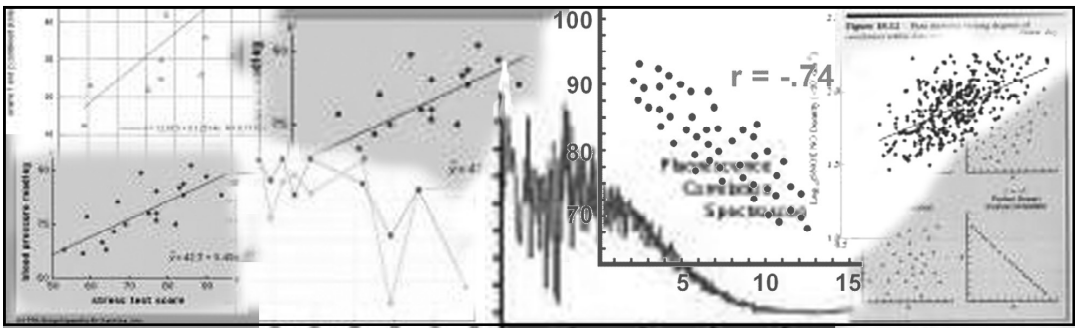
CHAPTER

7

Correlation



Studying this chapter should enable you to:

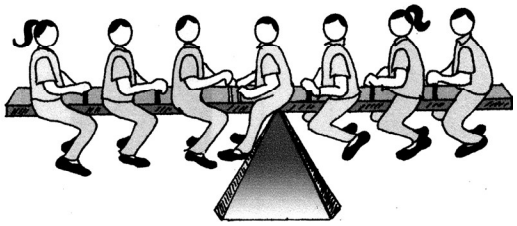
- *understand the meaning of the term correlation;*
- *understand the nature of relationship between two variables;*
- *calculate the different measures of correlation;*
- *analyse the degree and direction of the relationships.*

1. INTRODUCTION

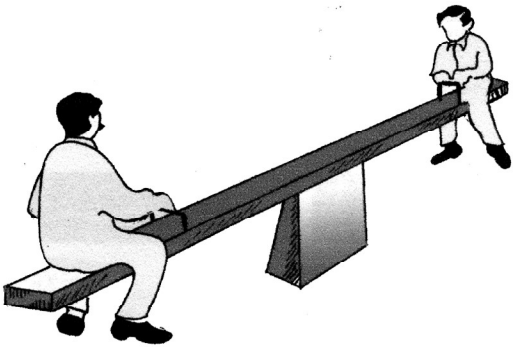
In previous chapters you have learnt how to construct summary measures out of a mass of data and changes among similar variables. Now you will learn how to examine the relationship between two variables.

As the summer heat rises, hill stations, are crowded with more and more visitors. Ice-cream sales become more brisk. Thus, the temperature is related to number of visitors and sale of ice-creams. Similarly, as the supply of tomatoes increases in your local *mandi*, its price drops. When the local harvest starts reaching the market, the price of tomatoes drops from a princely Rs 40 per kg to Rs 4 per kg or even less. Thus supply is related to price. Correlation analysis is a means for examining such relationships systematically. It deals with questions such as:

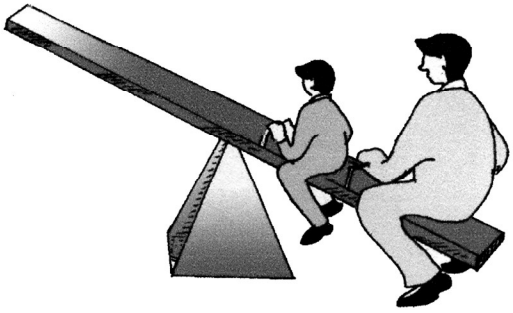
- Is there any relationship between two variables?



- If the value of one variable changes, does the value of the other also change?



- Do both the variables move in the same direction?



- How strong is the relationship?

2. Types of Relationship

Let us look at various types of relationship. The relation between movements in quantity demanded and the price of a commodity is an integral part of the theory of demand, which you will study in Class XII. Low agricultural productivity is related to low rainfall. Such examples of relationship may be given a cause and effect interpretation. Others may be just coincidence. The relation between the arrival of migratory birds in a sanctuary and the birth rates in the locality cannot be given any cause and effect interpretation. The relationships are simple coincidence. The relationship between size of the shoes and money in your pocket is another such example. Even if relationship exist, they are difficult to explain it.

In another instance a third variable's impact on two variables may give rise to a relation between the two variables. Brisk sale of ice-creams may be related to higher number of deaths due to drowning. The victims are not drowned due to eating of ice-creams. Rising temperature leads to brisk sale of ice-creams. Moreover, large number of people start going to swimming pools to beat the heat. This might have raised the number of deaths by drowning. Thus temperature is behind the high correlation between the sale of ice-creams and deaths due to drowning.

What Does Correlation Measure?

Correlation studies and measures the direction and intensity of relationship among variables. Correlation measures covariation, not causation. Correlation should never be

interpreted as implying cause and effect relation. The presence of correlation between two variables X and Y simply means that when the value of one variable is found to change in one direction, the value of the other variable is found to change either in the same direction (i.e. positive change) or in the opposite direction (i.e. negative change), but in a definite way. For simplicity we assume here that the correlation, if it exists, is linear, i.e. the relative movement of the two variables can be represented by drawing a straight line on graph paper.

Types of Correlation

Correlation is commonly classified into negative and positive correlation. The correlation is said to be positive when the variables move together in the same direction. When the income rises, consumption also rises. When income falls, consumption also falls. Sale of ice-cream and temperature move in the same direction. The correlation is negative when they move in opposite directions. When the price of apples falls its demand increases. When the prices rise its demand decreases. When you spend more time in studying, chances of your failing decline. When you spend less hours in your studies, chances of scoring low marks/grades increase. These are instances of negative correlation. The variables move in opposite direction.

3. Techniques for Measuring Correlation

Three important statistical tools used to measure correlation are scatter diagrams, Karl Pearson's coefficient of correlation and Spearman's rank correlation.

A scatter diagram visually presents the nature of association without giving any specific numerical value. A numerical measure of linear relationship between two variables is given by Karl Pearson's coefficient of correlation. A relationship is said to be linear if it can be represented by a straight line. Spearman's coefficient of correlation measures the linear association between ranks assigned to indiviual items according to their attributes. Attributes are those variables which cannot be numerically measured such as intelligence of people, physical appearance, honesty, etc.

Scatter Diagram

A scatter diagram is a useful technique for visually examining the form of relationship, without calculating any numerical value. In this technique, the values of the two variables are plotted as points on a graph paper. From a scatter diagram, one can get a fairly good idea of the nature of relationship. In a scatter diagram the degree of closeness of the scatter points and their overall direction enable us to examine the relation-

ship. If all the points lie on a line, the correlation is perfect and is said to be unity. If the scatter points are widely dispersed around the line, the correlation is low. The correlation is said to be linear if the scatter points lie near a line or on a line.

Scatter diagrams spanning over Fig. 7.1 to Fig. 7.5 give us an idea of the relationship between two variables. Fig. 7.1 shows a scatter around an upward rising line indicating the movement of the variables in the same direction. When X rises Y will also rise. This is positive correlation. In Fig. 7.2 the points are found to be scattered around a downward sloping line. This time the variables move in opposite directions. When X rises Y falls and vice versa. This is negative correlation. In Fig.7.3 there is no upward rising or downward sloping line around which the points are scattered. This is an example of no correlation. In Fig. 7.4 and Fig. 7.5 the points are no longer scattered around an upward rising or downward falling line. The points themselves are on the lines. This is referred to as perfect positive correlation and perfect negative correlation respectively.

Activity

- Collect data on height, weight and marks scored by students in your class in any two subjects in class X. Draw the scatter diagram of these variables taking two at a time. What type of relationship do you find?

A careful observation of the scatter diagram gives an idea of the nature and intensity of the relationship.

Karl Pearson's Coefficient of Correlation

This is also known as product moment correlation and simple correlation coefficient. It gives a precise numerical value of the degree of linear relationship between two variables X and Y. The linear relationship may be given by

$Y = a + bX$

This type of relation may be described by a straight line. The intercept that the line makes on the Y-axis is given by *a* and the slope of the line is given by *b*. It gives the change in the value of Y for very small change in the value of X. On the other hand, if the relation cannot be represented by a straight line as in

$Y = X^2$

the value of the coefficient will be zero. It clearly shows that zero correlation need not mean absence of any type of relation between the two variables.

Let $X_1, X_2, \ldots, X_N$ be N values of X and $Y_1, Y_2, \ldots, Y_N$ be the corresponding values of Y. In the subsequent presentations the subscripts indicating the unit are dropped for the sake of simplicity. The arithmetic means of X and Y are defined as

$$\bar{X} = \frac{\Sigma X}{N}; \qquad \bar{Y} = \frac{\Sigma Y}{N}$$

and their variances are as follows

$$\sigma^2_X = \frac{\Sigma(X - \bar{X})^2}{N} = \frac{\Sigma X^2}{N} - \bar{X}^2$$

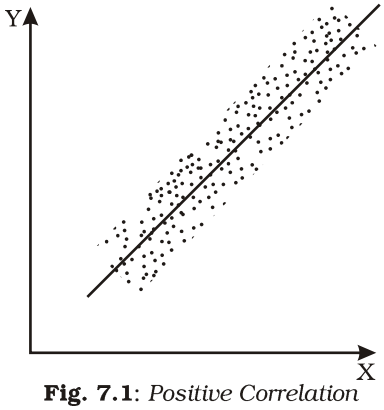


Fig. 7.1: *Positive Correlation*

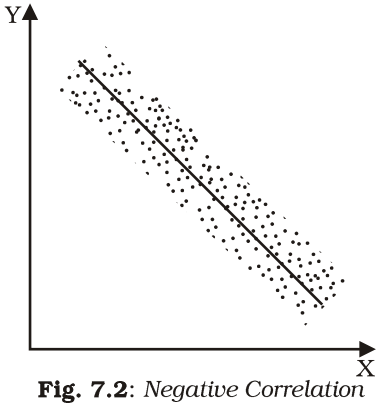


Fig. 7.2: *Negative Correlation*

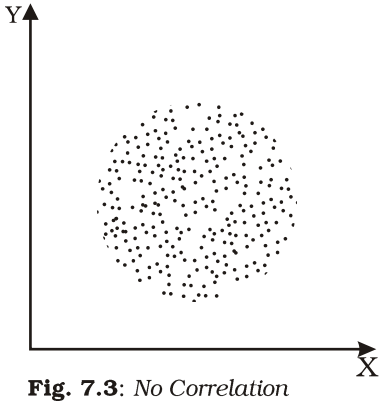


Fig. 7.3: *No Correlation*

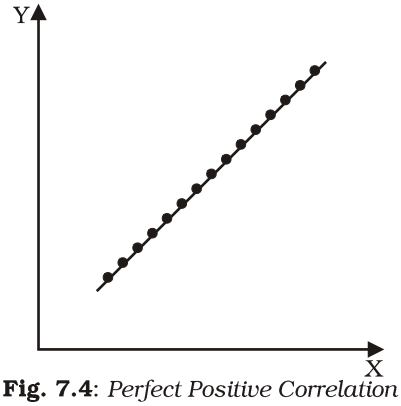


Fig. 7.4: *Perfect Positive Correlation*

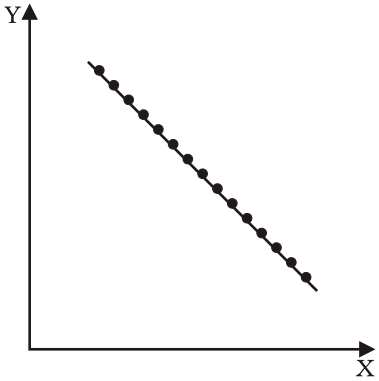


Fig. 7.5: Perfect Negative Correlation

and $\sigma^2_y = \frac{\Sigma(Y - \bar{Y})^2}{N} = \frac{\Sigma Y^2}{N} - \bar{Y}^2$

The standard deviations of X and Y respectively are the positive square roots of their variances. Covariance of X and Y is defined as

$$\text{Cov}(X,Y) = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{N} = \frac{\Sigma xy}{N}$$

Where $x = X - \bar{X}$ and $y = Y - \bar{Y}$ are the deviations of the i^{th} value of X and Y from their mean values respectively.

The sign of covariance between X and Y determines the sign of the correlation coefficient. The standard deviations are always positive. If the covariance is zero, the correlation coefficient is always zero. The product moment correlation or the Karl Pearson's measure of correlation is given by

$$r = \Sigma xy \Big/ N\sigma_x\sigma_y \quad ...(1)$$

or

$$r = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma(X - \bar{X})^2}\sqrt{\Sigma(Y - \bar{Y})^2}} \quad ...(2)$$

or

$$r = \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N}}{\sqrt{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}\sqrt{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}}} \quad ...(3)$$

or

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{N\Sigma X^2 - (\Sigma X)^2} \bullet \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}} \quad ...(4)$$

Properties of Correlation Coefficient

Let us now discuss the properties of the correlation coefficient

- *r* has no unit. It is a pure number. It means units of measurement are not part of *r*. *r* between height in feet and weight in kilograms, for instance, could be say 0.7.
- A negative value of *r* indicates an inverse relation. A change in one variable is associated with change in the other variable in the opposite direction. When price of a commodity rises, its demand falls. When the rate of interest rises the demand for funds also falls. It is because now funds have become costlier.



- If *r* is positive the two variables move in the same direction. When the price of coffee, a substitute of tea, rises the demand for tea also rises. Improvement in irrigation facilities is associated with higher yield. When temperature rises the sale of ice-creams becomes brisk.

- The value of the correlation coefficient lies between minus one and plus one, $-1 \leq r \leq 1$. If, in any exercise, the value of r is outside this range it indicates error in calculation.
- If $r = 0$ the two variables are uncorrelated. There is no linear relation between them. However other types of relation may be there.
- If $r = 1$ or $r = -1$ the correlation is perfect. The relation between them is exact.
- A high value of r indicates strong linear relationship. Its value is said to be high when it is close to +1 or –1.
- A low value of r indicates a weak linear relation. Its value is said to be low when it is close to zero.
- The magnitude of r is unaffected by the change of origin and change of scale. Given two variables X and Y let us define two new variables.

$$U = \frac{X - A}{B}; V = \frac{Y - C}{D}$$

where A and C are assumed means of X and Y respectively. B and D are common factors and of same sign. Then

$$r_{xy} = r_{uv}$$

This property is used to calculate correlation coefficient in a highly simplified manner, as in the step deviation method.

As you have read in chapter 1, the statistical methods are no substitute for common sense. Here, is another example, which highlights the need for understanding the data properly before correlation is calculated and interpreted. An epidemic spreads in some villages and the government sends a team of doctors to the affected villages. The correlation between the number of deaths and the number of doctors sent to the villages is found to be positive. Normally the health care facilities provided by the doctors are expected to reduce the number of deaths showing a negative correlation. This happened due to other reasons. The data relate to a specific time period. Many of the reported deaths could be terminal cases where the doctors could do little. Moreover, the benefit of the presence of doctors becomes visible only after some time. It is also possible that the reported deaths are not due to the epidemic. A tsunami suddenly hits the state and death toll rises.

Let us illustrate the calculation of r by examining the relationship between years of schooling of farmers and the annual yield per acre.

Example 1

No. of years of schooling of farmers	*Annual yield per acre in '000 (Rs)*
0	4
2	4
4	6
6	10
8	10
10	8
12	7

Formula 1 needs the value of $\Sigma xy, \sigma_x, \sigma_y$

From Table 7.1 we get,

$\Sigma xy = 42,$

$$\sigma_x = \sqrt{\frac{\Sigma(X - \bar{X})^2}{N}} = \sqrt{\frac{112}{7}},$$

$$\sigma_y = \sqrt{\frac{\Sigma(Y - \bar{Y})^2}{N}} = \sqrt{\frac{38}{7}}$$

Substituting these values in formula (1)

$$r = \frac{42}{7\sqrt{\frac{112}{7}}\ \sqrt{\frac{38}{7}}} = 0.644$$

The same value can be obtained from formula (2) also.

$$r = \frac{\Sigma (X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma (X - \bar{X})^2}\ \sqrt{\Sigma (Y - \bar{Y})^2}} \quad ...(2)$$

$$r = \frac{42}{\sqrt{112}\ \sqrt{38}} = 0.644$$

Thus years of education of farmers and annual yield per acre are positively correlated. The value of *r* is also large. It implies that more the number of years farmers invest in education, higher will be the yield per acre. It underlines the importance of farmers' education.

To use formula (3)

$$r = \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N}}{\sqrt{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}\ \sqrt{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}}} \quad ...(3)$$

the value of the following expressions have to be calculated i.e. $\Sigma XY, \Sigma X^2, \Sigma Y^2$.

Now apply formula (3) to get the value of *r*.

Let us know the interpretation of different values of *r*. The correlation coefficient between marks secured in English and Statistics is, say, 0.1. It means that though the marks secured in the two subjects are positively correlated, the strength of the relationship is weak. Students with high marks in English may be getting relatively low marks in statistics. Had the value of *r* been, say, 0.9, students with high marks in English will invariably get high marks in Statistics.

TABLE 7.1
Calculation of *r* between years of schooling of farmers and annual yield

Years of Education (X)	$(X-\bar{X})$	$(X-\bar{X})^2$	*Annual yield per acre in '000 Rs (Y)*	$(Y-\bar{Y})$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
0	–6	36	4	–3	9	18
2	–4	16	4	–3	9	12
4	–2	4	6	–1	1	2
6	0	0	10	3	9	0
8	2	4	10	3	9	6
10	4	16	8	1	1	4
12	6	36	7	0	0	0
$\Sigma X=42$		$\Sigma (X-\bar{X})^2=112$	$\Sigma Y=49$		$\Sigma (Y-\bar{Y})^2=38$	$\Sigma (X-\bar{X})(Y-\bar{Y})=42$

An example of negative correlation is the relation between arrival of vegetables in the local mandi and price of vegetables. If r is –0.9, vegetable supply in the local mandi will be accompanied by lower price of vegetables. Had it been –0.1 large vegetable supply will be accompanied by lower price, not as low as the price, when *r* is –0.9. The extent of price fall depends on the absolute value of *r*. Had it been zero there would have been no fall in price, even after large supplies in the market. This is also a possibility if the increase in supply is taken care of by a good transport network transferring it to other markets.

Activity

- Look at the following table. Calculate *r* between annual growth of national income at current price and the Gross Domestic Saving as percentage of GDP.

Step deviation method to calculate correlation coefficient.

When the values of the variables are large, the burden of calculation can be considerably reduced by using a property of *r*. It is that *r* is independent of change in origin and scale. It is also known as step deviation method. It involves the

TABLE 7.2

Year	*Annual growth of National Income*	*Gross Domestic Saving as percentage of GDP*
1992–93	14	24
1993–94	17	23
1994–95	18	26
1995–96	17	27
1996–97	16	25
1997–98	12	25
1998–99	16	23
1999–00	11	25
2000–01	8	24
2001–02	10	23

Source: *Economic Survey, (2004–05) Pg. 8,9*

transformation of the variables X and Y as follows:

$$U = \frac{X - A}{B}; V = \frac{Y - C}{D}$$

where A and B are assumed means, h and k are common factors and have same signs.

Then $r_{UV} = r_{XY}$

This can be illustrated with the exercise of analysing the correlation between price index and money supply.

Example 2

Price index (X)	120	150	190	220	230
Money supply in Rs crores (Y)	1800	2000	2500	2700	3000

The simplification, using step deviation method is illustrated below. Let A = 100; h = 10; B = 1700 and k = 100

The table of transformed variables is as follows:

Calculation of *r* between price index and money supply using step deviation method

TABLE 7.3

U $\left(\frac{X-100}{10}\right)$	V $\left(\frac{Y-1700}{100}\right)$	U^2	V^2	UV
2	1	4	1	2
5	3	25	9	15
9	8	81	64	72
12	10	144	100	120
13	13	169	169	169

$\Sigma U = 41$; $\Sigma U = 35$; $\Sigma U^2 = 423$;
$\Sigma V^2 = 343$; $\Sigma UV = 378$

Substituting these values in formula (3)

$$r = \frac{\Sigma UV - \frac{(\Sigma U)(\Sigma U)}{N}}{\sqrt{\Sigma U^2 - \frac{(\Sigma U)^2}{N}}\sqrt{\Sigma V^2 - \frac{(\Sigma V)^2}{N}}} \quad (3)$$

$$= \frac{378 - \frac{41 \times 35}{5}}{\sqrt{423 - \frac{(41)^2}{5}}\sqrt{343 - \frac{(35)^2}{5}}}$$

$= 0.98$

The strong positive correlation between price index and money supply is an important premise of monetary policy. When the money supply grows the price index also rises.

Activity

- Using data related to India's population and national income, calculate the correlation between them using step deviation method.

Spearman's rank correlation

Spearman's rank correlation was developed by the British psychologist C.E. Spearman. It is used when the variables cannot be measured meaningfully as in the case of price, income, weight etc. Ranking may be more meaningful when the measurements of variables are suspect. Consider the situation where we are required to calculate the correlation between height and weight of students in a remote village. Neither measuring rods nor weighing scales are available. The students can be easily ranked in terms of height and weight without using measuring rods and weighing scales.

There are also situations when you are required to quantify qualities such as fairness, honesty etc. Ranking may be a better alternative to quantification of qualities. Moreover, sometimes the correlation coefficient between two variables with extreme values may be quite different from the coefficient without the extreme values. Under these circumstances rank correlation provides a better alternative to simple correlation.

Rank correlation coefficient and simple correlation coefficient have the same interpretation. Its formula has

been derived from simple correlation coefficient where individual values have been replaced by ranks. These ranks are used for the calculation of correlation. This coefficient provides a measure of linear association between ranks assigned to these units, not their values. The Spearman's rank correlation formula is

$$r_a = 1 - \frac{6\Sigma D^2}{n^3 - n} \quad \ldots(4)$$

where *n* is the number of observations and D the deviation of ranks assigned to a variable from those assigned to the other variable. When the ranks are repeated the formula is

$$r_s = 1 - \frac{6\left[\Sigma D^2 + \frac{(m_1^3 - m_1)}{12} + \frac{(m_2^3 - m_2)}{12} + \ldots\right]}{n(n^2 - 1)}$$

where m_1, m_2, ..., are the number of repetitions of ranks and $\frac{m^3_1 - m_1}{12}$ ..., their corresponding correction factors. This correction is needed for every repeated value of both variables. If three values are repeated, there will be a correction for each value. Every time m_1 indicates the number of times a value is repeated.

All the properties of the simple correlation coefficient are applicable here. Like the Pearsonian Coefficient of correlation it lies between 1 and –1. However, generally it is not as accurate as the ordinary method. This is due the fact that all the information concerning the data is not utilised. The first differences of the values of items in the series, arranged in order of magnitude, are almost never constant. Usually the data cluster around the central values with smaller differences in the middle of the array. If the first differences were constant, then *r* and r_k would give identical results. The first difference is the difference of consecutive values. Rank correlation is preferred to Pearsonian coefficient when extreme values are present. In general

r_k is less than or equal to *r*.

The calculation of rank correlation will be illustrated under three situations.

1. The ranks are given.
2. The ranks are not given. They have to be worked out from the data.
3. Ranks are repeated.

Case 1: When the ranks are given

Example 3

Five persons are assessed by three judges in a beauty contest. We have to find out which pair of judges has the nearest approach to common perception of beauty.

	Competitors				
Judge	1	2	3	4	5
A	1	2	3	4	5
B	2	4	1	5	3
C	1	3	5	2	4

There are 3 pairs of judges necessitating calculation of rank correlation thrice. Formula (4) will be used.

$$r_s = 1 - \frac{6\Sigma D^2}{n^3 - n} \quad \ldots(4)$$

The rank correlation between A and B is calculated as follows:

A	*B*	*D*	D^2
1	2	–1	1
2	4	–2	4
3	1	2	4
4	5	–1	1
5	3	2	4
Total			14

Substituting these values in formula (4)

$$r_s = 1 - \frac{6\Sigma D^2}{n^3 - n} \quad \ldots(4)$$

$$= 1 - \frac{6 \times 14}{5^3 - 5} = 1 - \frac{84}{120} = 1 - 0.7 = 0.3$$

The rank correlation between A and C is calculated as follows:

A	*C*	*D*	D^2
1	1	0	0
2	3	–1	1
3	5	–2	4
4	2	2	4
5	4	1	1
Total			*10*

Substituting these values in formula (4) the rank correlation is 0.5. Similarly, the rank correlation between the rankings of judges B and C is 0.9. Thus, the perceptions of judges A and C are the closest. Judges B and C have very different tastes.

Case 2: When the ranks are not given

Example 4

We are given the percentage of marks, secured by 5 students in Economics and Statistics. Then the ranking has to be worked out and the rank correlation is to be calculated.

Student	*Marks in Statistics (X)*	*Marks in Economics (Y)*
A	85	60
B	60	48
C	55	49
D	65	50
E	75	55

Student	*Ranking in Statistics* (R_x)	*Ranking in Economics* (R_y)
A	1	1
B	4	5
C	5	4
D	3	3
E	2	2

Once the ranking is complete formula (4) is used to calculate rank correlation.

Case 3: When the ranks are repeated

Example 5

The values of X and Y are given as

X	25	45	35	40	15	19	35	42
Y	55	60	30	35	40	42	36	48

In order to work out the rank correlation, the ranks of the values are worked out. Common ranks are given to the repeated items. The

common rank is the mean of the ranks which those items would have assumed if they were slightly different from each other. The next item will be assigned the rank next to the rank already assumed. The formula of Spearman's rank correlation coefficient when the ranks are repeated is as follows

$$r_s = 1 - \frac{6\left[\Sigma D^2 + \frac{(m_1^3 - m_1)}{12} + \frac{(m_2^3 - m_2)}{12} + \ldots\right]}{n(n^2 - 1)}$$

where m_1, m_2, ..., are the number of repetitions of ranks and $\frac{m^3_1 - m_1}{12}$ ..., their corresponding correction factors.

X has the value 35 both at the 4th and 5th rank. Hence both are given the average rank i.e.,

$$\frac{4+5}{2}\text{th} = 4.5^{\text{th}} \text{ rank}$$

X	*Y*	*Rank of XR'*	*Rank of YR''*	*Deviation in Ranking D=R'–R''*	D^2
25	55	6	2	4	16
45	60	1	1	0	0
35	30	4.5	8	3.5	12.25
40	35	3	7	–4	16
15	40	8	5	3	9
19	42	7	4	3	9
35	36	4.5	6	–1.5	2.25
42	48	2	3	–1	1
Total				$\Sigma D^2 = 65.5$	

The necessary correction thus is

$$\frac{m^3 - m}{12} = \frac{2^3 - 2}{12} = \frac{1}{2}$$

Using this equation

$$r_s = 1 - \frac{6\left[\Sigma D^2 + \frac{(m^3 - m)}{12}\right]}{n^3 - n} \quad \ldots(5)$$

Substituting the values of these expressions

$$r_s = 1 - \frac{6(65.5 + 0.5)}{8^3 - 8} = 1 - \frac{396}{504}$$
$$= 1 - 0.786 = 0.214$$

Thus there is positive rank correlation between X and Y. Both X and Y move in the same direction. However, the relationship cannot be described as strong.

Activity

- Collect data on marks scored by 10 of your classmates in class IX and X examinations. Calculate the rank correlation coefficient between them. If your data do not have any repetition, repeat the exercise by taking a data set having repeated ranks. What are the circumstances in which rank correlation coefficient is preferred to simple correlation coefficient? If data are precisely measured will you still prefer rank correlation coefficient to simple correlation? When can you be indifferent to the choice? Discuss in class.

4. CONCLUSION

We have discussed some techniques for studying the relationship between

two variables, particularly the linear relationship. The scatter diagram gives a visual presentation of the relationship and is not confined to linear relations. Karl Pearson's coefficient of correlation and Spearman's rank correlation measure linear relationship among variables. When the variables cannot be measured precisely, rank correlation can be used. These measures however do not imply causation. The knowledge of correlation gives us an idea of the direction and intensity of change in a variable when the correlated variable changes.

Recap

- Correlation analysis studies the relation between two variables.
- Scatter diagrams give a visual presentation of the nature of relationship between two variables.
- Karl Pearson's coefficient of correlation r measures numerically only linear relationship between two variables. r lies between –1 and 1.
- When the variables cannot be measured precisely Spearman's rank correlation can be used to measure the linear relationship numerically.
- Repeated ranks need correction factors.
- Correlation does not mean causation. It only means covariation.

EXERCISES

1. The unit of correlation coefficient between height in feet and weight in kgs is
   (i) kg/feet
   (ii) percentage
   (iii) non-existent
2. The range of simple correlation coefficient is
   (i) 0 to infinity
   (ii) minus one to plus one
   (iii) minus infinity to infinity
3. If r_{xy} is positive the relation between X and Y is of the type
   (i) When Y increases X increases
   (ii) When Y decreases X increases
   (iii) When Y increases X does not change

4. If $r_{xy} = 0$ the variable X and Y are
   (i) linearly related
   (ii) not linearly related
   (iii) independent
5. Of the following three measures which can measure any type of relationship
   (i) Karl Pearson's coefficient of correlation
   (ii) Spearman's rank correlation
   (iii) Scatter diagram
6. If precisely measured data are available the simple correlation coefficient is
   (i) more accurate than rank correlation coefficient
   (ii) less accurate than rank correlation coefficient
   (iii) as accurate as the rank correlation coefficient
7. Why is r preferred to covariance as a measure of association?
8. Can *r* lie outside the –1 and 1 range depending on the type of data?
9. Does correlation imply causation?
10. When is rank correlation more precise than simple correlation coefficient?
11. Does zero correlation mean independence?
12. Can simple correlation coefficient measure any type of relationship?
13. Collect the price of five vegetables from your local market every day for a week. Calculate their correlation coefficients. Interpret the result.
14. Measure the height of your classmates. Ask them the height of their benchmate. Calculate the correlation coefficient of these two variables. Interpret the result.
15. List some variables where accurate measurement is difficult.
16. Interpret the values of *r* as 1, –1 and 0.
17. Why does rank correlation coefficient differ from Pearsonian correlation coefficient?
18. Calculate the correlation coefficient between the heights of fathers in inches (X) and their sons (Y)

X	65	66	57	67	68	69	70	72
Y	67	56	65	68	72	72	69	71

   (Ans. r = 0.603)
19. Calculate the correlation coefficient between X and Y and comment on their relationship:

X	–3	–2	–1	1	2	3
Y	9	4	1	1	4	9

   (Ans. r = 0)

20. Calculate the correlation coefficient between X and Y and comment on their relationship

X	1	3	4	5	7	8
Y	2	6	8	10	14	16

(Ans. r = 1)

Activity

- Use all the formulae discussed here to calculate r between India's national income and exports taking at least ten observations.