 를 이용하여, $b_1 = 350$ 이므로

$$\therefore y_i = 350 \quad (\text{단, } \hat{e} = 0)$$

4-4. $x_i = 10,000$ 일 경우,

$$\text{식(8)} \quad y_i = 0.15 \times 10,000 + 350 = 1,850$$

$$\therefore y_i = 1,850 \quad (\text{단, } \hat{e} = 0)$$

5. 만일 표본을 5명만 한정해서 문제 4에서 같이 한계소비성향을 추정했을 경우 과연 이 추정값을 신뢰할 수 있는가에 대해 논하시오.

- 신뢰할 수 있는 통계가 되기 위해서는 충분한 양의 표본이 존재하여야 한다.