



# SNS COLLEGE OF ENGINEERING



Kurumbapalayam(Po), Coimbatore – 641 107

Accredited by NAAC-UGC with 'A' Grade

Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

## Department of Information Technology

Course Name – 23ADT202 Fundamental of Data  
science and Analytics

II Year / IV Semester

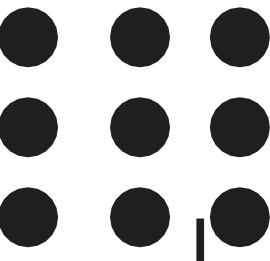
Unit 2 – Descriptive Analytics

Multiple regression equations





# What is Multiple Regression?



- Multiple regression predicts a dependent variable ( $Y$ ) using multiple independent variables ( $X_1, X_2, \dots$ ).
- It extends simple linear regression to handle more than one independent variable.
- The goal is to model the relationship between the variables and make predictions.
- It's widely used in fields like business, healthcare, and real estate.



# Multiple Regression Equation?



The general form of a multiple regression equation is:

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_nX_n + \varepsilon$$

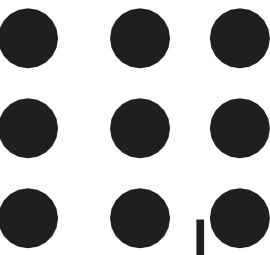
Y is the dependent variable;  $X_1, X_2, \dots$  are independent variables.

$\beta_0$  is the intercept, and  $\beta_1, \beta_2, \dots$  are coefficients for each independent variable.

$\varepsilon$  represents the error term or residuals.



# Assumptions of Multiple Regression



- **Linearity:** The relationship between  $Y$  and  $X$  is linear.
- **Independence:** Observations should be independent of each other.
- **Homoscedasticity:** Residuals have constant variance.
- **Normality:** Residuals follow a normal distribution.
- **No Multicollinearity:** Independent variables should not be highly correlated.



# Steps to Perform Multiple Regression



**Prepare the Data:** Clean the dataset, identify variables.

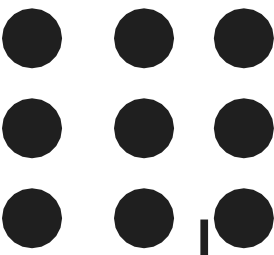
**Fit the Model:** Estimate coefficients using least squares method.

**Interpret the Coefficients:** Understand how each independent variable influences Y.

**Evaluate the Model:** Use metrics like  $R^2$ , Adjusted  $R^2$ , and MSE.



# Steps to Perform Multiple Regression



**Dataset:** Predict house prices ( $Y$ ) based on size ( $X_1$ ) and number of bedrooms ( $X_2$ ).

Sample Data:

- Size ( $X_1$ ): 1500, 2000, 2500, 3000 sqft.
- Bedrooms ( $X_2$ ): 3, 4, 4, 5.
- Price ( $Y$ ): \$300,000, \$400,000, \$450,000, \$500,000.
- Objective: Build a regression model to predict prices.



# Python Implementation (Code)



```
import numpy as np
import pandas as pd
from sklearn.linear_model import LinearRegression
# Dataset
data = {"Size": [1500, 2000, 2500, 3000], "Bedrooms": [3, 4, 4, 5], "Price": [300000, 400000, 450000, 500000]}
df = pd.DataFrame(data)
# Independent and dependent variables
X = df[["Size", "Bedrooms"]]
Y = df["Price"]
# Create and fit the model
model = LinearRegression()
model.fit(X, Y)
# Coefficients and intercept
print("Intercept:", model.intercept_)
print("Coefficients:", model.coef_)
```

This code uses Python to create and fit a multiple regression model.





# Interpreting Results



- Intercept ( $\beta_0$ ):** The predicted price when Size and Bedrooms are 0 ( $Y = 87,500$ ).
- **Coefficients ( $\beta_1, \beta_2$ ):**
    - Size ( $\beta_1$ ): For each extra sqft, price increases by \$100.
    - Bedrooms ( $\beta_2$ ): For each extra bedroom, price increases by \$25,000.
  - **Predictions:** The model can predict house prices based on Size and Bedrooms.





# Model Evaluation Metrics



**$R^2$** : The proportion of variance in Y explained by X variables.

**Adjusted  $R^2$** : Adjusts  $R^2$  for the number of predictors.

**MSE**: Mean Squared Error measures prediction accuracy.

These metrics help assess how well the model fits the data.



# Application & Exercise



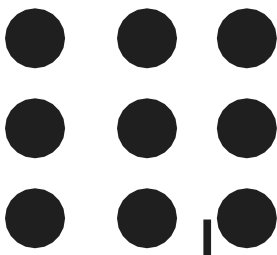
Applications: Multiple regression is used in business, healthcare, and real estate.

Exercise:

Calculate the regression equation for a given dataset.

Predict revenue for  $X_1 = 2500$ ,  $X_2 = 18$ .

By applying the steps, you can make predictions and evaluate models in various domains.



**THANK YOU**