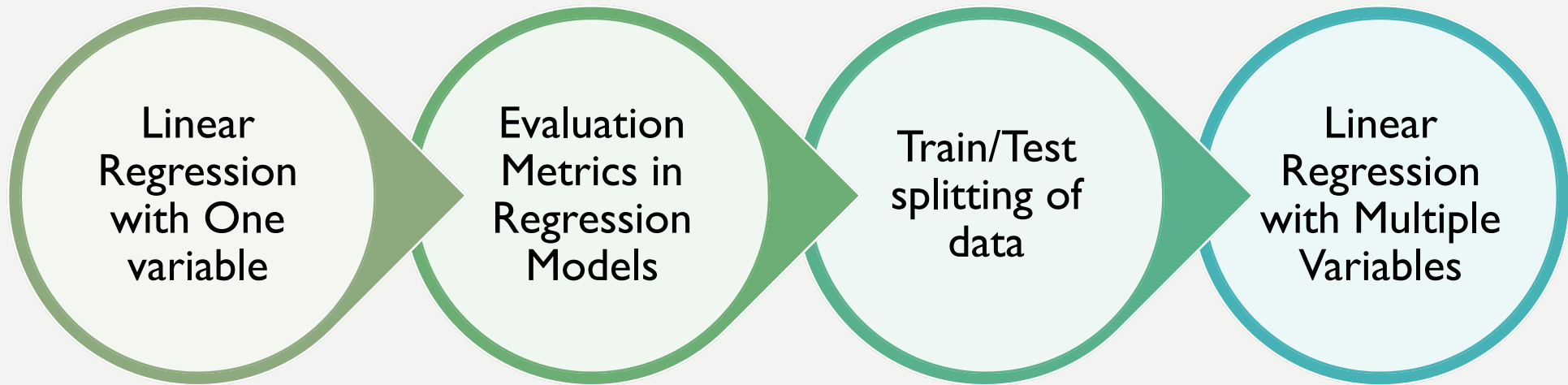




# Linear Regression using Machine Learning

# DAY2 AGENDA



# MACHINE LEARNING TYPES

## Supervised Learning

- Makes machine Learn explicitly
- Data with clearly defined output is given
- Direct feedback is given
- Predicts outcome/future
- Resolves classification and regression problems



## Unsupervised Learning

- Machine understands the data (Identifies patterns/structures)
- Evaluation is qualitative or indirect
- Does not predict/find anything specific



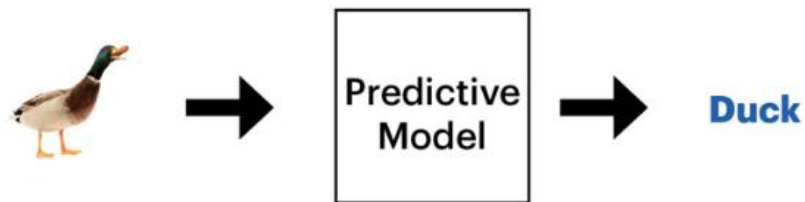
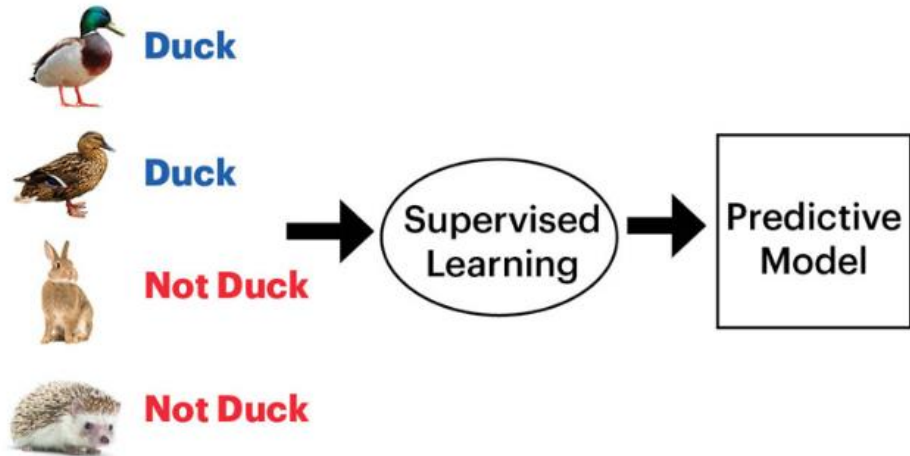
## Reinforcement Learning

- An approach to AI
- Reward based learning
- Learning form +ve & +ve reinforcement
- Machine Learns how to act in a certain environment
- To maximize rewards

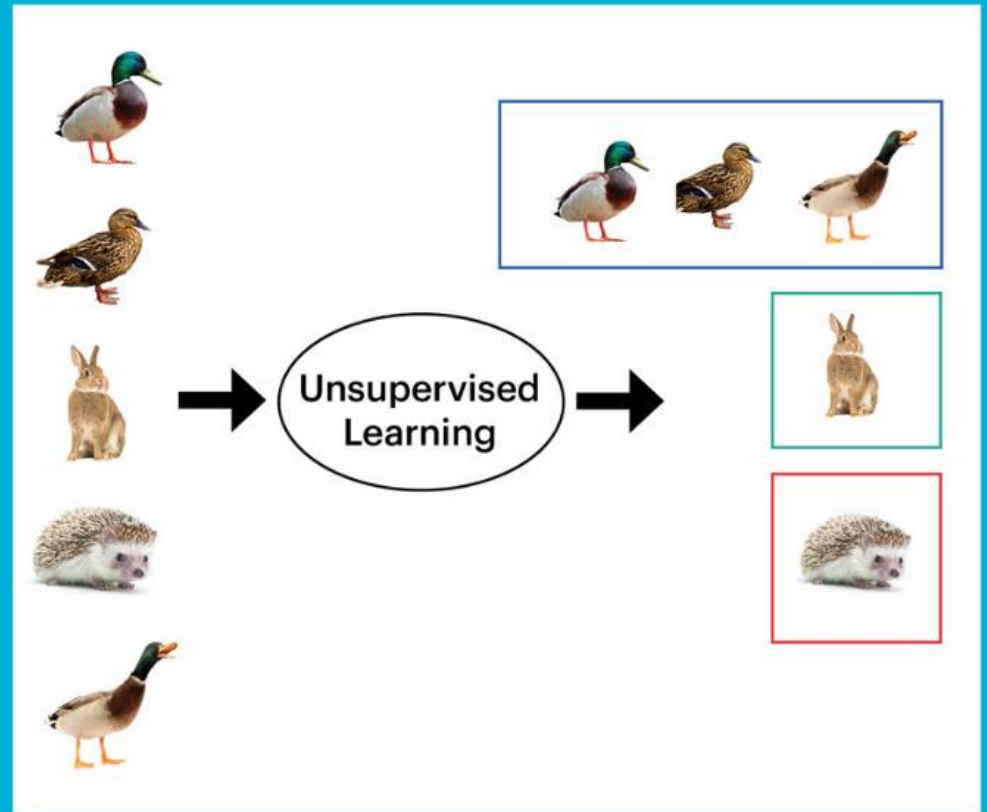


# SUPERVISED VS UNSUPERVISED

## Supervised Learning (Classification Algorithm)

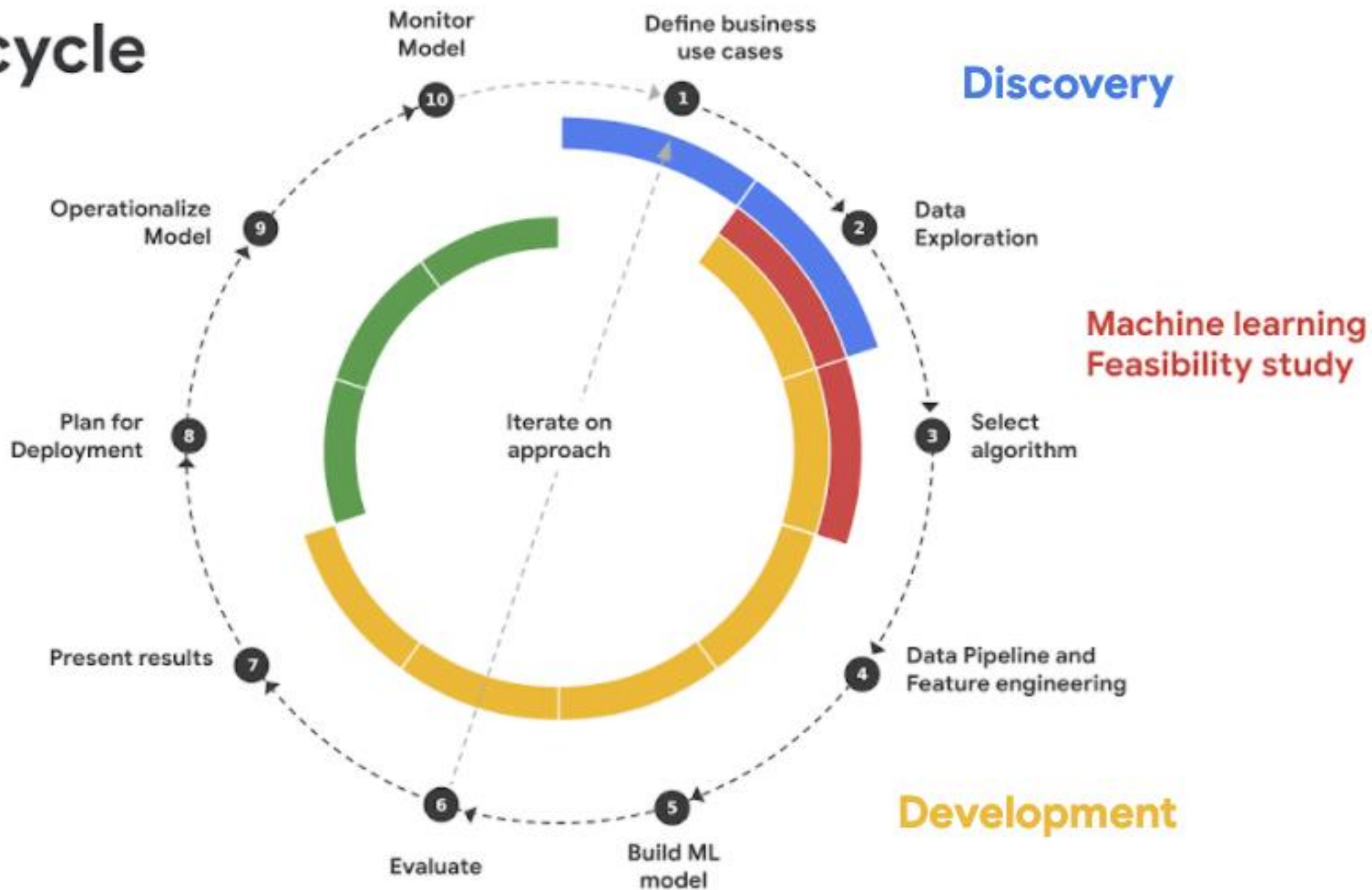


## Unsupervised Learning (Clustering Algorithm)

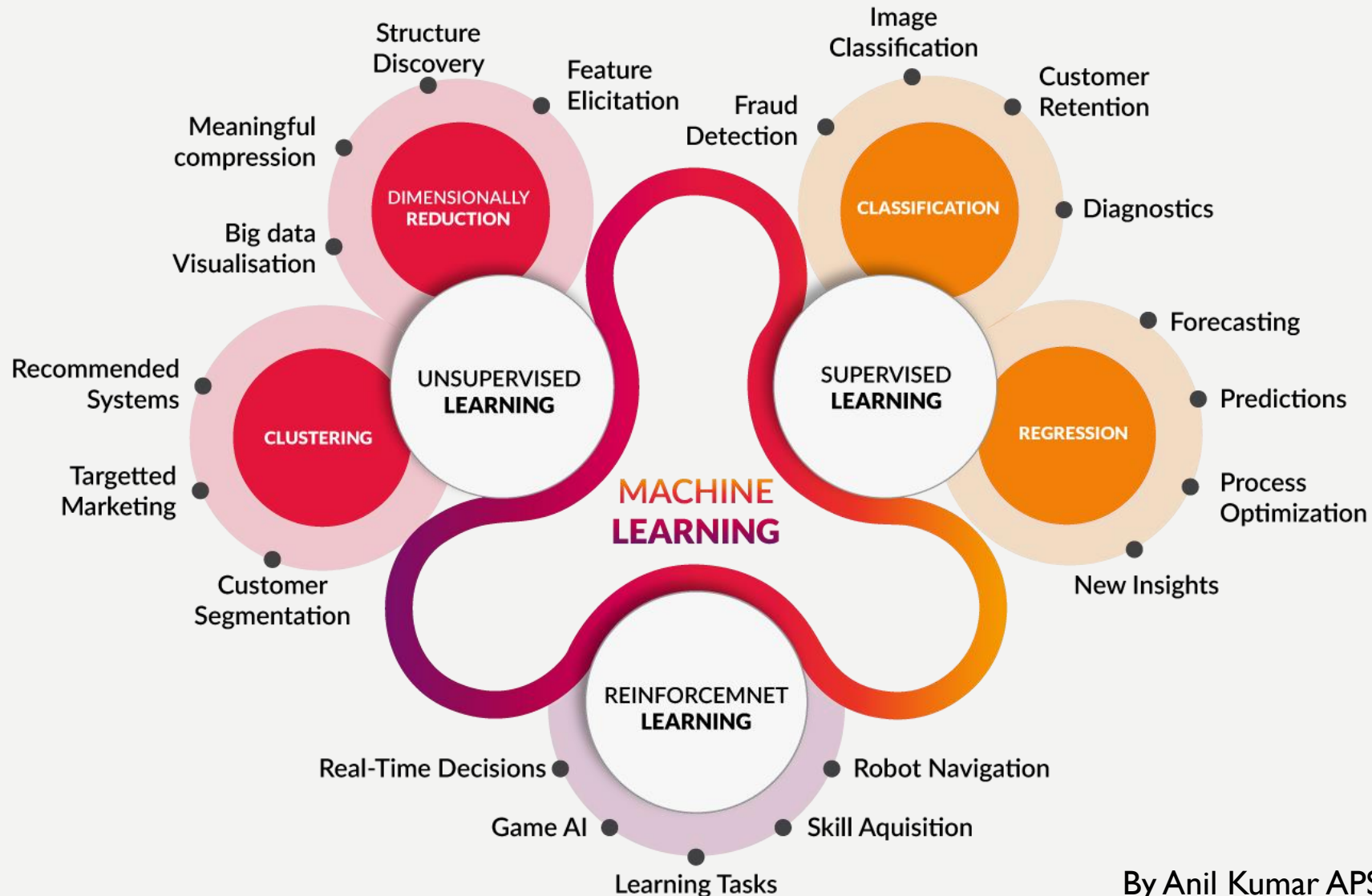


# ML Lifecycle

Deployment



# MACHINE LEARNING CATEGORIES



# MACHINE LEARNING ALGORITHMS

## SUPERVISED

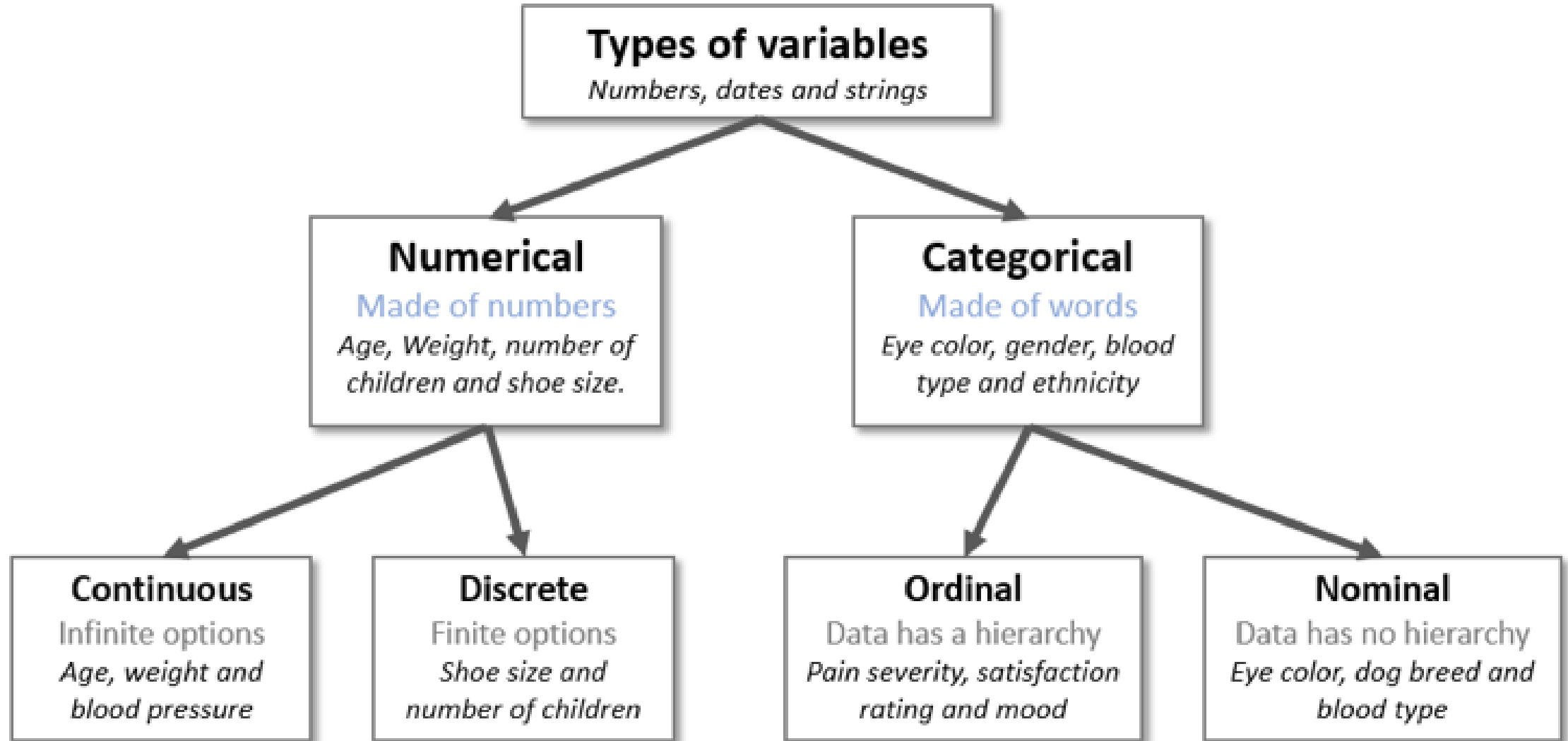
### Regression

- Linear Regression
  - Simple Linear Regression
  - Multi Linear Regression
- Polynomial Regression
  - Polynomial Regression
  - Multi Polynomial Regression

### Classification

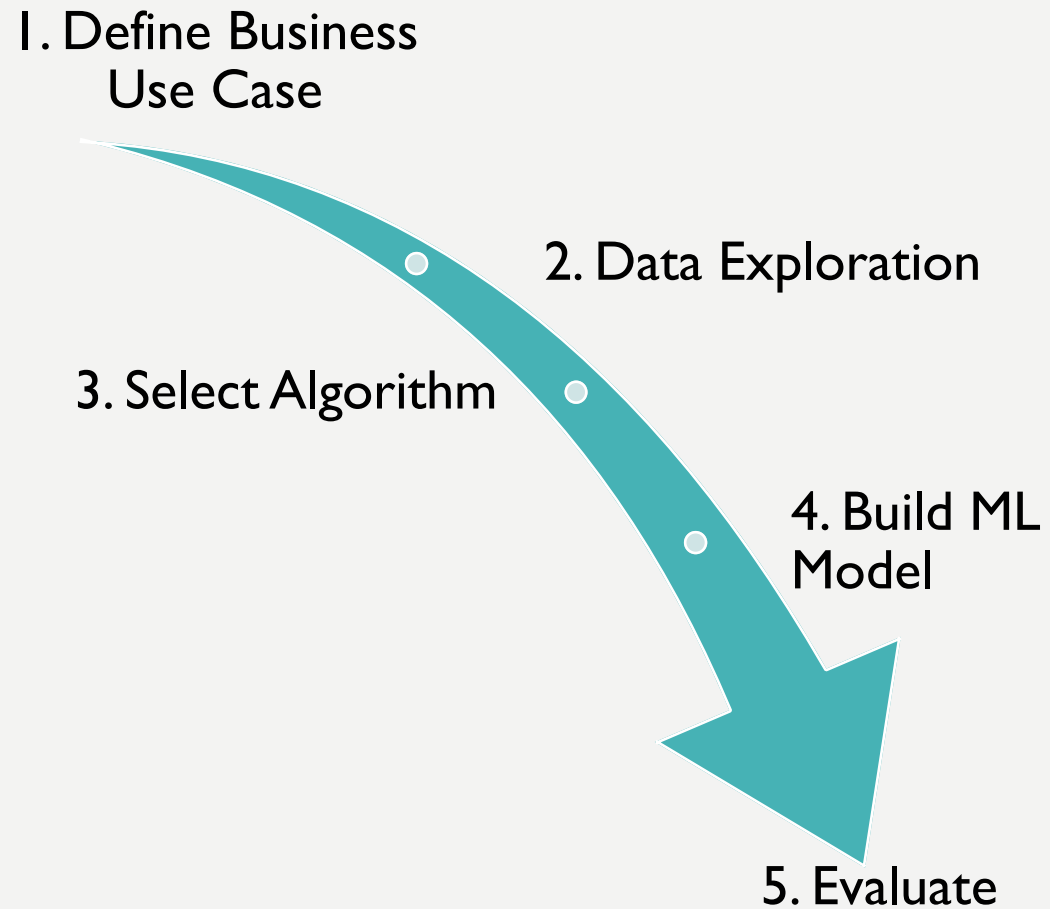
- Linear Classifiers
  - Logistic Regression
- K - Nearest Neighbors
- Decision Trees
- Random Forest
- Support Vector Machines

# TYPES OF STATISTICAL DATA





# ML MODEL DEVELOPMENT LIFE CYCLE



# FEATURES / ATTRIBUTES

- **Features (aka attributes)** are used to train an **ML system**.

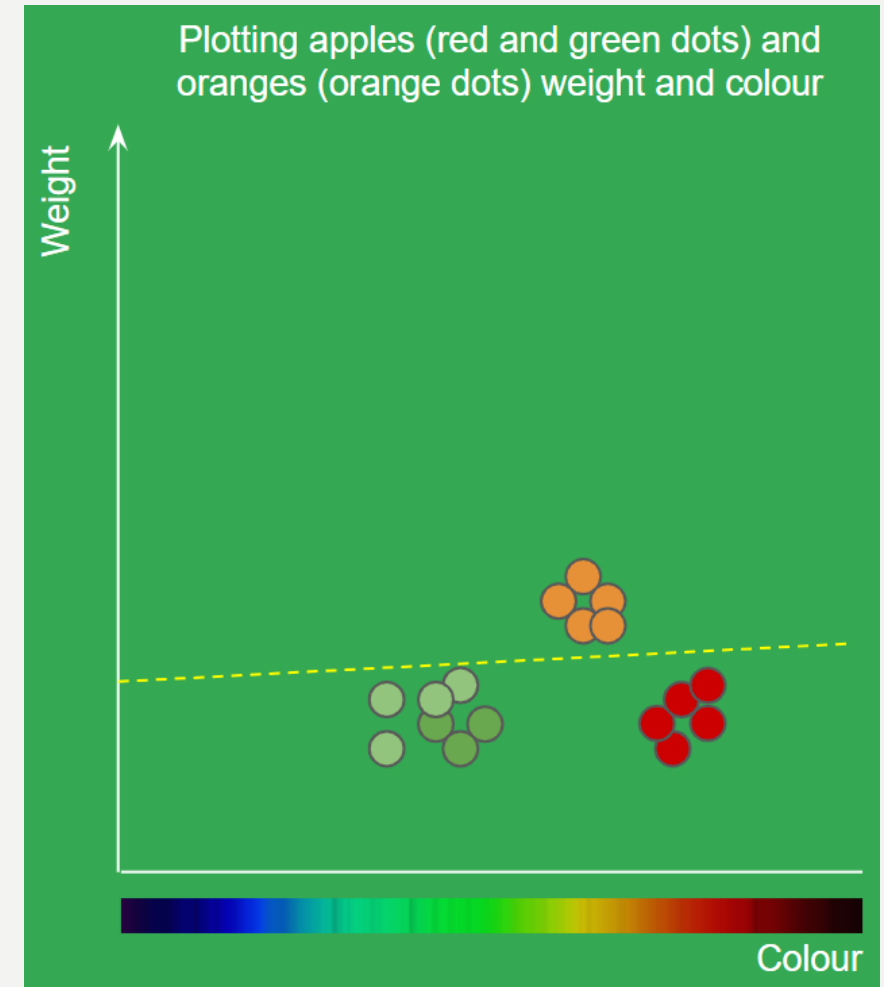
They are the properties of the things you are trying to learn about.



# FEATURES / ATTRIBUTES

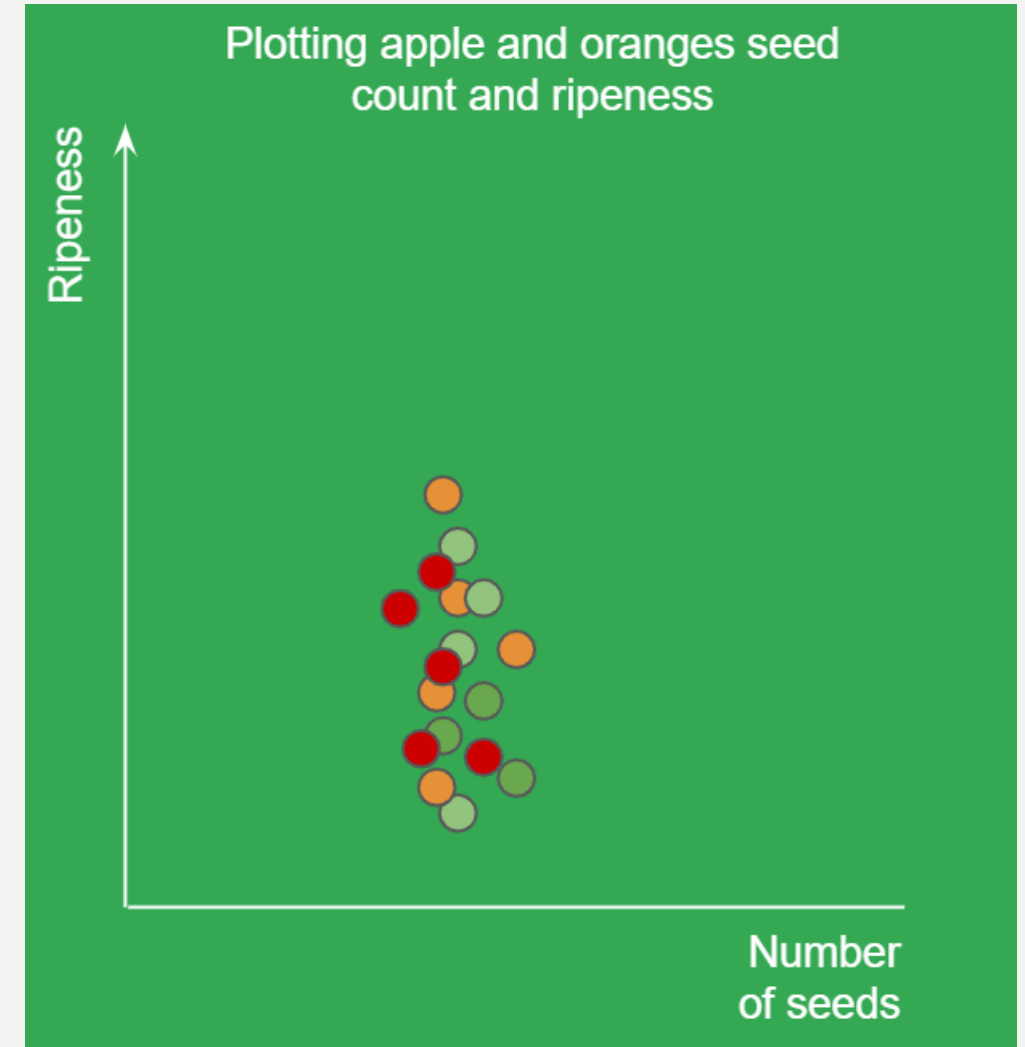
Taking fruit as an example. Features of a fruit might be weight and color. 2 features, would mean there are 2 dimensions. A 2D system may be plotted on a graph if features are represented in a numerical way.

In the plot on the right, the ML system can learn to split the data up with a line to separate apples from oranges. This **can now be used to make future classifications when we plot new points the system has not seen** (anything above is orange, below is apple)



# FEATURES / ATTRIBUTES

- **Choosing useful features can have a big impact on the quality of the ML system.** Some features may not be useful enough to separate the data points.
- In this example we take bad features of fruits(ripeness and seed count) that do not allow us to learn any distinguishing factors for the fruit.

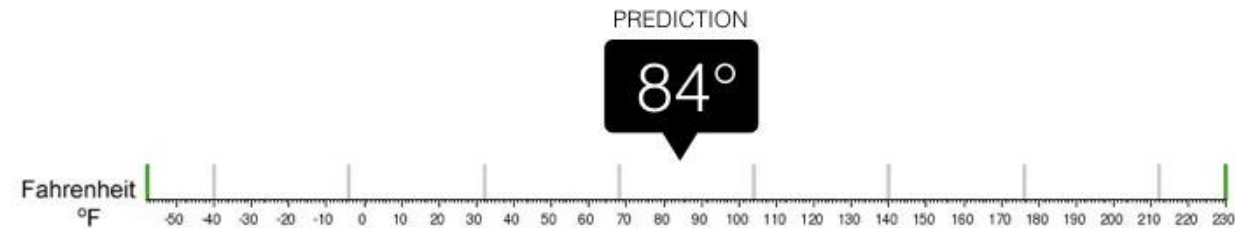


# REGRESSION VS CLASSIFICATION



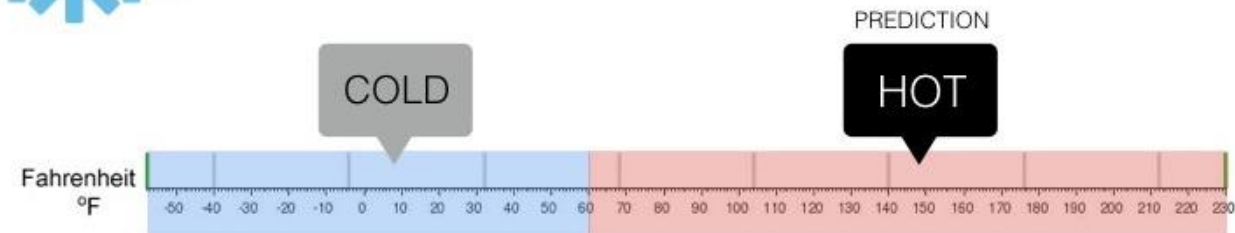
## Regression

What is the temperature going to be tomorrow?



## Classification

Will it be Cold or Hot tomorrow?





# Linear Regression in Machine Learning

By Anil Kumar  
APSSDC

# What is Regression?

- Function: a mathematical relationship enabling us to predict what values of one variable ( $Y$ ) correspond to given values of another variable ( $X$ ).
- $Y$ : is referred to as the **dependent variable**, the **response variable** or the **predicted variable**.
- $X$ : is referred to as **the independent variable**, the **explanatory variable** or the **predictor variable**.

Thus Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent and independent variable.

- ▶  $Y = MX + C$
- ▶  $\text{Temp} = M(\text{Humidity}) + C$



# Example

- ▶ Finding relationship between the features and target
- ▶ **Humidity, moisture, light**, → Temperature

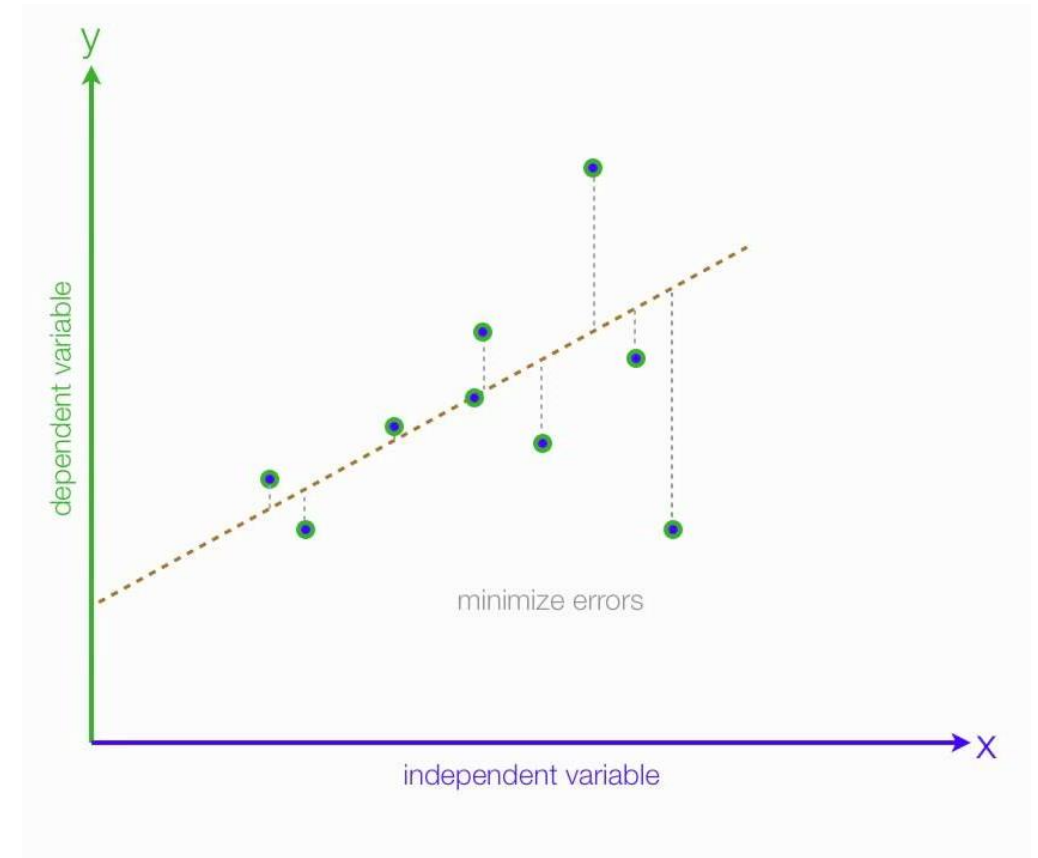
# Regression

- ▶ **Linear Regression**
  - ▶ **Linear Regression with one variable**
  - ▶ **Linear Regression with multiple variable**
- ▶ Non-Linear Regression/Polynomial Regression
  - ▶ Non-Linear Regression with one variable
  - ▶ Non-Linear Regression with multiple variables
- ▶ SGD
- ▶ Ridge
- ▶ Lasso
- ▶ Elastic Net

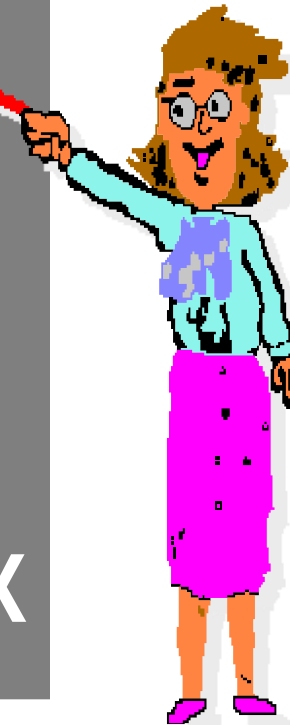
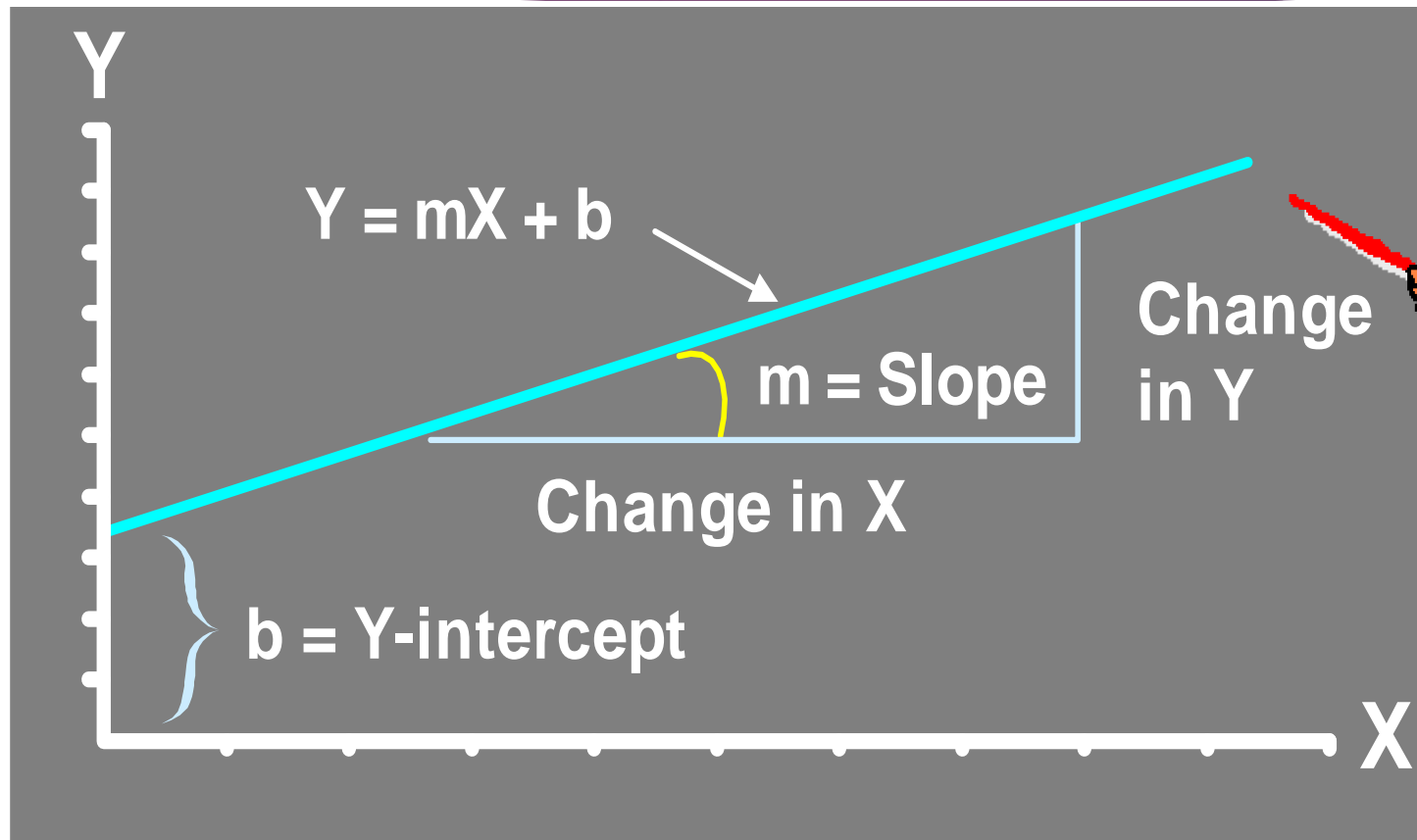
# Contd.

A typical Linear Regression model can be represented in the form :

$y = b_1x + b_0$  where  $b_1$  is slope and  $b_0$  is the intercept.



# Linear Equations



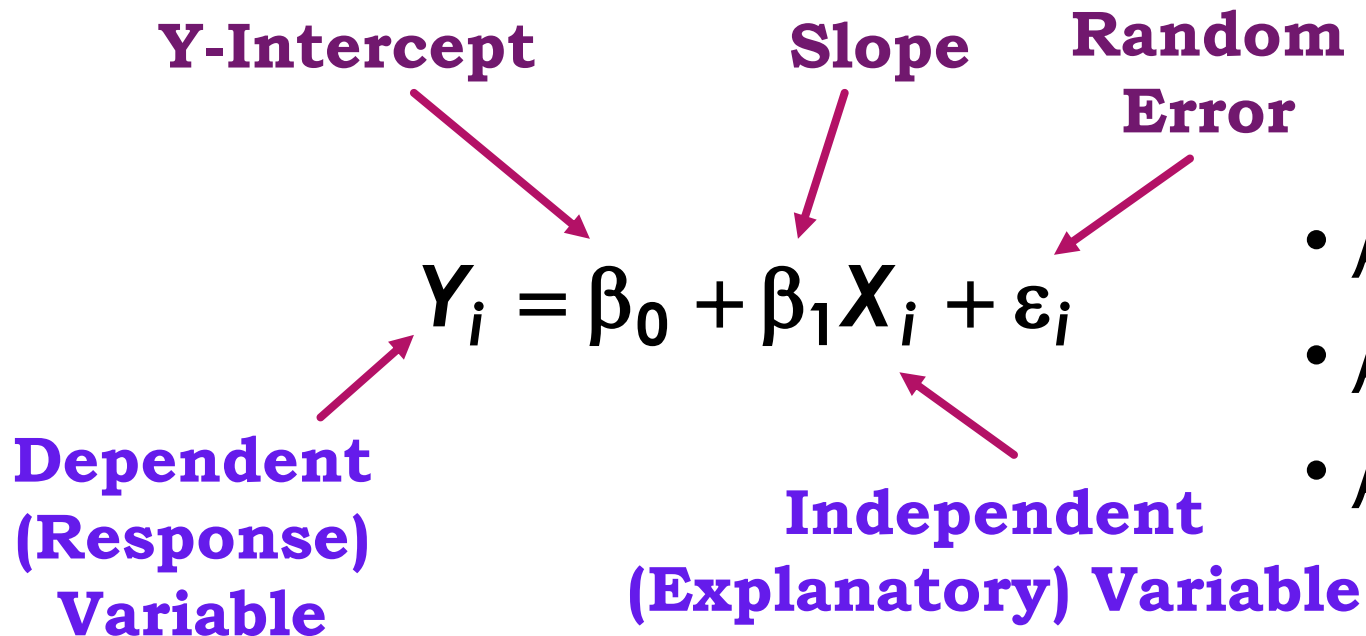
# Linear Regression Model

**Relationship Between Variables Is a Linear Function**

**Y-Intercept**      **Slope**      **Random Error**

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

**Dependent (Response) Variable**      **Independent (Explanatory) Variable**

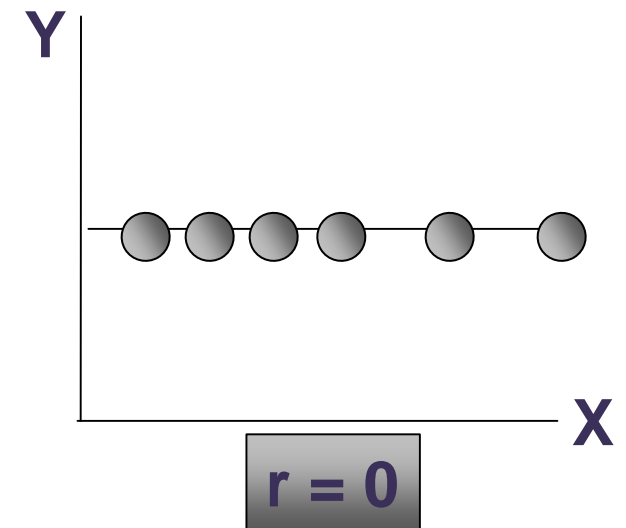
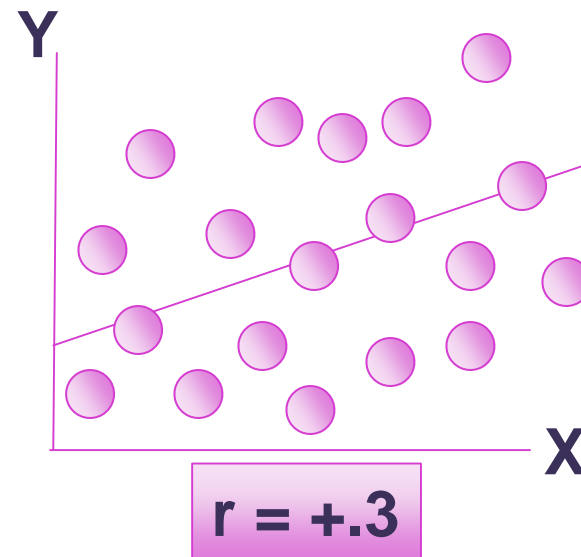
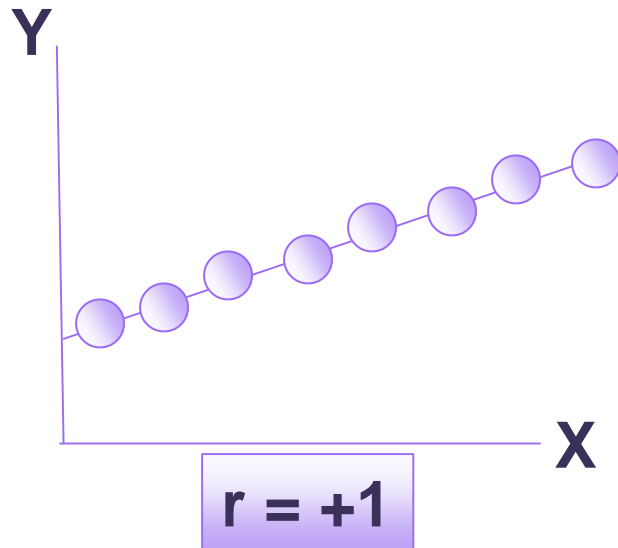
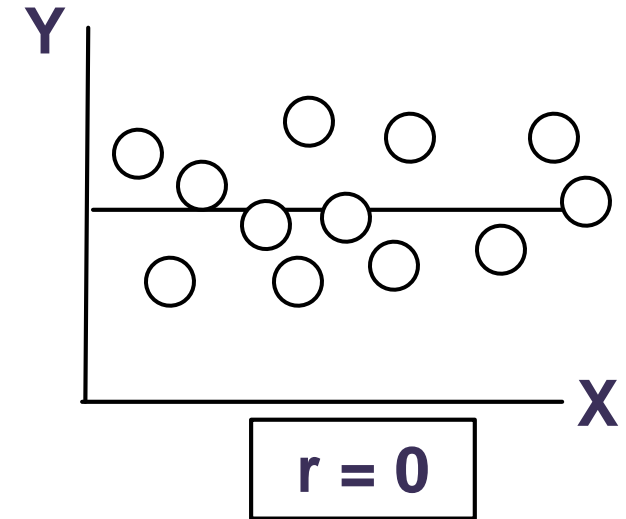
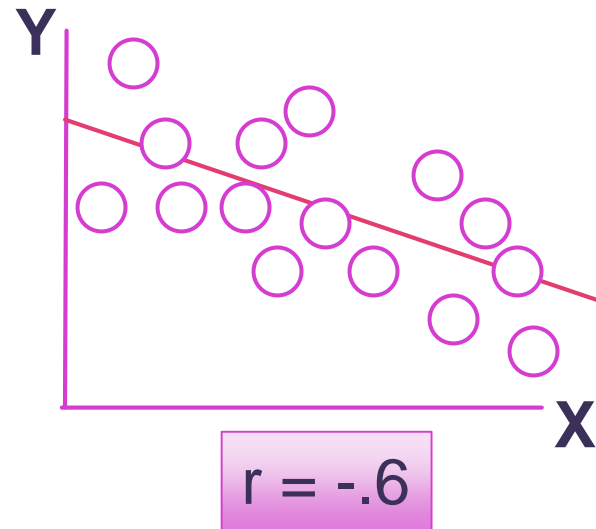
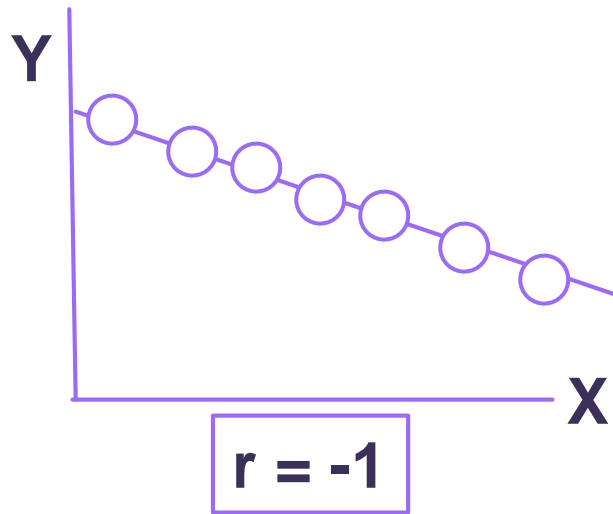


- $\beta_1 > 0 \Rightarrow$  Positive Association
- $\beta_1 < 0 \Rightarrow$  Negative Association
- $\beta_1 = 0 \Rightarrow$  No Association

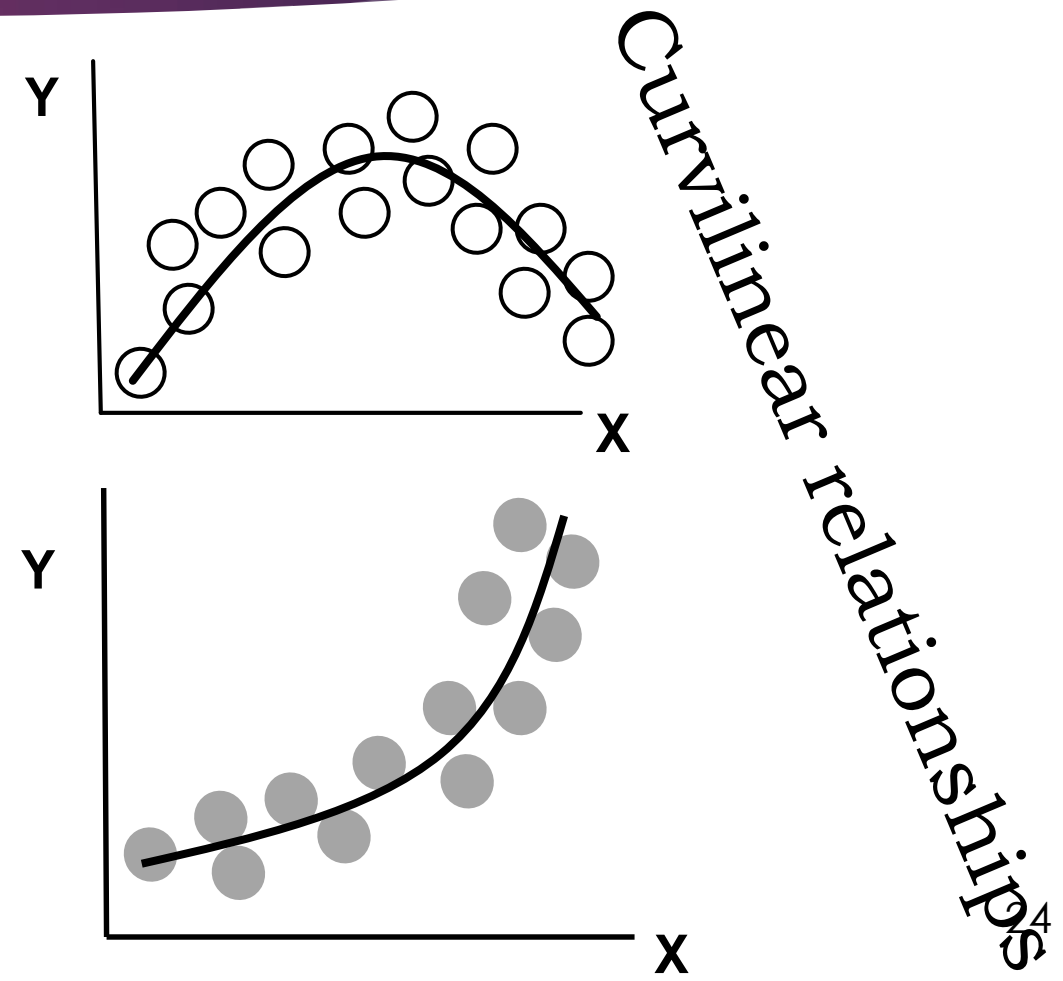
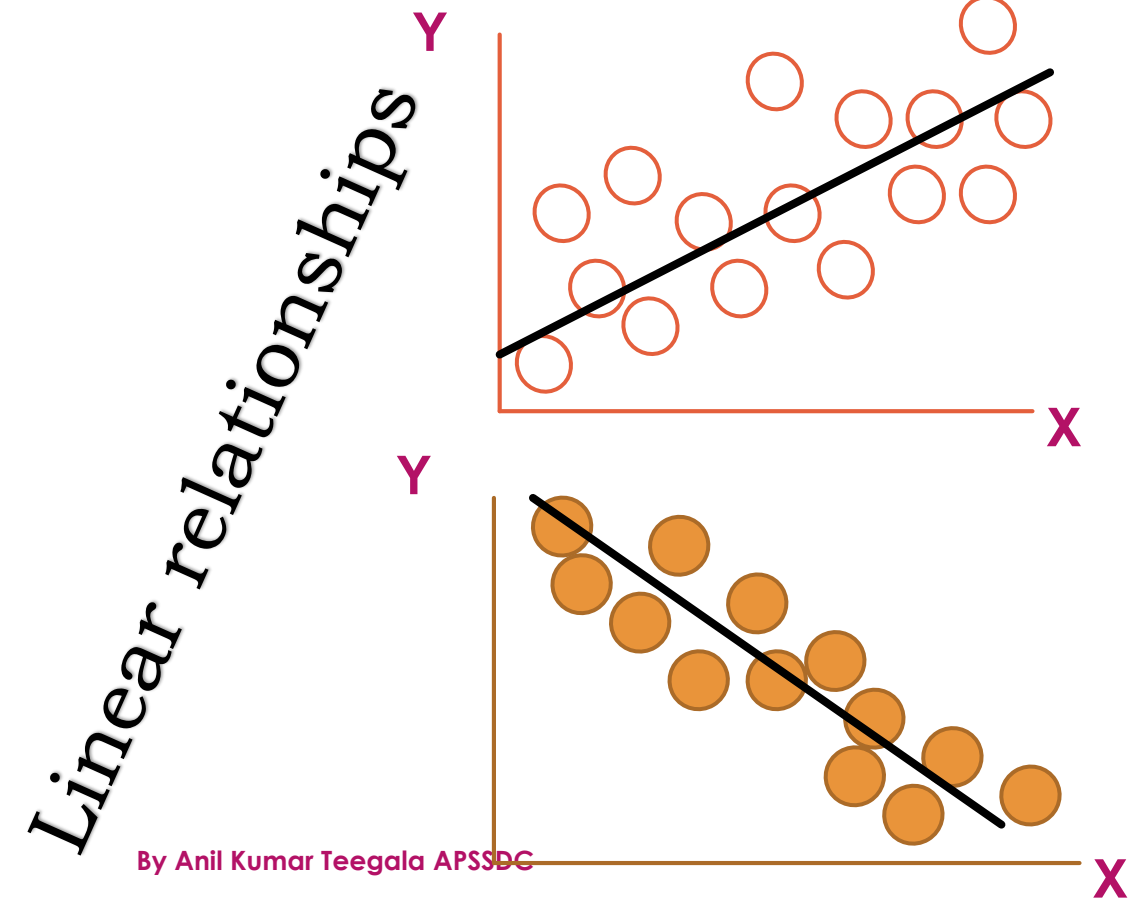
# Correlation

- ▶ Measures the relative strength of the *linear* relationship between two variables Unit-less
- ▶ Ranges between  $-1$  and  $1$
- ▶ The closer to  $-1$ , the stronger the negative linear relationship
- ▶ The closer to  $1$ , the stronger the positive linear relationship
- ▶ The closer to  $0$ , the weaker any positive linear relationship

# Scatter Plots of Data with Various Correlation Coefficients



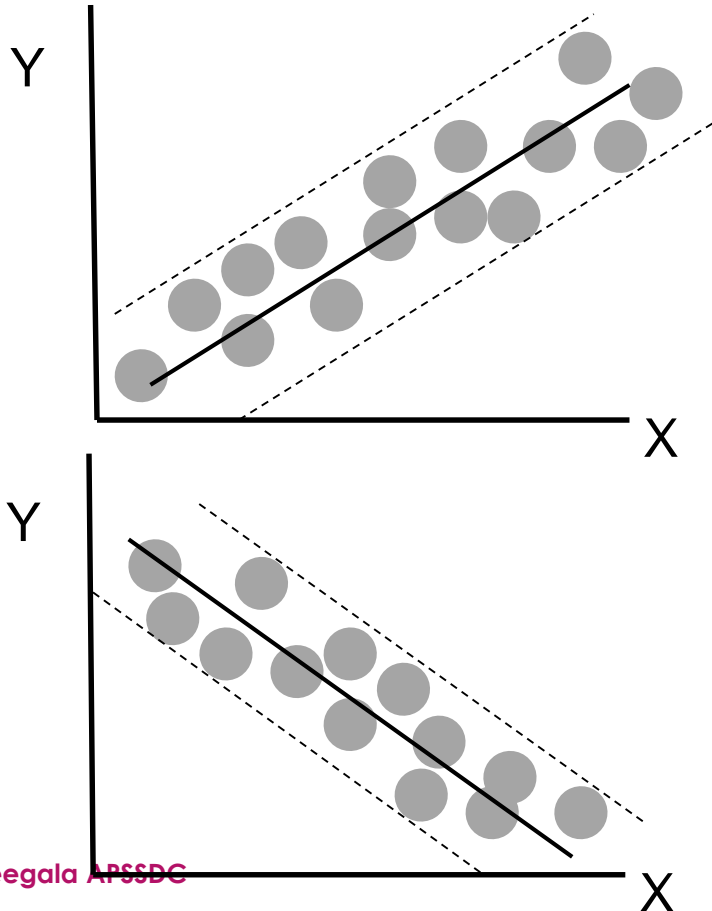
# Linear Correlation





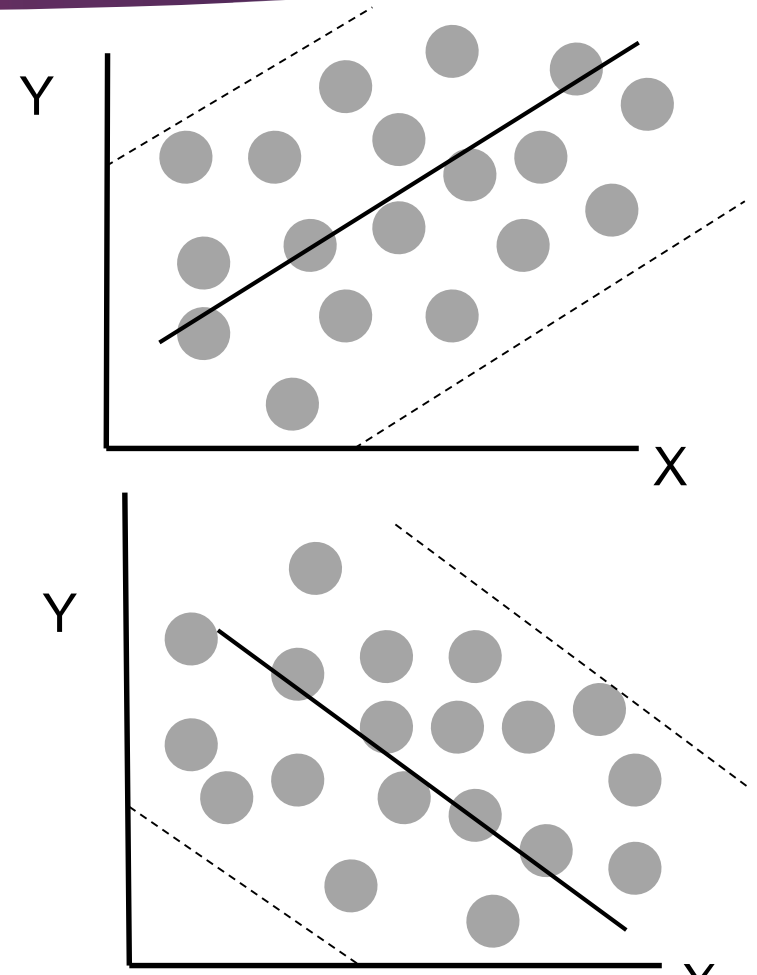
# Linear Correlation

**Strong  
relationships**



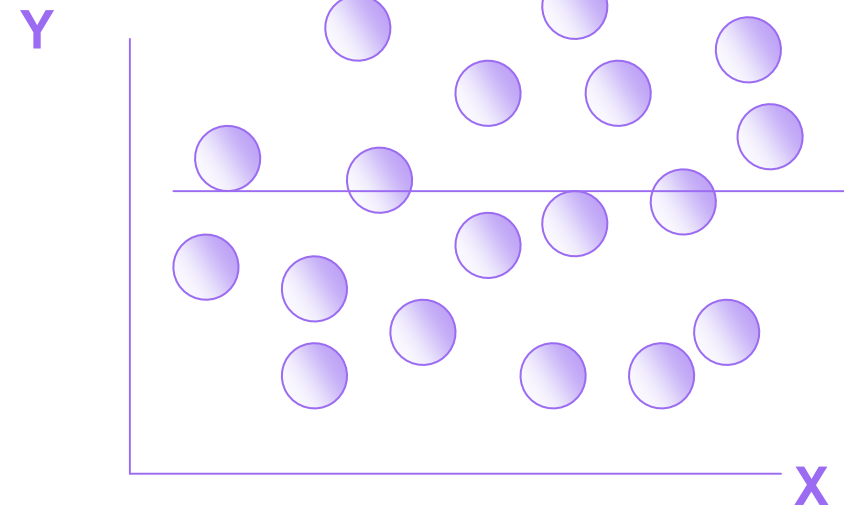
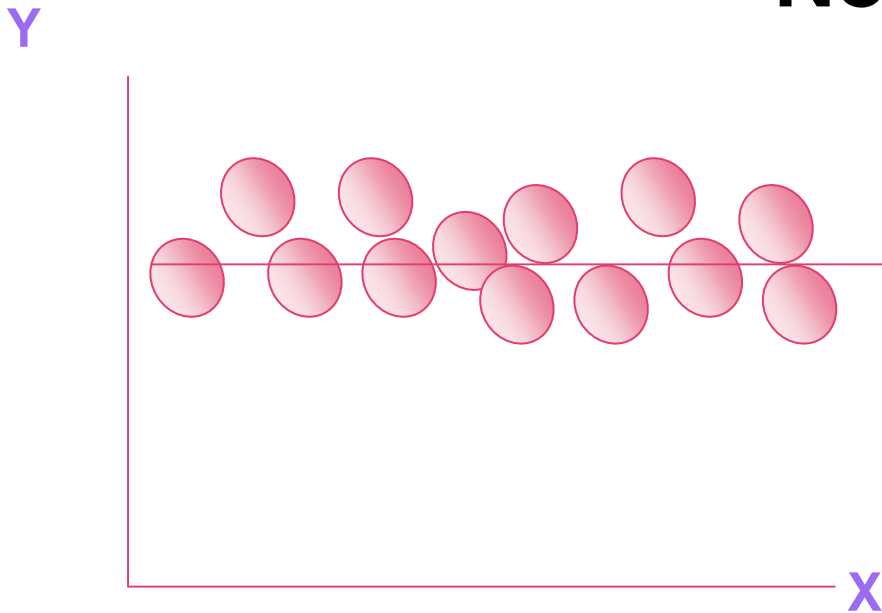
By Anil Kumar Teegala APSSDC

**Weak  
relationships**

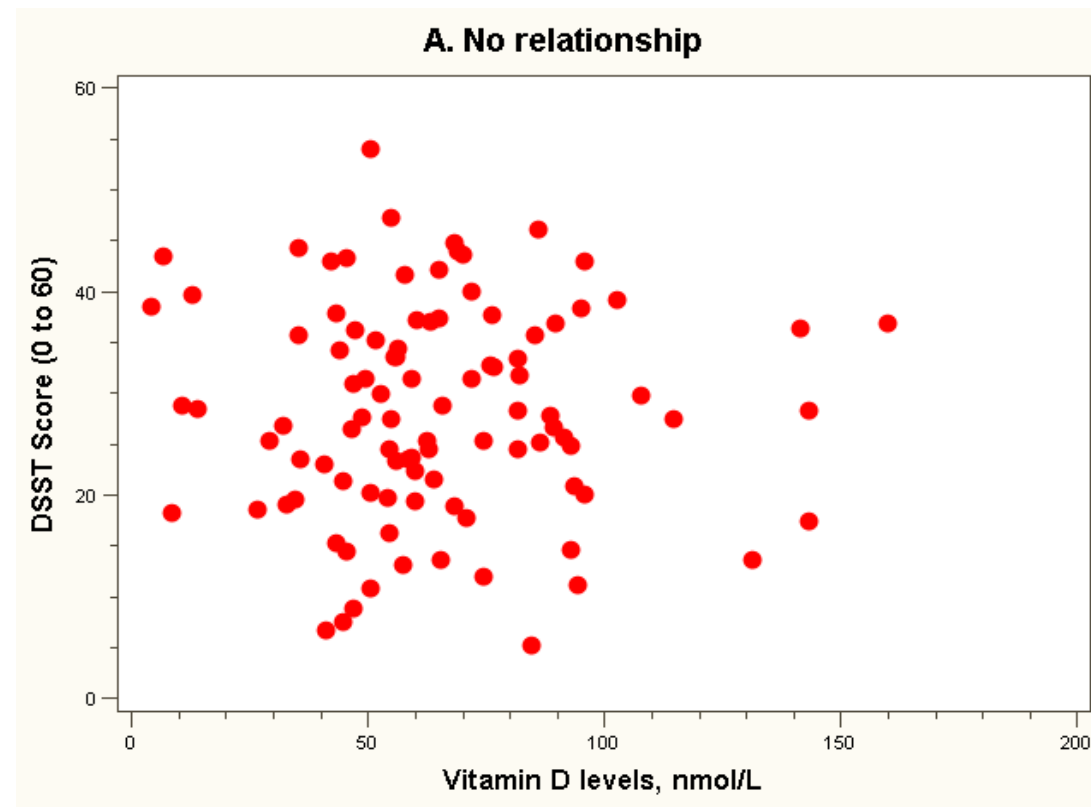


# Linear Correlation

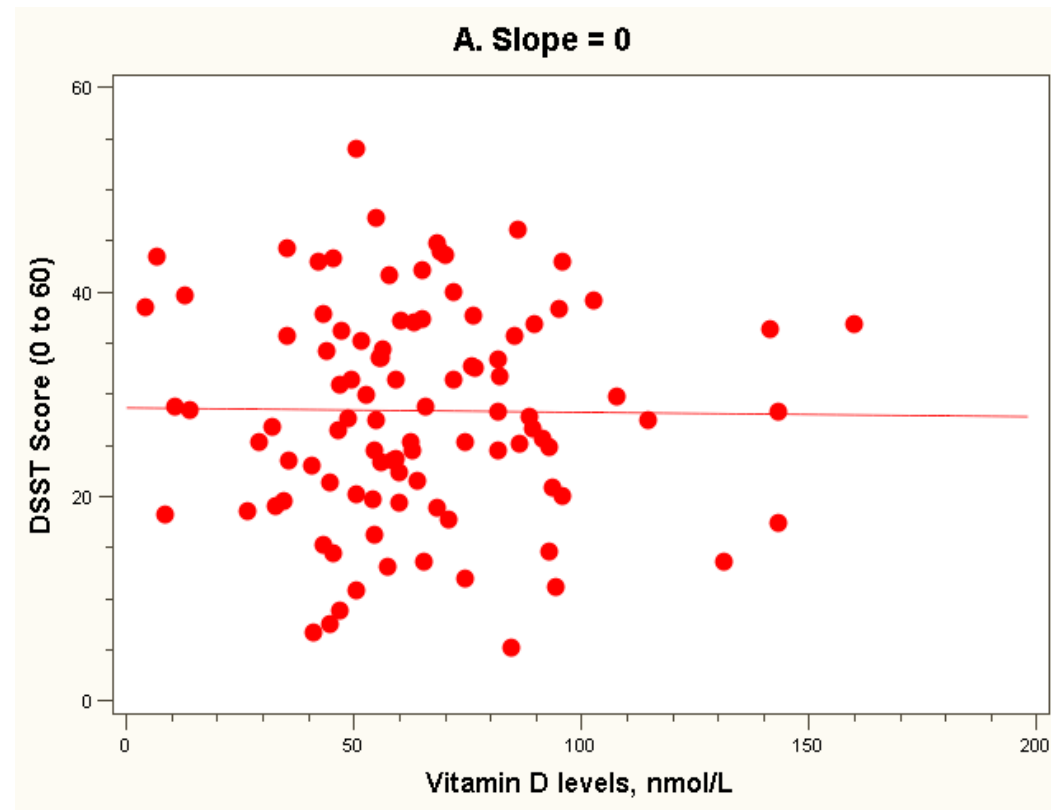
## No relationship



# Dataset 1: No Relationship



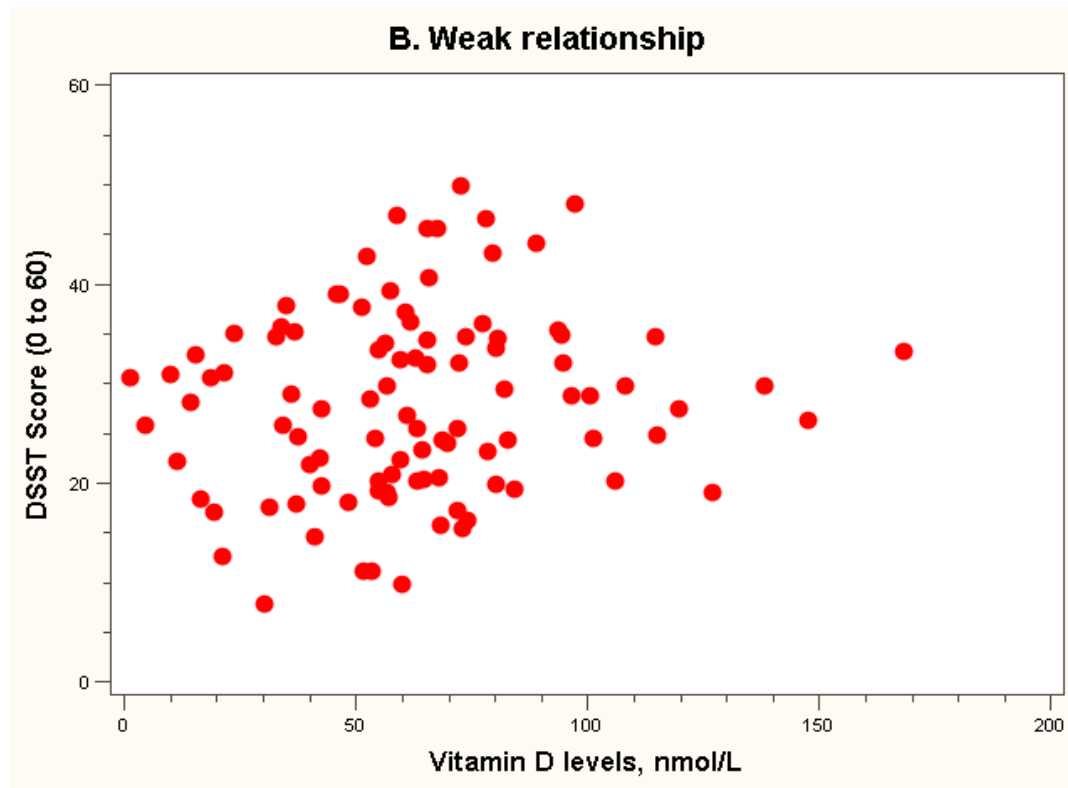
# The “Best fit” line



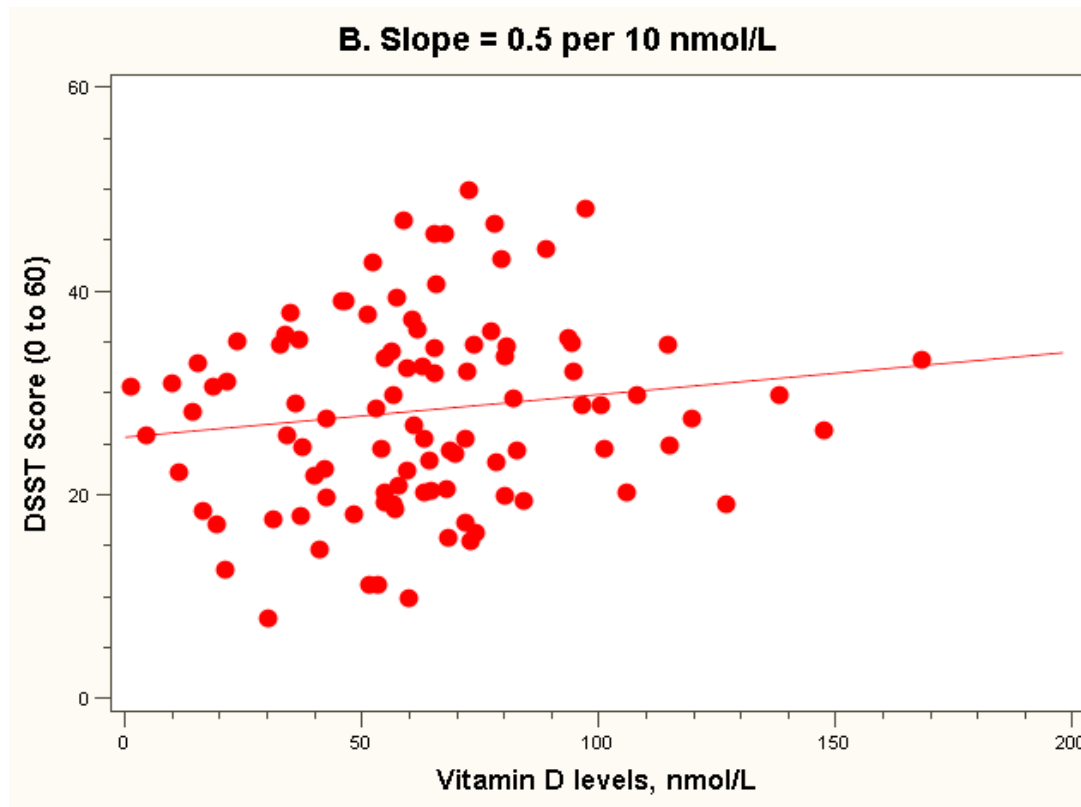
Regression  
equation:

$$E(Y_i) = 28 + 0 \cdot \text{vit} D_i \text{ (in 10 nmol/L)}$$

# Dataset 2: weak relationship



# The “Best fit” line

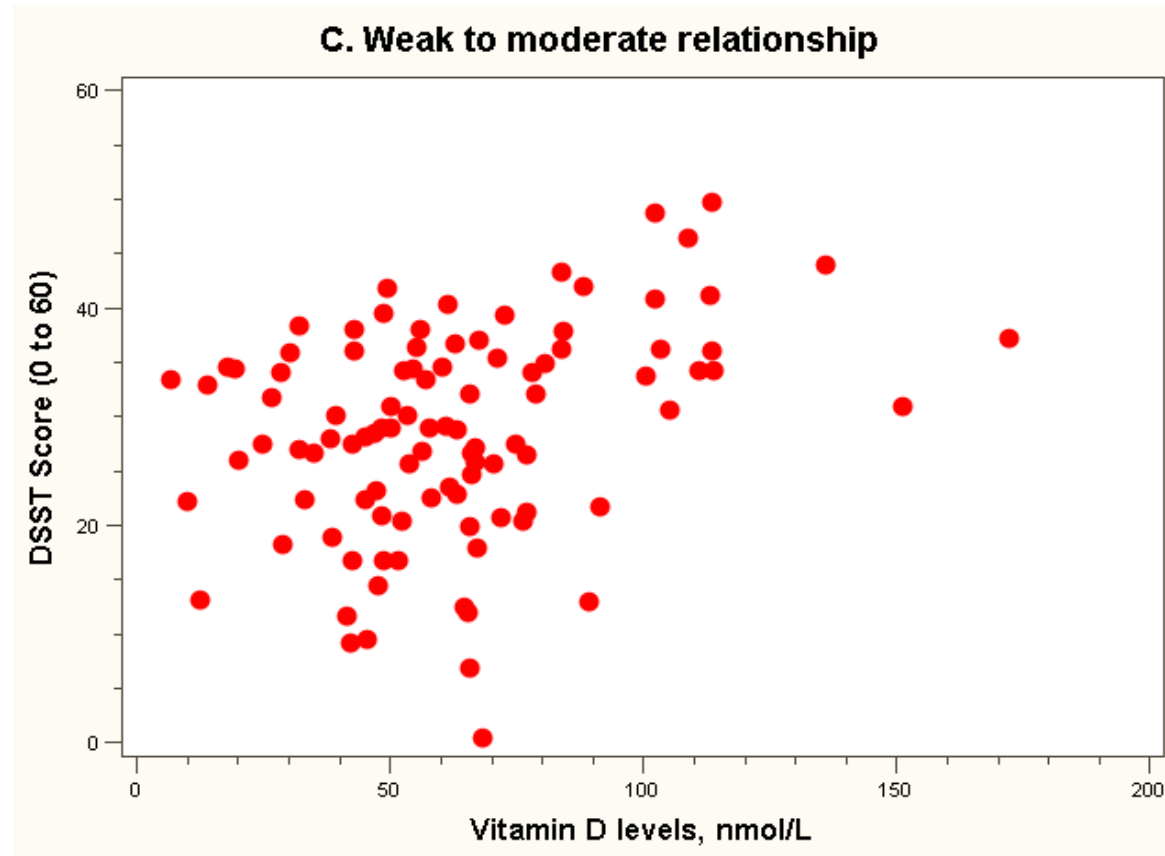


Note how the line is a little deceptive; it draws your eye, making the relationship appear stronger than it really is!

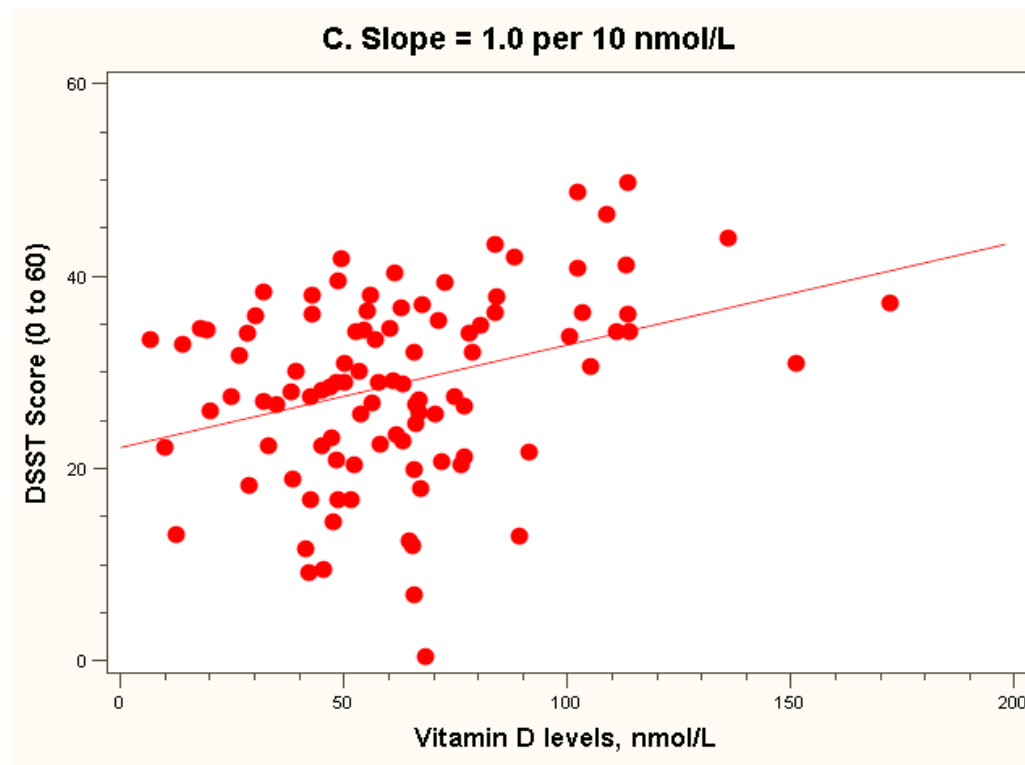
Regression equation:

$$E(Y_i) = 26 + 0.5 \cdot \text{vit Di (in 10 nmol/L)}$$

# Dataset 3: weak to moderate relationship



# The “Best fit” line

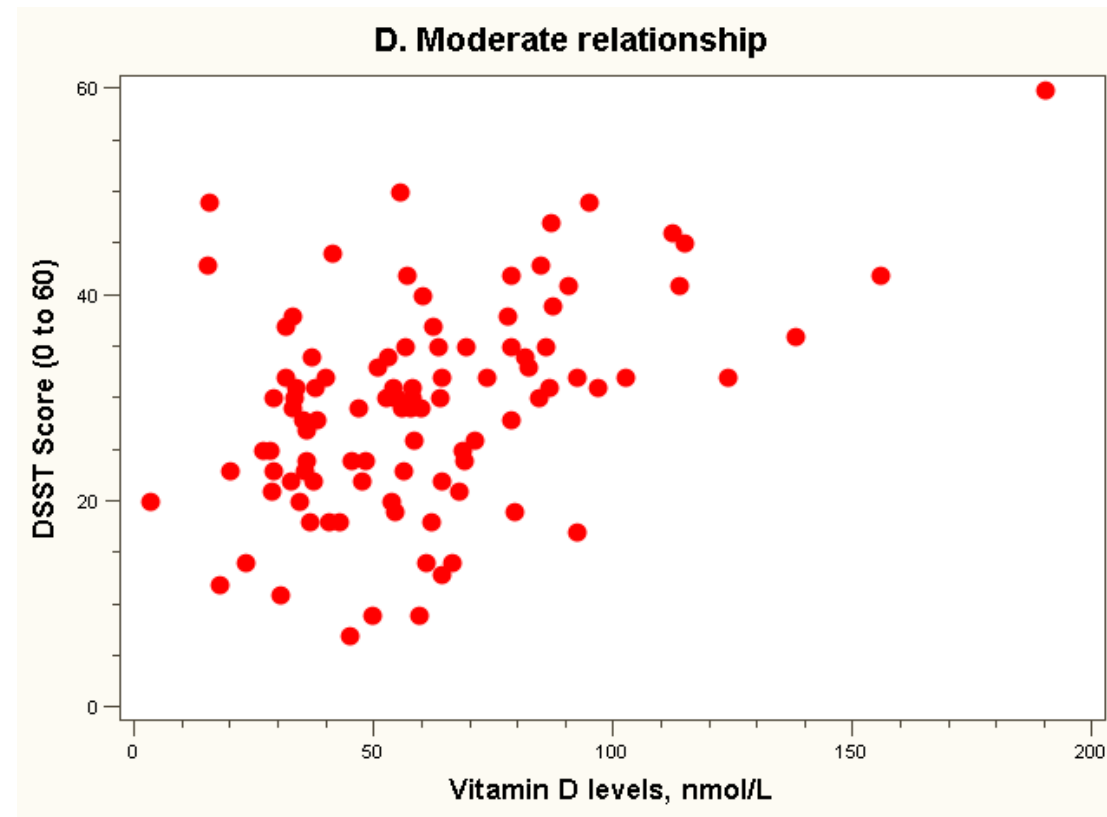


Regression  
equation:

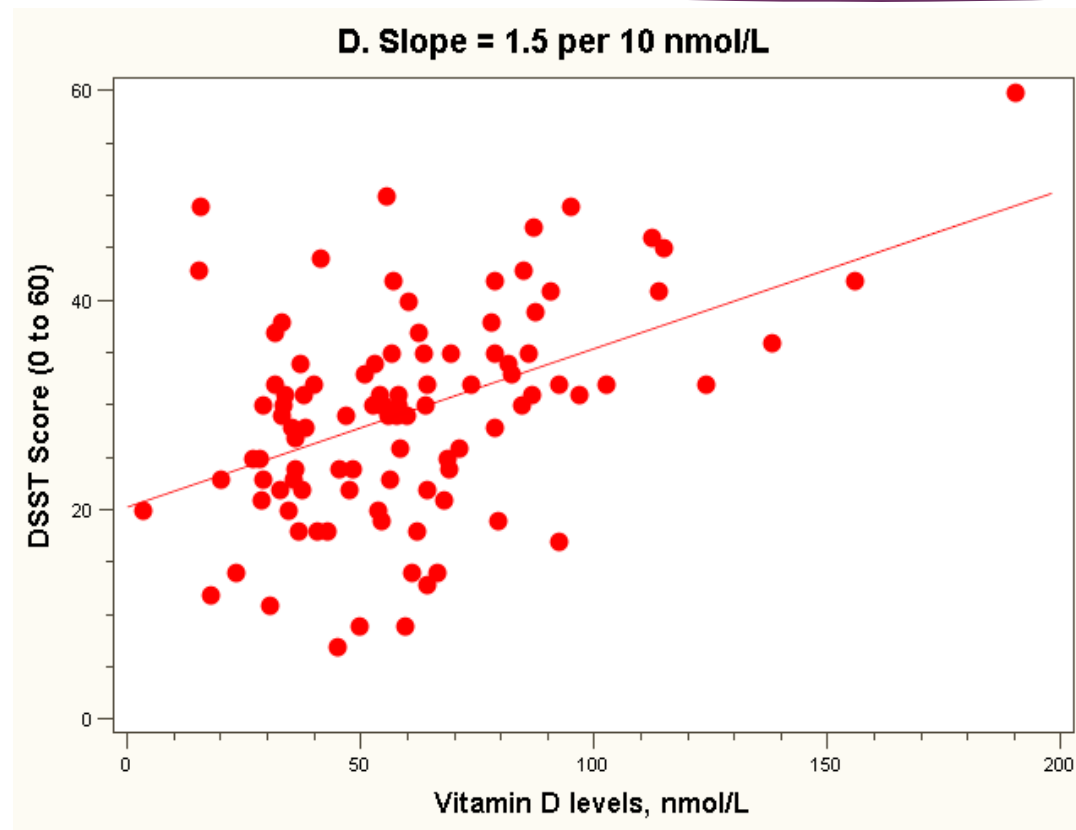
$$E(Y_i) = 22 + 1.0 \cdot \text{vit Di (in 10 nmol/L)}$$



# Dataset 4: moderate relationship



# The “Best fit” line



Regression equation:

$$E(Y_i) = 20 + 1.5 \cdot \text{vit Di (in 10 nmol/L)}$$

Note: all the lines go through the point (63, 28)!

# Steps in Regression Analysis

- Examine the scatterplot of the data.
  - I. Does the relationship look linear?
  - II. Are there points in locations they shouldn't be?
  - III. Do we need a transformation?
- Assuming a linear function looks appropriate, estimate the regression parameters.
  - I. How do we do this? (Method of Least Squares)
- If there is a significant linear relationship, estimate the response,  $Y$ , for the given values of  $X$ , and compute the residuals

# Regression Analysis

- Thus we have the regression formula as :

$$Y = MX + C + \text{error}(e).$$

Initially we calculate the value for slope and predict the values of Y for any given X values we have.

$$\text{Slope}(M) = \sum_{i=0}^{\text{len}(X)} \frac{(X_i - X_{\text{mean}}) * (Y_i - Y_{\text{mean}})}{(X_i - X_{\text{mean}})^2}$$

Thus we calculate the C value and find out the “Line of Regression”.

# Regression Analysis

- Next our job is to reduce the distance between the actual value and the predicted value or in other words reduce the error between the actual and predicted value. Thus the line with least error will be the “**Best Fit Line**”.
- In order to check it out we calculate the “Coefficient of Determination”.

$$\text{Mean Squared value } (R^2) = \sum_{i=0}^{\text{len}(X)} \frac{(Y_{\text{pred}} - Y_{\text{mean}})^2}{(Y - Y_{\text{mean}})^2}$$

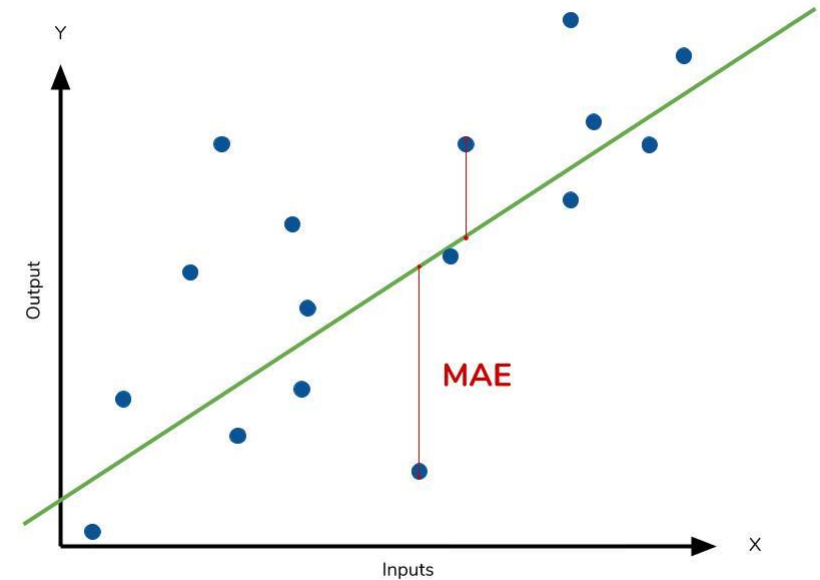
- Thus our ultimate aim is to reduce the error i.e. distance between the actual and predicted values.

# Evaluation Metrics

## 1. Mean Absolute Error

$$MAE = \frac{1}{n} \sum |y - \hat{y}|$$

Divide by the total number of data points (points to  $\frac{1}{n}$ )  
 Actual output value (points to  $y$ )  
 Predicted output value (points to  $\hat{y}$ )  
 Sum of (points to  $\sum$ )  
 The absolute value of the residual (points to  $|y - \hat{y}|$ )

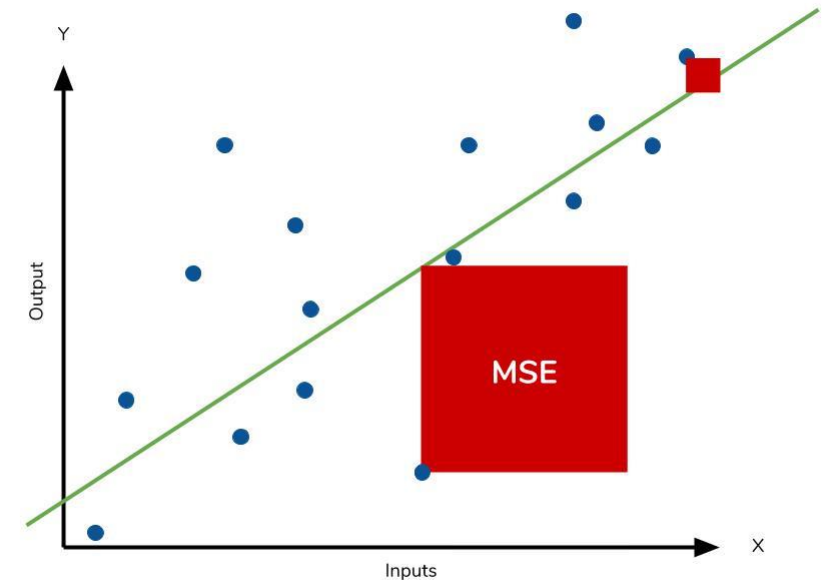


Contd..

## 2. Mean Square Error

$$MSE = \frac{1}{n} \sum \left( y - \hat{y} \right)^2$$

The square of the difference  
between actual and  
predicted

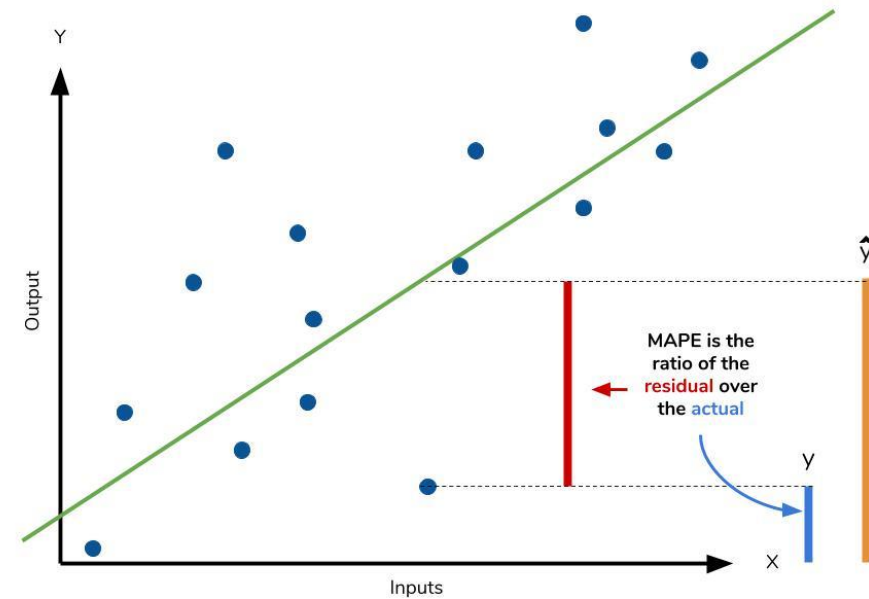


Contd..

### 3. Mean Absolute Percentage Error

$$MAPE = \frac{100\%}{n} \sum \left| \frac{\overbrace{y - \hat{y}}^{\text{The residual}}}{\underbrace{y}_{\text{Each residual is scaled against the actual value}}} \right|$$

Multiplying by 100% converts to percentage

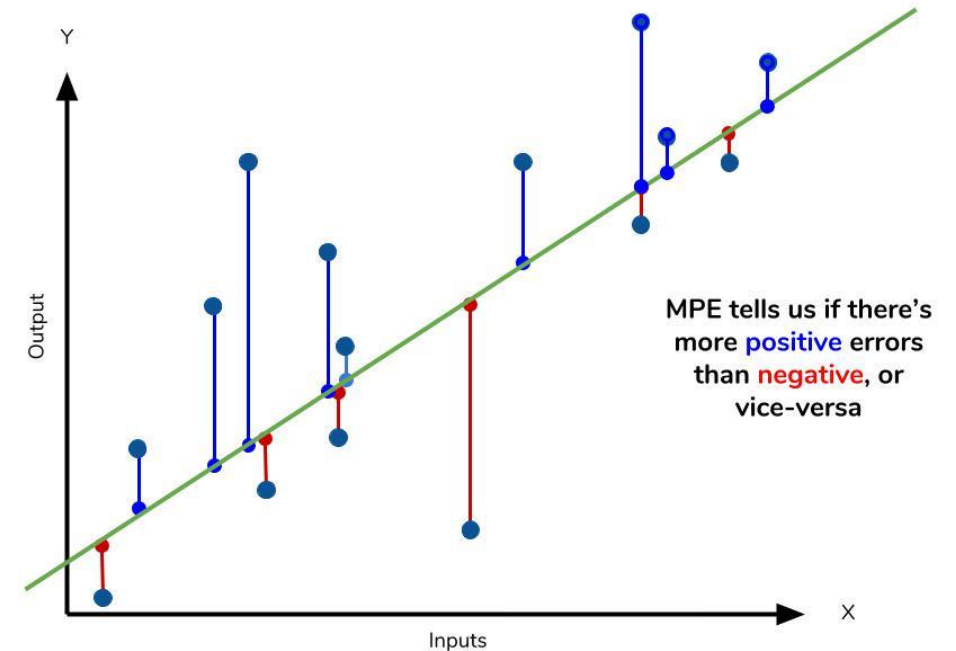




Contd..

### 3. Mean Percentage Error

$$MPE = \frac{100\%}{n} \sum \left( \frac{y - \hat{y}}{y} \right)$$



# Conclusion

Acroynm	Full Name	Residual Operation?	Robust To Outliers?
MAE	Mean Absolute Error	Absolute Value	Yes
MSE	Mean Squared Error	Square	No
RMSE	Root Mean Squared Error	Square	No
MAPE	Mean Absolute Percentage Error	Absolute Value	Yes
MPE	Mean Percentage Error	N/A	Yes