Least Squares Estimator in Simple and Multiple Linear Regression

Uhm Yoonhee

1 Simple Linear Regression

In simple linear regression, we model the relationship between a dependent variable y_i and an independent variable x_i with the equation:

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

where:

- y_i : dependent variable
- x_i : independent variable
- β_0 : intercept (the expected value of y when $x = 0$)
- β_1 : slope (the rate of change of y with respect to x)
- ϵ_i : error term

Now we want to find the values of β_0 and β_1 that minimize the sum of squared errors(SSE). The error for each data point is the difference between the observed value y_i and the predicted value $\hat{y_i} = \beta_0 + \beta_1 x_i$. Thus,

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

We minimize this sum by taking partial derivatives with respect to β_0 and β_1, setting them to zero to obtain the normal equations.

First, we differentiate with respect to β_0:

$$\frac{\partial SSE}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0$$

This simplifies to:

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

Solving for β_0, we get:

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

Next, we differentiate with respect to β_1:

$$\frac{\partial SSE}{\partial \beta_1} = -2 \sum_{i=1}^{n} x_i \left(y_i - \beta_0 - \beta_1 x_i \right) = 0$$

This simplifies to:

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

Substitute $\beta_0 = \bar{y} - \beta_1 \bar{x}$ into this equation:

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

Solving for β_1:

$$\beta_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Thus, the least squares estimators for β_0 and β_1 are:

$$\beta_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \quad \beta_0 = \bar{y} - \beta_1 \bar{x}$$

2 Multiple Linear Regression

In multiple linear regression, we model the relationship between a dependent variable y_i and multiple independent variables $x_{1i}, x_{2i}, \ldots, x_{ki}$:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \epsilon_i$$

where:

- y_i : dependent variable
- $x_{1i}, x_{2i}, \ldots, x_{ki}$: independent variables
- β_0 : intercept
- $\beta_1, \beta_2, \ldots, \beta_k$: coefficients for the independent variables
- ϵ_i : error term

We aim to estimate the parameters $\beta_0, \beta_1, \ldots, \beta_k$ using the method of least squares, which minimizes SSE. The error for each observation is the difference between the observed value and the predicted value. Thus:

$$J(\beta_0, \beta_1, \ldots, \beta_k) = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ji} \right)^2$$

Here, β_0 is the intercept, and $\beta_1, \beta_2, \ldots, \beta_k$ are the coefficients of the independent variables $x_1, x_2, \ldots, x_k$. We want to find the values that minimize the sum of squared errors.

In matrix form, the multiple linear regression model can be represented as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where:

- $\mathbf{y}$: $n \times 1$ column vector of the observed values of the dependent variable
- $\mathbf{X}$: $n \times (k+1)$ matrix of the independent variables
- $\boldsymbol{\beta}$: $(k+1) \times 1$ column vector of $\beta_0, \beta_1, \ldots, \beta_k$
- $\boldsymbol{\epsilon}$: $n \times 1$ column vector of ϵ_i

The sum of squared errors in matrix form is:

$$\begin{aligned} J(\boldsymbol{\beta}) &= \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \end{aligned}$$

Next step is to minimize the SSE with respect to the parameters $\boldsymbol{\beta}$. The derivative of the SSE with respect to $\boldsymbol{\beta}$ is:

$$\begin{aligned} \frac{\partial J}{\partial \boldsymbol{\beta}} &= -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \\ &= -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \end{aligned}$$

We set this equal to zero to find the value of $\boldsymbol{\beta}$ that minimizes the SSE:

$$\mathbf{X}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0$$

Simplifying this:

$$\mathbf{X}^T\mathbf{y} = \mathbf{X}^T\mathbf{X}\boldsymbol{\beta}$$

The solution for $\boldsymbol{\beta}$ is given by:

$$\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$