Lesson Documentation: Introduction to Simple Linear Regression

Lesson Title:

Understanding the Basic Principles of Simple Linear Regression

---

Lesson Overview

This lesson introduces students to the basic theory of **Simple Linear Regression** — a statistical method used to describe and model the relationship between two variables. It helps us determine whether and how one variable (the independent variable, X) influences another variable (the dependent variable, Y).

Regression also allows us to **predict** future or unobserved values of Y based on known values of X.

Learning Objectives

By the end of this lesson, students should be able to:

1. Explain what regression means and its main objectives.
2. Identify **dependent** and **independent** variables in a regression model.
3. Write and interpret the **simple linear regression equation**.
4. Understand that regression produces a **straight line** when plotted on a graph.

Concept Explanation

1. What is Simple Linear Regression?

Simple Linear Regression is a method used to:

- **Establish a relationship** between two variables — showing how one variable changes in response to the other.

- **Forecast new observations** — using the established relationship to predict unknown values of the dependent variable (Y).

In essence, regression helps answer two key questions:

1. *Does one variable affect the other?*

2. *If yes, can we predict one variable when we know the other?*

---

2. Roles of Variables in Regression Models

There are two main variables involved:

- **Independent Variable (X):** The factor we think influences the outcome.

- **Dependent Variable (Y):** The outcome we are trying to explain or predict.

In this relationship, the value of **Y** largely depends on the value of **X**.

Example:
In a study of *Income (Y)* and *Expense (X)*, we can say that a person's **income** might depend on or be related to their **expenses**, or vice versa.

---

3. The Linear Equation

The relationship between X and Y in simple linear regression follows a **straight-line equation**:

$$[Y = a + bX]$$

or equivalently,

$$[Y = mx + b]$$

Where:

- (Y) = Dependent variable (predicted outcome)
- (X) = Independent variable (input or predictor)
- (a) = Intercept (the value of Y when X = 0)
- (b) = Slope or coefficient of X (the rate at which Y changes for every unit increase in X)

Because the relationship is linear, plotting Y against X gives a **straight line**, called the **line of best fit**.

Example Question

Using the dataset below for 20 individuals, analyze the relationship between **Income (Y)** and **Expense (X)**.

Person	Income (Y)	Expense (X)
1	5,766.00	3,178.00
2	2,956.00	2,048.00
3	1,833.00	1,240.00
4	9,265.00	5,031.00
5	7,637.00	4,318.00
6	9,483.00	5,303.00
7	5,138.00	2,997.00
8	10,364.00	5,481.00
9	7,783.00	3,879.00
10	2,019.00	1,198.00
11	7,684.00	3,957.00
12	3,025.00	2,110.00
13	6,498.00	3,248.00

14	9,836.00	5,412.00
15	2,448.00	1,625.00
16	8,312.00	4,650.00
17	5,749.00	3,320.00
18	10,597.00	5,709.00
19	4,236.00	2,701.00
20	7,195.00	3,824.00

Task:

1. Plot a scatter diagram of *Income (Y)* against *Expense (X)*.
2. Draw the line of best fit.
3. Write the regression equation of Income on Expense in the form (Y = a + bX).
4. Interpret the slope and intercept in your own words.

Key Takeaways

- Simple linear regression studies the relationship between **two variables**.
- The **dependent variable (Y)** depends on the **independent variable (X)**.
- The model is based on a **linear equation**, producing a **straight line** when graphed.
- Regression serves two main purposes:
    1. Establishing relationships
    2. Forecasting new values