# Simple Linear Regression

## Simple Linear Regression:

Simple Linear Regression is a fundamental statistical method used to analyze the relationship between two continuous variables. In Simple Linear Regression, we aim to fit a straight line to the data that best represents the relationship between the independent variable (X) and the dependent variable (Y).
This line is represented by the equation:
$Y=\beta0+\beta1\times X+\epsilon$

Where:
Y is the dependent variable.
X is the independent variable.
β0 is the intercept (the value of Y when X=0). It's the value of Y when there is no influence from the independent variable.
β1 is the slope (the change in Y for a one-unit change in X). It quantifies the strength and direction of the relationship between the variables.
ϵ represents the error term, which captures the difference between the observed and predicted values of Y.

## Assumptions:

1. Linearity: The relationship between X and Y is linear.
2. Independence: Observations are independent of each other.
3. Homoscedasticity: The variance of the residuals is constant across all values of X.
4. Normality: The residuals follow a normal distribution.
5. No Multicollinearity: The independent variable X is not highly correlated with other variables.

## Model Fitting:

The goal in Simple Linear Regression is to estimate the coefficients β0 and β1 that minimize the sum of squared differences between the observed and predicted values of Y. This is typically done using the method of least squares.

## Evaluation:

Common metrics for evaluating the performance of a Simple Linear Regression model include the coefficient of determination (R-square), R-square helps us see how much of the changes in the thing we're trying to predict (Y-values) can be explained by the changes in what we're using to predict it (X-values). Range for coefficient of determination is between 0 to 1.

Below are formulas to calculate coefficient of determination (R-square):

### Formula 1: Using the correlation coefficient

Formula 1:

$$R^2 = (r)^2$$

Where $r$ = Pearson correlation coefficient

### Formula 2: Using the regression outputs

Formula 2:

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}$$

Where:

- RSS = sum of squared residuals
- TSS = total sum of squares

## Scenario to use Simple Linear Regression:

Here are a few scenarios where simple linear regression might be applied:
Predictive Modeling: Suppose you want to predict a person's weight based on their height. You can use simple linear regression to model the relationship between height (independent variable) and weight (dependent variable) and make predictions for new individuals.

**Sales Forecasting:** In retail or sales industries, you might want to understand how changes in advertising spending affect sales revenue. Simple linear regression can help you analyze this relationship and forecast future sales based on advertising expenditure.

**Real Estate**: In real estate, you might want to predict housing prices based on features like square footage, number of bedrooms, or location. Simple linear regression can be used to model the relationship between these features and the sale price of houses.

**Customer Satisfaction Analysis**: Suppose you want to analyze how the time spent waiting affects customer satisfaction ratings. Simple linear regression can help you determine if there's a significant relationship between wait time (independent variable) and customer satisfaction scores (dependent variable).

**Employee Performance Evaluation:** In a workplace setting, you might want to understand how the number of hours spent training employees affects their performance ratings. Simple linear regression can help you assess this relationship and make decisions about training investments.

**Crop Yield Prediction:** In agriculture, you might want to predict crop yields based on factors like rainfall, temperature, and soil nutrients. Simple linear regression can be used to model the relationship between these factors and crop yields, helping farmers make informed decisions about planting and harvest schedules.

Below is the file link for code of Simple Linear Regression:
Simple Linear Regression: