CHAPTER 6

Correlation Analysis

OBJECTIVES

After studying the material in this chapter, you should be able to :

- *Understand the concept of correlation between two variables.*
- *Identify different types of correlation.*
- *Understand the notion of correlation coefficient.*
- *Interpret the value of coefficient of correlation.*
- *Discuss various methods of computing the correlation coefficient.*
- *Compute correlation coefficient for bivariate frequency distribution.*
- *Appreciate properties of correlation coefficient.*
- *Know the merits and demerits of different methods of studying correlation.*
- *Calculate probable error and interpret its value.*
- *Understand the concept of coefficient of determination.*

6.1 INTRODUCTION

Thus far we have examined numerical methods used to describe various characteristics of a univariate data, *i.e.*, the data involving only one variable. The reader may recall that in univariate data only one variable is associated with each unit of observation. However, we may have data in which more than one variable can be associated with each unit of observation. For example, for providing information about the marks obtained by the students of a class in two subjects, say Statistics and Economics, we can associate two variables, one representing the marks in Statistics and the other marks in Economics to each unit of observation, namely, a student in the class. When we have two variables for which values are being observed for each unit of observation, we say that we have *bivariate data*. In general, the study of those data which involve more than two variables are termed as *multivariate data*.

Two variables are said to be *correlated* if the change in one variable is accompanied by a change in the other. For example, if X represents the price of a product and Y represents the demand for that product, than we would expect large values of X to correspond to small values of Y and small values of X to correspond to large values of Y. Hence we can say that price and demand of a product are correlated. *Correlation analysis* is a statistical procedure by which we determine the degree of association or relationship between two or more variables. That is, in correlation analysis, the purpose is to measure the strength or closeness of the relationship between the variables. For example, we might find a high degree of relationship between the price of a product and consumer demand for that product. Correlation is said to be *linear* if the amount of change in one variable tends to bear a constant ratio to the amount of change in the other variable.

In this chapter we shall consider the problem of measuring the linear relationship involving two variables only. The study of such a problem is called the *simple linear correlation*.

6.2 CORRELATION : SOME DEFINITIONS

In the following we shall give some important definitions of correlation.

1. *If two or more quantities vary in sympathy so that movements in one tend to be accompanied by corresponding movements in the other (s), then they are said to be correlated.*
   — L.R. Connor
2. *Correlation analysis attempts to determine the degree of relationship between variables.*
   — Ya Lun Chou
3. *Correlation is an analysis of the covariation between two or more variables.*
   — A.M. Tuttle
4. *Correlation analysis deals with the association between two or more variables.*
   — Simpson and Kafka
5. *When the relationship is of quantitative nature, the appropriate statistical tool for discovering and measuring the relationship and expressing it in a brief formula is known as correlation.*
   — Croxton and Cowden
6. *Correlation means that between two series or group of data there exists some casual connection.*
   — W.I. King
7. *When a group of items are recorded with respect to the values of two distinct variables and it is found that pairs of values tend to be associated, the two variables are said to be correlated.*
   — Wessel and Willet

6.3 TYPES OF CORRELATION

The following are different types of correlation :

(i) *Positive and Negative Correlation.*

(ii) *Simple, Partial and Multiple Correlation.*

(iii) *Linear and Non-linear Correlation.*

(i) Positive and Negative Correlation

The correlation between two variables is said to be **positive or direct** if an increase (or a decrease) in one variable corresponds to an increase (or a decrease) in the other.

For example, if X represents the amount of money spent annually on advertising by a firm and Y represents the total annual sales, then we might expect an increase (or a decrease) in the advertising budget to be accompanied by an increase (or decrease) in the annual sales. Thus we can say that the correlation between the advertising budget and the total sales is positive.

The correlation between two variables is said to be **negative or inverse** if an increase (or a decrease) in one variable corresponds to a decrease (or an increase) in the other.

For example, if X represents the price of a product and Y represents the demand for that product, then we would expect large values of X to correspond to small values of Y and small values of X to correspond to large values of Y. Hence we can say that the correlation between the price and demand of a product is negative.

(ii) Simple, Partial and Multiple Correlation

The study of simple, partial and multiple correlation is based upon the number of variables involved.

Simple Correlation : It involves the study of only two variables. That is, in simple correlation, we measure the degree of association or relationship between two variables only. For example, when we study the correlation between the price and demand of a product, it is a problem of simple correlation.

Partial Correlation : It involves the study of three or more variables, but consider only two variables to be influencing each other, the effect of other influencing variables being kept constant. Thus in partial correlation we measure the degree of relationship between the variable Y and one of the variables $X_1, X_2, \ldots, X_n$ with the effect of all the other variables removed. For example, if we consider three variables, namely yield of wheat, amount of rainfall and amount of fertilizers and limit our correlation analysis to yield and rainfall, with the effect of fertilizers removed, it becomes a problem relating to partial correlation only.

Multiple Correlation : It involves the study of three or more variables simultaneously. Thus in multiple correlation we measure the degree of relationship between the variable Y and all the variables $X_1, X_2, \ldots, X_n$ taken together. For example, if we study the relationship between the yield of wheat per acre and both amount of rainfall and the amount of fertilizers used, it becomes a problem relating to multiple correlation.

(iii) Linear and Non-linear Correlation

Linear Correlation : The correlation between two variables is said to be *linear* if the amount of change in one variable tends to bear a constant ratio to the amount of change in the other variable. For example, consider the following data:

X	:	10	20	30	40	50
Y	:	40	80	120	160	200

We observe that the ratio of changes between the two variables is same and hence the correlation between X and Y is linear. It may be remarked that if the values of two variables, which are linearly correlated, are plotted on a graph paper all the plotted points would be on a straight line.

Definition : The quantitative measure of strength in the linear relationship between two variables is called the **correlation coefficient.** It is denoted by *r*.

Thus the correlation coefficient *r* measures the extent to which the points cluster about a straight line. The correlation coefficient ranges from +1 to – 1. If two variables have no linear relationship, the correlation between them is zero. Consequently, the more correlation differs from zero, the stronger the linear relationship between the two variables.

The following table shows degrees of correlation according to various values of *r*:

TABLE 6.1

Degree of Correlation	*Positive*	*Negative*
Perfect correlation	+ 1	– 1
Very high degree of correlation	+ 0.9 to + 1	– 0.9 to – 1
Fairly high degree of correlation	+ 0.75 to + 0.9	– 0.75 to – 0.9
Moderate degree of correlation	+ 0.50 to + 0.75	– 0.50 to – 0.75
Low degree of correlation	+ 0.25 to + 0.50	– 0.25 to – 0.50
Very low degree of correlation	0 to + 0.25	– 0.25 to 0
No correlation	0	0

6.6 METHODS OF STUDYING CORRELATION

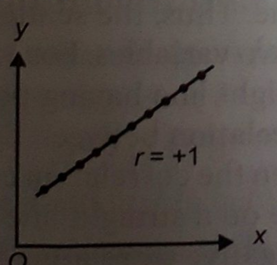
We shall discuss the following methods of measuring the linear relationship between two variables:

(i) Scatter Diagram Method,
(ii) Karl Pearson's Coefficient of Correlation,
(iii) Rank Correlation Method, and
(iv) Concurrent Deviation Method.

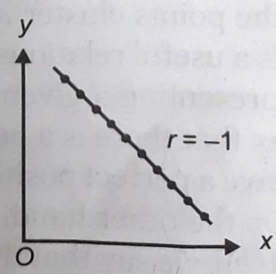
6.7 SCATTER DIAGRAM METHOD

A *scatter diagram* is a graphical presentation of bivariate data $\{(X_i, Y_i) : i = 1, 2, ..., n\}$ on two quantitative variables *X* and *Y* that allows us to show two variables together, one on each axis, each pair being represented by a point on the graph as in coordinate geometry.

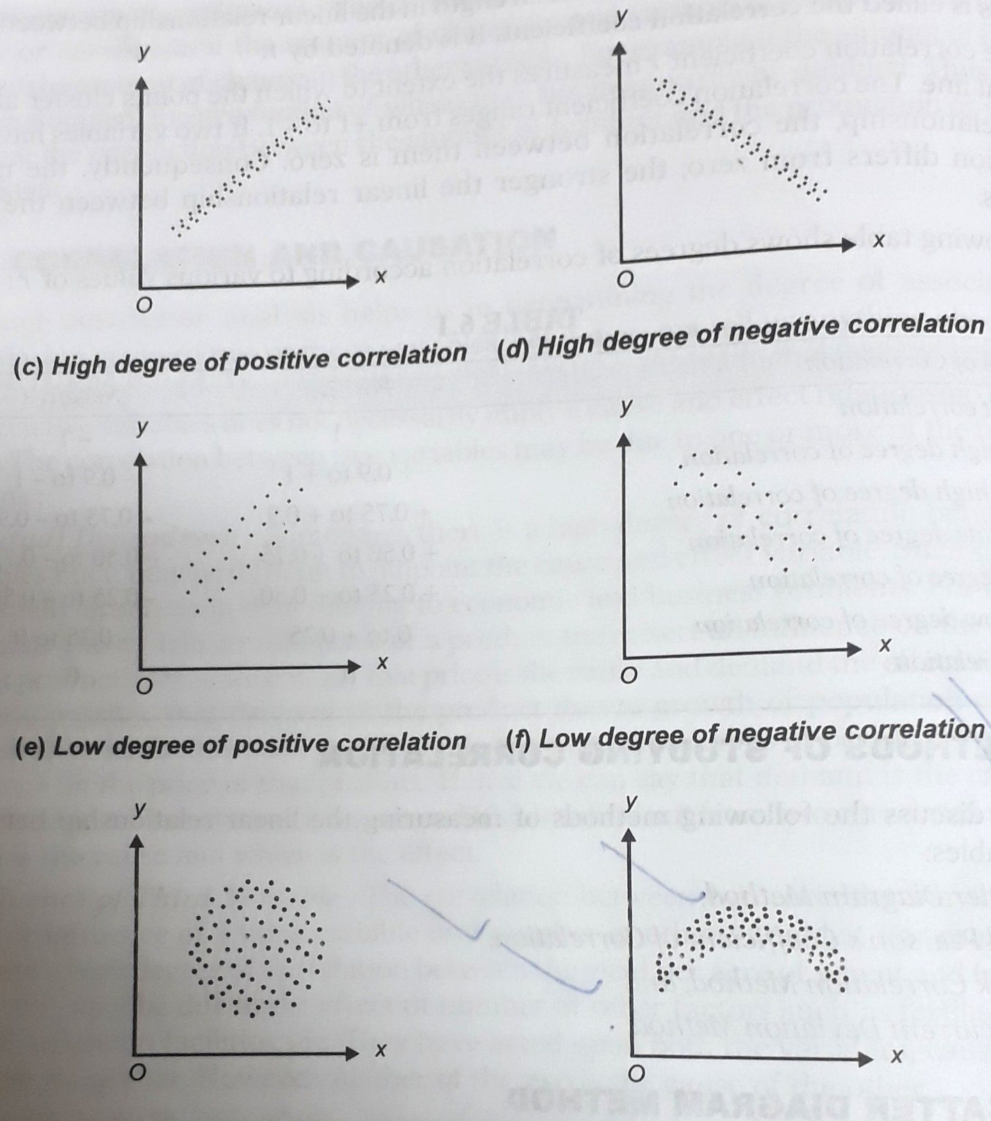
SCATTER DIAGRAMS SHOWING VARIOUS DEGREES OF CORRELATION



(a) Perfect positive correlation



(b) Perfect negative correlation



(c) High degree of positive correlation (d) High degree of negative correlation

(e) Low degree of positive correlation (f) Low degree of negative correlation

(g) No correlation (h) No correlation

Fig. 6.1

The scatter diagram is the simplest method of measuring the linear relationship between two variables. By constructing a scatter diagram for the n pairs of observations (X_1, Y_1), (X_2, Y_2), . . ., (X_n, Y_n) on two variables, we can draw certain conclusions concerning the extent to which the points cluster about a straight line. Thus, the scatter diagram helps us to see if there is a useful relationship between the two variables. For example, if all the plotted points representing a given data lie on a straight line having positive slope [*see* Fig. 6.1 (*a*)], we say that there is a **perfect positive correlation** between the two variables. If two variables have a perfect positive correlation, then the correlation coefficient would be equal to + 1. On the other hand, if all the points lie on a straight line having negative slope [*see* Fig. (6.1(*b*)], we say that there is a **perfect negative correlation** between the two variables. If two variables have a perfect negative correlation, then the correlation coefficient would be equal to –1. If the plotted points are all close to a straight line having positive slope [*see* Fig. 6.1(*c*)], we say that there is a **high degree of positive correlation** between the two variables. Similarly, if the plotted points are all close to a straight line having negative slope [*see* Fig. 6.1 (*d*)], we say that there is a **high degree of negative correlation.** In fact, the closer the correlation coefficient is to ±1, the stronger the linear relationship between the variables. If the plotted points are widely scattered over a straight line having positive slope [*see* Fig. 6.1 (*e*)], we say that there a low degree of positive correlation between the two variables. Similarly, if the plotted points are widely scattered over a straight line have negative slope [*see* Fig. 6.1 (*f*)], we say that there is a low degree of negative correlation. If the points follow a strictly random pattern as in Fig. 6.1 (*g*) and Fig. 6.1 (*h*), we have a zero correlation and conclude that no linear relationship exists between the two variables.

EXAMPLE 1. *Draw a scatter diagram to represent the following data and interpret it.*

X :	4	5	6	7	8	9	10	11	12	13	14	15
Y :	78	72	66	60	54	48	42	36	30	24	18	12

[Delhi Univ. B.Com. 1978]

SOLUTION. The scatter diagram is shown in Fig. 6.2.

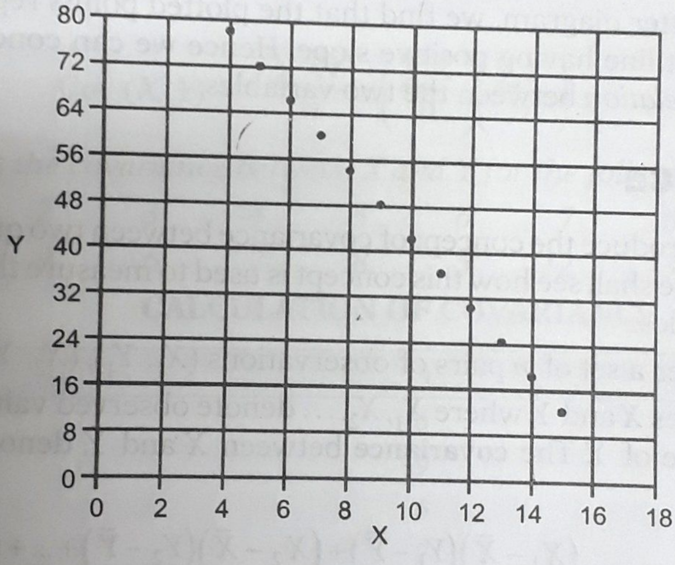


Fig. 6.2. Scatter Diagram

From the above scatter diagram, we find that the plotted points representing the given data lie on a straight line having negative slope. Hence we can conclude that there is a *perfect negative correlation* between the two variables.

EXAMPLE 2. *Draw a scatter diagram for the following data:*

X :	8	10	12	11	9	7	13	14	15	17	16
Y :	5	7	9	8	6	4	10	11	12	14	13

Also describe the relationship between X and Y. **[Delhi Univ. B.Com. 1979]**

SOLUTION. The scatter diagram is shown in Fig. 6.3.

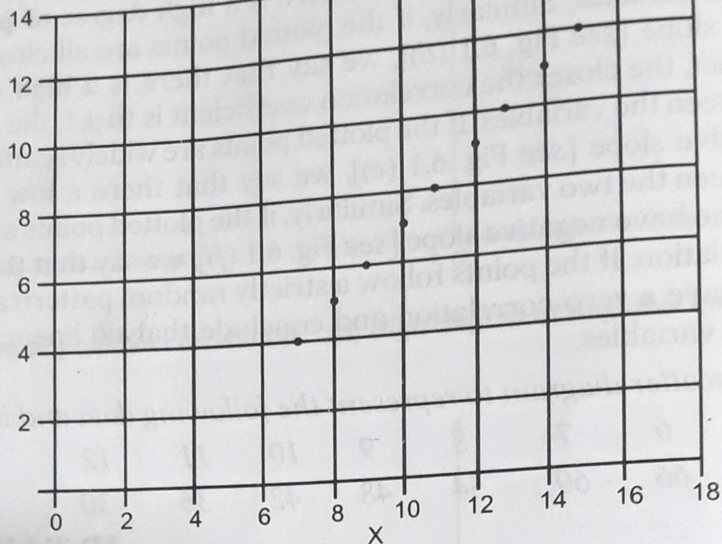


Fig. 5.3. Scatter Diagram

From the above scatter diagram, we find that the plotted points representing the given data lie on a straight line having positive slope. Hence we can conclude that there is a *perfect positive correlation* between the two variables.

6.8 COVARIANCE

In this section we introduce the concept of covariance between two quantitative variables. In the next section we shall see how this concept is used to measure the linear relationship between two variables.

Definition : Consider a set of n pairs of observations $(X_1, Y_1), (X_2, Y_2), \ldots (X_n, Y_n)$ on two quantitative variables X and Y, where $X_1, X_2, \ldots$ denote observed values of the variable X, and $Y_1, Y_2, \ldots$ those of Y. The **covariance** between X and Y, denoted by Cov (X, Y), is defined as

$$\text{Cov}(X, Y) = \frac{(X_1-\bar{X})(Y_1-\bar{Y})+(X_2-\bar{X})(Y_2-\bar{Y})+\ldots+(X_n-\bar{X})(Y_n-\bar{Y})}{n}$$

$$= \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{n} = \frac{\Sigma xy}{n}$$

where $\bar{X} = \frac{\Sigma X}{n}$, $\bar{Y} = \frac{\Sigma X}{n}$, $x = X - \bar{X}$ and $y = Y - \bar{Y}$

EXAMPLE 3. *Find Cov (X, Y) between X and Y if*

X	:	3	4	5	6	7
Y	:	8	7	6	5	4

SOLUTION.

CALCULATION OF COVARIANCE

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	$xy = (X - \bar{X})(Y - \bar{Y})$
3	8	−2	2	−4
4	7	−1	1	−1
5	6	0	0	0
6	5	1	−1	−1
7	4	2	−2	−4
$\sum X = 25$	$\sum Y = 30$			$\sum xy = -10$

Here $n = 5$, $\bar{X} = \dfrac{\sum X}{n} = \dfrac{25}{5} = 5$ and $\bar{Y} = \dfrac{\sum Y}{n} = \dfrac{30}{5} = 6$

$$\therefore \quad \text{Cov}(X, Y) = \frac{\sum xy}{n} = \frac{-10}{5} = -2.$$

Another Formula for Cov (X, Y) : We now give a slightly different formula (proof omitted) for calculating the covariance. This formula is particularly useful when $\bar{X}$ or $\bar{Y}$ is not an integer. The formula is :

$$\text{Cov}(X, Y) = \frac{\sum XY}{n} - \left(\frac{\sum X}{n}\right)\left(\frac{\sum Y}{n}\right) \quad ...(1)$$

EXAMPLE 4. *Calculate the covariance between X and Y for the following data :*

X :	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Y :	*6*	*9*	*6*	*7*	*8*	*5*	*12*	*3*	*17*	*1*

SOLUTION.

CALCULATION OF COVARIANCE

X	Y	XY
1	6	6
2	9	18
3	6	18
4	7	28
5	8	40
6	5	30
7	12	84
8	3	24
9	17	153
10	1	10
$\sum X = 55$	$\sum Y = 74$	$\sum XY = 411$

We have $n = 10$, $\sum X = 55$, $\sum Y = 74$ and $\sum XY = 411$

$$\therefore \quad \text{Cov}(X, Y) = \frac{\sum XY}{n} - \left(\frac{\sum X}{n}\right)\left(\frac{\sum Y}{n}\right) = \frac{411}{10} - \frac{55}{10} \times \frac{74}{10}$$

$$= 41.1 - 5.5 \times 7.4 = 41.1 - 40.7 = 0.4.$$

EXAMPLE 5. *Find the covariance between X and Y, given that $\sum X = 60$, $\sum Y = 90$, $\sum XY = 574$, and $n = 10$.*

SOLUTION. We have
$$\text{Cov}(X, Y) = \frac{\sum XY}{n} - \left(\frac{\sum X}{n}\right)\left(\frac{\sum Y}{n}\right) = \frac{574}{10} - \left(\frac{60}{10}\right)\left(\frac{90}{10}\right)$$

$$= 57.4 - 6 \times 9 = 57.4 - 54 = 3.4.$$

REMARK. It may be remarked that formula (1) for computing covariance is effective if the values of X or/and Y are small. However, if the values of X or/and Y are large, the calculation of covariance by means or Formula (1) is quite tedious and time consuming. In such a case, we use the following method, called *step deviation method.*

Let $\quad u = X - A \quad$ and $\quad v = Y - B$

where A and B are arbitrary constants. Then,

$$\text{Cov}(X, Y) = \frac{\sum uv}{n} - \left(\frac{\sum u}{n}\right)\left(\frac{\sum v}{n}\right) \quad \text{...(2)}$$

This formula, in fact, shows that covariance is independent of change of origin.

EXAMPLE 6. *Find the covariance between X and Y for the following data :*

X :	66	67	68	69	70	71	72
Y :	68	65	70	70	69	70	69

SOLUTION. We shall use the following formula :

$$\text{Cov}(X, Y) = \frac{\sum uv}{n} - \left(\frac{\sum u}{n}\right)\left(\frac{\sum v}{n}\right)$$

where $u = X - A$ and $v = Y - B$. Taking $A = 69$ and $B = 70$, we prepare the following table :

TABLE 6.2. CALCULATION OF COVARIANCE

X	$u = X - 69$	Y	$v = Y - 70$	uv
66	– 3	68	– 2	6
67	– 2	65	– 5	10
68	– 1	70	0	0
69	0	70	0	0
70	1	69	– 1	– 1
71	2	70	0	0
72	3	69	– 1	– 3
	$\sum u = 0$		$\sum v = -9$	$\sum uv = 12$

$$\therefore \quad \text{Cov}(X, Y) = \frac{\sum uv}{n} - \left(\frac{\sum u}{n}\right)\left(\frac{\sum v}{n}\right) = \frac{12}{7} - \left(\frac{0}{7}\right)\left(\frac{-9}{7}\right) = 1.7.$$

EXERCISE 6.1

1. Define correlation. Explain various types of correlation with suitable examples. **[*Delhi Univ. B.Com. 2000*]**
2. Explain the meaning and significance of the concept of correlation. How would you interpret the value of the coefficient of correlation? **[*Delhi Univ. B.Com. 1987*]**
3. Does correlation signify the existence of cause and effect relationship between two variables? **[*Delhi Univ. B.Com. 1977, 2004*]**
4. What is scatter diagram and how is it useful in the study of correlation? **[*Delhi Univ. B.Com. 1995, 2005*]**
5. How does a scatter diagram help in ascertaining the degree of correlation between two variables? **[*Delhi Univ. B.Com. 1994*]**
6. Distinguish between positive and negative correlation with the help of a scatter diagram. **[*Delhi Univ. B.Com. 1973*]**
7. Illustrate a perfect negative correlation on a scatter diagram. **[*Delhi Univ. B.Com. 1998*]**
8. Draw a scatter diagram for the following data:

Height (in inches) :	62	72	70	60	67	70	64	65	60	70
Weight (in lbs) :	50	65	63	52	56	60	59	58	54	65

Also indicate whether correlation is positive or negative. **[*Delhi Univ. B.Com. 1977*]**

9. Following data gives the heights and weights of 10 students in a class:

Height (in inches) :	72	60	63	66	70	65	58	78	72	62
Weight (in kgs) :	65	54	55	61	60	54	50	63	65	50

Draw a scatter diagram and indicate whether the correlation is positive or negative.

10. Following data represents the marks obtained by 10 students in Mathematics and Statistics

Mathematics :	15	18	21	24	27	30	36	39	42	48
Statistics :	25	25	27	27	31	33	35	41	41	45

Draw a scatter diagram and indicate whether the correlation is positive or negative.

11. Find Cov (X, Y) between X and Y if

X :	1	2	3	4	5
Y :	2	4	6	8	10

12. Find the covariance between X and Y for the data:

X :	64	65	66	67	68	69	70
Y :	66	67	65	68	70	68	72

13. Find the covariance between X and Y for the data:

X :	1	2	3	4	5	6	7	8	9	10
Y :	10	9	8	8	6	12	4	3	18	1

14. Find the covariance between X and Y for the data:

Mamo

X	:	1	2	3	4	5	6	7	8	9	10
Y	:	2	4	8	7	10	5	14	16	2	20

15. Find Cov (X, Y) between X and Y if : $\sum X = 55, \sum Y = 74, \sum XY = 411, n = 10.$

16. Find Cov (X, Y) between X and Y if : $\sum X = 50, \sum Y = 50, \sum XY = 280, n = 10.$

17. Find Cov (X, Y) between X and Y if : $\sum X = 50, \sum Y = -30, \sum XY = -115, n = 10.$

18. Find Cov (X, Y) between X and Y if : $\sum X = 15, \sum Y = 36, \sum XY = 110, n = 15.$

19. State, giving reasons, whether the following statements are true or false.

(*i*) Negative correlation in two series means that as the value of one of the variables decreases, the value of the other variable would also decrease. **[*Delhi Univ. B.Com.* 1984]**

(*ii*) Coefficient of correlation between the two variables must be in the same units as the original data. **[*Delhi. Univ. B.Com.* 1983]**

ANSWERS

8. Positive	**9.** Positive	**10.** Positive	**11.** 4
12. 3.57	**13.** – 2.45	**14.** 10.2	**15.** 0.4
16. 3	**17.** 3.5	**18.** 4.93	
19. (*i*) False	(*ii*) False		

6.9 KARL PEARSON'S COEFFICIENT OF CORRELATION

The **Karl Pearson's coefficient of correlation**, also called the *Pearson's product - moment correlation coefficient*, is the most widely used method of measuring the linear correlation between two variables.

Definition. The ***Karl Pearson's coefficient of correlation*** **between** two variables X and Y, denoted by $\rho(X, Y)$ or r, is defined by

$$\rho(X, Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var } X}.\sqrt{\text{Var } Y}} = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y},$$

where Var X and Var Y are the variances of the values of X and Y respectively, while σ_X and σ_Y are their standard deviations.

REMARK : It may be remarked that the correlation coefficient r is a measure of the linear relationship between the two variables X and Y. That is, r measures the extent to which the points of the scatter diagram cluster about a straight line. Hence a value of $r = 0$ implies a lack of linearity between the two variables and not a lack of association.

Characteristics of the Correlation Coefficient

1. The correlation coefficient r is independent of change of origin and scale. That is, if

the variables X and Y are replaced by the variables $U = aX + b$ and $V = cY + d$, then we have

$$\rho(U, V) = \rho(aX + b, cY + d) = \rho(X, Y),$$

where a and c are positive constants. However, if a and c are arbitrary constants, then

$$\rho(aX + b, cY + d) = \frac{a \times c}{|a| \times |c|}\, \rho(X, Y)$$

2. The value of r ranges from -1 to $+1$. That is, $-1 \le r \le +1$.
3. If $r = 1$, then all the points of the scatter diagram lie on a straight line having positive slope and we say that a perfect positive linear relationship exists between the two variables. Similarly, if $r = -1$, then all the points of the scatter diagram lie on a straight line having negative slope and we say that a perfect negative linear relationship exists between the two variables.
4. If r is close to $+1$, then all the points of the scatter diagram follow closely a straight line having positive slope and we say that a high positive correlation exists between the two variables. Similarly, if r is close to -1, then all the points of the scatter diagram follow closely a straight line having negative slope and we say that a high negative correlation exists between the two variables.
5. If r is close to 0, the linear relationship between the two variables is weak or perhaps non-existent.

EXAMPLE 7. (a) *"If the correlation coefficient between two variables X and Y is positive, then the coefficient of correlation between – X and – Y is also positive". Comment.*

[Delhi Univ. B.A. (Econ. Hons) 1996]

(b) *The correlation coefficient between two variables X and Y is found to be 0.4. What is the correlation coefficient between 2X and (– Y)?*

[Delhi Univ. B.A. (Econ. Hons) 1997]

(c) *"If the coefficient of correlation between two variables X and Y is 0.8, then the coefficient of correlation between – X and – Y is – 0.8" Comment.*

[Delhi Univ. B.Com. (H) 1997, 2010]

SOLUTION. We know that

$$\rho(aX, cY) = \frac{a \times c}{|a| \times |c|}\, \rho(X, Y) \quad \text{...(1)}$$

(a) We are given : $\rho(X, Y) > 0$. Using (1), we get

$$\rho(-X, -Y) = \frac{(-1) \times (-1)}{|-1| \times |-1|}\, \rho(X, Y) = \rho(X, Y)$$

Thus, if $\rho(X, Y)$ is positive, then so is $\rho(-X, -Y)$.

(b) We are given : $\rho(X, Y) = 0.4$. Using (1), we get

$$\rho(2X, -Y) = \frac{2\times(-1)}{|2|\times|-1|}\rho(X, Y) = \frac{-2}{2\times 1}\rho(X, Y)$$

$$= -\rho(X, Y) = -0.4.$$

Thus, the correlation coefficient between $2X$ and $-Y$ is -0.4.

(*c*) We are given : $\rho(X, Y) = 0.8$. Using (1), we get

$$\rho(-X, -Y) = \frac{(-1)\times(-1)}{|-1|\times|-1|}\rho(X, Y) = \rho(X, Y) = 0.8$$

Thus, the coefficient of correlation between $-X$ and $-Y$ is also 0.8.

EXAMPLE 8. *The covariance between the length and weight of five items is 6 and their standard deviations are 2.45 and 2.61 respectively. Find the coefficient of correlation between length and weight.* **[Delhi Univ. B.Com. (H) 2000]**

SOLUTION. If we let X denote the length and Y denote the weight, then we are given :

$$\text{Cov}(X, Y) = 6, \quad \sigma_X = 2.45 \quad \text{and} \quad \sigma_Y = 2.61$$

$\therefore$ The coefficient of correlation, r, between X and Y is given by

$$r = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{6}{2.45\times 2.61} = \frac{6}{6.3945} = 0.9383.$$

EXAMPLE 9. *If covariance of 10 pairs of items is 7, variance of X is 36, $\sum(Y-\bar{Y})^2 = 90$. find out r.* **[Delhi Univ. B.Com. 2004]**

SOLUTION. We are given : $\text{Cov}(X, Y) = 7$

$$\text{Var}(X) = 36 \Rightarrow \sigma_X = \sqrt{\text{Var } X} = 6$$

$$\sum(Y-\bar{Y})^2 = 90 \quad \text{and} \quad n = 10$$

$$\therefore \quad \sigma_Y = \sqrt{\frac{\sum(Y-\bar{Y})^2}{n}} = \sqrt{\frac{90}{10}} = 3$$

$$\therefore \quad r = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{7}{6\times 3} = \frac{7}{18} = 0.39 \text{ (app.)}$$

EXAMPLE 10. *The coefficient of correlation between two variables X and Y is 0.3 and their covariance is 9. If the variance of X series is 16, find the standard deviation of Y series.*

SOLUTION. The coefficient of correlation between X and Y is given by

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

Given $r = 0.3$, $\text{Cov}(X, Y) = 9$, $\text{Var}(X) = 16 \Rightarrow \sigma_X = \sqrt{\text{Var } X} = 4$

$$\therefore \quad 0.3 = \frac{9}{4 \cdot \sigma_Y} \Rightarrow \sigma_Y = \frac{9}{1.2} = 7.5.$$

EXAMPLE 11. *The coefficient of correlation between two variables X and Y is 0.4 and their covariance is 10. If the variance of X series is 9, find the second moment about mean of Y series.* **[Delhi Univ. B.Com. (H) 1996]**

SOLUTION. In terms of usual notations, we are given

$$r = 0.4, \quad \text{Cov}(X, Y) = 10, \quad \text{Var}(X) = 9 \Rightarrow \sigma_X = \sqrt{\text{Var } X} = 3$$

The coefficient of correlation between X and Y is given by

$$r = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} \Rightarrow 0.4 = \frac{10}{3 \times \sigma_Y} \Rightarrow \sigma_Y = \frac{10}{1.2} = 8.33$$

Thus the second moment about mean of Y series is :

$$\mu_2 = \text{Var } Y = \sigma_Y^2 = (8.33)^2 = 69.39.$$

EXAMPLE 12. *Given the following information :*

$$r = 0.8 \quad \sum xy = 60, \quad \sigma_Y = 2.5 \quad \text{and} \quad \sum x^2 = 90$$

where x and y are the deviations from the respective means, find the number of items (n). **[Delhi Univ. B.Com. (H) 1993, 2002]**

SOLUTION. We have $\quad \text{Cov}(X, Y) = \dfrac{\sum(X - \bar{X})(Y - \bar{Y})}{n} = \dfrac{\sum xy}{n} = \dfrac{60}{n}$

$$\sigma_X = \sqrt{\frac{\sum (X - \bar{X})^2}{n}} = \sqrt{\frac{\sum x^2}{n}} = \sqrt{\frac{90}{n}}$$

Applying the following formula

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

for the coefficient of correlation, we obtain

$$0.8 = \frac{60/n}{\sqrt{\frac{90}{n}} \times 2.5} \Rightarrow 2 = \frac{60}{n} \times \frac{\sqrt{n}}{\sqrt{90}} = \frac{60}{\sqrt{n}\sqrt{90}}$$

$$\Rightarrow \quad \sqrt{n}\sqrt{90} = 30 \Rightarrow n \times 90 = 900 \Rightarrow n = 10.$$

EXAMPLE 13. *Calculate Karl Pearson's coefficient of correlation from the following information and comment on the result :*

Standard deviation of X series	= 10
Standard deviation of Y series	= 12
Arithmetic mean of X series	= 25
Arithmetic mean of Y series	= 35
Summation of product of deviation from actual arithmetic means of two series	= 24
Number of observations	= 20

[Delhi Univ. B.Com. 2004]

SOLUTION. We are given : $n = 20, \quad \bar{X} = 25, \quad \bar{Y} = 35,$

$$\Sigma(X-\bar{X})(Y-\bar{Y}) = 24, \quad \sigma_X = 10 \quad \text{and} \quad \sigma_Y = 12$$

$$\therefore \qquad \text{Cov}(X, Y) = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{n} = \frac{24}{20} = 1.2$$

The Karl Pearson's coefficient of correlation is given by

$$r = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{1.2}{10 \times 12} = +0.01.$$

Comment. Since the value of r is close to 0, the correlation between X and Y series is negligible.

6.10 COMPUTING THE CORRELATION COEFFICIENT

We will now present various formulas for computing the correlation coefficient between the two variables X and Y.

The correlation coefficient, r, between two variables X and Y is given by

$$r = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}\,X}.\sqrt{\text{Var}\,Y}} \qquad \text{... (1)}$$

If the variable X takes on the values $X_1, X_2, ..., X_n$ and the variable Y takes on the values $Y_1, Y_2, ..., Y_n$, then we know that

$$\text{Cov}(X, Y) = \frac{1}{n}\Sigma(X-\bar{X})(Y-\bar{Y})$$

$$\text{Var}(X) = \frac{1}{n}\Sigma(X-\bar{X})^2 \quad \text{and} \quad \text{Var}(Y) = \frac{1}{n}\Sigma(Y-\bar{Y})^2$$

Substituting these values in (1), we obtain another formula for r:

$$r = \frac{\frac{1}{n}\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\frac{1}{n}\Sigma(X-\bar{X})^2}.\sqrt{\frac{1}{n}\Sigma(Y-\bar{Y})^2}} = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2}.\sqrt{\Sigma(Y-\bar{Y})^2}} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}} \qquad \text{... (2)}$$

where $x = X - \bar{X}$ and $y = Y - \bar{Y}$.

Normally, we do not use formula (2) for computing the correlation coefficient r. The calculations for computing r are rendered much easier by using the following formula:

$$r = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{n}}{\sqrt{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}.\sqrt{\Sigma Y^2 - \frac{(\Sigma Y)^2}{n}}} = \frac{n\Sigma XY - \Sigma X . \Sigma Y}{\sqrt{n\Sigma X^2 - (\Sigma X)^2} . \sqrt{n\Sigma Y^2 - (\Sigma Y)^2}} \qquad \text{.... (3)}$$

We now give yet another formula for computing r which is very useful when the values of X or/and Y are very large. This formula uses the fact that the correlation coefficient is independent of the change of origin and scale. Thus if we define

$$U = \frac{X-A}{h} \quad \text{and} \quad V = \frac{Y-B}{k}$$

where A and B are arbitrary constants, h and k are positive constants, then

$$r = \rho(X, Y) = \rho(U, V)$$

and hence by making use of formula (3), we have

$$r = \frac{\sum UV - \frac{\sum U \sum V}{n}}{\sqrt{\sum U^2 - \frac{(\sum U)^2}{n}} \cdot \sqrt{\sum V^2 - \frac{(\sum V)^2}{n}}} = \frac{n\sum UV - \sum U . \sum V}{\sqrt{n\sum U^2 - (\sum U)^2} \cdot \sqrt{n\sum V^2 - (\sum V)^2}} \quad \text{.... (4)}$$

EXAMPLE 14. *From the following data, compute coefficient of correlation between X and Y:*

	X Series	*Y Series*
No. of items	*15*	*15*
Arithmetic Mean	*25*	*18*
Sum of the squares of deviations from Arithmetic Mean	*136*	*138*
Summation of products of deviations of X and Y series from their respective arithmetic means	*122*	

[Delhi Univ. B.Com. 1989, 2005, 2006]

SOLUTION. In terms of usual notations, we are given :

$$n = 15, \quad \bar{X} = 25, \quad \bar{Y} = 18, \quad \sum(X-\bar{X})^2 = 136,$$

$$\sum(Y-\bar{Y})^2 = 138, \quad \text{and} \quad \sum(X-\bar{X})(Y-\bar{Y}) = 122$$

OR

$$\therefore \quad r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2 \times \sum(Y-\bar{Y})^2}} = \frac{122}{\sqrt{136 \times 138}} = \frac{122}{136.996} = 0.89$$

Hence the coefficient of correlation is 0.89.

EXAMPLE 15. *Calculate coefficients of correlation from the following results :*

$$n = 10, \quad \sum X = 100, \quad \sum Y = 150, \quad \sum(X-10)^2 = 180,$$

$$\sum(Y-15)^2 = 215 \quad and \quad \sum(X-10)(Y-15) = 60$$

SOLUTION. The arithmetic means of X and Y series are given by

$$\bar{X} = \frac{\sum X}{n} = \frac{100}{10} = 10 \quad \text{and} \quad \bar{Y} = \frac{\sum Y}{n} = \frac{150}{10} = 15$$

$\therefore$ If we let $x = X - \bar{X} = X - 10$ and $y = Y - \bar{Y} = Y - 15$, then

$$\sum x^2 = \sum(X - \bar{X})^2 = \sum(X - 10)^2 = 180 \quad \text{(given)}$$

$$\sum y^2 = \sum(Y - \bar{Y})^2 = \sum(Y - 15)^2 = 215 \quad \text{(given)}$$

$$\sum xy = \sum(X - \bar{X})\sum(Y - \bar{Y}) = \sum(X - 10)\ \sum(Y - 15) = 60 \quad \text{(given)}$$

$$\therefore \qquad r = \frac{\sum xy}{\sqrt{\sum x^2 \times \sum y^2}} = \frac{60}{\sqrt{180 \times 215}} = \frac{60}{196.72} = 0.305.$$

EXAMPLE 16. *Calculate correlation coefficient from the following results :*

$$n = 10,\ \sum X = 140,\ \sum Y = 150,\ \sum(X - 10)^2 = 180,$$

$$\sum(Y - 15)^2 = 215 \ \text{and} \ \sum(X - 10)(Y - 15) = 60$$

[Delhi Univ. B.Com.(H) 2007]

SOLUTION. Let $U = X - 10$ and $V = Y - 15$. Then

$$\sum U = \sum(X - 10) = \sum X - 10\ n = 140 - 10 \times 10 = 40$$

$$\sum V = \sum(Y - 15) = \sum Y - 15\ n = 150 - 15 \times 10 = 0$$

$$\sum U^2 = \sum(X - 10)^2 = 180, \quad \sum V^2 = \sum(Y - 15)^2 = 215$$

$$\sum UV = \sum(X - 10)(Y - 15) = 60$$

The coefficient of correlation, r, is given by

$$r = \rho(X, Y) = \rho(U, V)$$

$$= \frac{n\sum UV - \sum U . \sum V}{\sqrt{n\sum U^2 - (\sum U)^2}\ .\sqrt{n\sum V^2 - (\sum V)^2}}$$

$$= \frac{10 \times 60 - 40 \times 0}{\sqrt{10 \times 180 - 1600}\ .\ \sqrt{10 \times 215 - 0}} = \frac{600}{\sqrt{200} \times \sqrt{2150}}$$

$$= \frac{600}{655.74} = 0.915.$$

EXAMPLE 17. *The deviations from the respective means of X and Y series are given below:*

x	:	–4	–3	–2	–1	0	1	2	3	4
y	:	3	–3	–4	0	4	1	2	–2	–1

Calculate the Karl Pearson's coefficient of correlation from the above data.

[Delhi Univ. B.Com. 1995]

SOLUTION. **CALCULATION OF r**

x	y	x^2	y^2	xy
– 4	3	16	9	– 12
– 3	– 3	9	9	9
– 2	– 4	4	16	8
– 1	0	1	0	0
0	4	0	16	0
1	1	1	1	1
2	2	4	4	4
3	– 2	9	4	– 6
4	– 1	16	1	– 4
		$\sum x^2 = 60$	$\sum y^2 = 60$	$\sum xy = 0$

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

$$= \frac{0}{\sqrt{60 \times 60}} = 0.$$

EXAMPLE 18. *From the following data, calculate Karl Pearson's coefficient of correlation:*

Height of Fathers (in inches) :	66	68	69	72	65	59	62	67	61	71
Height of Sons (in inches) :	65	64	67	69	64	60	59	68	60	64

[*Delhi Univ. B.Com. 2005*]

SOLUTION. **CALCULATION OF COEFFICIENT OF CORRELATION**

Height of father X	$X - \bar{X}$ x	x^2	*Height of son* Y	$Y - \bar{Y}$ y	y^2	xy
66	0	0	65	1	1	0
68	2	4	64	0	0	0
69	3	9	67	3	9	9
72	6	36	69	5	25	30
65	– 1	1	64	0	0	0
59	– 7	49	60	– 4	16	28
62	– 4	16	59	– 5	25	20
67	1	1	68	4	16	4
61	– 5	25	60	– 4	16	20
71	5	25	64	0	0	0
$\sum X = 660$		$\sum x^2 = 166$	$\sum Y = 640$		$\sum y^2 = 108$	$\sum xy = 111$

Here $n = 10$, $\bar{X} = \frac{\sum X}{n} = \frac{660}{10} = 66$ and $\bar{Y} = \frac{\sum Y}{n} = \frac{640}{10} = 64$

$$\therefore \quad r = \frac{\sum xy}{\sqrt{\sum x^2 \times \sum y^2}} = \frac{111}{\sqrt{166 \times 108}} = \frac{111}{133.89} = 0.829.$$

EXAMPLE 19. *Making use of the data summarized below calculate the coefficient of correlation,* r_{12} :

Case	X_1	X_2	Case	X_1	X_2
A	10	9	E	12	11
B	6	4	F	13	13
C	9	6	G	11	8
D	10	9	H	9	4

[Delhi Univ. B.Com. 1982]

SOLUTION. **CALCULATION OF COEFFICIENT OF CORRELATION**

Case	X_1	$X_1-\bar{X}_1$ x_1	x_1^2	X_2	$X_2-\bar{X}_2$ x_2	x_2^2	x_1x_2
A	10	0	0	9	1	1	0
B	6	− 4	16	4	− 4	16	16
C	9	− 1	1	6	− 2	4	2
D	10	0	0	9	1	1	0
E	12	2	4	11	3	9	6
F	13	3	9	13	5	25	15
G	11	1	1	8	0	0	0
H	9	− 1	1	4	− 4	16	4
	$\sum X_1 = 80$		$\sum x_1^2 = 32$	$\sum X_2 = 64$		$\sum x_2^2 = 72$	$\sum x_1x_2 = 43$

Here $n=8$, $\bar{X}_1 = \frac{\sum X_1}{n} = \frac{80}{8} = 10$ and $\bar{X}_2 = \frac{\sum X_2}{n} = \frac{64}{8} = 8$

$$\therefore \qquad r_{12} = \frac{\sum x_1 x_2}{\sqrt{\sum x_1^2 \times \sum x_2^2}} = \frac{43}{\sqrt{32 \times 72}} = \frac{43}{48} = 0.896.$$

EXAMPLE 20. *Calculate the correlation coefficient from the following data :*

X :	12	9	8	10	11	13	7
Y :	14	8	6	9	11	12	3

Let now each value of X be multiplied by 2 and then 6 be added to it. Similarly, multiply each value of Y be 3 and subtract 2 from it. What will be the correlation coefficient between the new series of X and Y?

[C.A. (Foundation), May 1997]

SOLUTION. **CALCULATION OF COEFFICIENT OF CORRELATION**

X	$X-\bar{X}$ x	x^2	Y	$Y-\bar{Y}$ y	y^2	xy
12	2	4	14	5	25	10
9	– 1	1	8	– 1	1	1
8	– 2	4	6	– 3	9	6
10	0	0	9	0	0	0
11	1	1	11	2	4	2
13	3	9	12	3	9	9
7	– 3	9	3	– 6	36	18
$\sum X = 70$		$\sum x^2 = 28$	$\sum Y = 63$		$\sum y^2 = 84$	$\sum xy = 46$

Here $n=7$, $\bar{X} = \dfrac{\sum X}{n} = \dfrac{70}{7} = 10$ and $\bar{Y} = \dfrac{\sum Y}{n} = \dfrac{63}{7} = 9$

$$\therefore \qquad r(X, Y) = \frac{\sum xy}{\sqrt{\sum x^2 \times \sum y^2}} = \frac{46}{\sqrt{28 \times 84}} = \frac{46}{48.5} = 0.948$$

We know that if the variables X and Y are replaced by the variables $U = aX + b$ and $V = cY + d$, where a, b, c and d are constants, $a > 0$, $c > 0$, then $r(U, V) = r(X, Y)$.

Using this fact, we find that the coefficient of correlation between the new series of X and Y is also 0.948.

EXAMPLE 21. *From the following data, calculate Karl Pearson's coefficient of correlation :*

X :	6	2	10	4	8
Y :	9	11	?	8	7

Arithmetic means of X and Y series are 6 and 8 respectively.

SOLUTION. First we find the missing value of Y series.

Given : $$\bar{Y} = 8 \Rightarrow \frac{\sum Y}{5} = 8 \Rightarrow \sum Y = 8 \times 5 = 40$$

$$\therefore \qquad \text{missing value} = \sum Y - (9 + 11 + 8 + 7) = 40 - 35 = 5$$

CALCULATION OF COEFFICIENT OF CORRELATION

X	$X-\bar{X}$ x	x^2	Y	$Y-\bar{Y}$ y	y^2	xy
6	0	0	9	+ 1	1	0
2	– 4	16	11	+ 3	9	– 12
10	+ 4	16	5	– 3	9	– 12
4	– 2	4	8	0	0	0
8	+ 2	4	7	– 1	1	– 2
		$\sum x^2 = 40$			$\sum y^2 = 20$	$\sum xy = -26$

$$\therefore \quad r = \frac{\sum xy}{\sqrt{\sum x^2 \times \sum y^2}} = \frac{-26}{\sqrt{40 \times 20}} = \frac{-26}{28.284} = -0.9192.$$

EXAMPLE 22. *Calculate Karl Pearson's coefficient of correlation between the variables X and Y using the following data :*

X :	25	40	30	25	10	5	10	15	30	20
Y :	10	25	40	15	20	40	28	22	15	5

[Delhi Univ. B.Com. 2006]

SOLUTION. **CALCULATION OF COEFFICIENT OF CORRELATION**

X	x $(X-\bar{X})$	x^2	Y	y $(Y-\bar{Y})$	y^2	xy
25	4	16	10	− 12	144	− 48
40	19	361	25	3	9	57
30	9	81	40	18	324	162
25	4	16	15	− 7	49	− 28
10	− 11	121	20	− 2	4	22
5	− 16	256	40	18	324	− 288
10	− 11	121	28	6	36	− 66
15	− 6	36	22	0	0	0
30	9	81	15	− 7	49	− 63
20	− 1	1	5	− 17	289	17
$\sum X = 210$		$\sum x^2 = 1090$	$\sum Y = 220$		$\sum y^2 = 1228$	$\sum xy = -235$

Here $n = 10$, $\bar{X} = \frac{\sum X}{n} = \frac{210}{10} = 21$ and $\bar{Y} = \frac{\sum Y}{n} = \frac{220}{10} = 22$

$$\therefore \quad r = \frac{\sum xy}{\sqrt{\sum x^2 \times \sum y^2}} = \frac{-235}{\sqrt{1090} \times \sqrt{1228}} = -0.203 \text{ (app.)}$$

ALITER : Using $r = \dfrac{n\sum UV - \sum U \cdot \sum V}{\sqrt{n\sum U^2 - (\sum U)^2} \cdot \sqrt{n\sum V^2 - (\sum V)^2}}$,

where $U = X - A$ and $V = Y - B$.

Taking $A = 20$, $B = 20$, we prepare the following table :

X	$U = X - A$	U^2	Y	$V = Y - B$	V^2	UV
25	5	25	10	– 10	100	– 50
40	20	400	25	5	25	100
30	10	100	40	20	400	200
25	5	25	15	– 5	25	– 25
10	– 10	100	20	0	0	0
5	– 15	225	40	20	400	– 300
10	– 10	100	28	8	64	– 80
15	– 5	25	22	2	4	– 10
30	10	100	15	– 5	25	– 50
20	0	0	5	– 15	225	0
	$\sum U = 10$	$\sum U^2 = 1100$		$\sum V = 20$	$\sum V^2 = 1268$	$\sum UV = -215$

$$r = \frac{n\sum UV - \sum U . \sum V}{\sqrt{n\sum U^2 - (\sum U)^2} . \sqrt{n\sum V^2 - (\sum V)^2}}$$

$$= \frac{10(-215) - 10 \times 20}{\sqrt{10 \times 1100 - 100} . \sqrt{10 \times 1268 - 400}}$$

$$= \frac{-2350}{\sqrt{10900}\sqrt{12280}} = \frac{-2350}{11569.44} = -0.203 \text{ (app.)},$$

which is same as before.

EXAMPLE 23. *The total of the multiplication of deviation of X and Y = 3044. No of pairs of observations = 10. Total of deviations of X = – 170. Total of deviations of Y = – 20. Total of the squares of deviations of X = 8288. Total of the squares of deviations of Y = 2264. Find out the coefficient of correlation when the arbitrary means of X and Y are 82 and 68 respectively.* **[Delhi Univ. B.Com. 2001]**

SOLUTION. In terms of usual notations, we are given :

$n = 10, \quad \sum UV = 3044, \quad \sum U = -170, \quad \sum V = -20, \quad \sum U^2 = 8288, \quad \sum V^2 = 2264.$

Applying the formula : $r = \dfrac{n\sum UV - (\sum U).(\sum V)}{\sqrt{n\sum U^2 - (\sum U)^2} . \sqrt{n\sum V^2 - (\sum V)^2}}$, we get

$$r = \frac{10 \times 3044 - (-170)(-20)}{\sqrt{10 \times 8288 - (-170)^2} . \sqrt{10 \times 2264 - (-20)^2}}$$

$$r = \frac{30440 - 3400}{\sqrt{82880 - 28900}\ \sqrt{22640 - 400}}$$

$$= \frac{27040}{\sqrt{53980}\ \sqrt{22240}} = +0.78.$$

EXAMPLE 24. *Calculate coefficient of correlation from the following data :*

X :	*10,000*	*20,000*	*30,000*	*40,000*	*50,000*	*60,000*	*70,000*
Y :	*0.3*	*0.5*	*0.6*	*0.8*	*1.0*	*1.1*	*1.3*

SOLUTION. Since coefficient of correlation is independent of change of origin and scale, therefore if we let

$$U = X/10{,}000 \quad \text{and} \quad V = 10Y$$

Then
$$r = \rho(X, Y) = \rho(U, V)$$

Making use of the above fact, we prepare the following table :

CALCULATION OF COEFFICIENT OF CORRELATION

X	$U = X/10{,}000$	U^2	Y	$V = 10Y$	V^2	UV
10,000	1	1	0.3	3	9	3
20,000	2	4	0.5	5	25	10
30,000	3	9	0.6	6	36	18
40,000	4	16	0.8	8	64	32
50,000	5	25	1.0	10	100	50
60,000	6	36	1.1	11	121	66
70,000	7	49	1.3	13	169	91
	$\sum U = 28$	$\sum U^2 = 140$		$\sum V = 56$	$\sum V^2 = 524$	$\sum UV = 270$

$$\therefore \quad r = \frac{n\sum UV - (\sum U).(\sum V)}{\sqrt{n\sum U^2 - (\sum U)^2}\ .\sqrt{n\sum V^2 - (\sum V)^2}}$$

$$= \frac{7 \times 270 - 28 \times 56}{\sqrt{7 \times 140 - 784}\ .\sqrt{7 \times 524 - 3136}}$$

$$= \frac{1890 - 1568}{\sqrt{196} \times \sqrt{532}} = \frac{322}{322.91} = +0.997.$$

EXAMPLE 25. *Find Karl Pearson's coefficient of correlation between the age and the playing habits of the people from the following information :*

Age groups (in years)	*No. of people*	*No. of players*
15 and less than 20	200	150
20 and less than 25	270	162
25 and less than 30	340	170
30 and less than 35	360	180
35 and less than 40	400	180
40 and less than 45	300	120

15-20
20-25
25-30
30-

[Delhi Univ. B.Com. (H) 2006 (C.C.)]

SOLUTION. It is required to find the coefficient of correlation between the age and the playing habits. We must, therefore, express the number of players in terms of percentage by applying :

$$\text{Percentage of players} = \frac{\text{No. of players}}{\text{No. of people}} \times 100$$

Let this be denoted by Y and mid-point of class interval by X.

CALCULATION FOR COEFFICIENT OF CORRELATION

Age group (in years)	*Mid-point X*	*No. of people*	*No. of players*	*% of players Y*	$U = \frac{X - 32.5}{5}$	$V = \frac{Y - 50}{5}$	U^2	V^2	UV
15 and less than 20	17.5	200	150	$\frac{150}{200} \times 100 = 75$	−3	5	9	25	−15
20 and less than 25	22.5	270	162	$\frac{162}{270} \times 100 = 60$	−2	2	4	4	−4
25 and less than 30	27.5	340	170	$\frac{170}{340} \times 100 = 50$	−1	0	1	0	0
30 and less than 35	32.5	360	180	$\frac{180}{360} \times 100 = 50$	0	0	0	0	0
35 and less than 40	37.5	400	180	$\frac{180}{400} \times 100 = 45$	1	−1	1	1	−1
40 and less than 45	42.5	300	120	$\frac{120}{300} \times 100 = 40$	2	−2	4	4	−4
					$\Sigma U = -3$	$\Sigma V = 4$	$\Sigma U^2 = 19$	$\Sigma V^2 = 34$	$\Sigma UV = -24$

Σfm

$x = n$

Karl Pearson's coefficient of correlation between the age and the playing habits is given by

$$r = \frac{n\sum UV - (\sum U)(\sum V)}{\sqrt{n\sum U^2 - (\sum U)^2} \cdot \sqrt{n\sum V^2 - (\sum V)^2}}$$

$$= \frac{6\times(-24)-(-3)\times 4}{\sqrt{6\times 19-(-3)^2} \cdot \sqrt{6\times 34-(4)^2}}$$

$$= \frac{-144+12}{\sqrt{114-9} \cdot \sqrt{204-16}} = \frac{-132}{\sqrt{105} \cdot \sqrt{188}} = -0.9395,$$

which indicates a high degree of negative correlation between the age and the playing habits.

EXAMPLE 26. *With the following data in 6 cities calculate the coefficient of correlation by Pearson's method between the density of population and death rate.*

Cities	*Area in Sq. miles*	*Population in '000*	*No. of deaths*
A	150	30	300
B	180	90	1440
C	100	40	560
D	60	42	840
E	120	72	1224
F	80	24	312

SOLUTION. It is required to find the coefficient of correlation between density of population and death rate :

(*i*) Density of Population $= \dfrac{\text{Population}}{\text{Area}} \times 1000$

(*ii*) Death Rate $= \dfrac{\text{No. of Deaths}}{\text{Population}}$

Using the above formulae, the given data assumes the following form :

Cities :	A	B	C	D	E	F
Density (X) :	200	500	400	700	600	300
Death Rate (Y) :	10	16	14	20	17	13

CALCULATION OF COEFFICIENT OF CORRELATION

Density X	Death Rate Y	$U = \frac{X-400}{100}$	$V = Y - 14$	U^2	V^2	UV
200	10	− 2	− 4	4	16	8
500	16	+ 1	+ 2	1	4	2
400	14	0	0	0	0	0
700	20	+ 3	+ 6	9	36	18
600	17	+ 2	+ 3	4	9	6
300	13	− 1	− 1	1	1	1
		$\sum U = 3$	$\sum V = 6$	$\sum U^2 = 19$	$\sum V^2 = 66$	$\sum UV = 35$

$$\therefore \quad r = \frac{n\sum UV - (\sum U)(\sum V)}{\sqrt{n\sum U^2 - (\sum U)^2}\cdot\sqrt{n\sum V^2 - (\sum V)^2}}$$

$$= \frac{6\times 35 - 3\times 6}{\sqrt{6\times 19 - 9}\cdot\sqrt{6\times 66 - 36}} = \frac{210-18}{\sqrt{105}\times\sqrt{360}} = \frac{192}{194.42} = 0.9875,$$

indicating a high degree of positive correlation between density of population and death rate.

EXAMPLE 27. *The following table gives the distribution of the total population and those who are wholly or partially blind among them. Find out if there is any relation between age and blindness.*

Age	*No. of Persons (in thousand)*	*Blind*	*Age*	*No. of Persons (in thousands)*	*Blind*
0 – 10	100	55	40 – 50	24	36
10 – 20	60	40	50 – 60	11	22
20 – 30	40	40	60 – 70	6	18
30 – 40	36	40	70 – 80	3	15

SOLUTION. For facilitating comparison we must determine the number of blinds in terms of a common denominator, say, 1 lakh. The first figure would remain the same because 55 persons are blind out of 100 thousand, *i.e.*, 1 lakh persons. The second value would be obtained as follows :

Out of 60,000 persons, number of blinds = 40

$\therefore$ Out of 1,00,000 persons, number of blinds $= \frac{40}{60,000}\times 1,00,000 = 67$ (app.)

Other values are obtained in a similar fashion.

CALCULATION OF CORRELATION COEFFICIENT

Age	Mid-value X	$U=\frac{X-35}{10}$	U^2	No. of Blinds per lakh Y	$V=Y-185$	V^2	UV
0 – 10	5	– 3	9	55	– 130	16900	390
10 – 20	15	– 2	4	67	– 118	13924	236
20 – 30	25	– 1	1	100	– 85	7225	85
30 – 40	35	0	0	111	– 74	5476	0
40 – 50	45	1	1	150	– 35	1225	– 35
50 – 60	55	2	4	200	15	225	30
60 – 70	65	3	9	300	115	13225	345
70 – 80	75	4	16	500	315	99225	1260
		$\sum U = 4$	$\sum U^2 = 44$		$\sum V = 3$	$\sum V^2 = 157425$	$\sum UV = 2311$

$$\therefore \quad r = \frac{n\sum UV - (\sum U)(\sum V)}{\sqrt{n\sum U^2 - (\sum U)^2}\cdot\sqrt{n\sum V^2 - (\sum V)^2}}$$

$$= \frac{(8\times 2311)-(4\times 3)}{\sqrt{8\times 44-16}.\sqrt{8\times 157425-9}} = \frac{18488-12}{\sqrt{336}.\sqrt{1259391}} = \frac{18476}{20570.74} = 0.898,$$

indicating a high degree of positive correlation between age and blindness.

EXAMPLE 28. *A computer while calculating the correlation coefficient between two variables X and Y obtained the following results :*

$$n = 25, \quad \sum X = 125, \quad \sum X^2 = 650, \quad \sum Y = 100, \quad \sum Y^2 = 460 \quad and \quad \sum XY = 508$$

It was, however, later discovered at the time of checking that it had copied down two pairs of observations as (6, 14) and (8, 6), while the correct values were (8, 12) and (6, 8). Obtain the correct value of the correlation coefficient between X and Y.

[Delhi Univ. B.Com. (H) 2008]

SOLUTION. To find the correct value of the correlation coefficient, we have to make corrections in all the sums given. Thus

Corrected $\sum X = 125-6-8+8+6 = 125$

Corrected $\sum Y = 100-14-6+12+8 = 100$

Corrected $\sum X^2 = 650-6^2-8^2+8^2+6^2 = 650$

Corrected $\sum Y^2 = 460-14^2-6^2+12^2+8^2 = 436$

Corrected $\sum XY = 508-6\times 14-8\times 6+8\times 12+6\times 8 = 520$

Substituting these corrected values in the formula

$$r = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{n\sum X^2 - (\sum X)^2}\cdot\sqrt{n\sum Y^2 - (\sum Y)^2}}$$

for computing the correlation coefficient, we obtain

$$(\text{corrected } r) = \frac{25\times 520 - 125\times 100}{\sqrt{25\times 650 - (125)^2}\cdot\sqrt{25\times 436 - (100)^2}}$$

$$= \frac{13000 - 12500}{\sqrt{16255 - 15625}\,.\,\sqrt{10900 - 10000}}$$

$$= \frac{500}{\sqrt{625}\,.\,\sqrt{900}} = \frac{500}{25\times 30} = \frac{2}{3} = 0.67 \text{ (app.)}$$

EXAMPLE 29. *Coefficient of correlation between X and Y for 20 items is 0.3. Mean of X is 15 and that of Y is 20 while standard deviations are 4 and 5 respectively. At the time of calculation, one item 27 has wrongly been taken as 17 in case of X series and 35 instead of 30 in case of Y series. Find the correct coefficient of correlation.*

[*Delhi Univ. B.Com. (H) 2007 (C.C.), 2011*]

SOLUTION. We are given :

$n = 20, \quad r = 0.3, \quad \bar{X} = 15, \quad \bar{Y} = 20, \quad \sigma_X = 4 \quad$ and $\quad \sigma_Y = 5$

$$\because \quad \bar{X} = \frac{\sum X}{n} \Rightarrow \sum X = n\bar{X} = 20\times 15 = 300$$

But this is not the correct $\sum X$, because one item was wrongly taken as 17 instead of 27.

$$\therefore \quad \text{Correct } \sum X = 300 - 17 + 27 = 310$$

$$\because \quad \bar{Y} = \frac{\sum Y}{n} \Rightarrow \sum Y = n\bar{Y} = 20\times 20 = 400$$

But this is not the correct $\sum Y$, because one item was wrongly taken as 35 instead of 30.

$$\therefore \quad \text{Correct } \sum Y = 400 - 35 + 30 = 395$$

$$\because \quad \sigma_X = 4 \Rightarrow \sigma_X^2 = 16 \text{ or } \frac{\sum X^2}{n} - \bar{X}^2 = 16$$

$$\Rightarrow \quad \sum X^2 = n\left(\bar{X}^2 + 16\right) = 20\,(225 + 16) = 4820$$

However, this is not the correct $\sum X^2$. In fact,

$$\text{Correct} \sum X^2 = 4820 - (17)^2 + (27)^2 = 4820 - 289 + 729 = 5260$$

$$\because \quad \sigma_Y = 5 \Rightarrow \sigma_Y^2 = 25 \text{ or } \frac{\sum Y^2}{n} - \bar{Y}^2 = 25$$

$$\Rightarrow \quad \sum Y^2 = n\left(\bar{Y}^2 + 25\right) = 20\,(400 + 25) = 8500$$

However, this is not the correct $\sum Y^2$. In fact,

$$\text{Correct} \sum Y^2 = 8500 - (35)^2 + (30)^2 = 8500 - 1225 + 900 = 8175$$

Finally,
$$r = 0.3 \Rightarrow \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = 0.3$$

$$\Rightarrow \quad \frac{\text{Cov}(X,Y)}{4 \times 5} = 0.3 \Rightarrow \text{Cov}(X, Y) = 20 \times 0.3 = 6$$

$$\Rightarrow \quad \frac{\sum XY}{n} - \left(\bar{X}\right)\left(\bar{Y}\right) = 6 \quad \left(\because \text{Cov}(X, Y) = \frac{\sum XY}{n} - \bar{X}.\bar{Y}\right)$$

or,
$$\frac{\sum XY}{n} = 6 + \left(\bar{X}\right)\left(\bar{Y}\right) = 6 + (15)\,(20) = 306$$

$$\therefore \quad \sum XY = n \times 306 = 20 \times 306 = 6120$$

Again, this is not the correct $\sum XY$. In fact,

$$\text{Correct} \sum XY = 6120 - 17 \times 35 + 27 \times 30 = 6335$$

Substituting the corrected values of $\sum X$, $\sum Y$, $\sum X^2$, $\sum Y^2$ and $\sum XY$ in the formula

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

for computing the coefficient of correlation, we obtain

$$\text{Corrected } r = \frac{20 \times 6335 - 310 \times 395}{\sqrt{20 \times 5260 - (310)^2}\sqrt{20 \times 8175 - (395)^2}}$$

$$= \frac{126700 - 122450}{\sqrt{105200 - 96100}.\sqrt{163500 - 156025}}$$

$$= \frac{4250}{\sqrt{9100}.\sqrt{7475}} = \frac{4250}{8247.58} = 0.515.$$

EXAMPLE 30. *A student while calculating the correlation coefficient between two variables X and Y obtained the following results :*

$$n = 50, \quad \bar{X} = 10, \quad \bar{Y} = 6, \quad \sigma_X = 3 \quad \sigma_Y = 2 \quad \text{and} \quad r = 0.3$$

It was, however, later discovered at the time of checking that he had copied down one value of X (= 10) and one value of Y (= 6) incorrectly and hence weeded them out. With the remaining 49 pairs of observations, how is the original value of correlation coefficient affected?

SOLUTION. We know $\quad \bar{X} = \dfrac{\sum X}{n} \Rightarrow \sum X = n\bar{X} = 50 \times 10 = 500$

If the incorrect value of $X (= 10)$ is weeded out, then

$$\sum X \text{ (for 49 items)} = 500 - 10 = 490 \quad \therefore \quad \bar{X} \text{ (for 49 items)} = \frac{490}{49} = 10$$

Similarly, $\quad \bar{Y} = \dfrac{\sum Y}{n} \Rightarrow \sum Y = n\bar{Y} = 50 \times 6 = 300$

If the incorrect value of $Y (= 6)$ is weeded out, then

$$\sum Y \text{ (for 49 items)} = 300 - 6 = 294 \quad \therefore \quad \bar{Y} \text{ (for 49 items)} = \frac{294}{49} = 6$$

Also, $\quad \sigma_X = 3 \Rightarrow \sigma_X^2 = 9 \quad \text{or} \quad \dfrac{\sum X^2}{n} - \bar{X}^2 = 9$

$$\Rightarrow \quad \sum X^2 = n\left(9 + \bar{X}^2\right) = 50\,(9 + 100) = 50 \times 109 = 5450$$

However, if the incorrect value of $X (= 10)$ is weeded out, then

$$\sum X^2 \text{ (for 49 items)} = 5450 - (10)^2 = 5350$$

and hence $\quad \sigma_X^2 \text{ (for 49 items)} = \dfrac{\sum X^2}{n} - \bar{X}^2 = \dfrac{5350}{49} - (10)^2$

$$= \frac{5350 - 4900}{49} = \frac{450}{49} \qquad \text{....(1)}$$

Also, $\quad \sigma_Y = 2 \Rightarrow \sigma_Y^2 = 4 \quad \text{or} \quad \dfrac{\sum Y^2}{n} - \bar{Y}^2 = 4$

$$\Rightarrow \quad \sum Y^2 = n\left(4 + \bar{Y}^2\right) = 50\,(4 + 36) = 50 \times 40 = 2000$$

However, if the incorrect value of $Y (= 6)$ is weeded out, then

$$\sum Y^2 \text{ (for 49 items)} = 2000 - 6^2 = 1964$$

and hence σ_Y^2 (for 49 items) $= \dfrac{\sum Y^2}{n} - \bar{Y}^2 = \dfrac{1964}{49} - (6)^2 = \dfrac{1964 - 49 \times 36}{49}$

$$= \frac{1964 - 1764}{49} = \frac{200}{49} \quad \text{.... (2)}$$

Finally, $r = 0.3 \Rightarrow \dfrac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = 0.3$

$$\Rightarrow \quad \text{Cov}(X, Y) = 0.3 \times \sigma_X \times \sigma_Y = 0.3 \times 3 \times 2 = 1.8$$

i.e., $\dfrac{\sum XY}{n} - (\bar{X})(\bar{Y}) = 1.8$

$$\Rightarrow \quad \sum XY = n\left(1.8 + \bar{X}\bar{Y}\right) = 50\,(1.8 + 10 \times 6) = 50 \times 61.8 = 3090$$

However, if one value of $X(= 10)$ and one value of $Y(= 6)$ are weeded out, then

$$\sum XY \text{ (for 49 pairs)} = 3090 - 10 \times 6 = 3030$$

$$\therefore \quad \text{Cov}(X, Y) \text{ (for 49 pairs)} = \frac{\sum XY}{n} - (\bar{X})(\bar{Y}) = \frac{3030}{49} - 10 \times 6$$

$$= \frac{3030 - 2940}{49} = \frac{90}{49} \quad \text{.... (3)}$$

Thus, the coefficient of correlation between X and Y for the remaining 49 pairs of observations is : $r = \dfrac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \dfrac{90/49}{\sqrt{\dfrac{450}{49}} \times \sqrt{\dfrac{200}{49}}}$ (Using (1), (2) and (3))

$$= \frac{90}{\sqrt{450 \times 200}} = \frac{90}{\sqrt{90000}} = \frac{90}{\sqrt{300}} = 0.3$$

Hence r is unaffected.

EXERCISE 6.2

1. Explain the concept of correlation between two variables.
   [*Delhi Univ. B.Com. 1977, 1980*]
2. Define Karl Pearson's coefficient of correlation. What is it intended to measure?
3. State any two of the properties of Karl Pearson's coefficient of correlation.
   [*Delhi Univ. B.Com. 2003*]
4. Find the coefficient of correlation between X and Y, when
   (*i*) Cov $(X, Y) = -16.5$, Var $(X) = 2.89$ and Var $(Y) = 100$
   (*ii*) Cov $(X, Y) = 10.2$, Var $(X) = 8.25$ and Var $(Y) = 33.96$
   (*iii*) Cov $(X, Y) = 16.5$, Var $(X) = 8.25$ and Var $(Y) = 33$.
5. Find the coefficient of correlation between X and Y if covariance between X and Y is 10 and the variance of X and Y are respectively 16 and 9. **[*Delhi Univ. B.Com. 1991*]**
6. The coefficient of correlation between two variables X and Y is 0.8 and their covariance is 20. If the variance of X series is 16, find the standard deviation of Y series.
   [*C.A. Foundation, June 1993*]
7. The coefficient of correlation between two variables X and Y is 0.64 and their covariance is 16. If the variance of X series is 9, find the standard deviation of Y series.
8. Calculate the coefficient of correlation between X and Y series from the following data :

	X series	Y series
No. of observations	15	15
Arithmetic mean	25	18
Standard deviation	5	5

$\sum(X-25)(Y-18) = 125$ **[Delhi Univ. B.Com. 1994]**

9. Calculate Karl Pearson's coefficient of correlation from the following data :

(i) *Sum of deviations of X*	=	5
(ii) *Sum of deviations of Y*	=	4
(iii) *Sum of squares of deviations of X*	=	40
(iv) *Sum of squares of deviations of Y*	=	50
(v) *Sum of the product of deviations of X and Y*	=	32
(vi) *No. of pairs of observations*	=	10

[Delhi Univ. B.Com. 1984]

10. From the following data, compute the coefficient of correlation between two variables X and Y:

	X series	Y series
No. of items	12	12
Sum of the squares of deviations from arithmetic mean	360	250
Sum of the products of deviations of X and Y series from their respective arithmetic means	225	

11. Find the coefficient of correlation between X and Y for the following data :

(i) $n=25,\ \sum X = 125,\ \sum Y = 100,\ \sum X^2 = 650,\ \sum Y^2 = 436,\ \sum XY = 520$

(ii) $n=10,\ \sum X = 55,\ \sum Y = 40,\ \sum X^2 = 385,\ \sum Y^2 = 192,\ \sum XY = 185$

12. The following table gives the supply and price figures for a commodity for 6 days. Calculate the correlation coefficient between price and supply.

Days :	*Mon*	*Tues*	*Wed*	*Thur*	*Fri*	*Sat*
Price :	22	30	25	20	15	8
Supply :	10	12	15	20	23	28

What conclusions do you draw from the result? **[Delhi Univ. B.Com. 1990]**

13. Calculate Karl Pearson's coefficient of correlation from the following data, using 20 as the working mean for price and 70 as the working mean for demand.

Price :	14	16	17	18	19	20	21	22	23
Demand :	84	78	70	75	66	67	62	58	60

[Delhi Univ. B.Com. 1999]

14. Calculate coefficient of correlation from the following data and comment on the result.

Experience (X) :	16	12	18	4	3	10	5	12
Performance (Y) :	23	22	24	17	19	20	18	21

[Delhi Univ. B.Com. 2003]

15. Calculate the coefficient of correlation between X and Y for the data :

X :	79	68	78	61	89	96	69	59
Y :	136	123	125	108	137	156	112	107

[Delhi Univ. B.Com. 1987]

16. Calculate Karl Pearson's coefficient of correlation from the following data :

X :	6	8	12	15	18	20	24	28	31
Y :	10	12	15	15	18	25	22	26	28

[Delhi Univ. B.Com. 1984]

17. Calculate Karl Pearson's coefficient of correlation between expenditure on advertising and sales from the data given below :

Advertising expenses ('000 Rs.) :	39	65	62	90	82	75	25	98	36	78
Sales (lakh Rs.) :	47	53	58	86	62	68	60	91	51	84

[Delhi Univ. B.Com. 1985]

18. Calculate coefficient of correlation from the following data and interpret the result : .

S. No of Students :	1	2	3	4	5	6	7	8	9	10
Marks in Statistics :	20	35	15	40	10	35	30	25	45	30
Marks in Accounts :	25	30	20	35	20	25	25	35	35	30

[Delhi Univ. B.Com. 1977]

19. Calculate Karl Pearson's coefficient of correlation from the following data :

No.	*Subject*	*Percentage of marks*	
		First term	*Second term*
1.	*Hindi*	75	62
2.	*English*	81	68
3.	*Economics*	70	65
4.	*Accounts*	76	60
5.	*Commerce*	77	69
6.	*Mathematics*	81	72
7.	*Statistics*	84	76
8.	*Costing*	75	72

ANSWERS

4. (i) – 0.97 (ii) 0.61 (iii) 1
5. 0.833
6. $\sigma_Y = 6.25$
7. 8.333
8. 0.333
9. $r = 0.704$
10. (i) $r = 0.667$ (ii) $r = -0.6812$
11. (i) 0.667 (ii) – 0.068
12. $r = -0.89$, fairly high degree of negative correlation
13. $r = -0.954$
14. $r = 0.951$, a very high degree of positive correlation
15. $r = 0.95$
16. $r = 0.9$
17. $r = 0.80$
18. $r = 0.76$
19. $r = 0.623$
20. 0.967
21. $r = -0.67$
22. $n = 10$
23. $r = 0.05$
24. $r = 0.578$
25. 0.7953
26. 0.49
27. $r = -0.948$, P.E. $(r) = 0.027$
28. No
29. 0.546 to 0.654

6.14 SPEARMAN'S COEFFICIENT OF RANK CORRELATION

Sometimes, we are given a series of items where no numerical measure can be made, but where best and worst or most favoured and least favoured can be identified. Rankings are often applied in these situations to put the series into an order. For example, the characteristics like beauty, intelligence, leadership ability, honesty, etc. cannot be measured numerically, but the individuals in the group can be arranged in order thereby obtaining for each individual a number indicating its rank in the group.

If we have a group of individuals ranked according to two different qualities, it is natural to ask the following question:

"Is there an association between the rankings?"

To answer this question, we need to use a formula known as **Spearman's coefficient of rank correlation:**

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

where D is the difference between the two ranks given to each individual and n is the number of observations.

The Spearman's correlation coefficient is nothing but Karl Pearson's correlation coefficient between the ranks and is interpreted in much the same way. As before, the value of ρ will range from – 1 to + 1. A value of + 1 indicates perfect association for identical rankings and a value of – 1 indicates perfect association for reverse rankings. This will be clear from the following illustration:

Rank R_1	Rank R_2	D $R_1 - R_2$	D^2	Rank R_1	Rank R_2	D $R_1 - R_2$	D^2
1	1	0	0	1	4	–3	9
2	2	0	0	2	3	–1	1
3	3	0	0	3	2	1	1
4	4	0	0	4	1	3	9
			$\sum D^2 = 0$				$\sum D^2 = 20$

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

$$= 1 - 0 = 1 \qquad \qquad = 1 - \frac{6 \times 20}{4(4^2-1)} = 1 - \frac{120}{60} = 1 - 2 = -1.$$

6.15 COMPUTING THE RANK CORRELATION COEFFICIENT

We shall consider the following three cases to compute the rank correlation coefficient.

Case I : When Actual Ranks Are Given

In this case the following steps are involved:

(*i*) Compute D, the difference between the two ranks given to each individual.

(*ii*) Compute D^2 and obtain the sum $\sum D^2$.

(*iii*) Apply the formula: $\rho = 1 - \frac{6\sum D^2}{n(n^2-1)}$, where n is the number of observations.

The following examples will illustrate.

EXAMPLE 43. *Ten competitors in a beauty contest are ranked by two judges in the following order:*

I Judge	:	*1*	*6*	*5*	*10*	*3*	*2*	*4*	*9*	*7*	*8*
II Judge	:	*6*	*4*	*9*	*8*	*1*	*2*	*3*	*10*	*5*	*7*

Calculate the Spearman's rank correlation coefficient. Is there an association between the rankings?

SOLUTION. **COMPUTATION OF RANK CORRELATION COEFFICIENT**

Rank by I Judge R_1	*Rank by II Judge* R_2	D $(R_1 - R_2)$	D^2
1	6	– 5	25
6	4	2	4
5	9	– 4	16
10	8	2	4
3	1	2	4
2	2	0	0
4	3	1	1
9	10	– 1	1
7	5	2	4
8	7	1	1
			$\sum D^2 = 60$

Rank correlation between the judgment of the two judges is given by

$$\rho = 1 - \frac{6\sum D^2}{n(n^2-1)} = 1 - \frac{6\times 60}{10(10^2-1)} = 1 - \frac{36}{99} = 1 - \frac{4}{11} = \frac{7}{11} = 0.636.$$

This answer suggests that there is some degree of association between the rankings given by the two judges.

EXAMPLE 44. *Ten competitors in a beauty contest are ranked by three judges in the following order :*

I Judge	:	1	4	8	9	6	10	7	3	2	5
II Judge	:	4	8	7	5	9	6	10	2	3	1
III Judge	:	6	7	1	8	10	5	9	2	3	4

Use the rank correlation method to determine which pair of judges has the nearest approach to common taste in beauty.

SOLUTION. In order to find out which pair of judges has the nearest approach to common taste in beauty, we compare the Rank Correlation between the judgments of :

(*i*) *I* Judge and *II* Judge

(*ii*) *I* Judge and *III* Judge

(*iii*) *II* Judge and *III* Judge

COMPUTATION OF RANK CORRELATION

Rank by I Judge R_1	*Rank by II Judge* R_2	*Rank by III Judge* R_3	D_{12} (R_1-R_2)	D_{13} (R_1-R_3)	D_{23} (R_2-R_3)	D_{12}^2	D_{13}^2	D_{23}^2
1	4	6	−3	−5	−2	9	25	4
4	8	7	−4	−3	+1	16	9	1
8	7	1	+1	+7	+6	1	49	36
9	5	8	+4	+1	−3	16	1	9
6	9	10	−3	−4	−1	9	16	1
10	6	5	+4	+5	+1	16	25	1
7	10	9	−3	−2	+1	9	4	1
3	2	2	+1	+1	0	1	1	0
2	3	3	−1	−1	0	1	1	0
5	1	4	+4	+1	−3	16	1	9
						$\sum D_{12}^2 = 94$	$\sum D_{13}^2 = 132$	$\sum D_{23}^2 = 62$

The rank correlation coefficient is given by :

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

Rank correlation between the judgments of first and second judges :

$$\rho_{12} = 1-\frac{6\times 94}{10\left(10^2-1\right)} = 1-\frac{564}{990} = 0.43$$

Rank correlation between the judgments of first and third judges :

$$\rho_{13} = 1-\frac{6\times 132}{10\left(10^2-1\right)} = 1-\frac{792}{990} = 0.2$$

Rank correlation between the judgments of second and third judges :

$$\rho_{23} = 1-\frac{6\times 62}{10\left(10^2-1\right)} = 1-\frac{372}{990} = 0.62$$

Since rank correlation coefficient is maximum in the judgment of the second and third judges, we conclude that they have the nearest approach to common tastes in beauty.

EXAMPLE 45. *Rankings of 10 trainees at the beginning and at the end of a certain course are given below:*

Trainees	:	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Rank at the beginning	:	*1*	*6*	*3*	*9*	*5*	*2*	*7*	*10*	*8*	*4*
Rank at the end	:	*6*	*8*	*3*	*7*	*2*	*1*	*5*	*9*	*4*	*10*

Calculate spearman's rank correlation coefficient.

[*I.C.W.A. (Intermediate), June 1995*]

SOLUTION. **COMPUTATION OF RANK CORRELATION COEFFICIENT**

Trainees	*Rank at the beginning*	*Rank at the end* R_1	*D* $R_1 - R_2$	D^2
A	1	6	– 5	25
B	6	8	– 2	4
C	3	3	0	0
D	9	7	2	4
E	5	2	3	9
F	2	1	1	1
G	7	5	2	4
H	10	9	1	1
I	8	4	4	16
J	4	10	– 6	36
				$\sum D^2 = 100$

The rank correlation coefficient is given by :

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

where $n = 10$ and $\sum D^2 = 100$. Substituting these values, we obtain

$$\rho = 1 - \frac{6 \times 100}{10(10^2 - 1)} = 1 - \frac{6 \times 100}{10 \times 99} = 1 - \frac{20}{33} = \frac{13}{33} = 0.394.$$

EXAMPLE 46. *If the sum of squares of the rank differences of 9 pairs of values is 80, find the correlation coefficient between them.* **[Delhi Univ. B.Com. 2003]**

SOLUTION. With the usual notations, we are given the following:

$$n = 9 \quad \text{and} \quad \sum D^2 = 80$$

Substituting these values in the formula :

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

$$\rho = 1 - \frac{6 \times 80}{9(9^2 - 1)} = 1 - \frac{480}{720} = 1 - \frac{2}{3} = \frac{1}{3} = 0.33 \text{ (app.)}$$

EXAMPLE 47. *In a bivariate data of n pairs of observations, the sum of square of differences between the ranks of observed values of two variables is 231 and the rank correlation coefficient is – 0.4. Find the value of n.* **[Delhi Univ. B.Com. (H) 2006 (C.C.)]**

SOLUTION. With the usual notations, we are given the following :

$$\sum D^2 = 231 \quad \text{and} \quad \rho = -0.4$$

Substituting these values in the formula : $\rho = 1 - \dfrac{6 \sum D^2}{n(n^2 - 1)}$, we get

$$-0.4 = 1 - \frac{6 \times 231}{n(n^2 - 1)}$$

$$\Rightarrow \quad n(n^2 - 1) = \frac{6 \times 231}{1.4} = 990 = 10(10^2 - 1) \quad \Rightarrow \quad n = 10.$$

EXAMPLE 48. *The coefficient of rank correlation between debenture prices and share prices of a company is found to be 0.143. If the sum of the squares of the difference in ranks is 48, find the value of n.*

SOLUTION. With the usual notations, we are given the following :

$$\sum D^2 = 48 \quad \text{and} \quad \rho = +0.143$$

Substituting these values in the formula :

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

$$0.143 = 1 - \frac{6 \times 48}{n(n^2-1)} = 1 - \frac{288}{n(n^2-1)} \Rightarrow \frac{288}{n(n^2-1)} = 1 - 0.143 = 0.857$$

$$\Rightarrow \quad n(n^2-1) = \frac{288}{0.857} = 336 = 7(7^2-1) \Rightarrow n = 7.$$

EXAMPLE 49. *The coefficient of rank correlation of the marks obtained by 10 students in Statistics and Accountancy was found to be 0.2. It was later discovered that the difference in ranks in the two subjects obtained by one of the students was wrongly taken as 9 instead of 7. Find the correct value of coefficient of rank correlation.*

[Delhi Univ. B.Com. (H) 2011, 2012 (Modified)]

SOLUTION. With the usual notations, we are given:

$$n = 10 \quad \text{and} \quad \rho = 0.2$$

Substituting these values in the formula :

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

$$0.2 = 1 - \frac{6\sum D^2}{10(10^2-1)} \Rightarrow \frac{6\sum D^2}{990} = 1 - 0.2 = 0.8$$

$$\Rightarrow \quad \sum D^2 = \frac{0.8 \times 990}{6} = 132$$

However, $\sum D^2$ obtained above is not correct as one difference in ranks was wrongly taken as 9 instead of 7. In fact,

$$\text{Corrected } \sum D^2 = 132 - 9^2 + 7^2 = 100$$

$$\therefore \quad \text{(Corrected) } \rho = 1 - \frac{6\sum D^2}{n(n^2-1)} = 1 - \frac{6 \times 100}{10 \times 99} = 0.3939.$$

Case II: When Ranks Are Not Given

Sometimes we are given the actual bivariate data on two variables and not the ranks. In such situations, it is necessary to assign the ranks. Ranks can be assigned by taking either the highest value as 1 or the lowest value as 1. The next highest or the next lowest value is given rank 2 and so on. But whether we start with the lowest value or the highest value we must follow the same method in case of both the variables.

This is best illustrated with the help of following examples.

EXAMPLE 50. The marks obtained by 9 students in Mathematics and Accountancy are as follows:

Marks in Mathematics (X) : 30 33 45 23 8 49 42 4 31

Marks in Accountancy (Y) : 35 23 47 17 10 43 9 6 28

Calculate Spearman's rank correlation coefficient.

SOLUTION. By assigning rank 1 to the highest value in both the series, we get the following table:

CALCULATION OF RANK CORRELATION COEFFICIENT

X	Y	Rank in X R_1	Rank in Y R_2	D $(R_1 - R_2)$	D^2
30	35	5.	3	2	4
33	23	3	5	-2	4
45	47	2	1	1	1
23	17	6	6	0	0
8	10	8	7	1	1
49	43	1	2	-1	1
12	9	7	8	-1	1
4	6	9	9	0	0
31	28	4	4	0	0
					$\sum D^2 = 12$

The Spearman's coefficient of rank correlation is given by:

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

where $n = 9$ and $\sum D^2 = 12$

Substituting these values, we obtain

$$\rho = 1 - \frac{6 \times 12}{9(9^2-1)} = 1 - \frac{72}{9 \times 80} = 1 - \frac{1}{10} = 1 - 0.1 = 0.9.$$

EXAMPLE 26. *A test in Statistics was taken by 7 students. The teacher ranked his pupils according to their academic achievements. The order of achievement from high to low, together with family income for each pupil, is given as follows:*

Rai (Rs. 8700), Bhatnagar (Rs. 4200), Tuli (Rs. 5700), Desai (Rs. 8200), Gupta (Rs. 20,000), Chaudhary (Rs, 18,000) and Singh (Rs. 17,500).

Compute the Spearman's coefficient of rank correlation between academic achievement and family income. **[Delhi Univ. B.Com. 1999]**

SOLUTION. **CALCULATION OF RANK CORRELATION COEFFICIENT**

Student	Academic achievement X	Family income (Rs.) Y	Rank in X R_1	Rank in Y R_2	D $R_1 - R_2$	D^2
Rai	1	8700	1	4	− 3	9
Bhatnagar	2	4200	2	7	− 5	25
Tuli	3	5700	3	6	− 3	9
Desai	4	8200	4	5	− 1	1
Gupta	5	20,000	5	1	+ 4	16
Chaudhary	6	18,000	6	2	+ 4	16
Singh	7	17,500	7	3	+ 4	16
$n = 7$						$\sum D^2 = 92$

Spearman's coefficient of rank correlation between academic achievement (X) and family income (Y) is given by:

$$\rho = 1 - \frac{6\sum D^2}{n(n^2-1)} = 1 - \frac{6 \times 92}{7 \times 48} = 1 - \frac{23}{14} = -\frac{9}{14} = -\ 0.6429.$$

Case 3. When Ranks are Equal

If there are two or more items with the same rank in either series, then it is customary to assign common rank to each repeated item. The common rank is the average of the ranks which these items would have got if they were different from each other and the next item will get the rank next to the rank used in computing the common rank. For example, suppose there are two items at rank 4, then the common rank assigned to each item is $\frac{4+5}{2} = 4.5$, the average of 4 and 5, the ranks which these items would have got if they were different. The next item will be assigned the rank 6. Similarly, if there are three items at rank 7, the common rank assigned to each item will be $\frac{7+8+9}{3} = 8$, the average of 7, 8 and 9. The next rank to be assigned will be 10.

If equal ranks are assigned to some items, an adjustment is made in the Spearman's rank correlation coefficient formula by adding the correction factor $\frac{1}{12}(m^3 - m)$ to the value $\sum D^2$, where m is the number of times an item is repeated. This correction factor is to be added for each repeated item in both the series. The formula can thus be written as follows:

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

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

where m_1 represents the number of times first item is repeated, m_2 represents the number of times second item is repeated and so on.

EXAMPLE 52. *Calculate Spearman's coefficient of rank correlation from the following data:*

X :	*57*	*16*	*24*	*65*	*16*	*16*	*9*	*40*	*33*	*48*
Y :	*19*	*6*	*9*	*20*	*4*	*15*	*6*	*24*	*13*	*13*

[*Delhi Univ. B.Com. 1997*]

SOLUTION. The following table shows the ranking from the highest value in both the series. Moreover, certain items in both the series are repeated, so ranking is done in accordance with suitable average.

CALCULATION OF RANK CORRELATION COEFFICIENT

X	*Rank assigned* R_1	*Y*	*Rank assigned* R_2	*D* $R_1 - R_2$	D^2
57	2	19	3	– 1	1
16	8	6	8.5	– 0.5	0.25
24	6	9	7	– 1	1
65	1	20	2	– 1	1
16	8	4	10	– 2	4
16	8	15	4	4	16
9	10	6	8.5	1.5	2.25
40	4	24	1	3	9
33	5	13	5.5	– 0.5	0.25
48	3	13	5.5	–2.5	6.25
$n = 10$		$n = 10$			$\sum D^2 = 41$

Note that in series *X*, the item 16 is repeated 3 times (*i.e.*, $m_1 = 3$). In series *Y*, the item 13 is repeated twice (*i.e.*, $m_2 = 2$) and 6 is also repeated twice (*i.e.*, $m_3 = 2$).

Thus, Spearman's coefficient of rank correlation is given by

$$\rho = 1 - \frac{6\left[\sum D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \frac{1}{12}(m_3^3 - m_3)\right]}{n(n^2-1)}$$

$$= 1 - \frac{6\left[41 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)\right]}{10(10^2-1)}$$

$$= 1-\frac{6[41+2+0.5+0.5]}{990} = 1-\frac{6\times 44}{990} = 1-\frac{4}{15} = \frac{11}{15} = 0.7373.$$

EXAMPLE 53. *The following data relate to the marks obtained by 10 students of a class in Statistics and Costing.*

Marks in Statistics :	30	38	28	27	28	23	30	33	28	35
Marks in Costing :	29	27	22	29	20	29	18	21	27	22

Calculate Spearman's rank correlation coefficient. **[Delhi Univ. B.Com. (H) 2001]**

SOLUTION. The following table shows the ranking from the highest value in both the series. Moreover, certain values in both the series are repeated, so ranking is done in accordance with suitable average.

CALCULATION OF RANK CORRELATION COEFFICIENT

Marks in statistics (*X*)	*Rank* R_1	*Marks in costing* (*Y*)	*Rank* R_2	*D* $R_1 - R_2$	D^2
30	4.5	29	2	2.5	6.25
38	1	27	4.5	– 3.5	12.25
28	7	22	6.5	0.5	0.25
27	9	29	2	7	49
28	7	20	9	– 2	4
23	10	29	2	8	64
30	4.5	18	10	– 5.5	30.25
33	3	21	8	– 5	25
28	7	27	4.5	2.5	6.25
35	2	22	6.5	– 4.5	20.25
$n = 10$		$n = 10$			$\sum D^2 = 217.5$

In series *X* the value 30 is repeated twice (*i.e.*, $m_1 = 2$) and the value 28 is repeated three times (*i.e.*, $m_2 = 3$). In series *Y*, the value 29 is repeated thrice (*i.e.* $m_3 = 3$), the value 27 is repeated twice (*i.e.*, $m_4 = 2$), and the value 22 is also repeated twice (*i.e.*, $m_5 = 2$).

Thus Spearman's coefficient of rank correlation is given by

$$\rho = 1-\frac{6\left[\sum D^2+\frac{1}{12}(m_1^3-m_1)+\frac{1}{12}(m_2^3-m_2)+\frac{1}{12}(m_3^3-m_3)+\frac{1}{12}(m_4^3-m_4)+\frac{1}{12}(m_5^3-m_5)\right]}{n(n^2-1)}$$

$$= 1-\frac{6\left[217.5+\frac{1}{12}(2^3-2)+\frac{1}{12}(3^3-3)+\frac{1}{12}(3^3-3)+\frac{1}{12}(2^3-2)+\frac{1}{12}(2^3-2)\right]}{10(10^2-1)}$$

$$= 1-\frac{6[217.5+0.5+2+2+0.5+0.5]}{990} = 1-\frac{6\times 223}{990} = 1-\frac{223}{165} = -\frac{58}{165} = -0.35.$$

EXAMPLE 54. *Find the Rank Correlation Coefficient from the following marks awarded by the examiners in Statistics :*

Roll Nos.	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
Marks awarded by Examiner A :	24	29	19	14	30	19	27	30	20	28	11
Marks awarded by Examiner B :	37	35	16	26	23	27	19	20	16	11	21
Marks awarded by Examiner C :	30	28	20	25	25	30	20	24	22	29	15

[Delhi Univ. B.Com. (H) 2005]

SOLUTION. The following table shows the ranking from the highest value in each series. Moreover, certain values in both the series are repeated, so ranking is done in accordance with suitable average.

CALCULATION OF RANK CORRELATION COEFFICIENT

Roll Nos.	*Marks awarded by Examiner A*	*Rank* R_1	*Marks awarded by Examiner B*	*Rank* R_2	*Marks awarded by Examiner C*	*Rank* R_3	D^2_{AB} $(R_1-R_2)^2$	D^2_{AC} $(R_1-R_3)^2$	D^2_{BC} $(R_2-R_3)^2$
1	24	6	37	1	30	1.5	25	20.25	0.25
2	29	3	35	2	28	4	1	1	4
3	19	8.5	16	9.5	20	9.5	1	1	0
4	14	10	26	4	25	5.5	36	20.25	2.25
5	30	1.5	23	5	25	5.5	12.25	16	0.25
6	19	8.5	27	3	30	1.5	30.25	49	2.25
7	27	5	19	8	20	9.5	9	20.25	2.25
8	30	1.5	20	7	24	7	30.25	30.25	0
9	20	7	16	9.5	22	8	6.25	1	2.25
10	28	4	11	11	29	3	49	1	64
11	11	11	21	6	15	11	25	0	25
							$\sum D^2_{AB} = 225$	$\sum D^2_{AC} = 160$	$\sum D^2_{BC} = 102.5$

$$\therefore \quad R_{AB} = 1 - \frac{6\left[\sum D^2_{AB} + \frac{\Sigma(m^3 - m)}{12}\right]}{n(n^2-1)}$$

$$= 1 - \frac{6\left[225 + \frac{1}{12}(2^3-2) + \frac{1}{12}(2^3-2)\right]}{11(11^2-1)} = 1 - \frac{6[225+0.5+0.5]}{1320}$$

$$= 1 - \frac{6 \times 226}{1320} = 1 - \frac{226}{220} = -\frac{6}{220} = -0.027$$

$$R_{AC} = 1 - \frac{6\left[\sum D_{AC}^2 + \frac{\sum(m^3 - m)}{12}\right]}{n(n^2-1)}$$

$$= 1 - \frac{6\left[160 + \frac{1}{12}(2^3-2) + \frac{1}{12}(2^3-2) + \frac{1}{12}(2^3-2) + \frac{1}{12}(2^3-2) + \frac{1}{12}(2^3-2)\right]}{11(11^2-1)}$$

$$= 1 - \frac{6[160 + 0.5 \times 5]}{1320} = 1 - \frac{6(162.5)}{1320} = 1 - \frac{162.5}{220} = \frac{57.5}{220} = 0.26136$$

$$R_{BC} = 1 - \frac{6\left[\sum D_{BC}^2 + \frac{\sum(m^3 - m)}{12}\right]}{n(n^2-1)}$$

$$= 1 - \frac{6\left[102.5 + \frac{1}{12}(2^3-2) + \frac{1}{12}(2^3-2) + \frac{1}{12}(2^3-2) + \frac{1}{12}(2^3-2)\right]}{11(11^2-1)}$$

$$= 1 - \frac{6[102.5 + 2]}{1320} = 1 - \frac{6 \times 104.5}{1320} = 1 - \frac{104.5}{220} = \frac{115.5}{220} = 0.525$$

REMARK. Sometimes we are given data in the form of ranks but the highest rank in the series exceeds the number of pairs of observations. In such situations ranks are treated as values and then fresh ranks are determined. This is illustrated in the following example.

EXAMPLE 55. *Calculate the coefficient of correlation from the following data by the method of rank differences :*

Rank of X :	*10*	*4*	*2*	*5*	*8*	*5*	*6*	*9*
Rank of Y :	*10*	*6*	*2*	*5*	*8*	*4*	*5*	*9*

[C.A. Foundation, May 1994]

SOLUTION. Though the data is given in the form of ranks but it cannot be used as ranks as the highest rank exceeds the number of pairs of observations. Treating the ranks as values, we assign the fresh ranks. Moreover, certain items in both the series are repeated, so ranking is done in accordance with suitable average.

CALCULATION OF RANK CORRELATION COEFFICIENT

X	Rank assigned R_1	Y	Rank assigned R_2	D $R_1 - R_2$	D^2
10	1	10	1	0	0
4	7	6	4	3	9
2	8	2	8	0	0
5	5.5	5	5.5	0	0
8	3	8	3	0	0
5	5.5	4	7	– 1.5	2.25
6	4	5	5.5	– 1.5	2.25
9	2	9	2	0	0
$n = 8$		$n = 8$			$\sum D^2 = 13.50$

Note that in series X, the item 5 is repeated twice (*i.e.*, $m_1 = 2$). In series Y, the item 5 is repeated twice (*i.e.*, $m_2 = 2$).

Thus, Spearman's coefficient of rank correlation is given by

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

$$= 1 - \frac{6\left[13.5 + \frac{1}{12}\left(2^3 - 2\right) + \frac{1}{12}\left(2^3 - 2\right)\right]}{8\left(8^2 - 1\right)} = 1 - \frac{6 \times 14.5}{8 \times 63}$$

$$= 1 - 0.173 = 0.827.$$

6.16 MERITS AND DEMERITS OF SPEARMAN'S RANK CORRELATION METHOD

Merits. The rank correlation method has the following merits:

1. *It is easy to understand and simple to apply.*
2. *The Spearman's rank correlation method is the only method that can be used to find correlation coefficient if we are dealing with data of qualitative characteristics like beauty, intelligence, honesty, etc.*
3. *This is the only method that can be used where we are given the ranks and not the actual bivariate data on two variables.*

Demerits. The rank correlation method has the following limitations.

1. *This method cannot be used for finding correlation in the case of bivariate frequency distribution.*
2. *This method is very difficult to apply when the number of items is more than 30.*

EXERCISE 6.3

1. What is rank correlation? Explain the Spearman's rank correlation method for computing the coefficient of correlation.
2. State the merits and demerits of Spearman's rank correlation method.
   [*Delhi Univ. B.Com. 1990*]
3. Define Rank Correlation coefficient. When is it preferred to Karl Pearson's coefficient of correlation?
4. Twelve competitors in a beauty contest are ranked by two judges in the following order:

I Judge	:	5	2	3	4	1	6	8	7	10	9	12	11
II Judge	:	4	5	2	1	6	7	10	9	11	12	3	8

   Calculate the Spearman's rank correlation coefficient.
5. Two judges in a driving competition assign the following ranks to 9 finalists:

Rank assigned by Judge 1	:	1	3	9	6	4	8	2	5	7
Rank assigned by Judge 2	:	2	3	8	5	1	9	4	6	7

   Calculate Spearman's rank correlation coefficient.[*Delhi Univ. B.A. (Eco. Hons.) 2002*]
6. **Ten competitors in a beauty contest are ranked by three judges in the following order:**

Judge A	:	1	6	5	10	3	2	4	9	7	8
Judge B	:	3	5	8	4	7	10	2	1	6	9
Judge C	:	6	4	9	8	1	2	3	10	5	7

Use the rank correlation method to determine which pair of judges has the nearest approach to common taste in beauty.

7. The ranks of the same 15 students in two subjects *A* and *B* are given below; the two numbers within the brackets denote the ranks of the same student in *A* and *B* respectively.

(1, 10), (2, 7), (3, 2), (4, 6),(5, 4), (6, 8), (7, 3), (8, 1), (9, 11), (10, 15), (11, 9), (12, 5), (13, 14), (14, 12), (15, 13)

Use Spearman's formula to find the rank correlation coefficient.

8. Two ladies ranked 7 different types of lipsticks in the following order:

Lipsticks	:	A	B	C	D	£	F	G
Lady X	:	2	1	4	3	5	7	6
Lady Y	:	1	3	2	4	5	6	7

Calculate Spearman's rank correlation coefficient.

9. The ranks according to two attributes in a sample are given below :

R_1	:	1	2	3	4	5
R_2	:	5	4	3	2	1

Find the Spearman's rank correlation coefficient. **[*Delhi Univ. B.A. (Eco. Hons.) 1996*]**

10. In a bivariate data, the sum of squares of differences between the ranks of observed values of two variables is 30 and the rank correlation coefficient is 9/11. Find the number of pairs of observations.

11. In a bivariate sample, the sum of squares of differences between marks of observed values of two variables is 33 and the rank correlation between them is 0.8. Find the number of pairs. **[*Delhi Univ. B.A. (Eco. Hons.) 1993*]**

12. In a bivariate distribution, Spearman's coefficient of correlation is – 0.25. If the sum of the squares of various ranks is 150, find out the number of pairs of items. **[*Delhi Univ. B.Com. 2004*]**

13. In a bivariate data, Spearman's coefficient of rank correlation is 0.6 and sum of squares of differences between the ranks of observed values is 66. Find the number of pairs of observations.

14. The coefficient of rank correlation of the marks obtained by 10 students in Statistics and Accountancy was found to be 0.8. It was later discovered that the difference in ranks in the two subjects obtained by one of the students was wrongly taken as 7 instead of 9. Find the correct value of coefficient of rank correlation. **[*Delhi Univ. B.A. (Eco. Hons.) 1992*]**

15. Ten student were ranked on the basis of two attributes : beauty (*X*) and intelligence (*Y*). The coefficient of rank correlation between *X* and *Y* was found to be 0.5. It was later discovered that the difference in ranks in the two attributes obtained by one of the students was wrongly taken as 3 instead of 7. Find the correct coefficient of rank correlation.

16. If the sum of squares of the rank differences of 10 pairs of values is 30, find the correlation coefficient between them.

17. The marks obtained by 10 students in Mathematics and Accountancy are as follows:

Marks in Mathematics (X)	:	80	38	95	30	74	84	91	60	66	40
Marks in Accountancy (Y)	:	85	50	92	58	70	65	88	56	52	46

Find Spearman's coefficient of rank correlation.

18. Calculate Spearman's coefficient of rank correlation for the following data:

X	:	70	65	71	62	58	69	78	64
Y	:	91	76	65	83	90	64	55	48

[*C.A. Foundation, Nov. 1996*]

19. For the following data, calculate the coefficient of rank correlation:

X	:	80	91	99	71	61	81	70	59
Y	:	123	135	154	110	105	134	121	106

[*C.A. Foundation, May 2001*]

20. Calculate Spearman's rank correlation coefficient between advertisement cost and sales from the following data:

Advertisement cost ('000 Rs.)	:	39	65	62	90	82	75	25	98	36	78
Sales (lakhs Rs.)	:	47	53	58	86	62	68	60	91	51	84

21. Calculate Spearman's coefficient of rank correlation from the following data:

X	:	48	33	40	9	16	16	65	24	16	57
Y	:	13	13	24	6	15	4	20	9	6	19

[*Delhi Univ. B.Com. 1988*]

22. The marks obtained by 8 students in Mathematics and Accountancy are as follows:

Marks in Mathematics	:	40	30	50	30	20	10	30	60
Marks in Accountancy	:	15	20	28	12	40	60	20	80

Find Spearman's coefficient of rank correlation.

23. Compute the coefficient of rank correlation between *X* and *Y* from the data given below:

X	:	8	10	7	15	3	20	21	5	10	14	8	16	22	19	6
Y	:	3	12	8	13	20	9	14	11	4	16	15	10	18	23	25

[*Delhi Univ. B.Com. 1998*]

24. Calculate the coefficient of correlation using the method of

ANSWERS

4. 0.46	**5.** 0.85	**6.** A and C	**7.** 0.51
8. 0.786	**9.** -1	**10.** 10	**11.** 10
12. 9	**13.** 10	**14.** 0.606	**15.** 0.258
16. 0.82	**17.** 0.82	**18.** -0.3095	**19.** 0.952
20. 0.82	**21.** 0.734	**22.** 0	**23.** 0.0357