Proof of OLSE

Jungmook Kang

Contents

1 About SLR

Introduction

Let us consider the relationship between two variables X and Y. The goal is to express Y as a function of X:

$$Y = f(X)$$

Definitions

- Y: Dependent variable (response variable)
- X: Independent variable (predictor variable)

Objective

The objective is to determine the relationship between Y and X as a function of X.

Simple Linear Regression Model

The simple linear regression model is given by:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, \ldots, n$$

where:

- β_0: Intercept of the regression line
- β_1: Slope of the regression line
- X_i: The i-th predictor value (known constant)
- ϵ_i: The i-th error term, assumed to be independent and identically distributed (i.i.d.) with $\mathcal{N}(0, \sigma^2)$ (Normal distribution with mean 0 and variance σ^2)
- Y_i: The i-th response value (random variable, derived from ϵ_i)

Expected Value and Variance

For the simple linear regression model:

$$\begin{aligned} \text{Expected Value:} \quad & E(Y_i) = \beta_0 + \beta_1 X_i \\ \text{Variance:} \quad & \text{Var}(Y_i) = \sigma^2 \end{aligned}$$

The error term ϵ_i defines the probability distribution for each level of X. Thus, the expected value of Y is:

$$E(Y) = \beta_0 + \beta_1 X$$

2 OLS Estimate in SLR

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

$$b_1 = \frac{S_{xy}}{S_{xx}}, \quad b_0 = \bar{y} - b_1 \bar{x}$$

3 SLR LSE Proof

First-Order Conditions

We start by minimizing

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

Take partial derivatives and set them to zero:

$$\left.\frac{\partial Q}{\partial \beta_0}\right|_{\beta_0=b_0,\beta_1=b_1} = -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i) = 0,$$

$$\left.\frac{\partial Q}{\partial \beta_1}\right|_{\beta_0=b_0,\beta_1=b_1} = -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)\,x_i = 0.$$

Deriving b_0 and b_1

From

$$\frac{\partial Q}{\partial \beta_0} = 0 \quad \Longrightarrow \quad \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i) = 0,$$

we have

$$\bar{y} - b_0 - b_1\bar{x} = 0 \quad \Longrightarrow \quad \bar{x}\bar{y} - b_0\bar{x} - b_1\bar{x}^2 = 0.$$

From

$$\frac{\partial Q}{\partial \beta_1} = 0 \quad \Longrightarrow \quad \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)\,x_i = 0,$$

we get

$$\sum_{i=1}^{n} x_i y_i - b_0\sum_{i=1}^{n} x_i - b_1\sum_{i=1}^{n} x_i^2 = 0 \quad \Longrightarrow \quad \frac{\sum_{i=1}^{n} x_i y_i}{n} - b_0\bar{x} - b_1\frac{\sum_{i=1}^{n} x_i^2}{n} = 0.$$

Combining the two results:

$$\bar{x}\bar{y} - b_1\bar{x}^2 = \frac{\sum_{i=1}^{n} x_i y_i}{n} - b_1\frac{\sum_{i=1}^{n} x_i^2}{n}.$$

Hence,

$$\frac{b_1}{n}\Big(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\Big) = \frac{1}{n}\Big(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\Big).$$

Therefore,

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}, \quad b_0 = \bar{y} - b_1\bar{x}.$$

Here,

$$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}), \quad S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2.$$

4 About MLR

Regression of Y on more than one (multiple) predictors.

First-Order Model with X_1, X_2:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i,$$

where ϵ_i represents the error term.

Assumptions:

$$E(\epsilon_i) = 0.$$

Expected Value:

$$E(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}.$$

Interpretation:

The regression function is a *plane* in a three-dimensional space defined by Y, X_1, and X_2.

Definitions:

- β_0: The Y-intercept of the regression plane when $X_1 = X_2 = 0$.
- β_1: The change in $E(Y)$ per unit increase in X_1, holding X_2 constant.
- β_2: The change in $E(Y)$ per unit increase in X_2, holding X_1 constant.

Partial Regression Coefficients:

- β_1 and β_2 are referred to as *partial regression coefficients*.
- Partial derivatives of the regression function:

$$\frac{\partial E(Y)}{\partial X_1} = \beta_1, \quad \frac{\partial E(Y)}{\partial X_2} = \beta_2.$$

4.1 Extension

The regression model can be expressed in matrix form as:

$$Y = X\beta + \varepsilon,$$

where:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix}.$$

Components:

- Y: $n \times 1$ vector of response variables,
- β: $p \times 1$ vector of unknown parameters,
- ε: $n \times 1$ vector of random errors,
- X: $n \times p$ matrix of predictors, where the first column is all ones for the intercept.

$$E(\varepsilon)_{n\times 1} = \mathbf{0}, \quad \mathrm{Cov}(\varepsilon)_{n\times n} = \sigma^2 \mathbf{I}_n.$$

Thus:

$$E(Y)_{n\times 1} = X\beta, \quad \mathrm{Cov}(Y)_{n\times n} = \sigma^2 \mathbf{I}_n.$$

5 OLS Estimate in MLR

The OLS estimate $b = \hat{\beta}$ of β minimizes the following objective function:

$$Q(\beta) = \sum_{i=1}^{n} \left(Y_i - (\beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p-1} X_{i,p-1})\right)^2.$$

In matrix form:

$$Q(\beta) = (Y - X\beta)'(Y - X\beta),$$

which should be minimized with respect to β.

The OLS estimate is obtained as:

$$b = \hat{\beta} = \arg\min_{\beta} (Y - X\beta)'(Y - X\beta) = (X'X)^{-1}X'Y.$$

6 MLR LSE proof

The least squares estimate $\mathbf{b}$ minimizes the residual sum of squares:

$$\mathbf{b} = \arg\min_{\boldsymbol{\beta}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}),$$

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

The residual sum of squares can be written as:

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}.$$

Since $\mathbf{Y}'\mathbf{X}\boldsymbol{\beta}$ is a scalar, it holds that:

$$\mathbf{Y}'\mathbf{X}\boldsymbol{\beta} = (\mathbf{Y}'\mathbf{X}\boldsymbol{\beta})' = \boldsymbol{\beta}'\mathbf{X}'\mathbf{Y}.$$

Thus, the residual sum of squares becomes:

$$\mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}.$$

6.1 Derivative of Linear Form

To verify:

$$\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}'\mathbf{A}) = \mathbf{A}, \quad \boldsymbol{\beta}, \mathbf{A} \text{ is vector.}$$

Reason: For $\mathbf{A} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$, we have:

$$\frac{\partial}{\partial \boldsymbol{\beta}}(\beta_1 a_1 + \beta_2 a_2 + \cdots + \beta_n a_n) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \mathbf{A}.$$

6.2 Derivative of Quadratic Form

To compute the derivative:

$$\frac{\partial}{\partial \boldsymbol{\beta}}[\boldsymbol{\beta}'\mathbf{A}\boldsymbol{\beta}] = (\mathbf{A} + \mathbf{A}')\boldsymbol{\beta}, \quad \mathbf{A} \text{ is matrix.}$$

Proof:

$$\boldsymbol{\beta}'\mathbf{A}\boldsymbol{\beta} = \sum_{i,j} \beta_i A_{ij} \beta_j = \sum_{j \neq k} \beta_j A_{kj} \beta_k + \sum_{i \neq k} \beta_i A_{ik} \beta_k + A_{kk}\beta_k^2 + ...$$

and differentiating term by term:

$$\frac{\partial \beta' A \beta}{\partial \beta_k} = \sum_{j \neq k} \beta_j A_{kj} + \sum_{i \neq k} \beta_i A_{ik} + 2A_{kk}\beta_k.$$

$$= \sum_{j=1}^{n} \beta_j (A_{kj} + A_{jk}) = [\mathbf{A}\boldsymbol{\beta}]_k + [\mathbf{A}'\boldsymbol{\beta}]_k,$$

and for all k, the result is:

$$\frac{\partial}{\partial \boldsymbol{\beta}}[\boldsymbol{\beta}'\mathbf{A}\boldsymbol{\beta}] = (\mathbf{A} + \mathbf{A}')\boldsymbol{\beta}.$$

If $\mathbf{A}$ is symmetric ($\mathbf{A} = \mathbf{A}'$):

$$\frac{\partial}{\partial \boldsymbol{\beta}}[\boldsymbol{\beta}'\mathbf{A}\boldsymbol{\beta}] = 2\mathbf{A}\boldsymbol{\beta}.$$

To minimize:

$$Q(\boldsymbol{\beta}) = \mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}.$$

Taking the derivative:

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}.$$

Setting to 0:

$$-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = 0 \implies \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$