# BUSINESS ANALYTICS

# Prediction Techniques

Process of making predictions of the future based on past and present data

# REGRESSION ANALYSIS

- Generates an equation to describe the statistical relationship among variables and to predict new observations
- How the typical value of the dependent variable (response) changes when any one or more independent variables (predictor)
- Dependent Vs one or more Independent variable
- Regression model relates *Y* to a function of X and β:

**Y~f(X, β)**

β=Unknown parameter

X=Independent Variable

Y=Dependent Variable

- Regression will form a line, wherein the line is best suited on that situation (Best Fit Line)
- It will give 2 information:

  $\beta_0$ - Intercept

  $\beta_1$ _ Slop of the line

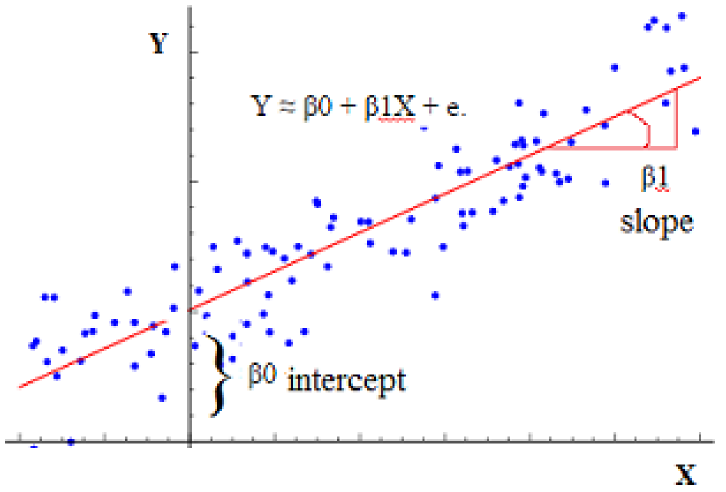
# Simple Linear Regression - Model

- Statistical method that allows us to summarize and study relationships between two continuous variable
- Y=dependent variable
- The model of Y will be:

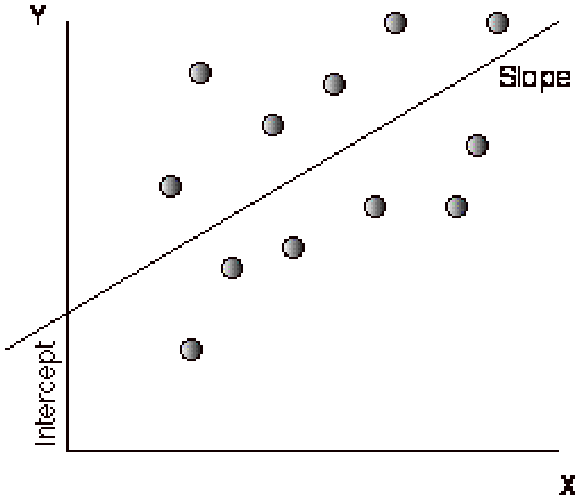
$$y= \beta_0 \pm \beta_1 x$$

- Slop of the line:

$$m = \frac{y_1 - y_2}{x_1 - x_2}$$

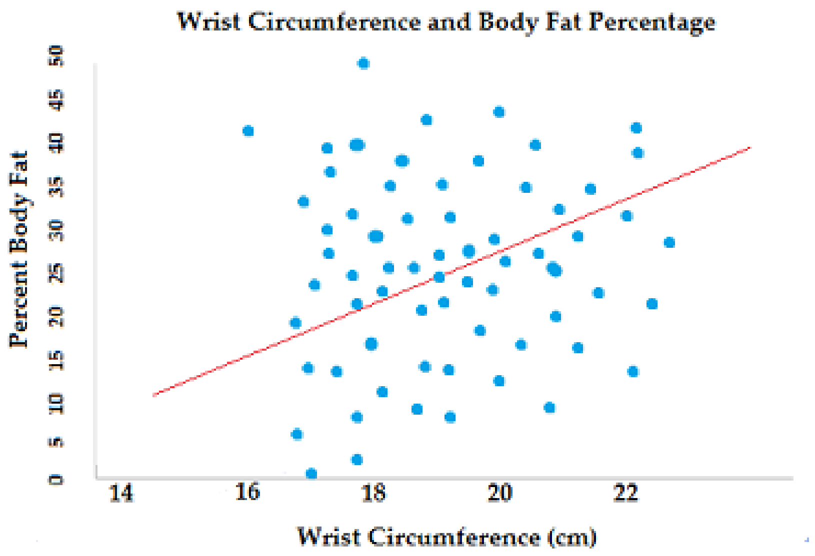


- Intercept - Is the expected mean value of Y when all X=0
- Model of Intercept – ($\beta_0$) $= \bar{y} - b(\bar{x})$



# Multiple Linear Regression - Model

- Attempts to **model** the relationship between two or more explanatory (independent) variables and a response variable by fitting a **linear** equation to observed data
- Every value of the independent variable x is associated with a value of the dependent variable y



# Multiple Regression Model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

- **$Y_i$ is the value of the response variable for the $i^{th}$ case**
- **$\beta_0$ is the intercept**
- **$\beta_1$, $\beta_2$, … , $\beta_{p-1}$ are the regression coefficients for the explanatory variables**

- **Multiple Regression with Two Predictor Variables:**

$$b_1 = \frac{\left(\sum x_2^2\right)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{\left(\sum x_1^2\right)\left(\sum x_2^2\right) - (\sum x_1 x_2)}$$

$$b_2 = \frac{\left(\sum x_1^2\right)(\sum x_2 y) - (\sum x_1 x_2)(\sum x_1 y)}{\left(\sum x_1^2\right)\left(\sum x_2^2\right) - (\sum x_1 x_2)}$$

$$a = b_0 = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2$$

- Residual Standard Error- Difference between the observed value of the dependent variable (y) and the predicted value (ŷ)
- R-square – Relationship between dependent and Independent
- Multiple R-square - Relationship between dependent and both significant and non-significant variable
- Adjusted R-square - Relationship between dependent and significant variable

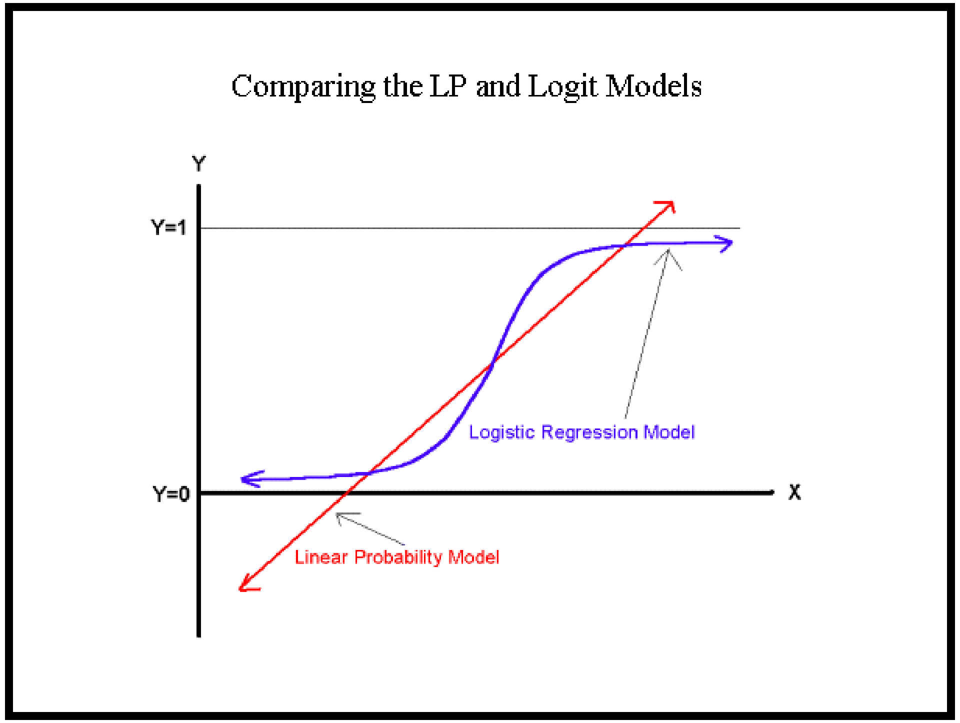
**Note:** Models with one predictor are referred to as simple regression. Models with more than one predictor are known as multiple linear regression.

# Logistic Regression

- Logistic regression is a predictive analysis
- Explain the relationship between one dependent binary variable (Category) and one or more independent variables
- Example of binary variable: pass/fail, win/lose, alive/dead or healthy/sick
- If the category variable is the dependent, then we cannot predict the future and cannot form a linear line

# Logistic Regression…..

- To draw linear line, convert the data points into linear format then convert into original probability



# Logistic Regression…..

- Range of P = 0 to 1
- Rang of continuous variable = -∞ to +∞
- Convert Probability (0 to 1) into (-∞ to +∞)

Step 1 : Odd ratio (p/1-p) = 0 to ∞

$\log(p/1-p) = \log(0/1-0) = \mathrm{Log}_{0} = -\infty$

$\log(p/1-p) = \log(1/1-1) = \mathrm{Log}_{\infty} = \infty$

Range = -∞ to +∞

Note - Anything divided by 0 is ∞

# Logistic Regression…..

- Through the range, calculate co-efficient:

$$z = \text{Log}(p/1\text{-}p) = \boldsymbol{\beta_0 \pm \beta_1 x_1 + \beta_2 x_{2+}} \ldots..$$

- To remove log use **($e^{power}$)**

**ie.** $\mathbf{E^{\log(p/1-p)} = e^z}$

**1.** $\mathbf{p/1\text{-}p = E^{\beta 0 \pm \beta 1x1 + \beta 2x2+\ldots..}}$

**2.** $\mathbf{p/1\text{-}p = e^z}$

**3.** $\mathbf{p = e^z - e^z p}$

**4.** $\mathbf{p + e^z p = e^z}$

**5.** $\mathbf{P(1 + e^z) = e^z}$

**6.** $\mathbf{p = e^z/1 + e^z}$

**ie.** $\mathbf{Z = \beta_0 \pm \beta_1 x_1 + \beta_2 x_{2+}\ldots..}$

# Types of Regression

- Linear Regression (Continuous Variable)
  - Simple linear regression
  - Multiple linear regression

- Logistic Regression (Category Variable)
  - Binary logistics
  - Ordered logistics
  - Multinomial logistics

- Simple linear regression – One dependent variable Vs one independent variable (between 2 continuous variable)

- Multiple linear regression – One dependent variable Vs multiple independent variable (only continuous variable)

- Binary logistics – Only 2 category variable (ie. Pass/fail, Yes/no, 0/1, etc.)

- Ordered logistics – Scale based measurements (ie. Good, better, best)

- Multinomial logistics – Multiple category variable (ie. Car types; Ice cream Flavour)

- All Regression Practice in R & SAS