

# **SNS COLLEGE OF TECHNOLOGY**

**An Autonomous Institution  
Coimbatore-35**

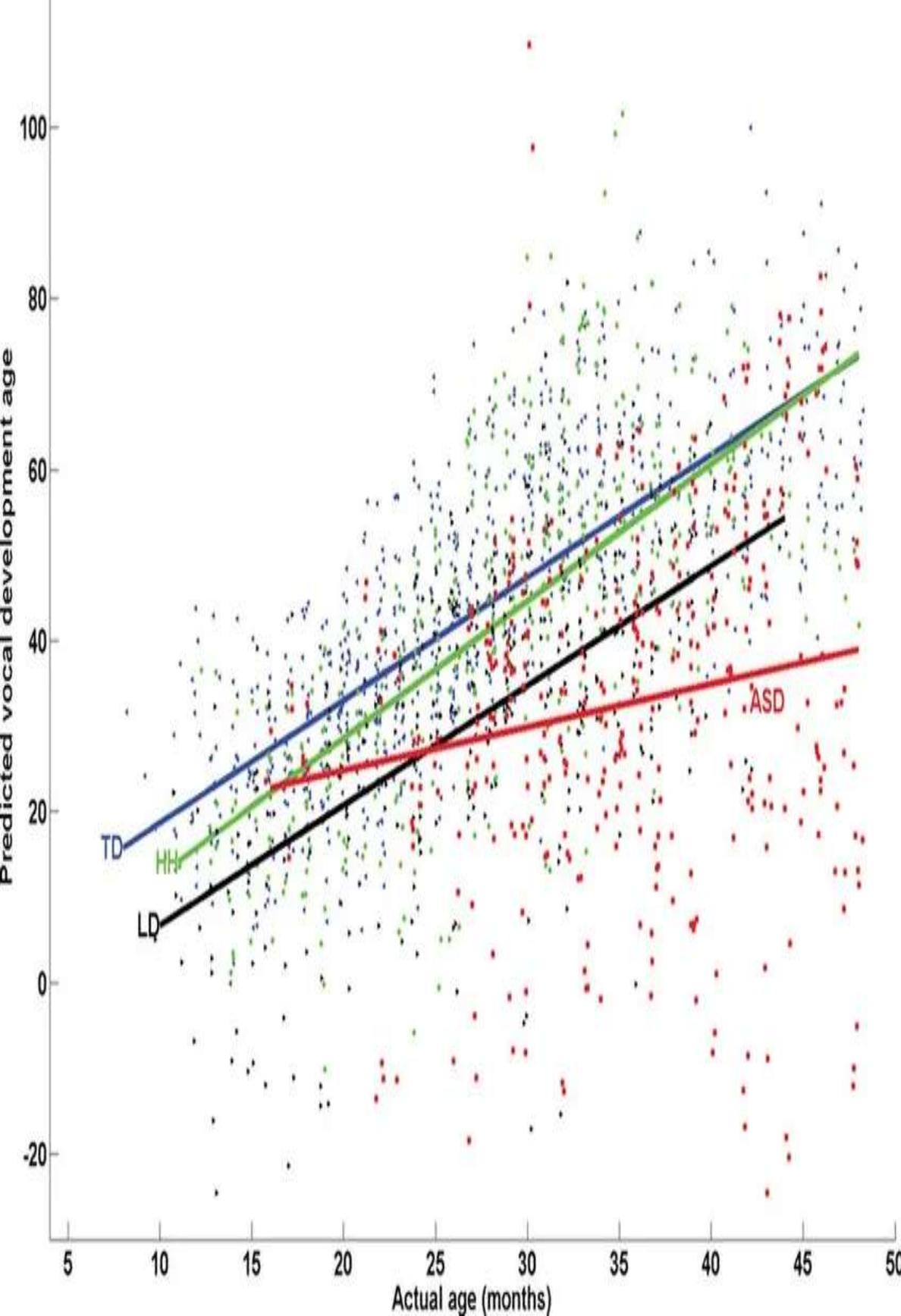


## **DEPARTMENT OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE**

### **23ADT202 – FUNDAMENTALS OF DATA SCIENCE AND ANALYTICS**

**II YEAR IV SEM**

#### **UNIT II – Multiple regression equations**



## EMPATHY:

### Multiple linear regression

- Multiple linear regression is an extension of linear regression, which allows a response variable,  $y$  to be modelled as a linear function of two or more predictor variables.
- In a multiple regression model, two or more independent variables, i.e. predictors are involved in the model.
- The simple linear regression model and the multiple regression model assume that the dependent variable is continuous.

# Difference between Simple and Multiple Regression

## DEFINE:

Sr. No.	Simple regression	Multiple regression
1.	One dependent variable Y predicted from one independent variable X.	One dependent variable Y predicted from a set of independent variable ( $X_1, X_2, \dots, X_k$ ).
2.	One regression coefficient.	One regression coefficient for each independent variables.
3.	$r^2$ : Proportion of variation in dependent variable Y predictable from X.	$R^2$ : Proportion of variation in dependent variable Y predictable by set of independent variables (X's).

## IDEATE:

### 1. Simple Linear Regression (SLR)

Used when there is **one independent variable** and **one dependent variable**.

**Equation:**

$$y = \beta_0 + \beta_1 x + \varepsilon$$

**Where:**

- $y \rightarrow$  Dependent variable (output)
- $x \rightarrow$  Independent variable (input)
- $\beta_0 \rightarrow$  Intercept (value of  $y$  when  $x = 0$ )
- $\beta_1 \rightarrow$  Slope (change in  $y$  for one unit change in  $x$ )
- $\varepsilon \rightarrow$  Error term (residual)

**Estimated form (using sample data):**

$$\hat{y} = b_0 + b_1 x$$

**PROTOTYPING:****2. Multiple Linear Regression (MLR)**

Used when there are **two or more independent variables**.

**Equation:**

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

**Where:**

- $y \rightarrow$  Dependent variable
- $x_1, x_2, \dots, x_p \rightarrow$  Independent variables
- $\beta_0 \rightarrow$  Intercept
- $\beta_1, \beta_2, \dots, \beta_p \rightarrow$  Regression coefficients
- $\varepsilon \rightarrow$  Error term

**Estimated form:**

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p$$

**TESTING:**

- Multiple regression is a type of regression where the dependent variable shows a linear relationship with two or more independent variables.
- It can also be non-linear, where the dependent and independent variables do not follow a straight line.

## Assumptions of Multiple Linear Regression

Multiple linear regression is based on the following assumptions:

- 1. A linear relationship between the dependent and independent variables**
- 2. The independent variables are not highly correlated with each other**
- 3. The variance of the residuals is constant**
- 4. Independence of observation**
- 5. Multivariate normality**