Simple Linear Regression

Shahkar

July 11, 2026

Variables and Symbols

- X: Independent variable (predictor)
- Y: Dependent variable (response)
- m: Slope of the regression line
- b: Intercept of the regression line
- $\bar{x}$: Mean of X values
- $\bar{y}$: Mean of Y values
- $\hat{y}_i$: Predicted Y for x_i
- e_i: Residual for observation i
- $SS_{residual}$: Sum of squared residuals
- SS_{total}: Total sum of squares
- R^2: Coefficient of determination

What is Linear Regression?

- Models the linear relationship between X and Y.
- The model:

$$Y = mX + b + \epsilon \quad \text{where} \quad \epsilon = \text{error term.}$$

Least Squares Method

- Best-fit line minimizes the sum of squared residuals.
- Slope (m) and intercept (b):

$$m = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \quad b = \bar{y} - m\bar{x}.$$

Goodness of Fit: R^2

- R^2 shows how well the line explains the variation in Y.
- Defined as:

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}} \quad \text{where}$$

- $$SS_{residual} = \sum (y_i - \hat{y}_i)^2, \quad SS_{total} = \sum (y_i - \bar{y})^2.$$

Example: Slope and Intercept

- Given: $X = [1, 2, 3]$, $Y = [2, 4, 5]$
- $\bar{x} = 2$, $\bar{y} = 3.67$
- Slope:

$$m = \frac{(-1)(-1.67) + (0)(0.33) + (1)(1.33)}{(-1)^2 + 0^2 + 1^2} = \frac{3.0}{2} = 1.5.$$

- Intercept:

$$b = \bar{y} - m\bar{x} = 3.67 - (1.5)(2) = 0.67.$$

- So, $\hat{Y} = 1.5X + 0.67$.

Example: R^2 Calculation

- Predictions:

$$\hat{Y} = [2.17,\ 3.67,\ 5.17]$$

- Residuals:

$$e = [2 - 2.17,\ 4 - 3.67,\ 5 - 5.17] = [-0.17,\ 0.33,\ -0.17]$$

- $SS_{residual}$:

$$SS_{residual} = (-0.17)^2 + (0.33)^2 + (-0.17)^2 = 0.03 + 0.11 + 0.03 = 0.17.$$

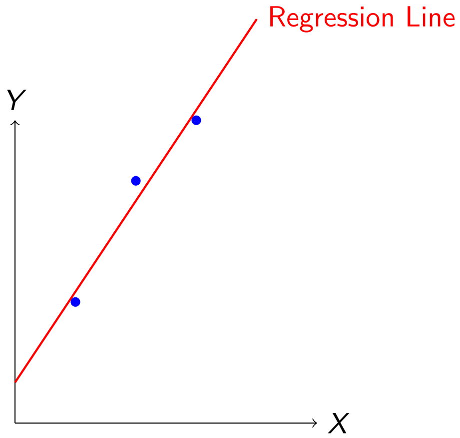
- SS_{total}:

$$SS_{total} = (2-3.67)^2 + (4-3.67)^2 + (5-3.67)^2 = 2.78 + 0.11 + 1.78 = 4.67.$$

- R^2:

$$R^2 = 1 - \frac{0.17}{4.67} = 0.964.$$

Sample Regression Plot



- Blue dots: actual data
- Red line: fitted line $\hat{Y} = 1.5X + 0.67$