CORRELATION ANALYSIS

1. Meaning of Correlation
2. Types of correlation
    - Positive correlation
    - Negative correlation
    - Zero correlation (No correlation)
3. Methods of studying correlation
    - Scatter Diagram
    - Karl Pearson's coefficient of correlation
    - Spearman's rank correlation

CORRELATION ANALYSIS

Correlation analysis is the study of the degree of relationship between two variables in a bivariate distribution.

MEANING OF CORRELATION

Two variables are said to be correlated if the change in the values of one variable affects the other.

Consider the following examples:

1. Consider the height and weight of a group of people. It can be seen that as the height increases weight also increases. So there is a correlation between the height and weight.

2. When the price of a commodity increases, the demand shows a tendency to decreases. Here also the variables are correlated.

TYEPS OF CORRELATION

There are 3 types of correlation.

1. Positive correlation
2. Negative correlation
3. No correlation (Zero correlation)

1. Positive correlation

Two variables are said to be positively correlated if an increase in one variable results an increase in the other. Here the two variables move in the same direction.

Eg:- Height and weight, Income and expenditure. Speed and probability of happening an accident etc.

2. Negative correlation

Two variables are said to be negatively correlated if an increase in one variable results a decrease in the other or a decrease in one variable results an increase in the other. Here the two variables move in the opposite directions.

Eg :- Pressure and volume, Intensity of light and distance from the source, Price and demand, etc.

3. No correlation (Zero correlation)

Two variable are said to be in no correlation if a variable is not affected by the change in the other

Eg:- Height and intelligence, Size of shoes and scores obtained in an examination, etc.

METHODS OF STUDYING CORRELATION

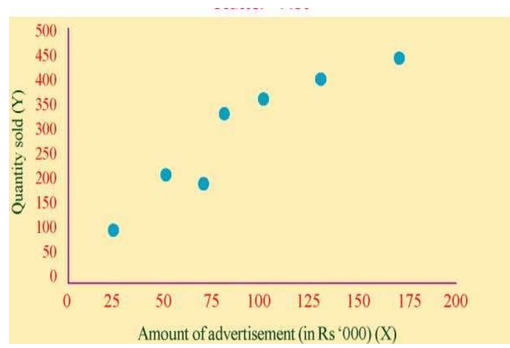
Here the three methods for studying correlation are discussed. They are:

1. Scatter Diagram
2. Karl Pearson's coefficient of correlation
3. Spearman's Rank correlation coefficient.

1. Scatter Diagram

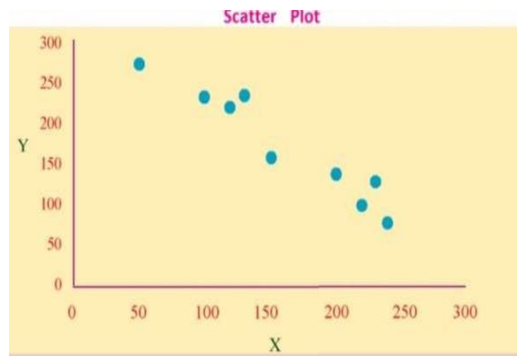
It is a graphical method for studying correlation. It is the simplest method for identifying the presence of correlation between two variables. A scatter diagram is prepared by plotting points in a XY plane using the given data.

Scatter diagram showing positive correlation

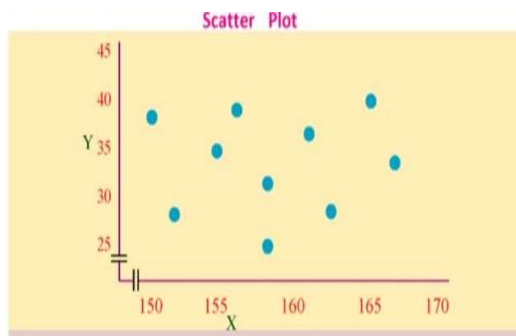


OGY Academy
Outstanding Guidance for Youth

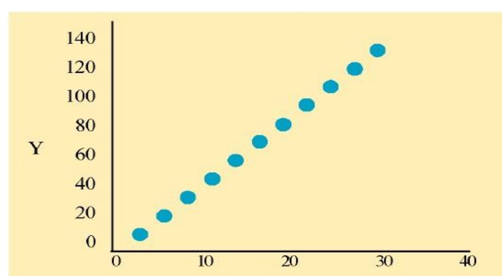
Scatter diagram showing negative correlation



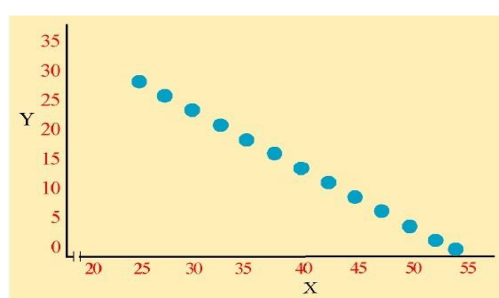
Scatter diagram showing no correlation



Scatter diagram showing perfect positive correlation



Scatter diagram showing perfect negative correlation



For Free Career Counselling :- +918891314091

Merits of scatter diagram

- Simple and attractive
- Easy to understand
- Gives a rough idea about the correlation at a glance
- Not influenced by extreme items.

Demerit of scatter diagram

- It does not give the exact degree of correlation.

2. Karl Pearson’s correlation coefficient

It is the most important method which gives an exact measure of correlation. This method was suggested by the famous statistician Karl Pearson.

The Karl Pearson’s coefficient of correlation between the variables X and Y is defined as:

$$r(X,Y) = \frac{Cov(X,Y)}{\sigma_X \times \sigma_Y}$$

$$= \frac{\frac{1}{n}\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\frac{1}{n}\sum\left(x-\bar{x}\right)^2 \times \frac{1}{n}\sum\left(y-\bar{y}\right)^2}}$$

The simplified form of this equation is given below.

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

(We can use this equation for problem solving when a bivariate data is given)

Properties of Karl Pearson’s coefficient of correlation

1. The correlation coefficient ta2he correlation coefficient takes any value from -1 to +1.

   That is, $-1 \le r \le 1$.

2. (i) The correlation coefficient is not affected if a constant is added to (subtracted from) the values of one (both) variable.

   (ii) The correlation coefficient is not affected if the values of one (both) variable are multiplied (divided) by a constant.

   That is, if x and y are two variables and $u = \frac{x-a}{c}$, $v = \frac{y-b}{d}$ then $r(x,y) = r(u,v)$.

3. $r(X,Y) = r(Y,X)$. That is correlation coefficient is symmetric with respect to variables.

4. Correlation coefficient between two independent variables is zero. But the converse need not be true.

Interpretation of Karl Pearson's coefficient of correlation

(i) If $r = +1$, the correlation is perfect positive.

(ii) If $r = -1$, the correlation is perfect negative.

(iii) If $r = 0$, the correlation is zero.

(iv) If $0 < r < +1$, the correlation is positive.

(v) If $-1 < r < 0$, the correlation is negative.

3. Spearman's Rank correlation coefficient

It is most useful when one or both the variables are qualitative or one or both variables are expressed in terms of ranks instead of the actual values. The values are converted into ranks before using this method. The formula for finding rank correlation coefficient is:

$$p = 1 - \frac{6\sum d^2}{n^3 - n}$$

Here d is the difference between ranks and n is the number of ranks.

The value of rank correlation coefficient also lies between -1 and +1. $ie, -1 \le p \le +1$. The interpretation of rank correlation coefficient is same as that of Karl Pearson's correlation coefficient.

Calculation of rank correlation coefficient when the ranks are repeated

There are situations in which two or more items to be assign equal ranks. Let us assume that the two items in a data are equal and they are in the second rank position. We cannot assume different ranks for them as they are equal. Also we cannot assign the same rank 2 to bath the items. Here we assume the average of the ranks to both of them. At the first we are assume ranks 2 and 3 to them and finally their ranks are equated to $\frac{2+3}{2} = 2.5$. Suppose three items are appearing in the position 5. Then each of them are given the rank $\frac{5+6+7}{3} = 6$. A correction factor (C. F) is to be added to the formula for finding rank correlation. The correction factor is given by

C. F = $\frac{\sum\left(m^3 - m\right)}{12}$, where m is the number of items with common rank.

If there are more than one such group of items with common rank, this value is added as many times as the number of such groups.

The formula for finding the rank correlation coefficient is:

$$p = 1 - \frac{6\left(\sum d^2 + C.F\right)}{n^3 - n}$$