Simple Linear Regression

Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,$$

where the errors satisfy

$$\mathbb{E}[\varepsilon_i] = 0 \quad \text{and} \quad \text{Var}(\varepsilon_i) = \sigma^2.$$

Derivation of the OLS Estimator for β_1

We start with the simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

and aim to minimize the sum of squared residuals:

$$S(b_0, b_1) = \sum_{i=1}^{n} \Big(Y_i - (b_0 + b_1 X_i) \Big)^2.$$

Our goal is to find the values b_0 and b_1 that minimize $S(b_0, b_1)$.

Step 1: Set Up the Normal Equations

To find the minimum, we take partial derivatives of $S(b_0, b_1)$ with respect to b_0 and b_1 and set them equal to zero.

(a) Partial derivative with respect to b_0:

$$\frac{\partial S}{\partial b_0} = -2 \sum_{i=1}^{n} \Big(Y_i - b_0 - b_1 X_i \Big) = 0.$$

Dividing by -2, we have:

$$\sum_{i=1}^{n} \Big(Y_i - b_0 - b_1 X_i \Big) = 0.$$

This can be rewritten as:

$$\sum_{i=1}^{n} Y_i - n b_0 - b_1 \sum_{i=1}^{n} X_i = 0,$$

which implies

$$b_0 = \bar{Y} - b_1 \bar{X},$$

where $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$ and $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

(b) Partial derivative with respect to b_1:

$$\frac{\partial S}{\partial b_1} = -2\sum_{i=1}^{n} X_i\Big(Y_i - b_0 - b_1 X_i\Big) = 0.$$

Dividing by -2, we get:

$$\sum_{i=1}^{n} X_i\Big(Y_i - b_0 - b_1 X_i\Big) = 0.$$

Step 2: Substitute b_0 into the Second Normal Equation

Replace b_0 with $\bar{Y} - b_1\bar{X}$ in the equation:

$$\sum_{i=1}^{n} X_i\Big(Y_i - (\bar{Y} - b_1\bar{X}) - b_1 X_i\Big) = 0.$$

Simplify the expression inside the summation:

$$Y_i - \bar{Y} + b_1\bar{X} - b_1 X_i = (Y_i - \bar{Y}) - b_1(X_i - \bar{X}).$$

Thus, the normal equation becomes:

$$\sum_{i=1}^{n} X_i\Big[(Y_i - \bar{Y}) - b_1(X_i - \bar{X})\Big] = 0.$$

Step 3: Simplify the Equation

Distribute X_i in the summation:

$$\sum_{i=1}^{n} X_i(Y_i - \bar{Y}) - b_1\sum_{i=1}^{n} X_i(X_i - \bar{X}) = 0.$$

Notice that the first term can be rewritten as:

$$\sum_{i=1}^{n} X_i(Y_i - \bar{Y}) = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) + \bar{X}\sum_{i=1}^{n}(Y_i - \bar{Y}).$$

Since

$$\sum_{i=1}^{n}(Y_i - \bar{Y}) = 0,$$

we have:

$$\sum_{i=1}^{n} X_i(Y_i - \bar{Y}) = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}).$$

Also, note that:

$$\sum_{i=1}^{n} X_i(X_i - \bar{X}) = \sum_{i=1}^{n}(X_i - \bar{X})^2 + \bar{X}\sum_{i=1}^{n}(X_i - \bar{X}) = \sum_{i=1}^{n}(X_i - \bar{X})^2,$$

since $\sum_{i=1}^{n}(X_i - \bar{X}) = 0$.

Thus, the equation simplifies to:

$$\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) - b_1\sum_{i=1}^{n}(X_i - \bar{X})^2 = 0.$$

Step 4: Solve for b_1

Rearrange the above equation to solve for b_1:

$$b_1\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}),$$

and therefore,

$$\hat{\beta}_1 = b_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Conclusion

We have derived that the OLS estimator for β_1 is given by:

$$\boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}.}$$

This completes the step-by-step proof.

Bias and Variance Computation

Step 1: Expressing $\hat{\beta}_1$ in Terms of ε_i

The ordinary least squares (OLS) estimator for β_1 is given by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Substitute the model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ into the estimator:

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(X_i - \bar{X})\left[(\beta_0 + \beta_1 X_i + \varepsilon_i) - \bar{Y}\right]}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \\ &= \frac{\sum_{i=1}^{n}(X_i - \bar{X})\left[(\beta_0 + \beta_1 X_i + \varepsilon_i) - (\beta_0 + \beta_1 \bar{X} + \bar{\varepsilon})\right]}{\sum_{i=1}^{n}(X_i - \bar{X})^2}.\end{aligned}$$

Notice that

$$\beta_0 + \beta_1 X_i - (\beta_0 + \beta_1 \bar{X}) = \beta_1 (X_i - \bar{X}),$$

and since $\bar{\varepsilon} = 0$ in expectation, we can write:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (X_i - \bar{X})\varepsilon_i}{\sum_{i=1}^n (X_i - \bar{X})^2}.$$

For convenience, define

$$s_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Thus, since $\sum_{i=1}^n (X_i - \bar{X})^2 = n s_X^2$, we have:

$$\hat{\beta}_1 = \beta_1 + \frac{1}{n s_X^2} \sum_{i=1}^n (X_i - \bar{X})\varepsilon_i.$$

Step 2: Proving Unbiasedness

Taking the expectation of $\hat{\beta}_1$:

$$\begin{aligned}
\mathbb{E}[\hat{\beta}_1] &= \mathbb{E}\left[\beta_1 + \frac{1}{n s_X^2} \sum_{i=1}^n (X_i - \bar{X})\varepsilon_i\right] \\
&= \beta_1 + \frac{1}{n s_X^2} \sum_{i=1}^n (X_i - \bar{X})\, \mathbb{E}[\varepsilon_i] \\
&= \beta_1 + \frac{1}{n s_X^2} \sum_{i=1}^n (X_i - \bar{X}) \cdot 0 \\
&= \beta_1.
\end{aligned}$$

Thus, $\hat{\beta}_1$ is an unbiased estimator of β_1.

Step 3: Computing the Variance

The variance of $\hat{\beta}_1$ is given by:

$$\mathrm{Var}(\hat{\beta}_1) = \mathrm{Var}\left(\beta_1 + \frac{1}{n s_X^2} \sum_{i=1}^n (X_i - \bar{X})\varepsilon_i\right).$$

Since β_1 is a constant, we have:

$$\mathrm{Var}(\hat{\beta}_1) = \mathrm{Var}\left(\frac{1}{n s_X^2} \sum_{i=1}^n (X_i - \bar{X})\varepsilon_i\right).$$

Because the errors ε_i are uncorrelated and each has variance σ^2, it follows that:

$$\mathrm{Var}\left(\frac{1}{n s_X^2} \sum_{i=1}^n (X_i - \bar{X})\varepsilon_i\right) = \frac{1}{(n s_X^2)^2} \sum_{i=1}^n (X_i - \bar{X})^2\, \mathrm{Var}(\varepsilon_i).$$

Substitute $\text{Var}(\varepsilon_i) = \sigma^2$ and use $\sum_{i=1}^{n}(X_i - \bar{X})^2 = ns_X^2$:

$$
\begin{aligned}
\text{Var}(\hat{\beta}_1) &= \frac{1}{(ns_X^2)^2} \cdot \sigma^2 \cdot (ns_X^2) \\
&= \frac{\sigma^2}{ns_X^2}.
\end{aligned}
$$

Summary

We have shown that:

$$
\mathbb{E}[\hat{\beta}_1] = \beta_1 \quad \text{and} \quad \text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{ns_X^2}.
$$