Derivation of Regression Coefficients Using the Least Squares Method

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

$$Y = \beta_0 + \beta_1 X + \epsilon$$

where:

- Y is the dependent variable.
- X is the independent variable.
- β_0 is the intercept of the regression line.
- β_1 is the slope of the regression line.
- ϵ is the error term.

Objective

The objective of the least squares method is to find the estimates of β_0 and β_1 that minimize the sum of squared residuals, defined as:

$$S = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 X_i))^2$$

where $\hat{Y}_i$ is the predicted value of Y_i.

Steps to Derive the Coefficients

1. Expand the Objective Function

We start by expanding the sum of squared residuals:

$$S = \sum_{i=1}^{n} (Y_i^2 - 2Y_i(\beta_0 + \beta_1 X_i) + (\beta_0 + \beta_1 X_i)^2)$$

2. Take Partial Derivatives

To find the values of β_0 and β_1 that minimize S, we take the partial derivatives of S with respect to β_0 and β_1 and set them to zero.

Partial Derivative with respect to β_0:

$$\frac{\partial S}{\partial \beta_0} = -2\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i) = 0$$

Simplifying this gives:

$$\sum_{i=1}^{n} Y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} X_i$$

Partial Derivative with respect to β_1:

$$\frac{\partial S}{\partial \beta_1} = -2\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)X_i = 0$$

Simplifying this gives:

$$\sum_{i=1}^{n} Y_i X_i = \beta_0 \sum_{i=1}^{n} X_i + \beta_1 \sum_{i=1}^{n} X_i^2$$

3. Solve the System of Equations

We now have a system of two equations:

1. $\sum_{i=1}^{n} Y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} X_i$
2. $\sum_{i=1}^{n} Y_i X_i = \beta_0 \sum_{i=1}^{n} X_i + \beta_1 \sum_{i=1}^{n} X_i^2$

Let's denote:

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

Substituting β_0 from the first equation into the second gives us:

$$\beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

This is the formula for the slope β_1.

4. Calculate the Intercept

Substituting β_1 back into the first equation allows us to solve for β_0:

$$\beta_0 = \bar{Y} - \beta_1\bar{X}$$

Final Formulas

The coefficients of the regression line can thus be expressed as:

$$\beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\beta_0 = \bar{Y} - \beta_1\bar{X}$$

These equations provide the least squares estimates of the regression coefficients β_0 and β_1.