

9주차 강의자료 및 과제 : 함수형태 선택

- 혹시 결정계수(R^2)가 예상외로 낮아서 모형의 설명력이 낮은 이유는 어디에 있을까?

우리가 지금까지 가정한 모형은 어떤 형태였는가?

식(9-1)에 따르면 독립변수가 오직 하나이다. 또 종속변수와 독립변수가 단순히 선형관계이다. 따라서 우리는 식(9-1)을 지금까지 표본회귀모형(표본회귀식)이라고 정의하였으나, 또 다른 표현으로 단순선형회귀모형(단순선형회귀식)이라고 부르기도 한다¹⁾.

$$(9-1) \quad y_i = b_1 + b_2x_i + \hat{e}_i.$$

- 문제는 종속변수와 독립변수간의 인과관계가 반드시 선형이라는 보장은 없다. 실례로 소비경제이론에 따르면 소득 증가에 따라 소비지출 증가가 체감한다. 이 경우 한계소비성향이 소득이 증가함에 따라 체감할 수 있도록 하기 위해서는 식(9-2)~(9-5)와 같이 비선형의 함수형태를 바꾸어야 한다. 또한 교과서에서 이미 배운 비용함수, 생산함수 등도 비선형이다. 이 경우 종속변수와 독립변수가 선형관계에서 비선형관계가 되도록 종속변수(y) 또는 독립변수(x)를 자연대수($\ln y$ 또는 $\ln x$)로 변환하거나 또는 독립변수 x 를 역변환($\frac{1}{x}$)하여 비선형회귀모형(비선형표본회귀식, 비선형회귀식)으로 바꾸어야 한다.

$$(9-2) \quad \ln y_i = b_1 + b_2 \ln x_i + \hat{e}_i \quad \ln \hat{y}_i = b_1 + b_2 \ln x_i : \text{log-log model, 대수-대수 모형, 양대수 모형}$$

$$(9-3) \quad \ln y_i = b_1 + b_2 x_i + \hat{e}_i \quad \ln \hat{y}_i = b_1 + b_2 x_i : \text{log-linear model, 대수-선형 모형}$$

$$(9-4) \quad y_i = b_1 + b_2 \ln x_i + \hat{e}_i \quad \hat{y}_i = b_1 + b_2 \ln x_i : \text{linear-log model, 선형-대수 모형}$$

$$(9-5) \quad y_i = b_1 + b_2 \frac{1}{x_i} + \hat{e}_i \quad \hat{y}_i = b_1 + b_2 \frac{1}{x_i} : \text{inverse transform model, 역변환모형}$$

- 하지만 우리는 지금부터 또 다른 문제에 봉착하게 된다.
 - 첫째, 사전적으로 회귀모형이 선형인지? 혹은 비선형인지?를 알 수 없는 경우가 대부분이다. 그리고 식(9-1)에서 식(9-5) 중 어느 모형을 택할 것인가?
 - 둘째, 종속변수 또는 독립변수 측정단위가 다를 경우 자료의 비율화 효과로 인해 계수 값이 다르게 된다. 또한 식(9-1)의 경우 b_2 와 식(9-2)의 경우 b_2 값도 다르다. 이 경우 어떻게 해야 할까?
 - 셋째, 만일 비선형회귀모형일 경우에도 지금까지 배웠던 최소제곱추정량, t 검정통계량, 결정계수 산출식을 그대로 적용해도 되는가?
 - 넷째, 식(9-1)~(9-5)이외에도 또 다른 비선형 모형이 있는지?

1) 식(9-1)에서 잔차항을 빼면 표본회귀식 또는 단순선형회귀식이다.

- 다섯째, 독립변수가 여러 개 투입되어 종속변수를 설명해야 하는 다중회귀모형의 경우까지 확장하게 될 경우 어떤 내용을 추가적으로 공부해야 할까?

- 첫째 질문에 대한 답은 다음과 같다.

- 어느 모형을 추정해야 할지를 우리는 사전적으로 잘 알 수 없다. 그렇기 때문에 모든 모형을 대상으로 최소제곱추정법에 의해 계수를 구한다. 계수의 부호가 경제이론에 부합하는지 따져본다(1차조건). 계속해서 t 검정통계량 산출식을 이용하여 계수의 통계적 유의성도 따져본다(2차조건). 끝으로 결정계수를 산출하여 모형의 적합성을 점검한다(3차조건). 최소제곱추정량으로 여러 모형을 추정한 후 그 중 1차 조건(계수의 부호 조건이 경제이론과 일치하는지?), 2차 조건(t 값이 커서 통계적으로 유의한지?), 3차 조건(모형의 적합성을 나타내는 결정계수가 가장 큰지?)을 가장 잘 만족하는 추정결과를 최종적으로 선택한다.

- 둘째 질문에 대한 답은 다음과 같다.

- 종속변수 또는 독립변수 측정단위가 다를 경우 계수 크기가 다르다. 또한 식(9-1)의 경우 b_2 와 식(9-2)의 경우 b_2 값도 다르다. 따라서 상호 비교하기 어려운 문제가 발생한다. 우리는 이와 같은 문제를 동시에 해결하기 위하여 탄력성을 산출하여 상호 비교한다²⁾,

- 식(9-1)~(9-4)에 따른 소득 탄력성 산출식은 다음과 같다. 양변을 편미분한 후 탄력성으로 전환하면 다음과 같다. 이때 \bar{x}, \bar{y} 는 x, y 의 평균이다.

$$(9-1)' \quad b_2 = \frac{\Delta y_i}{\Delta x_i}, \quad \epsilon_{b_2} = b_2 \frac{\bar{y}}{\bar{x}}$$

$$(9-2)' \quad b_2 = \frac{\Delta y_i}{\Delta x_i} \frac{x_i}{y_i}, \quad \epsilon_{b_2} = b_2$$

$$(9-3)' \quad b_2 = \frac{\Delta y_i}{y_i} \frac{1}{\Delta x_i}, \quad \epsilon_{b_2} = b_2 \bar{x}$$

$$(9-4)' \quad b_2 = \Delta y_i \frac{x_i}{\Delta x_i}, \quad \epsilon_{b_2} = b_2 \frac{1}{y}$$

- 첫째 질문의 답에서 3차조건까지 잘 만족하는 추정결과를 제시할 때 항상 탄력성까지 제시해야 한다. 탄력성 ϵ_{b_2} 를 제시함으로써 독립변수 x 가 1% 변화면 종속변수 y 는 몇 % 변화하는지를 알려줘야 한다.

- 셋째, 네째 질문에 대한 답은 다음과 같다.

- 비선형회귀모형일 경우에도 지금까지 배웠던 최소제곱추정량, t 검정통계량, 결정계수 산출

2) 종속변수 또는 독립변수 측정단위가 다를 경우에도 탄력성은 항상 같다. 그리고 모형이 다른 경우 계수는 다르지만 탄력성 크기로 비교가 가능하다.

식을 그대로 적용해도 되는가? 적용해도 된다.

- 식(9-2)는 양대수모형으로 비선형이다. 하지만 $\ln x_i, \ln y_i$ 를 X_i, Y_i 로 치환하면 식(9-6)이 된다. 치환된 식(9-6)은 선형이다. 비선형일지라도 변수를 치환할 경우 단순선형회귀식이 된다.
- 식(9-1)~(9-5)이외에도 비선형함수형태의 총비용함수를 식(9-7)과 같이 선형회귀모형으로 치환한 후 최소제곱추정량, t 검정통계량, 결정계수 산출식을 그대로 적용한다. 이때 조심할 것은 탄력성(ϵ_{b_2})이 어떻게 바뀌는가를 꼼꼼히 따져봐야 한다. 식(9-7)'은 생산량이 1%씩 늘 어날 때 총비용이 몇%씩 늘어나는가를 알려주는 비용탄력성이다.

· (9-2) $\ln y_i = b_1 + b_2 \ln x_i + e_i$

(9-6) $Y_i = b_1 + b_2 X_i + e_i \quad Y_i = \ln y_i \quad X_i = \ln x_i$

(9-7) $Y_i = b_1 + b_2 X_i + e_i \quad Y_i = TC_i \quad X_i = Q_i^2$

(9-7)' $b_2 = \frac{\Delta TC_i}{\Delta Q_i} \frac{1}{2Q_i} \quad , \quad \epsilon_{b_2} = b_2 * \frac{2Q_i^2}{TC_i}$

- 다섯째 질문은 10주차 다중회귀모형에서 공부를 하기로 한다. .

9주차 과제

다음과 같이 20개 가구의 연간 가구소득과 소비지출액이 있다. 다음 질문에 답하시오. 단, 과제 제출시 엑셀파일에 의해 계산했던 증빙자료를 첨부하시오.

개인구분(i) 소득(x_i , 만원) 소비지출액(y_i , 만원)

1.	600	500
2.	2000	700
3	3000	1000
4	4000	1200
5	5000	1200
6	1000	600
7	2000	800
8	3000	1000
9	4000	1100
10	5000	1300
11	6000	1200
12	10000	1500
13	7000	1400
14	500	500
15	2050	750
16	3000	900
17	4000	1000
18	500	450
19	1000	600
20	300	300

1. 단순선형모형을 추정한 후 b_2 , $se(b_2)$, t , R^2 , ϵ_{b_2} 를 구하시오.
2. 선형대수모형을 추정한 후 b_2 , $se(b_2)$, t , R^2 , ϵ_{b_2} 를 산출하시오.
3. 두 모형 중 어느 쪽이 더 우수한가? 3차 조건을 만족할 경우 그 중 어느 쪽을 택할 것인가? 이를 종합적으로 평가하시오.