Edunotes
WHERE PASSION MEETS EDUCATION

LEARN KWNIY
Where passion meets with educations

Chapter 10
Correlation

MEANING OF CORRELATION

The relationship between variables are measured by correlation analysis. The correlation indicates the relationship between two such variables in which change in the value of one variables, the value of other variable also change.

CORRELATION AND CAUSATION

The nature of dependence of one variable on the other is different in different cases. There can be correlation between two variables due to any one or more of the following reasons: -

1. Cause and effect: There is a cause and effect relationship between two variables. For example, heat and temperature may be concerned with correlated because heat affects temperature.

2. Effect of third variables: Sometimes, both the correlated variables are being affected by a third Variable or by more than one variable. For example: we may find high degree of correlation between yield of rice and yield of wheat. In reality it may be found that yield of both these Commodities have been affected either by rainfall and other factor like fertilizers, quality of seeds, Irrigation facilities etc.

2. One cause affecting two variances: - When we find that there is increase or decrease together in two variables, it is related to common cause. For example when the Rainfall is scanty (very little) the yield of rice as well as wheat is low and vice versa

4. Chance: - A high degree of correlation may derived due to sheer coincidence and such cases, these variable are not in any way related to each other and therefore, in fact, do not have any relationship.

KINDS OF CORRELATION

On the basis of nature of relationship between the variables correlation may be

1. Positive and Negative correlation

2. Linear and Curvilinear Correlation

3. Simple, multiple and Partial Correlation

1. Positive and negative correlation

When two variables move in the same direction, that is when one increases the other also increases and when one decreases the other also decreases, such a relation is called positive correlation. When two variables change in different directions, it is called negative correlation. When one is increasing and the other is decreasing and Vice-versa.

2. Linear and Curvilinear (Non-linear) correlation

When two variable change in the constant proportion, it is called Linear Correlation. Show on graph paper their relationship will be indicated by straight line.

In case of non- linear relationship (curvilinear), the amount of Change in one variable does not bear a constant ratio to the amount of change in the other variable. Such relationship will form a curve on graph.

3. Simple, Multiple and Partial Correlation

Simple Correlation implies the study of relationship between two variable only.

When the relationship among three or more than three variables are studied simultaneously, it is called multiple Correlation.

When more than two variables are involved and out of these the relationship between the two variables is studied treating other variable as constant, then such correlation is partial.

DEGREE OF CORRELATION

1. Perfect correlation - Perfect Correlation is that where changes in two related variable are exactly proportional. If equal proportional changes are in the same direction, there is perfect positive correlation between the two values described as + 1; and if equal proportional Changes are in the reverse direction, there is perfect negative correlation described as -1.

2. Zero Correlation - The value of the Coefficient of correlation may be zero. It means that there is zero correlation. It does not mean the absence of any type of relation between the two variables.

3. Limited Correlation: Correlation is said to be limited positive. When there is unequal changes in the two variables in the same direction and correlation is limited negative when there are unequal changes in the reverse direction. The limited degree of correlation can be high (between ± .75 to 1): moderate (± .25 to .75) or low (± 0 to 0.25).

METHOD OF STUDYING CORRELATION

a) Scatter Diagram

b) Karl Pearson's coefficient of correlation

c) Spearman's Rank Correlation

A Scatter Diagram

Scatter diagram offer a graphic expression of the diagram and degree of Correlation. To make a scatter diagram, data are plotted on a graph paper. A dot is marked for each value. The course of these dots would indicates direction and closeness of the variables.

From the Scattered diagram, closeness of the dots towards each other in a Particular direction indicates high degree of Correlation. If the dots are scattered, it is an indication of low degree of correlation when plotted Points show some trends - upward or downward. When the trend is upward the correlation is positive, when it is downward the correlation is negative.

Merits of Scatter diagram

1. It is very easy to draw a scatter diagram.

2. It is not affected by extreme values.

3. Scattered diagram is an attractive and easy method to find the nature of Correlation between two Variables.

4. Scattered diagram also indicates whether the relation is positive or negative.

Demerits of Scatter Diagram

1. The degree of Correlation between the two variable cannot be known in numerical terms by scattered diagram.

2. It gives only an approximate idea of the relationship.

3. If there are more than two variables, then it is not possible to draw a scattered diagram.

KARL PEARSON'S COEFFICIENT OF CORRELATION

Karl Pearson has given a quantitative method of calculating correlation. It is an important and widely used method of studying correlation. Karl Pearson's Coefficient of correlation is generally written as " r " Karl Pearson's co-efficient of correlation (r) of two variables is obtained by dividing the sum of the products of the corresponding deviation of the various items of two series from their respective means by the product of their stand and deviation and number of pairs of observation.

$$r = \frac{\sum XY}{N\sigma x \sigma y}$$

r = **Coefficient of correlation**

X = **X** - $\overline{X}$

Y = **Y** - $\overline{Y}$

σx = **Standard Deviation of X series** $\sqrt{\frac{\sum x2}{N}}$

σy = **Standard Deviation of X series** $\sqrt{\frac{\sum y2}{N}}$

N = Number of Observation

There is no need to calculate standard deviation of X and Y. Coefficient of Correlation can be calculated by using direct formula.

$$r = \frac{\sum XY}{\sqrt{\sum x^2 \, X \, \sum y^2}}$$

Method of Calculating coefficient of correlation

(A) Actual Mean Method

Steps

1. Calculate arithmetic means of X and Y series.

2. Find out the deviations of X series and denote these deviation by x

3. Square these deviation and obtain total $\sum x^2$

4. Find out the deviation of Y series and denote these deviation by y

5. Square these deviation and obtain the total $\sum y^2$

6. Multiply these calculated deviation of X and Y series and Calculate

7. Apply the formula

$$r = \frac{\sum XY}{\sqrt{\sum x^2 \, X \, \sum y^2}}$$

x = X - $\overline{X}$ $\qquad$ y = Y - $\overline{Y}$

(B) Direct Method

The coefficient of correlation can be calculated without finding out the deviation from mean of the series. It is obtained by using the following formula:

$$r = \frac{\sum XY - \frac{(\sum X)\,.\,(\sum Y)}{N}}{\sqrt{\sum x^2 - \frac{\sum X^2}{N}\ X \sum y^2 - \frac{\sum Y^2}{N}}}$$

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

If the Arithmetic Mean of X and Y are not in fraction then the following formula is used:

$$r = \frac{\sum XY - N\,.\bar{X}\,.\bar{Y}}{\sqrt{\sum X^2 - N(\bar{X})^2}\ X\ \sqrt{\sum Y^2 - (\bar{Y})^2}}$$

C) Assumed Mean Method

$$r = \frac{\sum dx.dy - \frac{(\sum dx).(\sum dy)}{N}}{\left[\sqrt{\sum dx^2} - \frac{(\sum dx)^2}{N}\right].\left[\sqrt{\sum dy^2} - \frac{(\sum dy)^2}{N}\right]}$$

Or

$$r = \frac{N.\sum dx.dy - (\sum dx).(\sum dy)}{\sqrt{N.\sum dx^2 - (\sum dx)^2}\ \sqrt{N.\sum dy^2 - (\sum dy)^2}}$$

d) Step Deviation Method

$$r = \frac{N.\sum dx'.dy' - (\sum dx').(\sum dy')}{\sqrt{N.\sum dx'^2 - (\sum dx')^2}\ \sqrt{N.\sum dy'^2 - (\sum dy')^2}}$$

ASSUMPTION OF KARL PEARSON COEFFICIENT OF CORRELATION

1. That the two variables are affected by a number of independent causes and form a normal distribution.

2. That the relationship between the two variables is linear.

3. The cause and effect relationship exists between the two variables.

4. If the error of measurement is reduced to the minimum the coefficient of Correlation is more reliable.

MERITS AND DEMERITS OF KARL PEARSON'S COEFFICIENT

Merits

1. It is the most important and popular method of measuring the relationship between variables, because it gives precise summary and quantitative figure which can be meaningfully interpreted.

2. Karl Pearson's Coefficient of Correlation gives direction (whether Positive or negative) as well as degree (high, moderate or low) of the relationship between the two variable.

Demerits

1. The value of the coefficient is affected by extreme items.

2. The Calculation process is time consuming.

3. Correlation lies between ±1. This yard stick needs a very careful implementation, otherwise it may be misinterpreted.

SPEARMAN'S RANK CORRELATION

Sometimes we come across statistical series in which the variables under consideration are not capable of quantitative measurement but can be arranged in serial order. This happens when we dealing with qualitative characteristic (attributes) such as honesty, beauty, character, morality, intelligence of student, leadership quality, Cooperation, etc.

They cannot be measured quantitatively (numerically) but can be arranged serially. In this situation Karl Pearson's coefficient of correlation cannot be used. Charles Edward Spearman, a British, Psychologist developed a formula in 1904 which consist of obtaining the correlation coefficient between ranks of N individuals in the two attributes under study called Coefficient of Correlation, by rank differences.

This method is applicable only to individual's observations rather than frequency distribution.

After assigning ranks to the various items, the differences of corresponding rank values are calculated and following formula is used.

$$rk = 1 - \frac{6\sum D^2}{N^3 - N}$$

rk = Coefficient of Rank correlation

$\sum D^2$ = The total of squares of the differences of corresponding ranks

N = The number of observation

Like Karl Pearson, the value of rk like between +1 and -1

MERITS AND DEMERITS OF THE RANK METHOD

Merits

1. It is easy to calculate and simple to understand as compared to Karl Pearson's Method.

2. When data are of qualitative nature like beauty, honesty, and intelligence etc., this method can be employed usefully.

3. If the values are not repeated, and answer obtained by Pearson's method and Rank Difference method will be same.

4. In case of actual data, rank method can be used to get rough degree of correlation.

Demerits

1. Rank correlation cannot be used for large series.

2. This method lacks precision as compared to Karl Pearson's method.

3. This method cannot be employed for finding Out Correlation in a grouped frequency distribution.

Another Formula for calculating when Rank Coefficient of correlation in case when ranks are equal and repeated.

$$rk = 1 - \frac{6\left[\sum D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \_\_\right]}{N^3 - N}$$

rk = Spearman's rank Correlation
D= Difference of ranks.
N: Number of pairs of observations.
m = Number of times the value repeated.