

단순선형회귀분석 (Simple Linear Regression)

기본 개념 및 추정 방법

Gemini

February 3, 2026

목차

- 1 단순선형회귀란?
- 2 회귀 모델
- 3 기본 가정
- 4 모수 추정 (OLS)
- 5 모델 평가
- 6 요약

단순선형회귀분석(Simple Linear Regression)이란?

- **정의:** 설명 변수(독립 변수) X 와 반응 변수(종속 변수) Y 간의 선형 관계를 모델링하는 기법입니다.
- **목적:**
 - 두 변수 간의 관계 이해
 - 독립 변수 값을 통한 종속 변수의 값 예측
- **예시:**
 - 키와 몸무게의 관계
 - 광고비 지출과 매출액의 관계

회귀 모델의 수식

모집단에 대한 단순선형회귀 모델은 다음과 같이 정의됩니다.

모델 식

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

여기서:

- y_i : i 번째 관측치의 종속 변수 (Dependent Variable)
- x_i : i 번째 관측치의 독립 변수 (Independent Variable)
- β_0 : y 절편 (Intercept)
- β_1 : 기울기 (Slope, 회귀 계수)
- ε_i : 오차항 (Error term, 노이즈)

기본 가정 (Assumptions)

신뢰할 수 있는 회귀 분석을 위해 다음 가정들이 필요합니다.

- ① **선형성 (Linearity)**: X 와 Y 의 관계는 선형이어야 합니다.
- ② **독립성 (Independence)**: 오차항(ε_i)들은 서로 독립이어야 합니다.
- ③ **등분산성 (Homoscedasticity)**: 오차항의 분산은 모든 X 값에 대해 일정해야 합니다 ($\text{Var}(\varepsilon_i) = \sigma^2$).
- ④ **정규성 (Normality)**: 오차항은 정규 분포를 따릅니다 ($\varepsilon_i \sim N(0, \sigma^2)$).

최소제곱법 (Ordinary Least Squares, OLS)

가장 적합한 회귀선은 잔차(Residuals)의 제곱합을 최소화하는 β_0 와 β_1 을 찾는 것입니다.

- **잔차 (Residual):** $e_i = y_i - \hat{y}_i$
- **목적 함수 (비용 함수):**

$$S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2$$

위 식을 β_0 와 β_1 에 대해 편미분하여 0이 되는 값을 찾습니다.

추정된 회귀 계수

최소제곱법을 통해 도출된 추정량($\hat{\beta}_0, \hat{\beta}_1$)은 다음과 같습니다.

기울기 ($\hat{\beta}_1$)

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\text{Cov}(x, y)}{\text{Var}(x)}$$

절편 ($\hat{\beta}_0$)

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

(여기서 \bar{x}, \bar{y} 는 각각 표본 평균입니다.)

결정 계수 (R^2)

모델이 데이터를 얼마나 잘 설명하는지 나타내는 지표입니다.

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

- $0 \leq R^2 \leq 1$
- 1에 가까울수록 회귀선이 데이터를 잘 설명함을 의미합니다.
- X가 Y의 변동을 몇 퍼센트 설명하는지로 해석합니다.

- 단순선형회귀는 하나의 독립 변수로 종속 변수를 예측하는 선형 모델입니다.
- 주로 최소제곱법(OLS)을 사용하여 파라미터를 추정합니다.
- 데이터가 선형성, 독립성, 등분산성, 정규성 가정을 만족할 때 가장 신뢰도가 높습니다.
- 결정 계수(R^2)를 통해 모델의 적합도를 평가할 수 있습니다.

참고 문헌: https://en.wikipedia.org/wiki/Simple_linear_regression