REGRESSION:

Regression is a statistical method to determine the relationship between one dependent variable and a series of other variables known as independent (explanatory) variables. A regression model is able to show whether changes observed in the dependent variable are associated with changes in one or more of the explanatory variables.

Applications of Regression.

* Forecasting Sales.
* Cash forecasting
* Analysing survey data
* Stock prediction.
* Predicting the behavior of consumers.

Types of Regression:

* Linear Regression
* Multiple Regression
* Non-Linear Regression.

1. Linear Regression:

Linear Regression is the most form of technique. It establishes the relationship between two variables based on the line of best fit. It is graphically depicted using a straight line with the slope defining how the change in one variable impacts

a change in another Variable. The y-intercept of a linear Regression represents the Value of one of Variable when the Value of the other is Zero.

Linear Regression models often use a least square approach to determine the line of best fit. The least squares technique is determined by minimizing the sum of squares created by a mathematical function. A square is, in turn, determined by squaring the distance between a data point and the regression line or mean Value of the data set.

Examples of Linear Regression:

* Impact of GPA on College Admission
* Impact of rainfall amount on crop yield.

Simple Linear Regression:

Simple Linear Regression is used to estimate the relationship between two Quantitative Variables. Simple linear Regression is a parametric test, meaning that it makes certain assumptions about the data.

Assumptions:

Homogeneity of Variance (homoscedasticity):

The size of the error in our prediction doesn't change significantly across the Value of the independent Variable.

Independent of Observation:

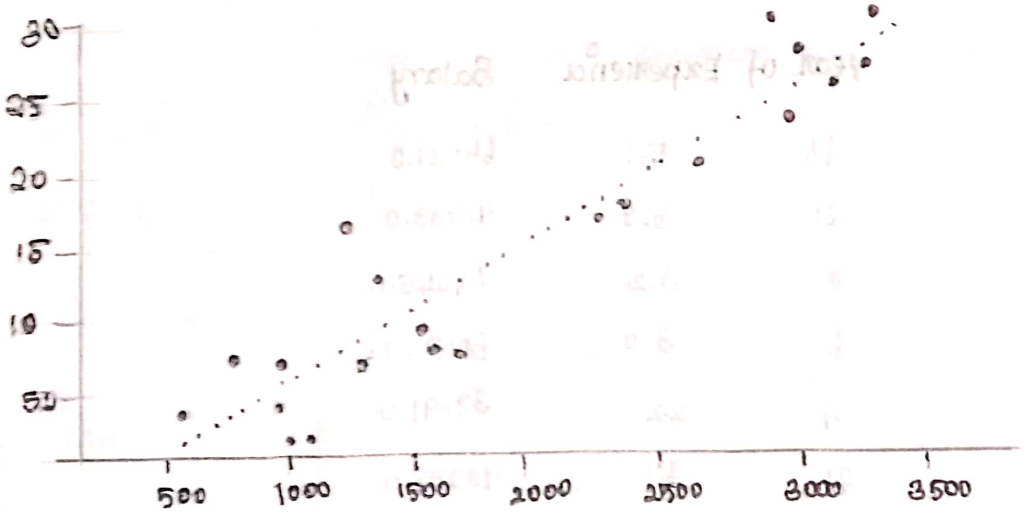
The observation in the data set were collected using statistically valid sampling methods, and there are no hidden relationship among observation.

Normality:

The data follows a normal distribution.

Linear Regression makes an additional Assumption:

The relationship between the independent and dependent variable is linear. The line of best fit through the data points is a straight line (rather than a curve or some sort of grouping factor).



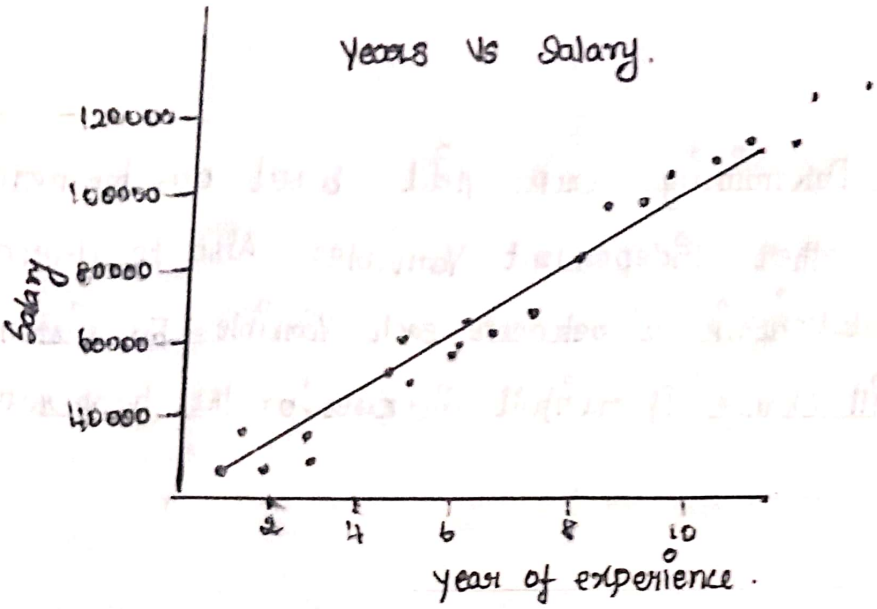
Linear Regression is applied in behavioral and social scenic, finance and economics, machine learning etc. The above observations of marketing campaigns, with the marketing spend amount and the achieved conversion rate in percentage points is shown in the scatter plot with the linear model.

Example:

Consider the data set shown in the figure having years of experience and Salary and the regression line for the same is shown in the scatter plot. Here dependent Variable is Salary and independent Variable is the years of Experience. If there is an increase in dependent Variable increases, then there is a positive correlation among them. If it decreases then there is a negative correlation among them.

The best fit line for the data is the one which produces Least error or least square approximation error among all regression line's that can be drawn. This method of finding best fit line is called Least Square Approximation Method.

Year of Experience	Salary.
16	5.1	66029.0
20	6.8	91738.0
8	3.2	64445.0
6	3.0	60150.0.
4	2.2	39891.0
21	7.1	98273.0
7	3.2	54445.0
29	10.5	121872.0
19	6.0	93940.0
11	4.0	55794.0
23	8.2	11812.0
1	1.3	46205.0
0	1.1	3943.0
25	9.0	105582.0.



From the above plot, it is observed that the regression line is far from some data points. This whole process is an iterable one and will be continued until the best line fit with least square approximation distance is obtained.

Multiple Regression:

Multiple linear Regression (MLR), often known as multiple Regression, is a statistical process that uses multiple explanatory factors to predict the outcome of a response Variable.

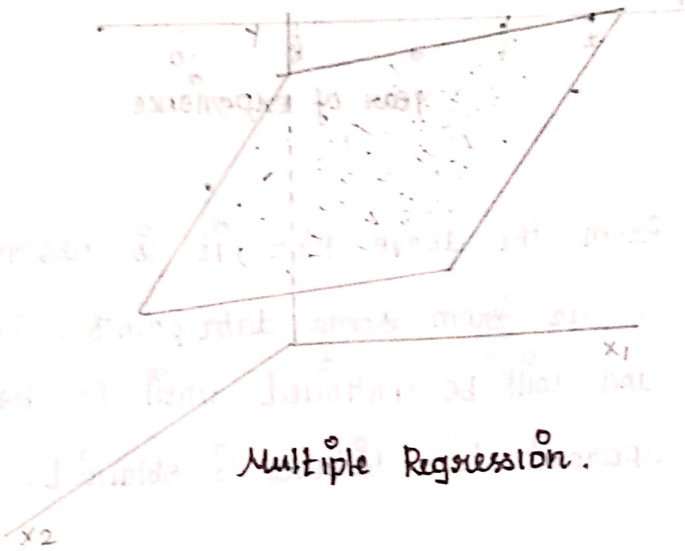
MLR is a method of representing the linear relationship between explanatory (independent) and response (dependent) Variables. It is used to explain a dependent Variable using more than one dependent Variables.

Uses:

The first is to determine the dependent Variable based on multiple independent Variables.

Example:

Determining Crop yield based on temperature, rainfall and other independent Variables. Also to determine how strong the relationship is between each Variable. For instance, how the Crop field will Change if rainfall increases or the temperature decreases.



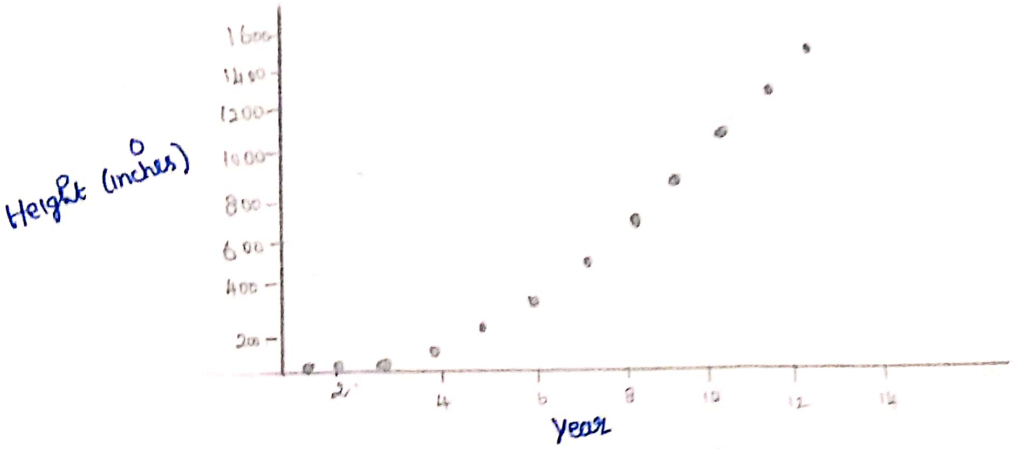
Multiple Regression.

Multiple linear Regression should be used when two or multiple independent Variables determine the outcome of a single dependent variable. This is often the case of forecasting more Complex relationships.

Nonlinear Regression:

Non linear Regression is a mathematical function that uses a generated line (typically a curve) to fit an equation to some data. A linear regression uses a straight line equation (such as $y = mx + c$) whereas non linear regression represents the association using a curve, making it nonlinear in parameter.

Nonlinear regression refers to the regression analysis where the regression model portrays a nonlinear relationship between dependent Variable and independent Variable.



The Sum of Squares is used to determine the fitness of a regression model, which is computed by calculating the difference between the mean and every point of data.

The Non-linear model is complex and at the same time, creates accurate results. The analysis develops a curve depicting the relationship between variables based on the dataset provided. The model offering great flexibility can create a curve that best suits the scenario.

Application of non-linear Regression :

a) The nonlinear regression models are used for prediction, financial modeling and forecasting purposes.

b) used in many fields and sectors like insurance, agriculture, finance, investing, machine learning AI and understanding broader markets.

c) The use of a nonlinear model in developing a wide range colorless gas, HCFC-22 formulation is an example from the field of Chemistry.

Example for Regression:

We find by using the method Least Square Regression

Least Square Regression equation $y' = bx + a$.

$$b = r\sqrt{\frac{SSy}{SSx}}$$

$$a = \bar{y} - b\bar{x}$$

Guidelines to find b.

1. calculate the correlation coefficient by using the formula $r = \sqrt{\frac{SPxy}{SSx \times SSy}}$

2. Determine the values of SSx, SSy and by using the formulas $SSx = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$, $SSy = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$

3. Apply the values of SSx, SSy and r in the formula to find b.

Guidelines to find a.

1. calculate the mean of the y ($\bar{y}$)
2. calculate the mean of the x ($\bar{x}$)
3. Apply the values of $\bar{x}$ and, $\bar{y}$ and b to find a

To find r:

x	y	xy	x^2	y^2
13	14	182	169	196
9	18	162	81	324
7	12	84	49	144
5	10	50	25	100
1	6	6	1	36
$\Sigma x = 35$	$\Sigma y = 60$	$\Sigma xy = 484$	$\Sigma x^2 = 325$	$\Sigma y^2 = 800$

$$SPxy = \Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n} = 484 - \frac{(35)(60)}{5} = 484 - 420 = 64$$

$$SS_x = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 305 - \frac{(35)^2}{5} = 325 - 245 = 80$$

$$SS_y = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 800 - \frac{(60)^2}{5} = 800 - 720 = 80$$

$$r = \sqrt{\frac{SPxy}{SS_x SS_y}} = \frac{64}{\sqrt{80 \times 80}} = \frac{64}{80}$$

$$\boxed{r = 0.80}$$

To find b.

$$b = r\sqrt{\frac{SS_x}{SS_y}}$$

$$b = 0.80\sqrt{\frac{80}{80}}$$

$$b = 0.80 \times 1$$

$$\boxed{b = 0.80}$$

To find a.

$$a = \bar{y} - b\bar{x}$$

~~$a = 60 - (0.80 \times 35)$~~

$$\bar{y} = 12$$

$$\bar{x} = 7.$$

$$a = \bar{y} - b\bar{x} = 12 - (0.80)(7)$$

$$= 12 - 5.60 = 6.40$$

$$\boxed{a = 6.40}$$

To find $\hat{y}$, $\hat{y} = (b)(x) + a$

$$\boxed{y' = (0.80)(x) + 6.40}$$