

9. 단순 선형 회귀분석

9.1 단순 선형 회귀분석의 이해

단순 선형 회귀분석

공분산과 상관계수가 설명하지 못하는 두 변수의 **원인과 결과** (인과관계)를 설명하고자 하는 경우,

이를 수식으로 표현하여 **두 변수의 관계를 설명**하고자 하는 분석방법

9.1 단순 선형 회귀분석의 이해

단순 선형 회귀분석 용어

1. 설명변수 (Explanatory Variable)

: 독립변수라고도 하며, 회귀분석에서 **원인**의 역할을 하는 변수로 통상적으로 X로 표현

2. 반응변수 (Response Variable)

: 종속변수라고도 하며, 원인에 따른 **결과**를 나타나는 변수로 통상적으로 Y로 표현

3. 선형모형 (Linear Model)

: 수학적 선형모형은 $Y = a + bX$ 또는 $Y \neq a + bX$, 즉 선형관계가 성립할지 안할지 결정

9.1 단순 선형 회귀분석의 이해

단순 선형 회귀모형

n 개의 자료가 쌍으로 관측된 경우 위의 모형은 다음과 같이 표현 (ϵ_i 는 오차항)

$$Y_i = a + bX_i + \epsilon_i, \quad i = 1, 2, \dots, n$$

9.1 단순 선형 회귀분석의 이해

단순 선형 회귀분석의 가정

1. 선형성

: 설명변수 X 와 반응변수 Y 의 관계는 선형식으로 표현되는 관계 (분석 모형이 선형식으로 표현)

2. 독립성

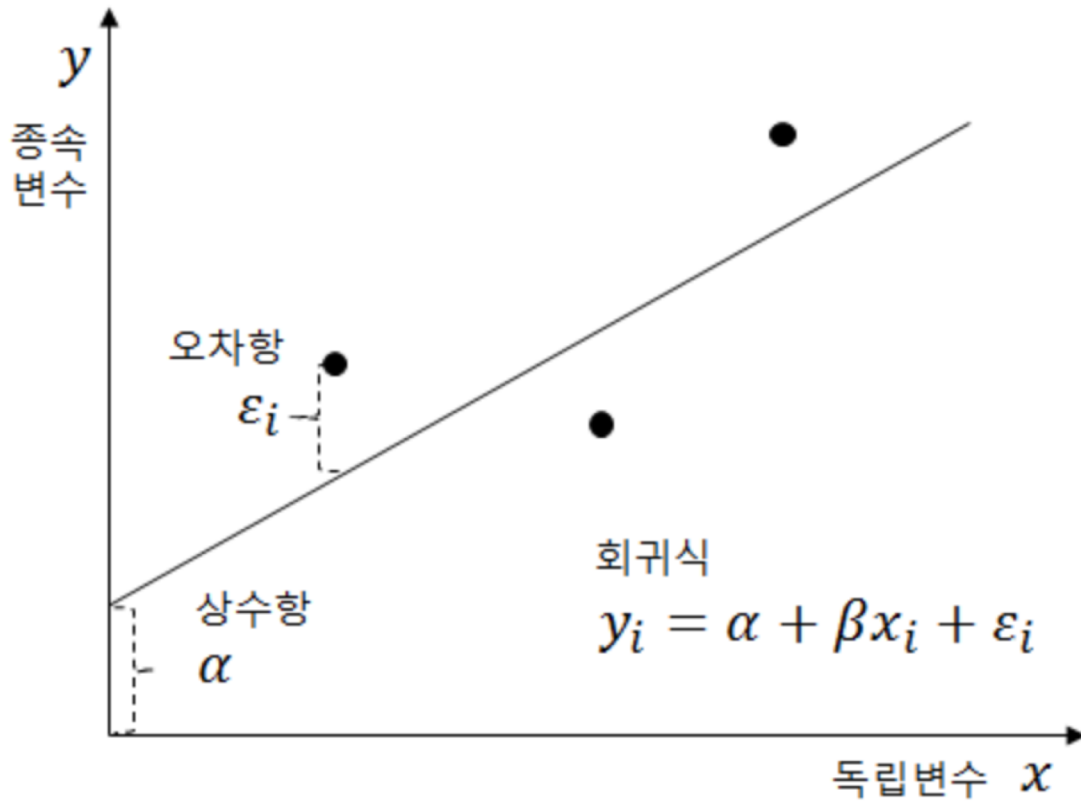
: 오차 ϵ_i 는 서로 독립 (쌍별로 관측되는 데이터는 상호작용을 하지 않음)

3. 정규성 및 등분산성

: 오차 ϵ_i 는 정규분포 $N(0, \sigma^2)$ 를 따름 (오차항은 0을 중심으로 동일한 분산을 가지는 확률변수)

9.2 단순 선형 회귀분석의 절차

모형식의 추정



절편과 기울기에 해당하는 α, β 를 추정해야 하는데 관측된 값과 예측하고자 하는 값의 차이인 **오차**를 **가장 최소화**하는 절편과 기울기를 찾게 됨

1. 절편(α) 추정량

$$: \hat{\alpha} = \frac{S_{xy}}{S_{xx}}, \quad S_{xy} = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}), \quad S_{xx} = \sum_{i=1}^n (X_i - \bar{X})^2$$

2. 기울기(β) 추정량

$$: \hat{\beta} = \bar{Y} - \hat{\alpha}\bar{X}$$

9.2 단순 선형 회귀분석의 절차

단순 선형 회귀분석의 검정 절차

가설의 설정	$H_0 : \beta = 0 \text{ vs } H_1 : \beta \neq 0$ (X, Y는 관련이 없다 (귀무가설))
모형의 분해	$(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)$
제곱합의 분해 (분산의 분해)	$SST = SSR + SSE$
자유도 계산	$n - 1 = 1 + (n - 2)$
제곱 평균의 계산	$MS_{Trt} = \frac{SSR}{1} , \quad MSE = \frac{SSE}{n - 2}$
검정 통계량의 계산과 의사결정	$F = \frac{MSR}{MSE} \sim F_{1,n-2}$

9.2 단순 선형 회귀분석의 절차

단순 선형 모형의 결과 확인

1. 설명력 확인 - 결정 계수 (Coefficient Determination, R^2)

: 결정 계수는 전체 변동 중에서 회귀 부분이 설명하는 변동의 비율

: 회귀식이 전체 변동 중에서 얼마만큼 설명하는지를 나타내는 설명력

$$: R^2 = \frac{SSR}{SST}$$

2. 선형모형 가정의 확인

: 최초 회귀분석에서 가정한 3가지 조건인 선형성, 독립성, 정규성 및 등분산성이 만족하는지 여부를

확인하는 과정이 필요