

응용계량경제학 8주차 과제 201547130 윤요섭

1. 모형의 적합성을 결정계수(R^2)로 검정하는 것과 독립변수의 통계적 유의성을 t 검정을 통해서 하는 것은 내용이 전혀 다르다. 왜 그럴까?

R^2 은 회귀모형 내에서 x 로 설명할 수 있는 y 의 변동비율을 의미한다. 반면 t -통계량 값은 추정값을 표준오차로 나눈 값을 의미하며, t 검정의 경우 구간추정을 이용하여 모수 β_2 가 몇 %의 확률로 주어진 범위에 있는지를 알아낸다. 따라서, 접근 방식이 다르다. t 검정의 경우

$$P(b_2 - t_{(1-\alpha/2, N-2)} \sqrt{\hat{\sigma}^2 / \Sigma(x_i - \bar{x})} \leq \beta_2 \leq b_2 + t_{(1-\alpha/2, N-2)} \sqrt{\hat{\sigma}^2 / \Sigma(x_i - \bar{x})}) = 1 - \alpha \text{ or } 100(1 - \alpha)\%$$

으로 β_2 가 일정구간에 있을 확률이 $100(1 - \alpha)\%$ 이라고 구하는 것이다. 그러나 결정계수 R^2 은

$$R^2 = \frac{\Sigma(\hat{y}_i - \bar{y})^2}{\Sigma(y_i - \bar{y})^2} = \frac{SSR}{SST} \text{으로, 회귀분석에 의해 설명되는 표본평균에 대한 } y \text{의 총변동부분을 표본}$$

평균에 대한 y 의 총변동 크기로 나누어 계산한다. 즉, y 의 자신의 평균에 대한 변동을 x 가 설명하는 비율을 의미한다. 따라서, t 검정을 통해 모수 β_2 에 대한 정보를 알아내는 것과 R^2 을 통해 표본 관측치가 추정된 회귀선상에 얼마나 있으며, 표본회귀선에서 추정된 \hat{y}_i 가 실제 관측값 y_i 를 얼마나 잘 설명하고 있는지 판단하는 것은 그 내용이 전혀 다르다고 할 수 있다.

2. 2. 식(7-2)의 양변을 제곱한 후 관측치 1부터 n 까지 모두 합하면 식(7-3)과 같이 됨을 증명하십시오.

$$(7-2) \quad y_i - \bar{y} = (\hat{y}_i - \bar{y}) + \hat{e}_i$$

$$(7-3) \quad \Sigma(y_i - \bar{y})^2 = \Sigma(\hat{y}_i - \bar{y})^2 + \Sigma\hat{e}_i^2$$

식(7-2)의 양변을 제곱하면 $(y_i - \bar{y})^2 = (\hat{y}_i - \bar{y})^2 + \hat{e}_i^2 + 2(\hat{y}_i - \bar{y})\hat{e}_i$ 가 된다. 이후 양변을 관측치 1부터 n 까지 모두 합하면 $\Sigma(y_i - \bar{y})^2 = \Sigma(\hat{y}_i - \bar{y})^2 + \Sigma\hat{e}_i^2 + 2\Sigma(\hat{y}_i - \bar{y})\hat{e}_i$ 가 된다. 여기서, $\Sigma(\hat{y}_i - \bar{y})\hat{e}_i = 0$ 이므로, 식(7-2)의 양변을 제곱한 후 관측치 1부터 n 까지 모두 합하면, $\Sigma(y_i - \bar{y})^2 = \Sigma(\hat{y}_i - \bar{y})^2 + \Sigma\hat{e}_i^2$ 이 된다.

3. 7주차 문제와 관련된다. 만일 갑이라는 학생이 개인소득 단위를 만원에서 천원으로 바꿔서 엑셀에 입력한 후 회귀분석을 했다고 하자. 단, 엑셀을 이용하여 직접 계산하십시오. 그리고 엑셀 파일을 답안지와 함께 동시에 제출하십시오.

3.1 갑이라는 학생이 입력한 자료를 이용하여 결정계수(R^2)를 산출하십시오.

$$R^2 = \frac{\Sigma(\hat{y}_i - \bar{y})^2}{\Sigma(y_i - \bar{y})^2} = \frac{SSR}{SST}, \quad R^2 = \frac{SSR}{SST} = 0.9949399$$

3.2. 갑이라는 학생이 엑셀에 입력한 자료에 기초하여 표본회귀선을 구하십시오.

$$y_i = b_1 + b_2x_i \text{이므로 표본회귀선은 } y_i = 350 + 0.0205x_i \text{이다.}$$

- 3.3. 3.2의 질문과 관련된다. 이때 결정계수와 한계소비성향(b_2)값은? 또한 b_2 의 표준오차($se(b_2)$)

와 t 값은 어떻게 되는가? 7주차 분석결과와 비교하여 설명하시오.

$b_2=0.0205$, $se(b_2) = 0.152983$ $t_{값} = 0.134$ 이므로, 7주차 분석결과와 비교했을 때, $b_2 = 0.205$, $se(b_2) = 0.011111$, $t_{값} = 18.451845$ 와 비교하여 달라진 것을 확인할 수 있다.

3.4. 3.3의 질문과 관련된다. 한계소비성향 β_2 가 0이라고 주장하는 이가 있다. 이에 대한 가설의 진위 여부를 검정하시오 단, 유의수준 α 는 0.05이다.

$H_0 : \beta_2 = 0$, $H_1 : \beta_2 \neq 0$ 을 설정하고, $t_c \leq t$ 혹은 $t \leq -t_c$ 일 때, 귀무가설을 기각하고 대립가설을 채택한다. 이 때, t 는 0.134이고, t_c 는 2.306이므로, 95%구간에서 구간추정범위는 $b_2 + 2.306se(b_2) = 0.0205 + (2.306 \times 0.152983) = 0.3732788$ 이기 때문에 ± 0.3732788 이다. 이 때, $t_{값}$ 은 0.134이므로 구간추정범위에 속해있다. 따라서 귀무가설이 채택되므로, 한계소비성향 β_2 가 0이라는 주장은 95%확률로 옳다고 할 수 있다.

7주차 과제 당시 조사자료			갑이 소득 단위를 바꿔서 입력한 자료	
개인구분(i)	소득(x_i ,만원)	소비지출액(y_i ,만원)	소득(x_i ,천원)	소비지출액(y_i ,만원)
1.	1000	500	10,000	500
2.	2000	700	20,000	700
3	3000	1000	30,000	1000
4	4000	1200	40,000	1200
5	5000	1400	50,000	1400
6	1000	600	10,000	600
7	2000	800	20,000	800
8	3000	1000	30,000	1000
9	4000	1100	40,000	1100
10	5000	1350	50,000	1350