Visit: iibf.info

CORRELATION & REGRESSION

IIBF ABM

CORRELATION & REGRESSION

- When we need to find out, how two variables or 2 items are correlated then we use correlation.
- Further if we want to conclude or predict one variable in reference to another, we use regression.
- For example: "A good education is essential to achieve success in life.'? Suppose we decide to test it.

SCATTER DIAGRAM

Visual representation of data can give us a good idea about the relationship between two variables.

We can conclude in 2-3 types of relations

- Direct relationship
- Inverse relationship
- No relationship
- When there is a linearity between the different variables like one variable is directly related to another one that means there is a linear relationship.

- Since we cannot draw straight line, we can draw a curve. Such relationship is called **curvilinear relationship**.
- More the linear relationship fits better to the curvilinear relationship more accurately predictions can be made
- Or simply we can say the variables are more closely related.
- Strength of our relationship is measured using co-efficient of correlation or correlation coefficient and is denoted by r

Mail: banishagupta2099@gmail.com
Website: https://iibf.info

CORRELATION

$$r = \frac{cov(X,Y)}{\sigma_x \sigma_y}$$

$$cov(X,Y) = \frac{1}{N}\sum (x-\bar{x})(y-\bar{y})$$

$$\sigma_x = \sqrt{\frac{(x-\bar{x})^2}{N}} \qquad \sigma_y = \sqrt{\frac{(y-\bar{y})^2}{N}}$$

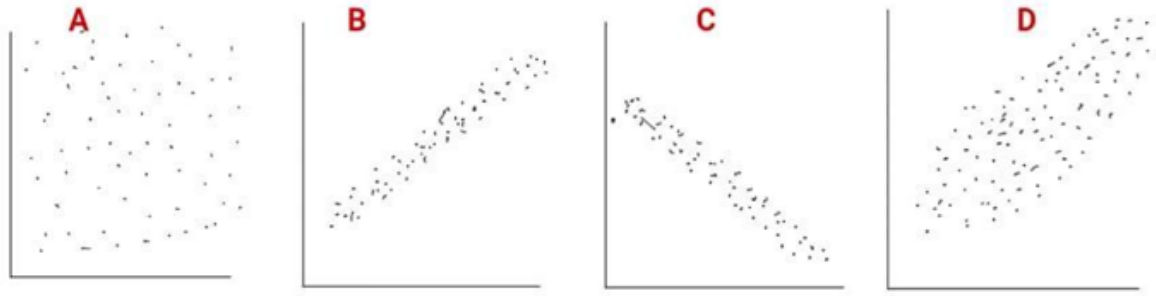
$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\left(\sqrt{N\sum x^2 - (\sum x)^2}\right)\left(\sqrt{N\sum y^2 - (\sum y)^2}\right)}$$

- Value of r lies between -1 and 1
- r=-1 indicates there is a **perfect inverse relationship** between the two variables
- r=1, indicates a perfect direct relationship with all the points in the scatter diagram lying on a line sloping up from left to right
- r=0, indicates that there is no linear relationship between the variables

- Thus, the correlation coefficient is a measure of linear relationship between two variables. Value of coefficient near + 1 or -1 indicates a strong relationship.
- A value closer to zero in the case of weak relationship

QUESTION AND ANSWER

Q1: Which of the following do represent the No relationship between the variables?



Ans: A

Q2. Calculate the correlation coefficient for the following data.

Export of raw materials	1	3	6	10	12
Import of finished goods	5	13	25	41	49

Ans: 1

Q3. Calculate the correlation coefficient for the following data.

X	1	1	3	5	5
Y	5	1	3	1	5

Ans: 0.375

LIMITATIONS

- Measures only linear relationship using correlation
- Data obtained should be homogenous
- Does not indicate any cause effect relationship between variables

REGRESSION

Once we get an idea of the kind and strength of the relationship between two variables from the scatter diagram and the value of the correlation coefficient, we try to draw a line which best represents the points.

We can predict the values if know the values of intercepts a and b in equation of line.

Q4. Calculate the regression coefficient and obtain the lines of regression for the following data and estimate *Y* when *X*=9 from the following

X	1	2	3	4	5	6	7
Y	9	8	10	12	11	13	14

Ans: 15.65

Covariance is an indicator of the extent to which 2 random variables are dependent on each other. A higher number denotes higher dependency. Correlation is a statistical measure that indicates how strongly two variables are related.