

* Regression -

Regression is the study of formulation & determination ~~off~~ ~~an~~ of algebraic ~~term~~ ~~for~~ the relationship between the variables.

- Also, it predict the value of one variable from that of the other
- One variable is independent & other is dependent.

for example -

$$Y = a + bX$$

where,

Y = denotes the estimated value of Y for given value of X

In this ' Y ' is dependent & X independent

* Methods to find regression lines -

1) Scatter diagram.

2) Method of least square.

* Multiple regression -

Multiple regression Analysis is used for predicting the unknown value of variable from the known value of two or more variables.

- It is model with one dependent & two or more independent variables.
- The variables whose known value is

to be predicted is known as dependent & whose values are used for prediction are called independent variables.

- It can be linear or non-linear.

- In general, Multiple regression eqn Y on $X_1, X_2, X_3, \dots, X_n$ is given by

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_nX_n$$

Here,

Y = dependent variable

a = Intercept

$b_1, b_2, b_3, \dots, b_n$ - are regression coefficient of $X_1, X_2, X_3, \dots, X_n$ - are independent variables

- It is widely used in biological, social & pharmaceutical sciences to study the possible relationships between variables.

eg - bat fat percentage

~~Standard error of regression~~

Also known as est standard error of estimate, It represents the average distance that the observed values fall from the regression line.

- It tells you how wrong the regression model is on average using units of the response variable.

- The smaller values are better because it indicates that the observations are closer to the fitted line.
- If the standard error is '0' that shows ~~the~~ correlation is perfect.
- * It is for both linear & non-linear regression models.
- It is imp in calculation of confidence & prediction intervals.

$$S_{y|x} = \sqrt{\frac{\sum (y - y_e)^2}{N}}$$

where, $S_{y|x}$ = std. error of estimate of y on x
 y_e = estimated value of y for given value of x .

Similarly, $S_{x|y} = \sqrt{\frac{\sum (x - x_e)^2}{N}}$

- The large the value of $S_{y|x}$ & $S_{x|y}$, the greater the scatter on the line of regression.