YBIGTA 기초통계 과제

윤희찬

January 2025

1 Simple Linear Regression LSE 도출 과정

단순 선형 모델식 : $y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$

잔차는 실제 값과 예측 값의 차이며 최소제곱 추정량은 이러한 잔차의 제곱합(SSR)을 최소화함으로써 구할 수 있다. 다음은 잔차와 잔차제곱합, 최소제곱법 적용(편미분한 뒤 두 미분식을 0으로 설정하여 정상방정식 도출), 회귀 계수 계산 과정이다.

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

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

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

$$\sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i y_i = \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2$$

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

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

- x_i: 독립 변수의 관측값 - β_0: 절편 (Intercept) - β_1: 기울기 (Slope) - n: 데이터의 총 개수 (관측값의 수) - $\hat{\beta}_0$: 최소제곱법으로 추정된 절편 - $\hat{\beta}_1$: 최소제곱법으로 추정된 기울기 - $\bar{x}$: x_i의 평균값 - $\bar{y}$: y_i의 평균값

2 Multiple Linear Regression LSE 도출 과정

다중 선형 모델식 :

$$\mathbf{y} = \mathbf{X}\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2\mathbf{I})$$

잔차 벡터는 실제 관측값과 예측값의 차이를 나타내며, 다음과 같다.

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\beta$$

잔차 벡터의 제곱합은 잔차 제곱합(Sum of Squared Residuals, SSR)으로 정의된다.

$$S(\beta) = \mathbf{e}^\top\mathbf{e} = (\mathbf{y} - \mathbf{X}\beta)^\top(\mathbf{y} - \mathbf{X}\beta)$$

이후 편미분을 통해 정상방정식을 도출하고 회귀 계수를 구하는 과정은 다음과 같다.

$$S(\beta) = (\mathbf{y} - \mathbf{X}\beta)^\top(\mathbf{y} - \mathbf{X}\beta)$$

$$\frac{\partial S}{\partial \beta} = -2\mathbf{X}^\top(\mathbf{y} - \mathbf{X}\beta)$$

$$\breve{})\breve{} : \quad \mathbf{X}^\top\mathbf{y} = \mathbf{X}^\top\mathbf{X}\beta$$

$$\hat{\beta} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$$

이 때, - $\mathbf{y}$: 종속 변수 벡터 $(n \times 1)$ - $\mathbf{X}$: 독립 변수 행렬 $(n \times p)$ - β: 회귀 계수 벡터 $(p \times 1)$ - $S(\beta)$: 잔차 제곱합 (Sum of Squared Residuals) - $\mathbf{X}^\top\mathbf{X}$: 독립 변수의 공분산 행렬