Regression Formulas

📘 1. Simple Linear Regression

Model Equation

$$y = a + bx$$

Where:

- y: Dependent variable
- x: Independent variable
- a: Intercept
- b: Slope

Slope (b)

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

Intercept (a)

$$a = \bar{y} - b\bar{x}$$

Predicted Value

$$\hat{y}_i = a + bx_i$$

Residual

$$e_i = y_i - \hat{y}_i$$

📕 2. Coefficient of Determination (R^2)

$$R^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}}$$

Where:

- SSR = Regression sum of squares
- SSE = Error sum of squares
- SST = Total sum of squares

📗 3. Multiple Linear Regression

Model Equation

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

- y: Response variable
- $x_1, x_2, \ldots, x_p$: Predictors
- β_0: Intercept
- $\beta_1, \ldots, \beta_p$: Coefficients
- ϵ: Error term

Matrix Form (for computation)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Solution via Least Squares

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

📕 4. Standard Error of Estimate

$$SE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - 2}}$$

📒 5. Assumptions of Linear Regression

- Linearity
- Independence of errors
- Homoscedasticity (constant variance)
- Normality of residuals
- No multicollinearity (in multiple regression)
