제 6 장 선형회귀모형의 응용

1. 자료의 측정단위

- 회귀모형의 모수들(α ,β)은 표본자료(sample)의 측정 단위에 따라 그 값이 달라질 수 있다

  → 종속변수와 설명변수의 단위변경: 추정값은 변하나 t-통계치와 R^2 값은 불변

  예) 소득에 대한 표본자료를 백만원 단위로 측정한 경우와 만원단위로 측정한 경우 모수들의 추정값들은 다르게 나타날 수 있다

  ⇨ 측정단위(백만원): income = -0.036 + 0.046 age (age: 1 년 측정단위)

  ⇨ 측정단위(만원): income = -3.6 + 4.6age (age: 1 년 측정단위)

2. 함수 형태

1) 모든 변수에 대해 선형인 기본모형

- 기본적인 다중회귀 함수, $Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$ 에서 종속변수는 독립변수 X_{1i} 와 X_{2i} 에 대해 선형을 이루고 있다.

  → 회귀 계수의 의미는 해당 독립변수의 한단위 변화에 따른 종속변수의 평

균변화율, 한계효과(marginal effect)/ 순효과(net effect)를 나타내는 것으로 일정한 값을 가지는 상수이다:

→ $\beta_1 = \partial E(Y_i)/ \partial X_{1i}$, $\beta_2 = \partial E(Y_i)/ \partial X_{2i}$

2) 한변수에는 비선형, 다른변수에는 선형인 모형

- 종속변수가 하나의 독립변수에 대해서는 선형적 관계를 갖고있으나 다른 독립변수에 대해서는 비선형적 관계를 가진 모형: $Y_i = \alpha + (\beta_1 X_i + \beta_2 X^2_i) + \beta_3 Z_i + \varepsilon_i$

→ 독립변수 Z 에 대한 한계효과는 $\beta_3 = \partial E(Y_i)/ \partial Z_i$

→ 독립변수 X 에 대한 한계효과는 $\partial E(Y_i)/ \partial X_i = \beta_1 + 2\beta_2 X_i$로 X_i 에 의존한다

→ 교과서 P134 (예 6.2) 참조

- 이 경우 회귀계수 추정방법은 $W_i = X^2_i$로 바꾸어 $Y_i = \alpha + \beta_1 X_i + \beta_2 W_i + \beta_3 Z_i + \varepsilon_i$ 형태의 선형함수를 변환시켜 모수와 변수들이 모두 선형인 상태에서 최소자승법을 이용하여 추정한다.

3) 변형된 선형함수의 여러가지 형태

가) 더블로그(double log) 모형

- 독립변수와 종속변수 모두 log 를 취한 경우: $\log(Y_i) = \alpha + \beta \log(X_i) + \varepsilon_i$

  → $\beta = \partial[\log E(Y_i)] / \partial \log X_i$: 독립변수의 상대적 1 단위 변화에 따른 종속변수의 상대적변화 (X 의 1%변화에 대한 변수 Y 의 β% 변화)

- 회귀함수가 지수함수(exponential function)일 경우 이를 선형함수로 변환시켜 최소자승법을 이용하여 추정하고자 할 때 유용하게 사용되는 모형

  → Cobb-Doulas 생산함수: $Y_i = \alpha K^{\beta}{}_i L^{\gamma}{}_i\ e^{\varepsilon i}$ ➔ $\log(Y_i) = \alpha + \beta_1 \log(K_i) + \gamma\beta_2 \log(L_i) + \varepsilon_i$

나) 로그-선형 모형

- 회귀함수에서 종속변수에 log 를 취하는 경우, $\log(Y_i) = \alpha + \beta X_i + \varepsilon_i$

  → $\beta = \partial[\log E(Y_i)] / \partial X_i$: 독립변수의 절대적인 1 단위변화에 대한 종속변수의 상대적 변화 (X 변수 1 단위변화에 대한 Y 변수 β%변화)

다) 선형-로그 모형

- 회귀함수에서 종속변수에 log 를 취하는 경우, $Y_i = \alpha + \beta \log(X_i) + \varepsilon_i$

  → $\beta = \partial E(Y_i) / \partial \log(X_i)$: 독립변수의 상대적인 1 단위변화에 대한 종속변수의 절대적 변화 (X 변수 1%변화에 대한 Y 변수 β 단위의 변화)

라) 역(inverse) 함수

- 회귀함수에서 종속변수와 독립변수간에 역관계를 가지는 경우: $Y_i = \alpha + \beta (1/X_i) + \varepsilon_i$

- 예) Phillips Curve: 실업률과 물가상승율 간에는 역관계가 존재한다

4) 비선형모형

- 회귀함수에서 모수 β에 대해 비선형인 함수 일 경우: $Y_i = \alpha + \beta_1 [1/(X_i - \beta_2)] + \varepsilon_I$

  → 이 경우 $Z_i = 1/(X_i - \beta_2)$로 변환하면, 회귀함수가 선형모형, $Y_i = \alpha + \beta Z_i + \varepsilon_i$을 갖게되나, Z_i에 추정해야할 모수 β_2가 포함되어 있어 선형모형이 되지 못한다

- 이러한 비선형모형에 대한 추정은 반복기법(iteration)을 이용한 비선형최소자승법(nonlinear least squares method)를 이용 한다.

3. 가변수 (모의변수: dummy variable)모형

- 수량(정량:quantitative)적 변수와 질(정성:qualitative)적 변수

  → 수량(정량:quantitative)적 변수: 변수값이 수치로 나타나며 관측 또는 예측이 가능하고, 변수의 크기가 중용한 의미를 갖는다

→ 질(정성:qualitative)적 변수: 변수들이 수량화가 불가능하여 변수값이 어떤 범주(category) 또는 특성을 나타내는 변수

- 가변수(dummy variable)의 개요

→ 질적 변수는 수량화가 불가능하여 기존의 회귀분석방법을 적용할 수 없으므로 이러한 질적변수의 영향을 고려하려면, 이들을 일종의 범주별 자료(categorical data)로 파악해야 한다. 따라서, 질적변수에 일정한 특징이 존재하는지의 여부에 따라 1 과 0 의 값을 부여하여 제한적인 수량변수로 전환시켜 회귀분석의 적용을 가능케한다.

→ 이들 변수들을 2 원변수(binary variable)이라고도 부르며, 임의의 두 값을 대입시켜 수량변수적 특성을 부여 회귀분석의 적용을 가능케함으로써 이들 2 원변수를 통상 가변수또는 모의변수라 한다.