Simplified notes for Msc semII Exam. (2018-20) on

Correlation Analysis

Correlation is the statistically study of the relationship between two or more variables.

According to A M Tuttle – '**An analysis of the co- variance of two or more variables is usually called Correlations**'

Correlation analysis is concern with the measurement of the strength or degree of relationship between variables. 'The measure of correlation is called the **correlation coefficient**. It tells about the direction and degree of correlation.

Types of Correlation

Positive and negative Correlation -

Positive Correlation or Direct Correlation - Both of the variables increase simultaneously. Eg Correlation between age & weight of child.

Negative or inverse Correlation-

Here one variable increases and other variable decreases. E.g. yield of consumer goods and their price.

Simple & Multiple Correlation -

When Correlation between two variables is studied it is called **simple correlation**. E.g. age and weight of the body.

When correlation among more than two variables is studied it is called **Multiple Correlation**. E.g. Study of diet, age and weight of the baby.

Partial and Total Correlation-

The study of the two variables excluding some other variables is called **Partial correlation**. Eg we study diet, age and weight of the baby eliminating the diet

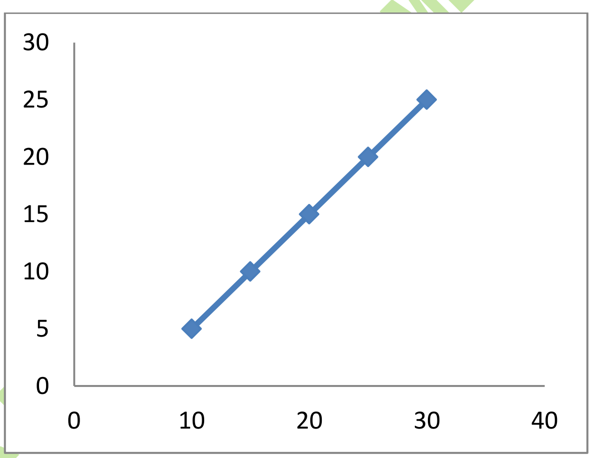
If all the variables are taken into an account it is called **Total correlation**.

Linear and Nonlinear Correlation -

In **linear correlation** straight line graph is found between the considered both variables. Here ratio of change between the variables is same.

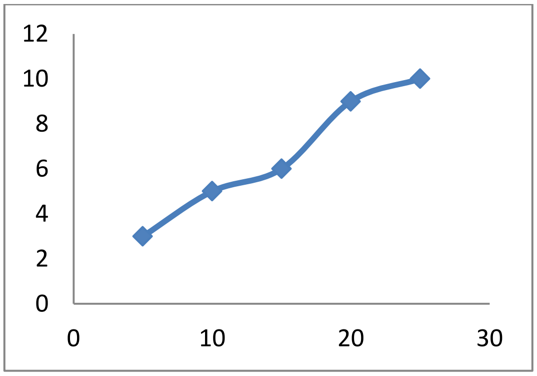
Example

X	10	15	20	25	30
Y	5	10	15	20	25



Nonlinear Correlation - Here graph is not in straight line

X	5	10	15	20	25
y	3	5	6	9	10



Degree of correlation

Degree of correlation	Positive Correlation	Negative Correlation
Perfect correlation	+1	-1
Very high degree of correlation	+0.9 or more	-0.9 or more
Fairly high degree of degree of correlation	From +0.75 to +0.9	From -0.75 to -0.9
Moderate degree of correlation	From +0.5 to +0.9	From -0.5 to -0.9
Low degree of correlation	From +0.25 to +0.5	From -0.25 to -0.5
Very low degree of correlation	Less than +0.25	Less than -0.25
No correlation	0	0

Use of Correlation-

i. To study the relationship between two variables.

ii. To measure the degree of the relationship between two variables.

iii. Sampling error may be calculated

iv. It is the basis of the concept of regression

v. With the use of correlation analysis, the relationship between variables can be verified and tested for significance.

Methods of studying correlation

1) Karl Pearson's coefficient of correlation

2) Spearman's rank coefficient of correlation

1) Karl Pearson's coefficient of correlation-

It is most widely used method. Its coefficient is known as **Pearsonian** coefficient of correlation.

It is denoted by the symbol 'γ'

If the value of 'γ' (Pearsonian coefficient) is +1 there is perfect positive correlation and if its value is -1 for perfect negative correlation. Its value is 0 for no correlation. Its value is varies from +1 to -1 but the value lies between +0.5 to -0.5.

Calculation

We can use any of the following formula to calculate 'γ'. This formula can be derived from each other.

1. **Based on covariance** *and standard deviation*

$$'\gamma' = \frac{\text{Cov}(x,y)}{\text{SD}(x)*\text{SD}(y)}$$

Where- $\text{Cov}(x, y) = \text{Covariance}(x, y) = \frac{\Sigma(x-\overline{x})(y-\overline{y})}{N}$

SD(x) = standard deviation of 'x' $= \delta_x = \sqrt{\frac{\Sigma(x-\overline{x})^2}{N}}$

SD(y) = standard deviation of 'y' $= \sqrt{\frac{\sum y-\overline{y}^2}{N}}$

N = Total number

2. **Based on actual Mean**

$$\gamma = \frac{\sum d_x d_y}{\sqrt{\sum(d_x)^2 \times \sum(d_y)^2}}$$

Where

d_x = *deviation from* $\overline{x} = (x - \overline{x})$

d_y = *deviation from* $\overline{y} = (y - \overline{y})$

$\sum d_x d_y$ = *summation (totaling) of the product of dx and dy*

$\Sigma (dx)^2$ = *sum of the square of the deviations of X*

Rakesh Mohan

$\Sigma (dy)^2$ = sum of the square of the deviations of Y

3. Based on assumed mean (A)

$$\gamma = \frac{\sum dxdy - \frac{\sum dx * \sum dy}{N}}{\sqrt{\sum dx^2 - \frac{\sum (dx)^2}{N}} * \sqrt{\sum dy^2 - \frac{\sum (dy)^2}{N}}}$$

Where

dx = deviation of the items of x series from A(assumed mean) $= (x - A)$

dy = deviation of the items of y series from assumed mean $= (y - A)$

N = number of items

Σdx = total of the all calculated dx from assumed mean

Σdy = total of the all calculated dy from assumed mean

$\sum dxdy$ = summation (totaling) of the product of dx and dy

$\Sigma (dx)^2$ = sum of the square of the deviations of X

$\Sigma (dy)^2$ = sum of the square of the deviations of Y

Merit of pearson's coefficient

i. It is based on the calculation and involved all of the items as such.
ii. It provide clear-cut and precise result .
iii. In this result is obtained in both direction +ve and –ve.
iv. It is very popular measure.

Demerit

i. Calculation is very tedious.

ii. Indicate the relation between series but not give cause.

iii. It can be unduly affected by extreme values.

II) Spearrman's rank correlation coefficient-

- It is given by Charles Edward Spearrman in 1904.
- It is denoted by '**ρ**'
- It is based on rank.
- Like the Pearsonian coefficient it also varies from -1to +1

Formula

$$\rho = \frac{6\sum D^2}{n(n^2-1)}$$

Where, D is differences of two ranks (R_x-R_y)

n – Number of paired observations

Assignment of rank

Rank is given by taking the highest as one and next to the highest as two and so on.

When two or more item have equal value then items are given the average of the ranks they would have received for example if two items are placed on fourth place then each given the 4.5 rank (average of 4 and 5). The next item is given sixth rank.

Merit of the Spearrman's coefficient

- It is simple to calculate.
- It is simple to understand.
- Also useful for qualitative items.
- Not to be affected by extreme values.

Demerit

- It cannot be used in the case of bi-variant distribution.
- If the number of items is greater, the calculation becomes tedious and requires a lot of time.

Rakesh Mohan rzoornc@rediffmail.com

Assignment

21 From the following data calculate the rank correlation coefficient after making adjustment for tied ranks.

X	49	34	41	10	17	17	66	25	17	58
Y	14	14	25	7	16	5	21	10	7	20

22 Calculate Pearson's correlation coefficient.

X	12	18	16	15	12	10	20-	17
Y	6	10	9	8	9	8	12	10

Ans- γ= +0.80

23 Calculate the coefficient between x and y for the values given below.

X	2	5	7	9	19	17
Y	25	27	26	29	34	35

Ans- γ= +0.96