SNS COLLEGE OF TECHNOLOGY

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

Department of Computer Science and Engineering

# Course Code & Title : 23ADO201 - Data Science Fundamentals

III YEAR / VI SEMESTER - EEE
Unit 2 - DESCRIPTIVE ANALYTICS

Topic : Regression, Regression Line and Multiple Regression Equations .

Mr.K.Karthikeyan AP /CSE

Empathy

Regression Analysis Overview

Multiple Regression Equations

Regression Line

Simple Linear Regression

Regression Analysis

1/26/2026   REGRESSION, REGRESSION LINE AND MULTIPLE REGRESSION EQUATIONS   / 23ADO201 - DATA SCIENCE FUNDAMENTALS / Mr.K.KARTHIKEYAN/CSE/SNSCT

2/16
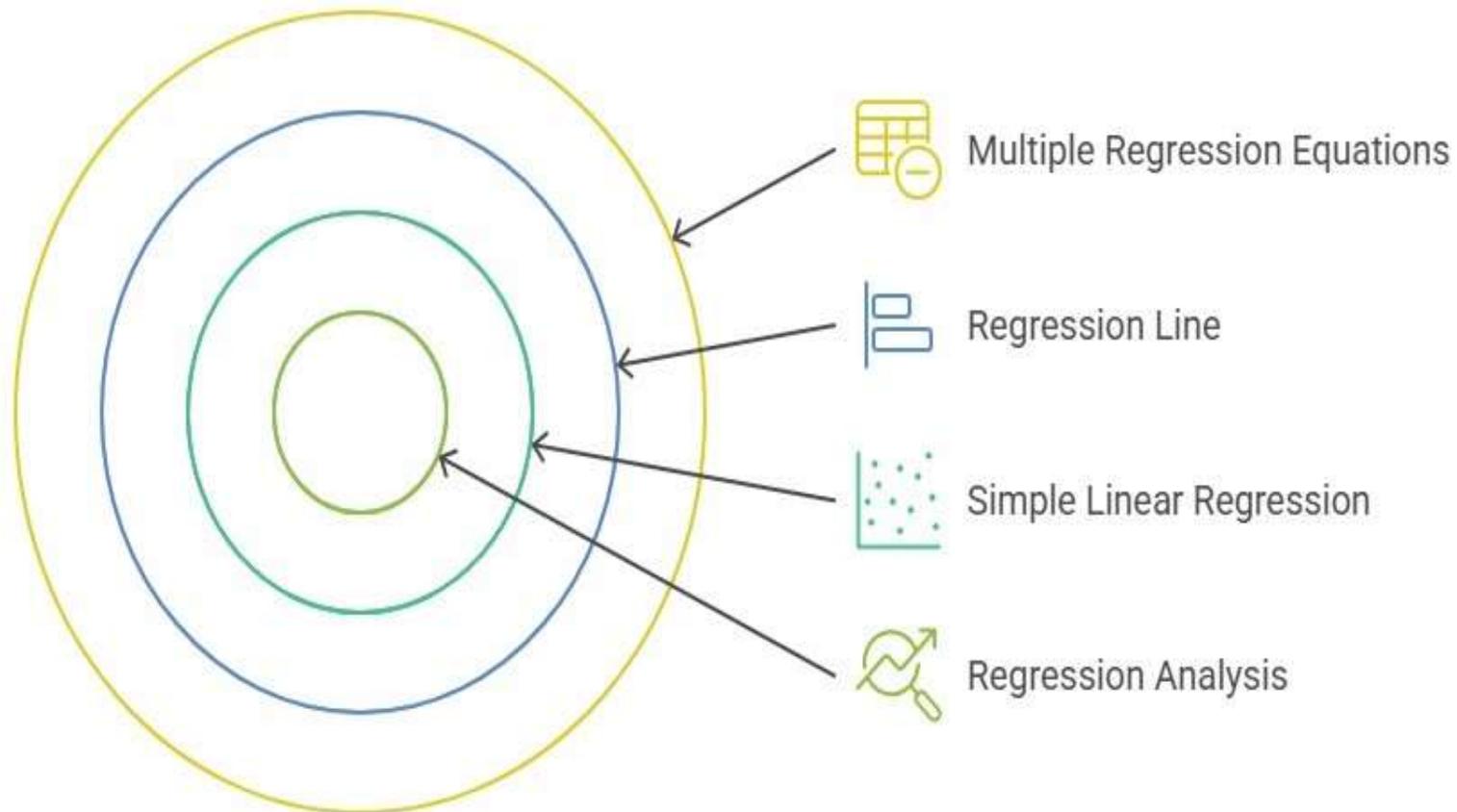
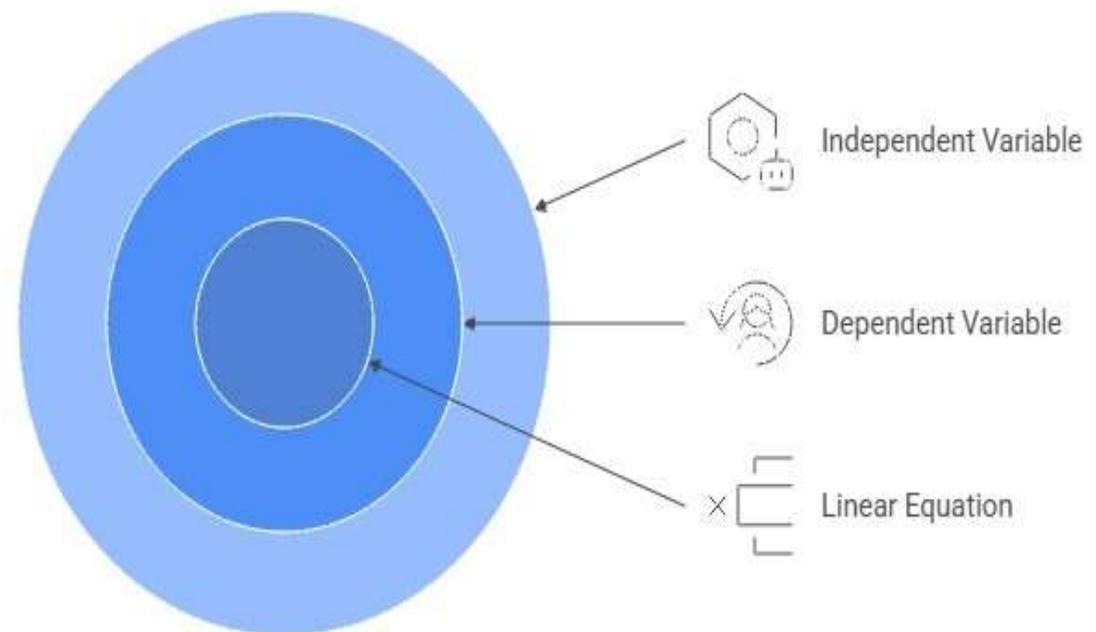## Simple Linear Regression

Simple linear regression is a statistical method used to model the relationship between two variables: an independent variable (predictor) and a dependent variable (response). The goal is to find a linear equation that best describes how the dependent variable changes as the independent variable changes.

Simple Linear Regression

Independent Variable

Dependent Variable

Linear Equation

REGRESSION, REGRESSION LINE AND MULTIPLE REGRESSION EQUATIONS   / 23ADO201 - DATA SCIENCE FUNDAMENTALS / Mr.K.KARTHIKEYAN/CSE/SNSCT

1.The Linear Equation

The linear equation in simple linear regression is represented as:

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

Where:

- Y is the dependent variable.
- X is the independent variable.
- β_0 is the y-intercept (the value of Y when X is 0).
- β_1 is the slope (the change in Y for a one-unit change in X).
- ε is the error term (representing the difference between the observed and predicted values).

2. Estimating the Coefficients

The coefficients β_0 and β_1 are estimated using the method of least squares, which minimizes the sum of the squared differences between the observed and predicted values of Y. The formulas for calculating these coefficients are:
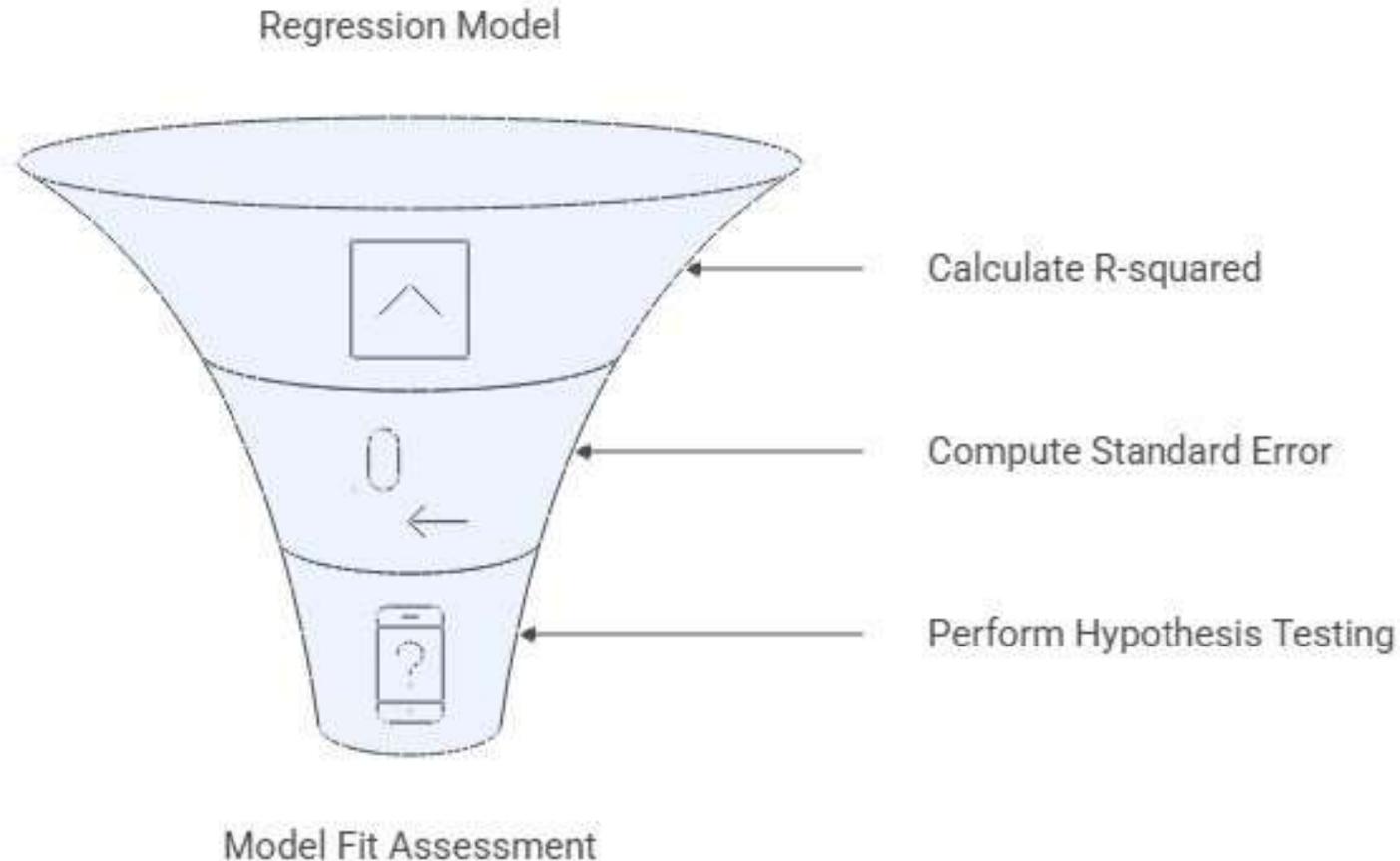
$$\beta_1 = \Sigma[(X_i - \bar{X})(Y_i - \bar{Y})] / \Sigma[(X_i - \bar{X})^2]$$

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

Where:

- X_i and Y_i are the individual data points.
- $\bar{X}$ and $\bar{Y}$ are the means of X and Y, respectively.

1/26/2026

REGRESSION, REGRESSION LINE AND MULTIPLE REGRESSION EQUATIONS / 23ADO201 - DATA SCIENCE FUNDAMENTALS / Mr.K.KARTHIKEYAN/CSE/SNSCT

4/16

Evaluating Regression Model Fit

Regression Model

Calculate R-squared

Compute Standard Error

Perform Hypothesis Testing

Model Fit Assessment

REGRESSION, REGRESSION LINE AND MULTIPLE REGRESSION EQUATIONS / 23ADO201 - DATA SCIENCE FUNDAMENTALS / Mr.K.KARTHIKEYAN/CSE/SNSCT

Regression Line

Pros
- Visual representation
- Minimizes error
- Indicates direction
- Shows magnitude

Cons
- Assumes linearity
- Sensitive to outliers
- Limited interpretation

Made with Napkin

1/26/2026

REGRESSION, REGRESSION LINE AND MULTIPLE REGRESSION EQUATIONS / 23ADO201 - DATA SCIENCE FUNDAMENTALS / Mr.K.KARTHIKEYAN/CSE/SNSCT

6/16

MULTIPLE REGRESSION – ACTIVITIES

5. Many Causes, One Effect 🔍

Time: 15 minutes

Mode: Groups

Example:

Marks depend on:

- Study hours (X_1)

- Attendance (X_2)

- Sleep hours (X_3)

Steps:

- Identify variables

- Write a multiple regression form:

- $Y = a + b_1X_1 + b_2X_2 + b_3X_3$

- Discuss effect of each variable.

✅ **Outcome:** Understanding multiple regression structure

1/26/2026   REGRESSION, REGRESSION LINE AND MULTIPLE REGRESSION EQUATIONS / 23ADO201 - DATA SCIENCE FUNDAMENTALS / Mr.K.KARTHIKEYAN/CSE/SNSCT

7/16

Multiple Regression Equations

Multiple regression is an extension of simple linear regression that allows for the modeling of the relationship between a dependent variable and two or more independent variables.
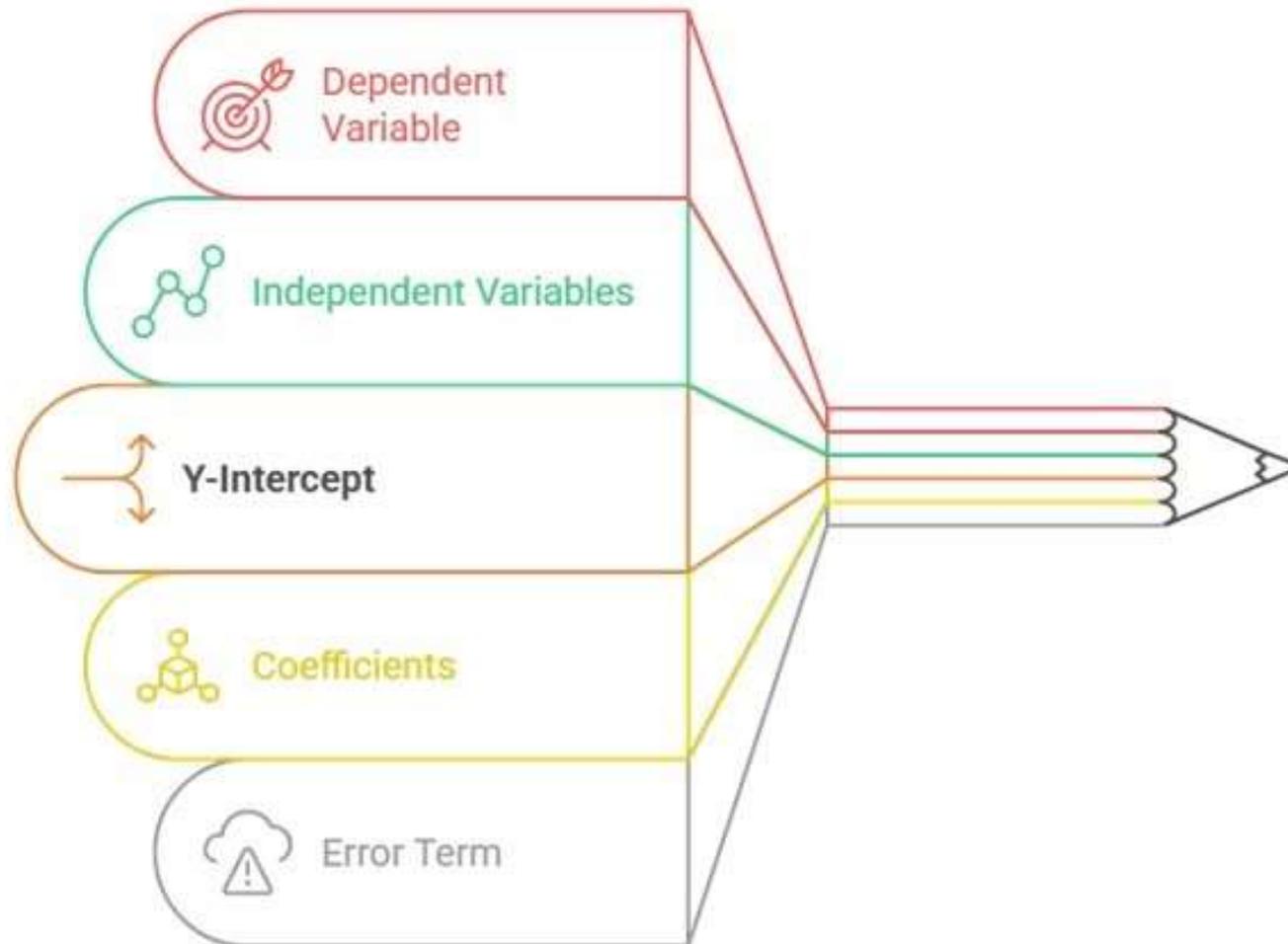
The Multiple Regression Equation

The multiple regression equation is represented as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta p X p + \varepsilon$$

Where:
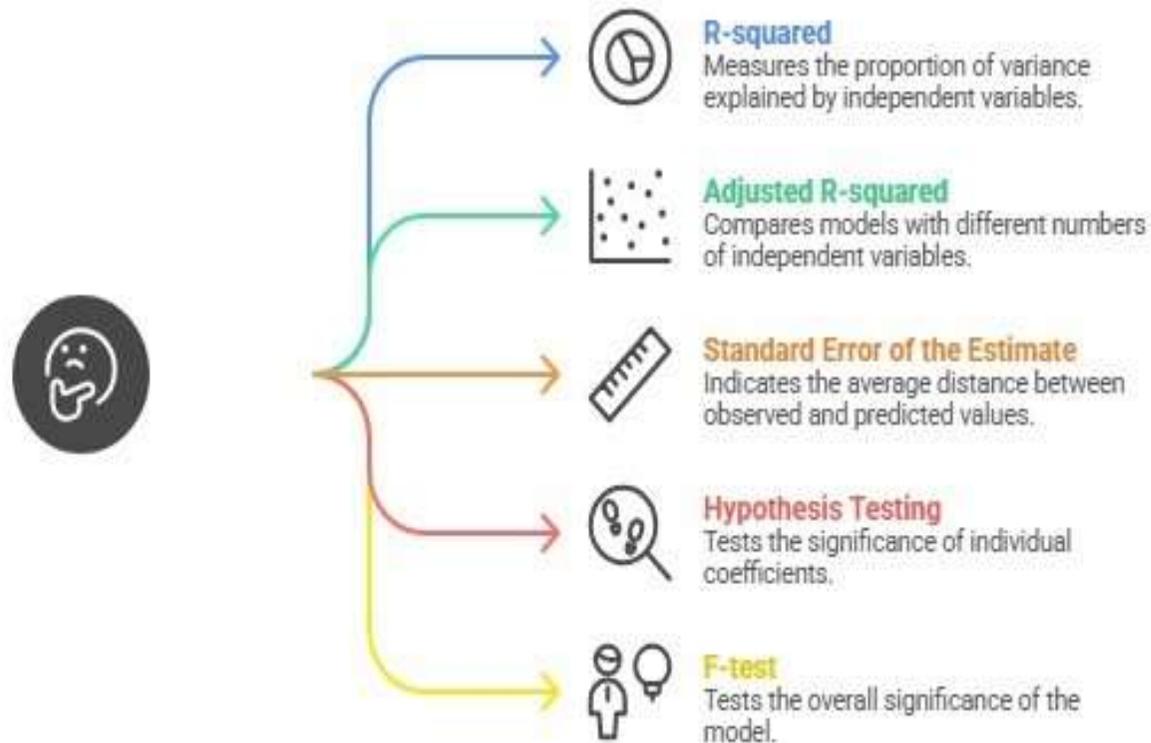- Y is the dependent variable.
- X_1, X_2, ..., Xp are the independent variables.
- β_0 is the y-intercept.
- β_1, β_2, ..., βp are the coefficients for each independent variable.
- ε is the error term.

1/26/2026

REGRESSION, REGRESSION LINE AND MULTIPLE REGRESSION EQUATIONS   / 23ADO201 - DATA SCIENCE FUNDAMENTALS / Mr.K.KARTHIKEYAN/CSE/SNSCT

8/16

Building a Predictive Equation

- Dependent Variable
- Independent Variables
- Y-Intercept
- Coefficients
- Error Term

1/26/2026

REGRESSION, REGRESSION LINE AND MULTIPLE REGRESSION EQUATIONS   /
23ADO201 - DATA SCIENCE FUNDAMENTALS / Mr.K.KARTHIKEYAN/CSE/SNSCT

9/16

How to evaluate the goodness of fit of a multiple regression model?

R-squared
Measures the proportion of variance explained by independent variables.

Adjusted R-squared
Compares models with different numbers of independent variables.

Standard Error of the Estimate
Indicates the average distance between observed and predicted values.

Hypothesis Testing
Tests the significance of individual coefficients.

F-test
Tests the overall significance of the model.

REGRESSION, REGRESSION LINE AND MULTIPLE REGRESSION EQUATIONS / 23ADO201 - DATA SCIENCE FUNDAMENTALS / Mr.K.KARTHIKEYAN/CSE/SNSCT
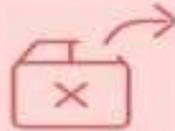
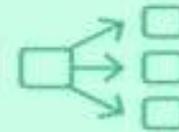# Which variable selection method should be used in multiple regression?

Forward Selection
Adds variables one by one based on significance, ensuring statistical relevance.

Backward Elimination
Removes insignificant variables, simplifying the model.
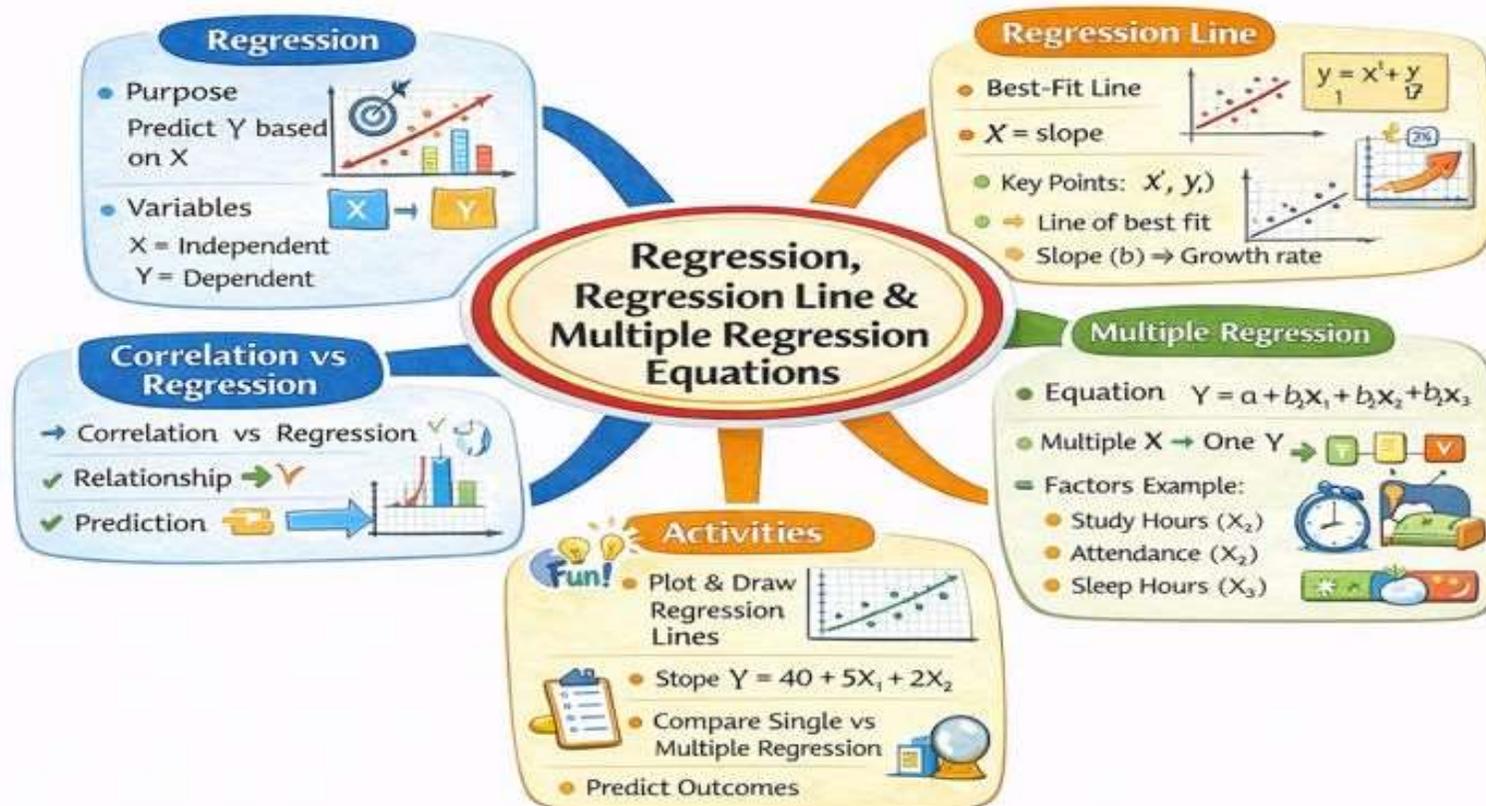
Stepwise Regression
Combines both methods for a balanced approach.

1/26/2026

REGRESSION, REGRESSION LINE AND MULTIPLE REGRESSION EQUATIONS / 23ADO201 - DATA SCIENCE FUNDAMENTALS / Mr.K.KARTHIKEYAN/CSE/SNSCT

11/16

Which field can benefit from multiple regression analysis?

Economics

Model relationships between economic indicators and consumer behavior

Finance

Predict stock prices based on financial ratios and market trends

Marketing

Analyze the impact of advertising campaigns on sales

Healthcare

Identify risk factors for diseases

Social Sciences

Study factors influencing social behavior

REGRESSION, REGRESSION LINE AND MULTIPLE REGRESSION EQUATIONS / 23ADO201 - DATA SCIENCE FUNDAMENTALS / Mr.K.KARTHIKEYAN/CSE/SNSCT

MIND MAP

1/26/2026
REGRESSION, REGRESSION LINE AND MULTIPLE REGRESSION EQUATIONS   / 23ADO201 - DATA SCIENCE FUNDAMENTALS / Mr.K.KARTHIKEYAN/CSE/SNSCT

13/16

ASSESSMENT

REGRESSION, REGRESSION LINE & MULTIPLE REGRESSION – MCQs

🔷 **A. REGRESSION – BASICS**

1. Regression analysis is used to:

A. Measure variability only

B. Study cause–effect relationship between variables

C. Classify data

D. Find frequency

✅ **Answer:** B

2. In simple linear regression, the number of independent variables is:

A. One

B. Two

C. Three

D. More than two

✅ **Answer:** A

1/26/2026

REGRESSION, REGRESSION LINE AND MULTIPLE REGRESSION EQUATIONS / 23ADO201 - DATA SCIENCE FUNDAMENTALS / Mr.K.KARTHIKEYAN/CSE/SNSCT

14/16

REFERENCE BOOKS

1.Allen B. Downey, "Think Stats: Exploratory Data Analysis in Python", Green Tea Press, 2014.
2.Sanjeev J. Wagh, Manisha S. Bhende, Anuradha D. Thakare, "Fundamentals of Data Science", CRC Press, 2022.
3.Chirag Shah, "A Hands-On Introduction to Data Science", Cambridge University Press, 2020.
4.Vineet Raina, Srinath Krishnamurthy, "Building an Effective Data Science Practice: A Framework to Bootstrap and Manage a Successful Data Science Practice", A press, 2021.

1/26/2026

REGRESSION, REGRESSION LINE AND MULTIPLE REGRESSION EQUATIONS / 23ADO201 - DATA SCIENCE FUNDAMENTALS / Mr.K.KARTHIKEYAN/CSE/SNSCT

15/16

REGRESSION, REGRESSION LINE AND MULTIPLE REGRESSION EQUATIONS   / 23ADO201 - DATA SCIENCE FUNDAMENTALS / Mr.K.KARTHIKEYAN/CSE/SNSCT