참조 : 충남대 백민희 교수님 경제통계분석 강의

단순회귀분석

▫ 상관계수

◦ 단순히 두 연속형 변수 사이에 어느 정도 밀접한 상관관계가 있는가를 분석하는 통계량

◦ $\gamma = S_{XY}/(S_X \cdot S_Y) = \sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y}) / \sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2 \sum_{i=1}^{n}(Y_i-\overline{Y})^2}$

$= \sum_{i=1}^{n} X_i Y_i - \overline{XY} / \sqrt{(\sum_{i=1}^{n} X_i^2) - n\overline{X}^2 (\sum_{i=1}^{n} Y_i^2) - n\overline{Y}^2}$

◦ 두 변수 사이의 관계에서 원인과 결과의 관계는 설명할 수 없음.

1. 단순선형회귀분석(linear regression analysis)

(1) 선형회귀분석

① 쌍으로 관찰된 두 연속형 변수들 사이에 선형관계가 있다고 전제한 후 한 변수를 원인으로 하고 다른 변수를 결과로 하여 두 변수 사이의 선형식을 구하는 통계분석 방법

② 설명변수(explanatory variable ; 또는 독립변수, independent variable) = 원인의 역할

③ 반응변수(response variable ; 또는 종속변수, dependent variable) = 결과의 역할

④ 단순선형회귀분석 = 하나의 설명변수와 하나의 반응변수 사이의 선형관계식을 구하는 것

◦ 단순(simple) = 설명변수가 하나인 것

◦ 선형회귀(linear regression) = 두 변수 사이의 관계가 선형식으로 표현될 수 있는 것

⑤ 선형식의 표현 : Y = a + b X

(2) 선형회귀모형

① 선형회귀모형

◦ 설명변수X , 반응변수Y 라 할 때 회귀모형은 다음과 같이 정의됨 : $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

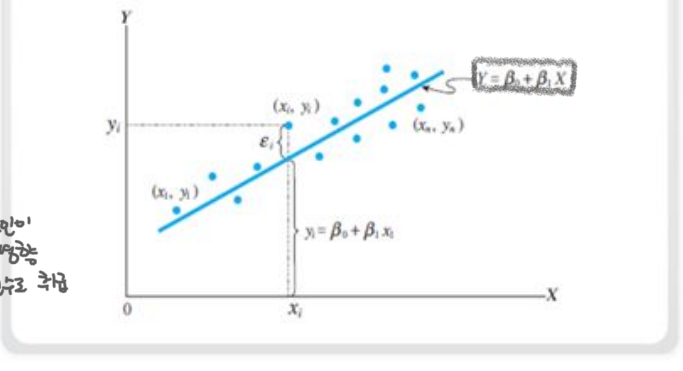
(피설명변수) 오차항

(OLS) 최소제곱법을 이용한 파라메터 $\hat{\beta}_0, \hat{\beta}_1$ 추정 → 회귀계수 (파라메터)

변수X의 영향이 변수Y에 주는 영향을 → 확률변수로 취급

추정된 회귀식 $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$

그림 10-6 X, Y 산포도와 회귀모형



② 단순선형회귀모형의 조건

(a) X와 Y의 관계는 선형식으로 표현할 수 있음.

(b) 잔차는 서로 독립이며 평균 0과 분산 σ^2 을 갖는 정규분포를 따름.

· $cov(\varepsilon_i,\ \varepsilon_j) = 0,\ i \neq j \quad ,\ \varepsilon_i \sim N(0,\ \sigma^2)$

③ Y_i 와 ε_i 의 확률모형

◦ $Y_i \mid X = X_i \sim N(\beta_0 + \beta_1 X_i,\ \sigma^2)$, i=1,2, …, n

◦ $\varepsilon_i \sim N(0,\ \sigma^2)$

⇨ Y_i **와** ε_i **의 분산이 같은 이유** $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

· $X_i = x_i$ 가 주어졌을 때 Y_i 의 평균인 $\beta_0 + \beta_1 X_i$ 는 **상수값**으로 분산 σ^2 을 구하는 데에는 영향을 미치지 않기 때문임.

· Y_i 와 Y_j , $i \neq j$, 즉 Y_i 와 Y_j 는 서로 독립이라고 가정함 $\rightarrow \varepsilon_i,\ \varepsilon_j,\ i \neq j$ 가 서로 독립임.

(a) Y_i 의 조건부 기댓값 = $E[Y_i \mid X = X_i]$ = PRF(population regression function)

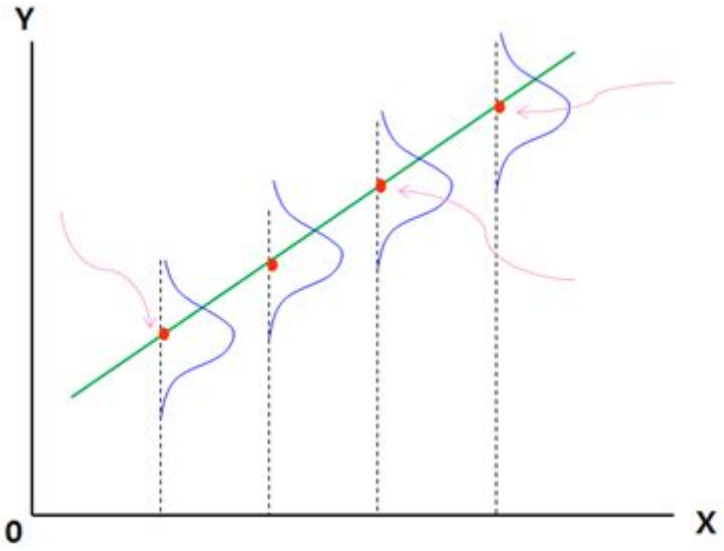
· X가 Y에 어떻게 '평균적으로' 어떤 영향을 미치는가? → X와 Y의 평균적 관계

· $E[Y_i \mid X = X_i] = \beta_0 + \beta_1 X_i$

(b) Y_i 와 ε_i 의 확률모형

◦ $Y_i \mid X = X_i \sim N(\beta_0 + \beta_1 X_i,\ \sigma^2)$, i=1,2, …, n

◦ $\varepsilon_i \sim N(0,\ \sigma^2)$



③ 단순선형회귀분석의 목표

(a) 두 계수 β_0, β_1 의 추정 $\rightarrow$ 선형식 $\beta_0 + \beta_1 X$ 구함

(b) 모형의 분산 σ^2 을 추정 $\rightarrow$ β_0, β_1 의 추정량 분포를 구한 후 β_0 또는 β_1 이 0인지 검정

(c) 선형식 $\beta_0 + \beta_1 X$ 가 자료에 얼마나 적합한가를 측정함.