

Part II 선형회귀분석

(Linear Regression Analysis)

제 3 장 계량모델과 회귀분석

1. 회귀분석(Regression Analysis)의 개념

- 기본개념: 한 변수(종속변수)가 다른 여러 변수(독립변수)에 의해 어떻게 영향을 받는가를 수학적 함수식으로 파악함으로써 상호관계를 추론, 분석하는 것
- 회귀분석은 인과관계가 있는 두 변수간의 함수식을 분석대상으로 하며, 경제학연구에 두 가지 측면에서 큰 역할을 담당하고 있다
 - 각종 경제행위에 관한 직관적 견해 또는 이론을 "가설"로 가정하고 이들의 현실타당성 여부를 실증적 분석을 통해 검토
 - 회귀분석이 바로 이 과정에서 경제변수 간의 상호관계를 수학적 함수식으로 정의하고 그 관계식의 현실타당성 여부를 판정하는 역할을 한다.
 - 두 변수간의 관계식이 타당시 될 때, 한 변수(독립변수) 값을 기초로 다른

변수(종속변수) 값을 추정, 예측하는 일이다.

- 협의적으로는 독립(설명)변수의 값이 주어졌을 때 종속변수의 조건부 기대치

(conditional expectation: $E(Y|X)$)를 추정, 예측하는 것이다.

- 예) 식료품 지출과 소득간의 관계

2. 모집단 회귀함수 (Population Regression Function)

- 모집단 자료로부터 각 독립변수 X 값에서 종속변수 Y 의 평균이 선형인 관계를 이루는 식을 뜻하며, 두 변수간에 성립하는 이론적 함수관계를 나타낸다.

$$\rightarrow E[Y|X] = \alpha + \beta X$$

- 모집단 회귀선(population regression line): 모집단 회귀함수를 직선으로 표현한 것.

$$\rightarrow Y_i = \alpha + \beta X_i$$

- 모집단 회귀 모형(population regression model)

$$\rightarrow Y_i = \alpha + \beta X_i + \varepsilon_i$$

$\rightarrow \alpha + \beta X_i$ 는 종속변수(Y)의 값중에서 독립변수(X)가 설명할 수 있는 체계적 부분(systematic part)이며, ε_i 는 독립변수에 의해 설명되지 못하는 종속변수의 비체계적 (non-systematic part)

$\rightarrow \alpha$ 와 β 는 모집단 회귀모형의 절편과 기울기를 나타내는 회귀계수 (regression coefficient)로, 모집단 회귀함수의 특징을 분석한다는 것은 회귀계수(α 와 β)를 구하는 것을 뜻한다.

- 회귀분석은 많은 종속변수 (Y)의 실제 측정값들중 일부를 추출한 표본관측값

에 의하여 통계적 추론을 하는것으로, 추출된 표본측정치를 이용하여 모회귀선($\alpha + \beta X_i$)의 특성치인 모수 (α 와 β)에 대한 추정치를 도출하여 추론하게 되는 과정이다.

→ α 와 β 값을 알면, 독립변수(X)값을 이용하여 종속변수(Y)에 대한 예측이 가능하다.

3. 오차항 (error term)에 대한 분석

- 현실적으로 관측할 수 있는 Y 의 실제 값들은 $E[Y|X]$ 값과 차이가 있으며, 실제로 관측된 Y 값들을 Y_i 라 하면 $E[Y|X]$ 와 Y_i 값의 차이는 이론적으로는 설정된 함수관계식의 현실 설명능력을 평가하는 것이며, 모형에 포함된 설명변수(독립변수)로서는 설명할 수 없는 요소가 된다.

→ 이러한 차이를 오차항(error term) 또는 교란항(disturbance term)이라 하고, " ϵ_i "로 표기한다.

- 오차항(error term)의 발생원인은,
 - a) 전혀 설명이 불가능한 확률오차(random factors),
 - b) 측정상의 오차(measurement errors),

- c) 설명력이 있는 다른독립변수가 함수에서 제외된 경우 (omitted variables),
- d) 두변수간의 관계가 1 차함수가 아닐경우 (specification error)

4. 표본회귀선 (sample regression line)

- 추정(estimation): 모집단 회귀선은 대상집단의 모든 자료가 획득가능 한 경우에만 추정할 수 있으나, 현실적으로 모집단의 모든 자료획득이 불가능함으로, 대신 모집단의 자료중 일부를 표본으로 추출하여 표본자료로부터 모집단 회귀선의 특성을 유추하는 것
- 표본회귀선 (sample regression line): 모집단으로부터 추출한 표본자료(sample)을 잘반영하는 회귀선.
 - ➔ 표본회귀선은 모집단의 일부인 표본자료로부터 구한 것으로 모집단 회귀선과 일치하지 않을 수 있다
 - ➔ 표본의 수를 크게 하고 모집단의 특성을 잘 반영하는 확률표본(random sample)을 추출한다면, 이 표본회귀선으로부터 모집단 회귀선에 대한 정확한 정보를 얻을 수 있다.
- 표본회귀모형 (sample regression model)

→ $\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i + e_i$

→ $\hat{\alpha}$ 와 $\hat{\beta}$ 은 모집단 회귀선의 모수인 α 와 β 의 추정량(estimator)

→ e_i 는 잔차항(residuals) 또는 회귀오차(regression error)이며, 추정치인 \hat{Y}_i 와 실제값인 Y_i 의 차이를 나타낸다. 또한 모집단회귀모형의 오차항인 ε_i 에 대한 추정치이기도 하다

→ 표본자료로부터 회귀선을 추정(estimation)한다 함은 회귀계수 α 와 β 의 추정량 $\hat{\alpha}$ 와 $\hat{\beta}$ 를 구하는 것이다.