In Ordinary Least Squares (OLS) regression, the coefficients b_0 (intercept) and b_1 (slope) are derived by minimizing the sum of squared residuals (errors). Let's go through the process of deriving the formulas for b_0 and b_1 step by step.

OLS Problem In a simple linear regression model, we aim to predict y_i based on an independent variable x_i for each data point i. The model can be written as:

$$\hat{y}_i = b_0 + b_1 x_i$$

Where: - y_i is the actual value. - $\hat{y}_i$ is the predicted value. - b_0 is the intercept. - b_1 is the slope.

The goal is to minimize the **sum of squared residuals (errors)**, which can be expressed as:

$$S(b_0, b_1) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

This is the function we want to minimize with respect to b_0 and b_1.

Step 1: Derivative with respect to b_0

To minimize $S(b_0, b_1)$, we first take the partial derivative of S with respect to b_0:

$$\frac{\partial S}{\partial b_0} = \frac{\partial}{\partial b_0} \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

Using the chain rule:

$$\frac{\partial S}{\partial b_0} = \sum_{i=1}^{n} 2 (y_i - b_0 - b_1 x_i)(-1)$$

Simplifying:

$$\frac{\partial S}{\partial b_0} = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)$$

Setting the derivative equal to zero for minimization:

$$\sum_{i=1}^{n} (y_i - b_0 - b_1 x_i) = 0$$

Expanding the sum:

$$\sum_{i=1}^{n} y_i - n b_0 - b_1 \sum_{i=1}^{n} x_i = 0$$

Solving for b_0:

$$b_0 = \frac{\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i}{n}$$

This equation implies that b_0 is the difference between the average of y and the scaled average of x. In fact, this can be simplified by using the means $\acute{y}$ and $\acute{x}$:

$$b_0 = \acute{y} - b_1 \acute{x}$$

Step 2: Derivative with respect to b_1

Next, take the partial derivative of $S(b_0, b_1)$ with respect to b_1:

$$\frac{\partial S}{\partial b_1} = \frac{\partial}{\partial b_1} \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

Using the chain rule:

$$\frac{\partial S}{\partial b_1} = \sum_{i=1}^{n} 2(y_i - b_0 - b_1 x_i)(-x_i)$$

Simplifying:

$$\frac{\partial S}{\partial b_1} = -2 \sum_{i=1}^{n} x_i (y_i - b_0 - b_1 x_i)$$

Setting the derivative equal to zero for minimization:

$$\sum_{i=1}^{n} x_i (y_i - b_0 - b_1 x_i) = 0$$

Expanding the sum:

$$\sum_{i=1}^{n} x_i y_i - b_0 \sum_{i=1}^{n} x_i - b_1 \sum_{i=1}^{n} x_i^2 = 0$$

Substitute $b_0 = \acute{y} - b_1 \acute{x}$ into the equation:

$$\sum_{i=1}^{n} x_i y_i - (\acute{y} - b_1 \acute{x}) \sum_{i=1}^{n} x_i - b_1 \sum_{i=1}^{n} x_i^2 = 0$$

Rearrange the equation:

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

This is the formula for the slope b_1.

Final Results

- $b_1 = \frac{\sum_{i=1}^{n} (x_i - \acute{x})(y_i - \acute{y})}{\sum_{i=1}^{n} (x_i - \acute{x})^2}$ - $b_0 = \acute{y} - b_1 \acute{x}$

Let's break down the process of solving for b_1 step by step in detail, starting from the key equation where $b_0 = \acute{y} - b_1 \acute{x}$ is substituted into the equation.

Step 1: The Initial Equation

We previously derived this equation by taking the derivative of the sum of squared residuals with respect to b_1:

$$\sum_{i=1}^{n} x_i (y_i - b_0 - b_1 x_i) = 0$$

Now, substitute $b_0 = \acute{y} - b_1 \acute{x}$ into this equation.

Step 2: Substitution of b_0

The expression for b_0 is:

$$b_0 = \acute{y} - b_1 \acute{x}$$

Substitute this into the equation:

$$\sum_{i=1}^{n} x_i (y_i - (\acute{y} - b_1 \acute{x}) - b_1 x_i) = 0$$

Simplify the terms inside the parentheses:

$$\sum_{i=1}^{n} x_i (y_i - \acute{y} + b_1 \acute{x} - b_1 x_i) = 0$$

Step 3: Distribute x_i and Rearrange Terms

Distribute x_i inside the sum:

$$\sum_{i=1}^{n} (x_i y_i - x_i \acute{y} + b_1 x_i \acute{x} - b_1 x_i^2) = 0$$

Now group the terms involving b_1:

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \acute{y} + b_1 \left(\sum_{i=1}^{n} x_i \acute{x} - \sum_{i=1}^{n} x_i^2 \right) = 0$$

Step 4: Simplify Using Averages

We can simplify the sums involving the averages $\acute{x}$ and $\acute{y}$ by using the following facts:

1. $\sum_{i=1}^{n} x_i \acute{y} = \acute{y} \sum_{i=1}^{n} x_i$ 2. $\sum_{i=1}^{n} x_i \acute{x} = \acute{x} \sum_{i=1}^{n} x_i$

Substitute these into the equation:

$$\sum_{i=1}^{n} x_i y_i - \acute{y} \sum_{i=1}^{n} x_i + b_1 \left(\acute{x} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2 \right) = 0$$

Step 5: Solve for b_1

Now we isolate b_1. Rearrange the equation:

$$\sum_{i=1}^{n} x_i y_i - \acute{y} \sum_{i=1}^{n} x_i = b_1 \left(\sum_{i=1}^{n} x_i^2 - \acute{x} \sum_{i=1}^{n} x_i \right)$$

Solve for b_1:

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

Step 6: Further Simplify Using Deviations

We can express this equation in terms of deviations from the mean. Recall that:

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

Using these identities, we get the final form for b_1:

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

Conclusion: The Formula for the Slope

Thus, the slope b_1 in OLS regression is given by:

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

This is the formula that minimizes the sum of squared residuals and gives the best fit line for simple linear regression.