Lesson – 10

Correlation Analysis

Summary

The concept of correlation analysis is based on the study of relationship between two or more variables. In other words we can say that correlation analysis is a method of statistical evaluation used to study the strength of a relationship between two, numerically measured, continuous variables. That is why correlation is a bivariate analysis that measures the strength of association between two variables and the direction of the relationship. In this chapter we will study about various methods to apply formulae to calculate the value degree and direction of coefficient of correlation. It also provides the basis to acquaint businessmen to make policies for the growth of business and understand the real – life situations.

Meaning of Correlation

The correlation between two variables measures the strength of the relationship between them.

- Correlation is a describing the degree to which two variables either in move in same direction or in opposite direction.
- If the two variables move in the same direction, then those variables are said to have a positive correlation.
- If they move in opposite directions, then they have a negative correlation.

Correlation and Causation

- Two or more variables considered to be related, in a statistical context, if their values change so that as the value of one variable increases or decreases so does the value of the other variable.
- Causation refers that one event is the result of the occurrence of the other event; i.e. there is a causal relationship between the two events. This is also referred to as cause and effect.

Types of correlation

Positive & Negative Correlation Correlation	Linear and Non Linear Correlation

Positive correlation

- If two variables change in the same direction (i.e. if one increases the other also increases, or if one decreases, the other also decreases), then this is called a positive correlation. For example: Advertising and sales.

Negative Correlation

- If two variables change in the opposite direction (i.e. if one increases, the other decreases and vice versa), then the correlation is called a negative correlation. For example: T.V. registrations cinema attendance.

Linear Correlation

- The linear correlation coefficient is a number calculated from given data that measures the strength of the linear relationship between two variables x and y. Relationship between X and Y is presented in the form of straight line equation as –

$$Y = a + b\ Y$$

Where, a and b are real numbers constant and X and Y are variables

Non – linear Correlation

- A non-linear equation is such which does not form a straight line. It looks like a curve in a graph and has a variable slope value. The equation is written as –

$$Y = a + bx + c^2$$

Degrees of Correlation

Correlation may be positive, negative or zero but lies with the limits ± 1.i. e. the value of r is such that -1 ≤ r ≤ +1.

Degrees	Positive	Negative
Absence of Correlation	**Zero**	**Zero**
Perfect Correlation	**+1**	**-1**
High Degree	**+0.75 to +1**	**-0.75 to -1**
Moderate Degree	**+0.25 to +0.75**	**-0.25 to -0.75**
Low Degree	**0 to + 0.25**	**0 to -0.25**

Properties of coefficient of correlation

- The correlation coefficient 'r' lies between -1 to +1.
- The correlation coefficient 'r' is the pure number and is independent of the units of measurement of the variables.
- The correlation coefficient 'r' is independent of change of origin i.e. Cont……..

Cont…………

The value of r is not affected even if each of the individual value of two variables is increased or decreased by some non-zero constant.

- The correlation coefficient ‘r’ is independent of change of scale i.e. the value of r is not affected even if each of the individual value of two variables is multiplied or divided by some non-zero constant.

Methods of Determining Correlation

- **Scatter Plot**
- **Karl Pearson`s Cefficient of Correlation**
- **Spearman`s cCoeffic[ent of Correlation**

Scatter Plot or Scatter diagram or dot diagram

- In this method, the values of the two variables are plotted on a graph paper. One is taken along the horizontal X-axis and the other along the vertical Y-axis.

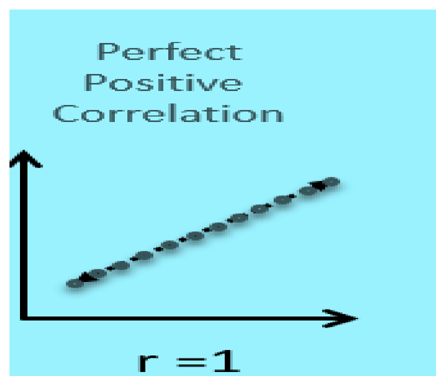
Cont………………

By plotting the data, we get points (dots) on the graph which are generally scattered and hence the name ‘Scatter Plot,’

- The manner in which these points are scattered, suggest the degree and the direction of correlation.

Perfect Positive Correlation

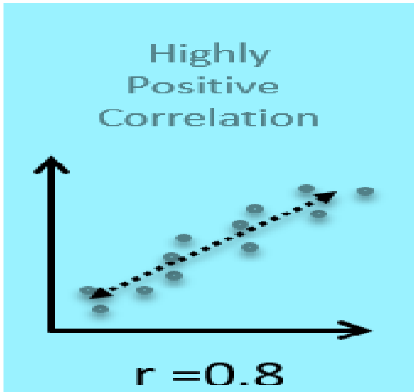
- If all points lie on a rising straight line, the correlation is perfectly positive and r = +1.



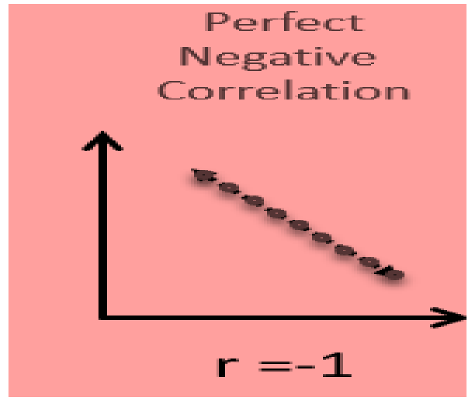
High Degree of Positive Correlation

- If the points lie in narrow strip, rising upwards, this is Known as high degree of positive correlation.

Cont….

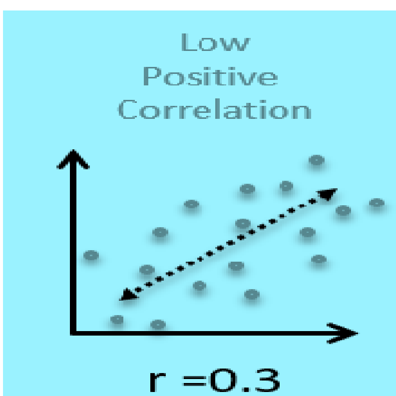


Perfect Negative Correlation



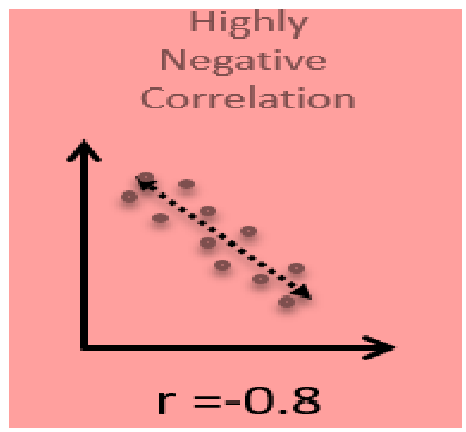
Low Degree of Positive Correlation

- If the points are spread widely over a broad strip, rising upwards, the correlation is low degree positive.



High Degree of Negative Correlation

- If the points lie in a narrow strip, falling downwards, the correlation is high degree of negative.

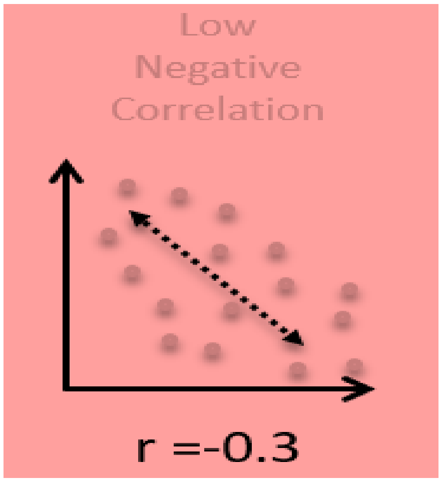


Perfect Negative Correlation

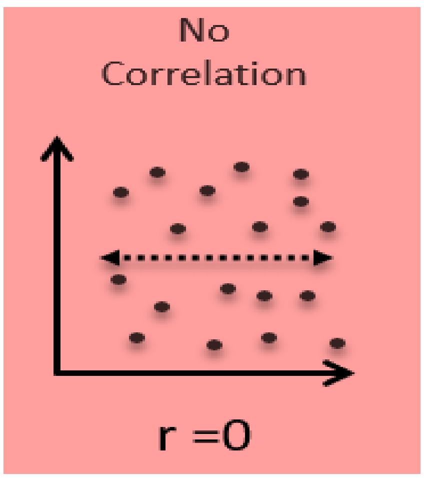
- If all points lie on a falling straight line the correlation is perfectly negative and r = -1

Low Degree of Negative Correlation

- If the points are spread widely over a broad strip, falling downward, the correlation is low degree negative.



No correlation



Karl Pearson's Coefficient of correlation

- It gives the precise numerical expression for the measure of correlation. It is denoted by 'r'. The value of 'r' gives the magnitude of correlation and its sign denotes its direction.

Formula for computing r

$$r = \frac{\sum xy}{N\sigma_x \sigma_y}$$

Where, $x = (X - \bar{X})$, $y = (Y - \bar{Y})$

$\sigma_x = S.D.\ X$, $\sigma_y = S.D.\ Y$

$$\sigma_x = \sqrt{\frac{\sum x^2}{N}} \quad \text{And} \quad \sigma_x = \sqrt{\frac{\sum y^2}{N}}$$

Above equation can be rewritten as-

$$r = \frac{\sum(xy)}{\sqrt{\sum x^2 \times \sum y^2}}$$

Actual Mean Method

$$r = \frac{\sum[(X - \bar{X})^2 \times (Y - \bar{Y})^2]}{\sqrt{\sum(X - \bar{X})^2 \times \sum(Y - \bar{Y})^2}}$$

Assumed Mean Method

$$r = \frac{\sum d_x d_y - \frac{(\sum d_x)\cdot(\sum d_y)}{N}}{\sqrt{\sum d_x^2 - \frac{\sum d_x^2}{N}} \times \sqrt{\sum d_y^2 - \frac{\sum d_y^2}{N}}}$$

Direct Method

$$r = \frac{N\sum XY - [\sum X][\sum Y]}{\sqrt{N\sum X^2 - (\sum X)^2}\sqrt{N\sum Y^2 - (\sum Y)^2}}$$

<u>Rank Correlation</u>

- The concept is based on the ranks of the items rather than on their actual values. The advantage of this method over the others in that it can be used even when the actual values of items are unknown.
- For example correlation between honesty and wisdom

Formula for computing Rank correlation

$$R = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$$

Where, R = Rank correlation coefficient

D = Difference between the ranks of two items

N = the number of observations.

- When R = +1 ⇒ Perfect positive correlation or complete agreement in the same direction
- When R = −1 ⇒ Perfect negative correlation or complete agreement in the opposite direction.
- When R = 0 ⇒ No Correlation.

Formula for computing Rank correlation in case of Repeated Rank

$$R = 1 - \frac{6\left[D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) \ldots..\right]}{N(N^2 - 1)}$$

Method of Calculation

- Give ranks to the values of items. Generally the item with the highest value is ranked 1 and then the others are given ranks 2, 3, 4 ... according to their values in the decreasing order
- Find the difference D = R1 – R2 where R1 = Rank of X and R2 = Rank of Y Note that ΣD = 0 (always)
- Calculate D2 and then find ΣD2
- Apply the formula.
- $\boldsymbol{R = 1 - \frac{6\left[D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) \ldots..\right]}{N(N^2 - 1)}}$

Evaluate Yourself

Q. Give a brief description about degrees of coefficient of correlation.

Q. Calculate coefficient of correlation for the following data by actual mean method.

X	2	4	5	6	8	11
Y	18	12	10	8	7	5

Q. Calculate coefficient of correlation by assumed mean method from the data Given above.

Q. Find out rank correlation coefficient of the given data.

S. No.	Marks in Math	Marks in English
1	15	36
2	17	46
3	13	35
4	16	24
5	6	12
6	11	18
7	14	27
8	9	22
9	7	2
10	12	8