11

MEASURES OF CORRELATION

LEARNING OBJECTIVES

11.1 INTRODUCTION

In the earlier chapters, we have studied the statistical problems and distributions relating to one variable. We discussed various measures of central tendency and dispersion, which are confined to a single variable. This kind of statistical analysis involving one variable is known as *univariate distribution.*

But, we may come across a number of situations with distributions having two variables. *For example,* we may have data relating to income and expenditure, price and demand, height and weight, etc. The distribution involving two variables is called is called *bivariate distribution.* In a **bivariate distribution,** we may be interested to find if there is any relationship between the two variables under study. In day-to-day life, we observe that there exists certain relationship between two variables, like between income and expenditure, price and demand and so on. ***Correlation** is a statistical tool which studies the relationship between two variables.*

Two Variables are Correlated

If the change in one variable results in a corresponding change in the other variable, then the two variables are said to be correlated.

For example:

- Decrease in *'Price'* of a commodity increases its *'Demand'*; or
- Rise in maximum *'Temperature'* leads to an increase in the *'Sale of Ice-creams'*.

Correlation analysis is a means for examining such relationships and aims to determine the extent of relationship between two variables, if any.

Meaning of Correlation

Correlation indicates the relationship between two variables of a series so that changes in the values of one variable are associated with changes in the values of the other variable.

Definitions of Correlation

In the words of L.R. Connor, "If two or more quantities vary in sympathy so that movements in one tend to be accompanied by corresponding movements in others, then they are said to be correlated."

In the words of A.M. Tuttle, "Correlation is an analysis of covariation between two or more variables."

In the words of Ya Lun Chou, "Correlation analysis attempts to determine the 'degree of relationship' between variables."

In the words of Croxton and Cowden, "When the relationship is of a quantitative nature, the appropriate statistical tool for discovering and measuring the relationship and expressing it in a brief formula, is known as correlation."

11.2 CORRELATION AND CAUSATION

Correlation analysis helps us in determining the degree of relationship between two or more variables. **However, it does not tell us anything about cause and effect relationship.** Even a high degree of correlation does not necessarily imply a relationship of cause and effect between the variables.

If there is correlation between two variables, it may be due to the following reasons:

1. Both variables being influenced by a third variable: It is possible that a high degree of correlation between the two variables may be due to the influence of a third variable not included in the analysis.

   For example, a high degree of correlation between the yield per acre of rice and of jute may be due to the fact that both are related to the amount of rainfall. But in reality, none of the two variables is the cause of the other.

2. Mutual Dependence (Cause and Effect): When two variables shows a high degree of correlation, then it may be difficult to explain from the two correlated variables, which is the cause and which is the effect because both may be reacting on each other.

   For example, increase in price of a commodity increases its demand. So, price is the cause and demand is the effect. But, it is also possible that increase in demand of a commodity (due to growth of population or other reasons) may force its price up. Now, the cause is the increased demand and effect is the price.

3. Pure Chance: The correlation between the two variables may be obtained due to sheer coincidence (or pure chance). Such a correlation is known as spurious. So, while interpreting the correlation coefficient, it is essential to see if there is likelihood of any relationship between variables under study.

   For example, we may get a high degree of correlation between two variables in a sample (say, size of shoes and income of people in a locality), when in fact there does not exist any relationship.

So, **correlation analysis provides only a quantitative measure and does not necessarily signify a cause and effect relationship between the variables.** *Therefore, it is necessary to ensure that the variables for correlation analysis are properly selected to make the analysis really purposeful.*

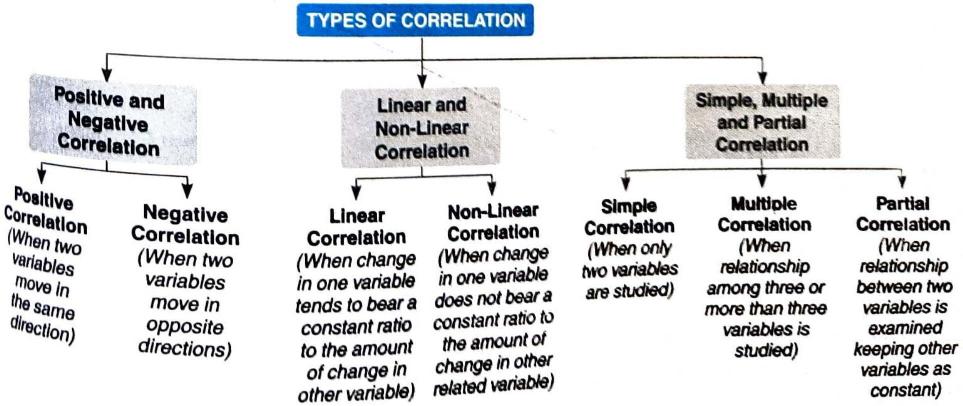
11.3 IMPORTANCE OR SIGNIFICANCE OF CORRELATION

The study of correlation is of great significance because of the following reasons:

1. The correlation coefficient helps in measuring the extent of relationship between two variables in one figure.
2. Correlation analysis facilitates understanding of economic behaviour and helps in locating the critically important variables on which others depend.
3. When two variables are correlated, then value of one variable can be estimated, given the value of another. This is done with the help of regression equations.
4. Correlation facilitates the decision-making in the business world. It reduces the range of uncertainty as predictions based on correlation are likely to be more reliable and near to reality.

11.4 TYPES OF CORRELATION

Correlation can be classified into various categories. The main categories are as follows:


TYPES OF CORRELATION
Positive and Negative Correlation
Linear and Non-Linear Correlation
Simple, Multiple and Partial Correlation
Positive Correlation (When two variables move in the same direction)
Negative Correlation (When two variables move in opposite directions)
Linear Correlation (When change in one variable tends to bear a constant ratio to the amount of change in other variable)
Non-Linear Correlation (When change in one variable does not bear a constant ratio to the amount of change in other related variable)
Simple Correlation (When only two variables are studied)
Multiple Correlation (When relationship among three or more than three variables is studied)
Partial Correlation (When relationship between two variables is examined keeping other variables as constant)

Positive and Negative Correlation

On the bases of direction of change in the values of two variables, correlation can be positive or negative. both ↑ / both ↓

1. Positive Correlation: *When two variables move in the same direction i.e. when one increases the other also increases and when one decreases the other also decreases, then such a relation is called positive correlation.* Examples: Relationship between height and weight, income and expenditure, age of husband and age of wife, etc.

Positive Correlation

Increase in both the variables (X and Y)		Decrease in both the variables (X and Y)	
X	Y	X	Y
20	10	30	20
25	12	25	18
28	15	20	16
30	18	15	14
35	20	10	12

2. Negative Correlation: *When two variables move in opposite directions, i.e. when one increases the other decreases and when one decreases the other increases, then such a relation is called negative correlation.* Examples: Relationship between price and demand, day temperature and sale of woollen garments, etc. opposite direction

Negative Correlation

Rise in value of one variable (X) and fall in other (Y)		Rise in value of one variable (Y) and fall in other (X)	
X	Y	X	Y
10	30	50	5
12	25	40	10
14	20	30	15
16	15	20	20
18	10	10	25

Linear and Non-Linear (Curvilinear) Correlation

On the basis of ratio of variations in the related variables, correlation can be linear or non-Linear.

1. Linear Correlation: *Linear correlation is said to exist if the amount of change in one variable tends to bear a constant ratio to the amount of change in the other variable.* The graph of variables having such a relationship will form a straight line.

X	10	20	30	40
Y	5	10	15	20

In the given schedule, there is a linear relationship between variable X and variable Y as ratio (2:1) of change between X and Y is the same.

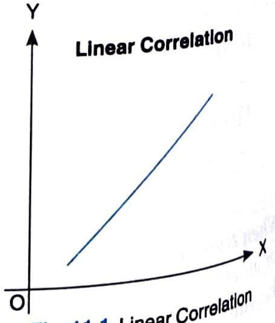


Fig. 11.1 Linear Correlation

2. Non-Linear (Curvilinear) Correlation: *In non-linear or curvilinear correlation, the amount of change in one variable does not bear a constant ratio to the amount of change in the other related variable.* For example, when we double the use of fertilizers, the production of rice would not necessarily double. When the values of the two variables are plotted graphically, these points would not give a straight line.

X	10	20	30	40
Y	5	7	12	20

From the above example, it is clear that the ratio of change between two variables (X and Y) is not the same.

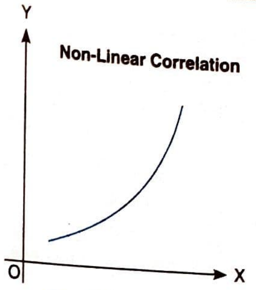


Fig. 11.2 Non-Linear Correlation

Simple, Multiple and Partial Correlation

On the basis of the number of variables involved, correlation may be of three types:

1. Simple Correlation: *When only two variables are studied, it is a problem of simple correlation.* For example, relationship between income and expenditure, price and demand, etc.
2. Multiple Correlation: *When the relationship among three or more than three variables is studied simultaneously, then such relationship is called multiple correlation.* For example, relationship of wheat output with fertilizers and rainfall.
3. Partial Correlation: *Under partial correlation, the relationship between two variables is examined keeping other variables as constant.*

   For example, production of wheat depends on many factors like rainfall, quality of seeds, manure, etc. If we study the relation between production of wheat and quality of seeds, keeping rainfall and manure constant, then the correlation is called partial.

11.5 DEGREE OF CORRELATION

The degree or intensity of relationship between two variables is measured with the help of coefficient of correlation. The degree of correlation can be expressed in the following ways:

Perfect Correlation

If the relationship between the two variables is such that the values of the two variables change (increase or decrease) in the same proportion, correlation between them is said to be perfect. Perfect correlation may be positive or negative.

- If the proportionate change in the values of two variables is in the same direction, it is called perfect positive correlation and the value is described as +1.
- However, if equal proportional changes are in the reverse direction, then the relationship is known as perfect negative correlation and described as –1.

Zero Correlation

When there is no relationship between the two variables, we say that there is zero correlation (or absence of correlation). So, a change in the value of one variable has no particular effect on the value of the other variable. In this case, the value of coefficient of correlation will be zero.

Limited Degree of Correlation

In real life, economic data do not indicate perfect positive or negative correlation. At the same time, the cases of zero correlation are also very limited.

- Generally, we find some limited degree of correlation between economic variables.
- The value of correlation coefficient (r) normally lies in between +1 and –1.
- If the changes in two variables are unequal and in the same direction, correlation is limited and positive.
- Correlation is limited negative when there are unequal changes in the reverse direction.
- *The limited degree of correlation can be high, moderate or low.*

Interpretation of Correlation Coefficient

According to Karl Pearson, the coefficient of correlation lies between two limits i.e. ± 1.

- If there is perfect positive relationship between two variables, the value of correlation would be +1.
- On the contrary, if there is perfect negative relationship between two variables, the value of the correlation will be –1.
- It means r lies between +1 and –1.

Within these limits, the value of correlation is interpreted as:

Degree of Correlation

Degree of Correlation	Positive Correlation	Negative Correlation
Perfect Correlation	+1	–1
Very High degree of Correlation	+0.9	–0.9
Fairly High degree of Correlation	Between + 0.75 and + 0.9	Between – 0.75 and – 0.9
Moderate degree of Correlation	Between + 0.5 and + 0.75	Between – 0.5 and – 0.75
Low degree of Correlation	Between + 0.25 and + 0.5	Between – 0.25 and – 0.5
Very low degree of Correlation	Between 0 and + 0.25	Between 0 and – 0.25
No Correlation	0	0

11.6 METHODS OF MEASUREMENTS OF CORRELATION

There are different methods for measuring correlation between two variables. Some of them are:

1. Scatter Diagram
2. Karl Pearson's Coefficient of Correlation
3. Spearman's Rank Correlation Coefficient

11.7 SCATTER DIAGRAM

Scatter diagram is a simple and attractive method of diagrammatic representation of a bivariate distribution to determine the nature of correlation between the variables.

- A scatter diagram gives a visual idea about the nature of association between the two variables.

Visually Present Correlation.

- It is the simplest method of studying the relationship between two variables, without calculating any numerical value.
- *How to Draw a Scatter Diagram:* Plot the values of the variables X and Y along the X-axis and Y-axis respectively. Show the values of two variables by dots on the graph. Each dot represents a pair of values.

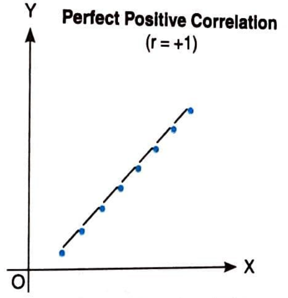
This graphical representation of the values of the variables is known as scatter diagram or dot diagram.

Interpretation of Scatter Diagram

By observing the pattern of dots, we can know the presence or absence of correlation and the type of correlation. *Inspection of the scatter diagram gives an idea of the nature and intensity of the relationship.*

The scatter diagram can be interpreted in the following ways:

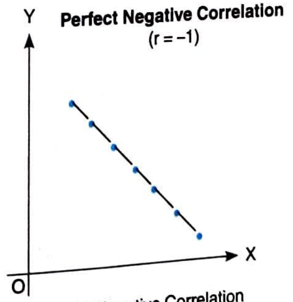
1. Perfect Positive Correlation: If all the points of scatter diagram fall on a straight line with positive slope (see Fig. 11.3), then the correlation is said to be perfectly positive (r = + 1)



Perfect Positive Correlation

Fig. 11.3

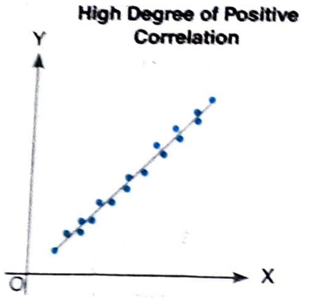
2. Perfect Negative Correlation: If all the points of scatter diagram fall on a straight line with negative slope (see Fig. 11.4), then the correlation is said to be perfectly negative (r = – 1).



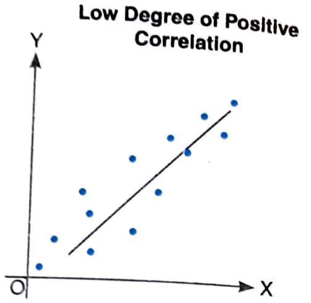
Perfect Negative Correlation

Fig. 11.4

3. Positive Correlation: When all the points of scatter diagram cluster around a straight line going upwards from left to right, the correlation is positive correlation (see Fig. 11.5 and 11.6).

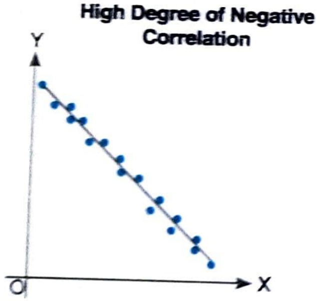


High Degree Positive Correlation
Fig. 11.5

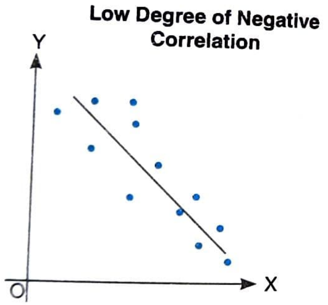


Low Degree Positive Correlation
Fig. 11.6

4. Negative Correlation: When all the points of scatter diagram cluster around a straight line with negative slope, the correlation is said to be negative as shown in (Fig. 11.7 and 11.8).

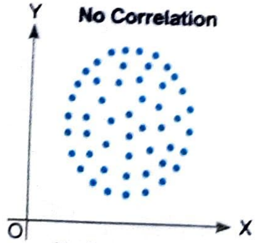


High Degree Negative Correlation
Fig. 11.7

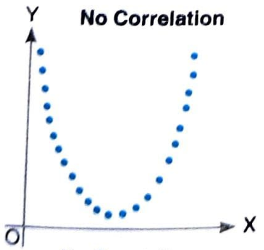


Low Degree Negative Correlation
Fig. 11.8

5. No Correlation: If the points are scattered in a haphazard manner, then it is a case of zero or no correlation (see Fig. 11.9 and 11.10).



No Correlation
Fig. 11.9



No Correlation
Fig. 11.10

How to interpret a Scatter Diagram?

Consider the following points, while interpreting a scatter diagram:

1. **Dense or scattered points:** If the points (dots) are close to each other, a high degree of correlation may be expected between the two variables. However, if the points are widely scattered, a poor correlation may be expected between them.
2. **Trend or no trend:** If the points on the scatter diagram shows any trend (either upward or downward), then variables are said to be correlated and if no trend is revealed, the variables are uncorrelated.
3. **Upward or Downward trend:** If there is an upward trend rising from lower left hand corner and going upward to the upper right hand corner, the correlation is positive (it means values of the two variables move in the same direction). On the other hand, if points depict a downward trend from the upper left hand corner to the lower right hand corner, the correlation is negative (it means values of the two variables move in the opposite directions).
4. **Perfect Correlation:** If the points on the scatter diagram lie on a straight line with a positive slope, then correlation is perfect and positive. On the other hand, if all the points lie on a straight line with a negative slope, then there is perfect negative correlation.

Practicals on Scatter Diagram

Example 1. Make a scatter diagram for the following data and state the type of correlation between X and Y.

X	10	20	30	40	50
Y	70	140	210	280	350

Solution:

The scatter diagram is obtained by plotting the values of Series X on the X-axis and values of Series Y on the Y-axis. Plotting the values (10, 70), (20, 140),.........(50, 350) on the graph paper, we get the scatter diagram (See Fig. 11.11):

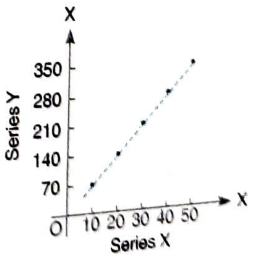


Fig. 11.11

It is obvious from the scatter diagram that ***there is perfect positive correlation between the values of Series X and Series Y.***

Example 2. Draw a scatter diagram to represent the following values of X and Y variables. Comment on the type and degree of correlation.

X	15	20	25	27	30
Y	7	10	12	16	18

Solution:

Plot the values of variable X on the X-axis and variable Y on the Y-axis.

A glance at the above scatter diagram shows that there is an upward trend of the dots from lower left-hand corner to the upper right-hand corner. It means, ***there is positive correlation between values of X and Y variables.***

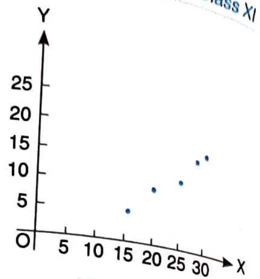


Fig. 11.12

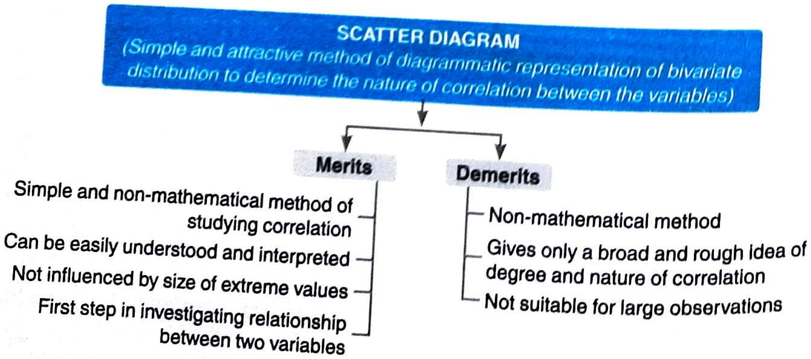
Merits and Demerits of Scatter Diagram

Merits of Scatter Diagram

1. Simplicity: It is a simple and a non-mathematical method of studying correlation between two variables.
2. Easily understandable: It can be easily understood and interpreted. It enables us to know the presence or absence of correlation at a single glance of the diagram.
3. Not affected by extreme items: It is not influenced by the size of extreme values, whereas most of the mathematical methods lack this quality.
4. First Step: It is a first step in investigating the relationship between two variables.

Demerits of Scatter Diagram

1. Non-mathematical method: This method does not indicate the exact numerical value of correlation which is possible by other mathematical methods of correlation.
2. Rough Measure: It gives only a broad and rough idea of the degree and nature of correlation between two variables. Thus, it is only a qualitative expression rather than a quantitative expression.
3. Unsuitable for large observations: It is not possible to draw a scatter diagram on a graph paper in case of more than two variables.


SCATTER DIAGRAM
(Simple and attractive method of diagrammatic representation of bivariate distribution to determine the nature of correlation between the variables)
Merits
Demerits
Simple and non-mathematical method of studying correlation
Can be easily understood and interpreted
Not influenced by size of extreme values
First step in investigating relationship between two variables
Non-mathematical method
Gives only a broad and rough idea of degree and nature of correlation
Not suitable for large observations

*Scatter diagram method shows the existence and direction of correlation. However, it does not quantify the extent of correlation, i.e., it does not indicate the exact numerical value of correlation. A mathematical method for measuring the correlation between the two variables was suggested by the famous British statistician, Karl Pearson, known as **Karl Pearson's Coefficient of Correlation.***

11.8 KARL PEARSON'S COEFFICIENT OF CORRELATION

Karl Pearson (1867–1936) was the first person to give a mathematical formula for measuring the degree of relationship between two variables in 1890.

Karl Pearson (1867 - 1936)

- The Karl Pearson's coefficient of correlation is also known as *'Product Moment Correlation'* or *'Simple Correlation Coefficient'*.
- It is the most popular and widely used method to calculate the correlation coefficient.
- It is denoted by the symbol 'r' (r is a pure number, i.e. it has no unit).

According to Karl Pearson, Coefficient of correlation is determined by dividing the sum of products of deviations from their respective means by their number of pairs and their standard deviations."

It means, *Karl Pearson's Coefficient of Correlation (r) is calculated as:*

$$r = \frac{\text{Sum of Products of Deviations from their respective means}}{\text{Number of Pairs} \times \text{Standard Deviations of Both Series}}$$

$$\text{i.e. } r = \frac{\Sigma xy}{N \times \sigma_x \times \sigma_y}$$

Where:

N = Number of Pair of observations

x = Deviation of X series from mean $(X - \overline{X})$

y = Deviation of Y series from mean $(Y - \overline{Y})$

σ_x = Standard deviation of X series, i.e., $\sqrt{\frac{\Sigma x^2}{N}}$

σ_x = Standard deviation of Y series, i.e., $\sqrt{\frac{\Sigma y^2}{N}}$

r = Coefficient of Correlation

Karl Pearson's Coefficient of Correlation and Covariance

Karl Pearson's method of calculating coefficient of correlation is based on covariance of the concerned variables.

Covariance is a statistical representation of the degree to which two variables vary together. Basically, it is a number that reflects the degree to which two variable vary together.

- Covariance Symbol of two variables X and Y is denoted by = COV (X, Y)

- Covariance of X and Y is defined as:

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

The formula mentioned for computing Pearsonian coefficient of correlation can be transformed into a much easier formula:

As stated earlier:

$$r = \frac{\Sigma xy}{N \times \sigma_x \times \sigma_y}$$

or, $$r = \frac{\Sigma xy}{N} \times \frac{1}{\sigma_x} \times \frac{1}{\sigma_y}$$

or, $$r = \frac{\Sigma xy}{N \times \sqrt{\frac{\Sigma x^2}{N}} \times \sqrt{\frac{\Sigma y^2}{N}}}$$

or, $$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

It must be noted that this method is to be applied only when the deviations of items are taken from actual means and not from assumed means.

Example 3. Find the coefficient of correlation between X and Y.

	X Series	*Y Series*
No. of items	30	30
Standard deviation	3	2

Summation of product of deviations of X and Y series from their respective means = 150

Solution:

Given, $N = 30$, $\sigma_x = 3$, $\sigma_y = 2$ and $\Sigma xy = 150$

$$r = \frac{\Sigma xy}{N \times \sigma_x \times \sigma_y} = \frac{150}{30 \times 3 \times 2} = \frac{150}{180} = 0.833$$

Ans. Coefficient of correlation = 0.833. There is a fairly high degree of positive correlation between X and Y.

Example 4. Find standard deviation of Y series, if: Coefficient of correlation = 0.25; Covariance between X and Y = 10; Variance of X = 64

Solution:

$$\text{Covariance between X and Y} = \frac{\Sigma xy}{N} = 10$$

$$\text{Variance of X} = \sigma_x^2 = 64$$

$$\sigma_x = \sqrt{64} = 8$$

$$\text{Coefficient of Correlation (r)} = \frac{\Sigma xy}{N \times \sigma_x \times \sigma_y}$$

$$r = \frac{\Sigma xy}{N} \times \frac{1}{\sigma_x} \times \frac{1}{\sigma_y}$$

$$0.25 = 10 \times \frac{1}{8} \times \frac{1}{\sigma_y}$$

$$\sigma_y = \frac{10}{8 \times 0.25}$$

Ans. Standard deviation of Y series = 5

Example 5. If covariance between two variables X and Y is 9.4 and variance of Series X and Series Y are 10.6 and 12.5 respectively. Calculate the coefficient of correlation.

Solution:

$$\text{Covariance between X and Y} = \frac{\Sigma xy}{N} = 9.4$$

$$\text{Variance of X} = \sigma_x^2 = 10.6$$

$$\sigma_x = \sqrt{10.6} = 3.25$$

$$\text{Variance of Y} = \sigma_y^2 = 12.5$$

$$\sigma_y = \sqrt{12.5} = 3.53$$

$$\text{Coefficient of Correlation (r)} = \frac{\Sigma xy}{N \times \sigma_x \times \sigma_y}$$

$$r = \frac{\Sigma xy}{N} \times \frac{1}{\sigma_x} \times \frac{1}{\sigma_y} = 9.4 \times \frac{1}{3.25} \times \frac{1}{3.53}$$

$r = 9.4 \times 0.307 \times 0.283 = 0.816$

Ans. Coefficient of correlation = 0.816. There is a fairly high degree of positive correlation between X and Y.

Features of Karl Pearson's Coefficient of Correlation

1. **Knowledge of direction of correlation:** It gives the knowledge about the direction of relationship, i.e., whether the relationship between two variables is positive or negative.
2. **Size of Correlation:** It indicates the size of relationship between the variables, i.e. correlation coefficient ranges between +1 and −1.
3. **Ideal Measure:** It is an appropriate measure of correlation as it is based on most important statistical measures like mean and standard deviation.
4. **Indicates magnitude and direction:** The coefficient of correlation not only specifies the magnitude of correlation but also its direction. If the two variables are directly related, then the correlation coefficient will be a positive value and in case of inverse relationship, we will get negative correlation coefficient.

Value of Correlation Coefficient

The value of the correlation coefficient shall always lie between ±1.

- When r = +1, it means there is perfect positive correlation:
- When r = –1, it shows that there exists perfect negative correlation between the two variables.
- If r = 0, it means there is no relationship between the two variables.

However, in practice, correlation coefficient normally lies in between +1 and –1.

11.9 CALCULATION OF KARL PEARSON'S COEFFICIENT OF CORRELATION

While calculating coefficient of correlation according to Karl Pearson's formula, we can use the following methods:

1. Actual Mean Method (*Example 6*)
2. Direct Method (*Example 9*)
3. Short-Cut Method/Assumed Mean Method/Indirect Method (*Example 10*)
4. Step Deviation Method (*Example 12*)

Actual Mean Method

Karl Pearson's method of computing correlation by the Actual Mean method involves the following steps:

Steps for Calculation:

Step 1. Calculate the means of the two series (X and Y), i.e., calculate $\overline{X}$ and $\overline{Y}$.

Step 2. Take the deviation of X series from $\overline{X}$ (mean of X) and denote the deviations by x.

Step 3. Square these deviations and obtain the total, i.e., Σx^2.

Step 4. Take the deviations of Y series from $\overline{Y}$ (mean of Y) and denote the deviations by y.

Step 5. Square these deviations and obtain the total, i.e., Σy^2.

Step 6. Multiply the respective deviations of X and Y series and obtain their total, i.e., Σxy

Step 7. Substitute the values of Σxy, Σx^2, Σy^2 in the following formula:

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

Example 6 will illustrate the computation of correlation by Actual Mean Method.

Example 6. Calculate the coefficient of correlation for the following data by the Actual Mean Method.

X	12	15	18	21	24	27	30
Y	6	8	10	12	14	16	18

Solution:

Calculation of Coefficient of Correlation (Actual Mean Method)

X-Series			Y-Series			
X	$x = X - \overline{X}$	x^2	Y	$y = Y - \overline{Y}$	y^2	xy
12	– 9	81	6	– 6	36	54
15	– 6	36	8	– 4	16	24
18	– 3	9	10	– 2	4	6
21	0	0	12	0	0	0
24	+ 3	9	14	+ 2	4	6
27	+ 6	36	16	+ 4	16	24
30	+ 9	81	18	+ 6	36	54
ΣX = 147		$\Sigma x^2 = 252$	ΣY = 84		$\Sigma y^2 = 112$	Σxy = 168

$$\overline{X} = \frac{\Sigma X}{N} = \frac{147}{7} = 21$$

$$\overline{Y} = \frac{\Sigma Y}{N} = \frac{84}{7} = 12$$

$$\text{Coefficient of Correlation (r)} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

$\Sigma xy = 168$; $\Sigma x^2 = 252$; $\Sigma y^2 = 112$

$$= \frac{168}{\sqrt{252 \times 112}} = \frac{168}{\sqrt{28{,}224}} = \frac{168}{168} = 1$$

Ans. Coefficient of Correlation = 1. There is perfect positive correlation between the values of Series X and Series Y.

Note: *Actual Mean method has a lengthy process because the actual means of both the series are to be calculated and then deviations are taken.*

Example 7. Calculate coefficient of correlation from the following data:

(i) Sum of squares of deviations of X values from mean (Σx^2) = 136.

(ii) Sum of squares of deviations of Y values from mean (Σy^2) = 138.

(iii) Sum of products of deviation of X and Y values from their means (Σxy) = 122.

Solution:

$$\text{Coefficient of Correlation (r)} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

$\Sigma xy = 122$; $\Sigma x^2 = 136$; $\Sigma y^2 = 138$

$$= \frac{122}{\sqrt{136 \times 138}} = \frac{122}{\sqrt{18{,}768}} = \frac{122}{137} = 0.89$$

Ans. Coefficient of Correlation = 0.89. There is a fairly high degree of positive correlation between X and Y.

Example 8. Calculate the number of items, when standard deviation of series X is 6 and coefficient of correlation is 0.75. Also given that sum of the product of deviations of X and Y from actual means is 180 and sum of squares of deviations of Y from actual means is 160.

Solution:

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

$$0.75 = \frac{180}{\sqrt{\Sigma x^2 \times 160}}$$

$$(0.75)^2 = \frac{(180)^2}{\Sigma x^2 \times 160}$$

$$0.5625 = \frac{32{,}400}{\Sigma x^2 \times 160}$$

$$\Sigma x^2 = \frac{32{,}400}{160 \times 0.5625} = \frac{32{,}400}{90} = 360$$

$$\Sigma x^2 = 360$$

$$\text{Standard Deviation } (\sigma_x) = \sqrt{\frac{\Sigma x^2}{N}}$$

$$6 = \sqrt{\frac{360}{N}}$$

$$(6)^2 = \frac{360}{N}$$

$$N = \frac{360}{36} = 10$$

Ans. Number of items = 10.

Direct Method

The coefficient of correlation can also be obtained ***without finding out the deviations from the actual means.*** The steps involved in the Direct Method are:

Steps for Calculation

Step 1. Calculate the sum of Series X and denote it by ΣX

Step 2. Calculate the sum of Series Y and denote it by ΣY

Step 3. Square the values of X series and obtain the total, i.e., ΣX^2

Step 4. Square the values of Y series and obtain the total, i.e., ΣY^2

Step 5. Multiply the values of series X and series Y and obtain their total, i.e., ΣXY

Step 6. Substitute the values of ΣXY, ΣX, ΣY, ΣX^2 and ΣY^2 in the following formula:

$$\text{Coefficient of Correlation (r)} = \frac{N\Sigma XY - \Sigma X.\Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2} \times \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

The following example will make the direct method more clear:

Example 9. Calculate the coefficient of correlation of the data given in *Example 6* by the Direct Method.

Solution:

Calculation of Coefficient of Correlation (Direct Method)

X-Series		*Y-Series*		
X	*X²*	*Y*	*Y²*	*XY*
12	144	6	36	72
15	225	8	64	120
18	324	10	100	180
21	441	12	144	252
24	576	14	196	336
27	729	16	256	432
30	900	18	324	540
ΣX = 147	**ΣX² = 3,339**	**ΣY = 84**	**ΣY² = 1,120**	**ΣXY = 1,932**

$$\text{Coefficient of Correlation (r)} = \frac{N\Sigma XY - \Sigma X.\Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2} \times \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

Here, ΣX = 147; ΣY = 84; ΣX^2 = 3,339; ΣY^2 = 1,120; ΣXY = 1,932 and N = 7

$$= \frac{7 \times 1{,}932 - 147 \times 84}{\sqrt{7 \times 3{,}339 - (147)^2} \times \sqrt{7 \times 1{,}120 - (84)^2}}$$

$$= \frac{13{,}524 - 12{,}348}{\sqrt{23{,}373 - 21{,}609} \times \sqrt{7{,}840 - 7{,}056}}$$

$$= \frac{1{,}176}{\sqrt{1{,}764} \times \sqrt{784}} = \frac{1{,}176}{42 \times 28} = \frac{1{,}176}{1{,}176} = 1$$

Ans. Coefficient of Correlation = 1. There is perfect positive correlation between the values of Series X and Series Y.

Short-Cut Method (Assumed Mean Method)

The method based on actual means is quite lengthy and is possible only when the mean values are whole numbers. But in practice, mean values are in fractions.

In order to avoid difficult calculations, it is better to use short-cut method (assumed mean method). In this method, deviations are taken from an assumed mean and the following formula is used:

$$r = \frac{N\Sigma dxdy - \Sigma dx \times \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2} \times \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

Where:

N	=	Number of pair of observations
Σdx	=	Sum of deviations of X values from assumed mean .
Σdy	=	Sum of deviations of Y values from assumed mean.
Σdx^2	=	Sum of squared deviations of X values from assumed mean.
Σdy^2	=	Sum of squared deviations of Y values from assumed mean.
$\Sigma dxdy$	=	Sum of the products of deviations dx and dy.

Steps of Short-cut method

Step 1. Take the deviations of X series from the assumed mean and denote them by dx and obtain their total, i.e., Σdx.

Step 2. Square the deviations of X series and obtain the total, i.e., Σdx^2.

Step 3. Take the deviations of Y series from the assumed mean and denote them by dy and obtain their total, i.e., Σdy

Step 4. Square the deviations of Y series and obtain the total, i.e., Σdy^2.

Step 5. Multiply dx with dy and obtain the total $\Sigma dxdy$.

Step 6. Substitute the values of Σdx, Σdy, Σdx^2, Σdy^2 and $\Sigma dxdy$ in the following formula:

$$r = \frac{N\Sigma dxdy - \Sigma dx \times \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2} \times \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

Example 10 will illustrate the calculation of coefficient of correlation by using short-cut method.

Example 10. Calculate the coefficient of correlation of the data given in *Example 6* by Short-Cut Method.

Solution:

Calculation of Coefficient of Correlation (Short-Cut Method)

X-Series			*Y-Series*			
X	*dx = X – A* *A = 18*	*dx²*	*Y*	*dy = Y – A* *A = 10*	*dy²*	*dxdy*
12	–6	36	6	–4	16	24
15	–3	9	8	–2	4	6
18 (A)	0	0	10 (A)	0	0	0
21	+3	9	12	+2	4	6
24	+6	36	14	+4	16	24
27	+9	81	16	+6	36	54
30	+12	144	18	+8	64	96
	$\Sigma dx = 21$	$\Sigma dx^2 = 315$		$\Sigma dy = 14$	$\Sigma dy^2 = 140$	$\Sigma dxdy = 210$

$$r = \frac{N\Sigma dxdy - \Sigma dx \times \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2} \times \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

$\Sigma dxdy = 210$; $\Sigma dx = 21$; $\Sigma dy = 14$; $N = 7$; $\Sigma dx^2 = 315$; $\Sigma dy^2 = 140$

$$= \frac{7 \times 210 - (21)(14)}{\sqrt{7 \times 315 - (21)^2} \times \sqrt{7 \times 140 - (14)^2}}$$

$$= \frac{1{,}470 - 294}{\sqrt{1{,}764} \times \sqrt{784}} = \frac{1{,}176}{42 \times 28} = \frac{1{,}176}{1{,}176} = 1$$

Ans. Coefficient of Correlation = 1. There is perfect positive correlation between the values of Series X and Series Y.

Example 11. Calculate the Karl Pearson's coefficient of correlation from the following data:

(i) Sum of deviation of X values (Σdx) = 5

(ii) Sum of deviation of Y values (Σdy) = 4

(iii) Sum of squares of deviations of X values (Σdx^2) = 40

(iv) Sum of squares of deviation of Y values (Σdy^2) = 50

(v) Sum of the product of deviations of X and Y values ($\Sigma dxdy$) = 32

(vi) No of pairs of observations (N) = 10

Solution:

$$r = \frac{N\Sigma dxdy - \Sigma dx \times \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2} \times \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

Here, $\Sigma dxdy = 32$; $\Sigma dx = 5$; $\Sigma dy = 4$; $N = 10$; $\Sigma dx^2 = 40$; $\Sigma dy^2 = 50$

$$= \frac{10 \times 32 - (5)(4)}{\sqrt{10 \times 40 - (5)^2} \times \sqrt{10 \times 50 - (4)^2}}$$

$$= \frac{320 - 20}{\sqrt{375} \times \sqrt{484}} = \frac{300}{426.028} = 0.704$$

Ans. Coefficient of Correlation = 0.704. There is a fairly high degree of positive correlation between X and Y.

Step Deviation Method

The use of step deviation method simplifies the method of calculating coefficient of correlation. Under this method, deviations of X and Y are taken from assumed means and are divided by a common factor.

Steps of Step Deviation Method

Step 1. Take the deviations of X series from the assumed mean and divide them by common factor (C) to get step deviations (dx′). Find out their total to get $\Sigma dx'$.

$$r = \frac{N\Sigma dxdy - \Sigma dx \times \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2} \times \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

Where:

N = Number of pair of observations

Σdx = Sum of deviations of X values from assumed mean .

Σdy = Sum of deviations of Y values from assumed mean.

Σdx^2 = Sum of squared deviations of X values from assumed mean.

Σdy^2 = Sum of squared deviations of Y values from assumed mean.

$\Sigma dxdy$ = Sum of the products of deviations dx and dy.

Steps of Short-cut method

Step 1. Take the deviations of X series from the assumed mean and denote them by dx and obtain their total, i.e., Σdx.

Step 2. Square the deviations of X series and obtain the total, i.e., Σdx^2.

Step 3. Take the deviations of Y series from the assumed mean and denote them by dy and obtain their total, i.e., Σdy

Step 4. Square the deviations of Y series and obtain the total, i.e., Σdy^2.

Step 5. Multiply dx with dy and obtain the total $\Sigma dxdy$.

Step 6. Substitute the values of Σdx, Σdy, Σdx^2, Σdy^2 and $\Sigma dxdy$ in the following formula:

$$r = \frac{N\Sigma dxdy - \Sigma dx \times \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2} \times \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

Example 10 will illustrate the calculation of coefficient of correlation by using short-cut method.

Example 10. Calculate the coefficient of correlation of the data given in *Example 6* by Short-Cut Method.

Solution:

Calculation of Coefficient of Correlation (Short-Cut Method)

X-Series			*Y-Series*			
X	$dx = X - A$ $A = 18$	dx^2	Y	$dy = Y - A$ $A = 10$	dy^2	dxdy
12	−6	36	6	−4	16	24
15	−3	9	8	−2	4	6
18 (A)	0	0	10 (A)	0	0	0
21	+3	9	12	+2	4	6
24	+6	36	14	+4	16	24
27	+9	81	16	+6	36	54
30	+12	144	18	+8	64	96
	$\Sigma dx = 21$	$\Sigma dx^2 = 315$		$\Sigma dy = 14$	$\Sigma dy^2 = 140$	$\Sigma dxdy = 210$

$$r = \frac{N\Sigma dxdy - \Sigma dx \times \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2} \times \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

$\Sigma dxdy = 210; \Sigma dx = 21; \Sigma dy = 14; N = 7; \Sigma dx^2 = 315; \Sigma dy^2 = 140$

$$= \frac{7 \times 210 - (21)(14)}{\sqrt{7 \times 315 - (21)^2} \times \sqrt{7 \times 140 - (14)^2}}$$

$$= \frac{1,470 - 294}{\sqrt{1,764} \times \sqrt{784}} = \frac{1,176}{42 \times 28} = \frac{1,176}{1,176} = 1$$

Ans. Coefficient of Correlation = 1. There is perfect positive correlation between the values of Series X and Series Y.

Example 11. Calculate the Karl Pearson's coefficient of correlation from the following data:

(i) Sum of deviation of X values $(\Sigma dx) = 5$

(ii) Sum of deviation of Y values $(\Sigma dy) = 4$

(iii) Sum of squares of deviations of X values $(\Sigma dx^2) = 40$

(iv) Sum of squares of deviation of Y values $(\Sigma dy^2) = 50$

(v) Sum of the product of deviations of X and Y values $(\Sigma dxdy) = 32$

(vi) No of pairs of observations $(N) = 10$

Solution:

$$r = \frac{N\Sigma dxdy - \Sigma dx \times \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2} \times \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

Here, $\Sigma dxdy = 32; \Sigma dx = 5; \Sigma dy = 4; N = 10; \Sigma dx^2 = 40; \Sigma dy^2 = 50$

$$= \frac{10 \times 32 - (5)(4)}{\sqrt{10 \times 40 - (5)^2} \times \sqrt{10 \times 50 - (4)^2}}$$

$$= \frac{320 - 20}{\sqrt{375} \times \sqrt{484}} = \frac{300}{426.028} = 0.704$$

Ans. Coefficient of Correlation = 0.704. There is a fairly high degree of positive correlation between X and Y.

Step Deviation Method

The use of step deviation method simplifies the method of calculating coefficient of correlation. Under this method, deviations of X and Y are taken from assumed means and are divided by a common factor.

Steps of Step Deviation Method

Step 1. Take the deviations of X series from the assumed mean and divide them by common factor (C) to get step deviations (dx'). Find out their total to get $\Sigma dx'$.

Step 2. Take the deviations of Y series from the assumed mean and divide them by common factor to get step deviations (dy′). Obtain their total to get Σdy′.

Step 3. Square the step deviations of X series and obtain the total, i.e., $\Sigma dx'^2$.

Step 4. Square the step deviations of Y series and find the total to get $\Sigma dy'^2$

Step 5. Multiply dx′ with dy′ and obtain the total Σdx′dy′.

Step 6. Substitute the values of Σdx′, Σdy′, $\Sigma dx'^2$, $\Sigma dy'^2$ and Σdx′dy′ in the following formula:

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

Where:

N = Number of pair of observations.

Σdx′ = Sum of step deviations of X values from assumed mean.

Σdy′ = Sum of step deviations of Y values from assumed mean.

$\Sigma dx'^2$ = Sum of squared step deviations of X values from assumed mean.

$\Sigma dy'^2$ = Sum of squared step deviations of Y values from assumed mean.

Σdx′dy′ = Sum of the products of step deviations dx′ and dy′.

Let us understand this method with the help of following example.

Example 12. Calculate the coefficient of correlation of the data given in *Example 6* by the Step Deviation Method.

Solution:

Calculation of Coefficient of Correlation (Step Deviation Method)

X-Series				Y-Series				
X	dx = X − A A = 18	$dx' = \frac{dx}{C}$ C = 3	dx'^2	Y	dy = Y − A A = 10	$dy' = \frac{dy}{C}$ C = 2	dy'^2	dx′dy′
12	−6	−2	4	6	−4	−2	4	4
15	−3	−1	1	8	−2	−1	1	1
18 (A)	0	0	0	**10 (A)**	0	0	0	0
21	+3	+1	1	12	+2	+1	1	1
24	+6	+2	4	14	+4	+2	4	4
27	+9	+3	9	16	+6	+3	9	9
30	+12	+4	16	18	+8	+4	16	16
		Σdx′ = 7	**$\Sigma dx'^2$ = 35**			**Σdy′ = 7**	**$\Sigma dy'^2$ = 35**	**Σdx′dy′ = 35**

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

Here, Σdx′dy′ = 35; Σdx′ = 7; Σdy′ = 7; N = 7; $\Sigma dx'^2 = 35$; $\Sigma dy'^2 = 35$

$$= \frac{7 \times 35 - (7) \times (7)}{\sqrt{7 \times 35 - (7)^2} \times \sqrt{7 \times 35 - (7)^2}}$$

$$= \frac{245 - 49}{\sqrt{196} \times \sqrt{196}} = \frac{196}{196} = 1$$

Ans. Coefficient of Correlation = 1. There is perfect positive correlation between values of Series X and Series Y.

Change of Scale and Origin

Coefficient of correlation is independent of change of scale and origin.

- Any constant added or subtracted (change of origin) does not affect the value of correlation coefficient.
- Similarly, any constant multiplied or divided (change of scale) will also not affect the coefficient of correlation.

In **Example 12,** values of deviation of X series (12, 15, 18, etc.) are divided by a common factor 3. Similarly, values of deviation of Y series (6, 8, 10, etc.) are divided by a common factor 2. But such a change in scale of X or Y series has not changed the value of r (value of r is 1). So, we can conclude that correlation coefficient is independent of change of scale of X and Y.

Refer Example 13 for more clarity on this point.

Example 13. Find coefficient of correlation from the following figures:

X	0.5	1.5	1.2	2.0	1.8	2.5	1.6	3.0	3.2	3.5
Y	100	400	300	600	500	700	400	800	900	800

Solution:

The coefficient of correlation in not affected by the change in scale and origin of variables. So, in order to make calculations easier, we can multiply series X by 10 and divide series Y by 100. By doing so, we get the following table:

Calculation of Coefficient of Correlation

Series X			Series Y			
X	dx = X − A A = 20	dx^2	Y	dy = Y − A A = 5	dy^2	dxdy
5	−15	225	1	−4	16	60
15	−5	25	4	−1	1	5
12	−8	64	3	−2	4	16
20 (A)	0	0	6	+1	1	0
18	−2	4	**5 (A)**	0	0	0
25	+5	25	7	+2	4	10
16	−4	16	4	−1	1	4
30	+10	100	8	+3	9	30
32	+12	144	9	+4	16	48
35	+15	225	8	+3	9	45
	Σdx = 8	**Σdx^2 = 828**		**Σdy = 5**	**Σdy^2 = 61**	**Σdxdy = 218**

$$r = \frac{N\Sigma dxdy - \Sigma dx \times \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2} \times \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

Here, $\Sigma dxdy = 218$; $\Sigma dx = 8$; $\Sigma dy = 5$; $N = 10$; $\Sigma dx^2 = 828$ and $\Sigma dy^2 = 61$

$$= \frac{10 \times 218 - (8)\,(5)}{\sqrt{10 \times 828 - (8)^2} \times \sqrt{10 \times 61 - (5)^2}}$$

$$= \frac{2{,}180 - 40}{\sqrt{8{,}216} \times \sqrt{585}} = \frac{2{,}140}{2{,}192.34} = 0.976$$

Ans. Coefficient of Correlation = 0.976. There is a very high degree of positive correlation between X and Y.

Example 14. The following table shows the heights of a sample of 10 fathers and their eldest sons:

Height of father (in cm)	170	167	162	163	167	166	169	171	164	165
Height of son (in cm)	168	167	166	166	168	165	168	170	165	168

Solution:

Calculation of Coefficient of Correlation

Height of father (in cm)			*Height of son (in cm)*			
X	*dx = X – A* *A = 165*	*dx²*	*Y*	*dy = Y – A* *A = 165*	*dy²*	*dxdy*
170	+ 5	25	168	+ 3	9	+ 15
167	+ 2	4	167	+ 2	4	+ 4
162	– 3	9	166	+ 1	1	– 3
163	– 2	4	166	+ 1	1	– 2
167	+ 2	4	168	+ 3	9	+ 6
166	+ 1	1	**165 (A)**	0	0	0
169	+ 4	16	168	+ 3	9	+ 12
171	+ 6	36	170	+ 5	25	+ 30
164	– 1	1	165	0	0	0
165 (A)	0	0	168	+ 3	9	0
	Σdx = 14	**Σdx² = 100**		**Σdy = 21**	**Σdy² = 67**	**Σdxdy = 62**

$$r = \frac{N\Sigma dxdy - \Sigma dx \times \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2} \times \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

Here, $\Sigma dxdy = 62$; $\Sigma dx = 14$; $\Sigma dy = 21$; $N = 10$; $\Sigma dx^2 = 100$; $\Sigma dy^2 = 67$

$$= \frac{10 \times 62 - (14)\,(21)}{\sqrt{10 \times 100 - (14)^2} \times \sqrt{10 \times 67 - (21)^2}}$$

$$= \frac{620 - 294}{\sqrt{804} \times \sqrt{229}} = \frac{326}{429.087} = 0.76$$

The given example can also be simplified with the help of Log Tables:

$$\text{Let } z = \frac{32.6}{\sqrt{80.4} \times \sqrt{22.9}}$$

$$\log z = \log\left[\frac{32.6}{\sqrt{80.4} \times \sqrt{22.9}}\right]$$ (Taking log on both sides)

$$= \log(32.6) - \log\left[\sqrt{80.4} \times \sqrt{22.9}\right]$$ $\left(\because \log\frac{m}{n} = \log m - \log n\right)$

$$= \log(32.6) - \log\left[(80.4 \times 22.9)^{\frac{1}{2}}\right]$$

$$= \log(32.6) - \frac{1}{2}\left[\log(80.4) + \log(22.9)\right]$$ $(\because \log m^n = n \log m)$

$$= 1.5132 - \frac{1}{2}\left[1.9053 + 1.3598\right] = 1.5132 - 1.6325 = -0.1193$$

$$= -1 + (1 - 0.1193)$$

$$\log z = \bar{1}.8807$$

$$z = \text{Antilog}\,(\bar{1}.8807) = 0.7598 = 0.76$$

Ans. Coefficient of Correlation = 0.76. It shows a fairly high degree of positive correlation between the heights of fathers and their eldest sons.

Example 15. Calculate the coefficient of correlation for the following data:

Husband's age	*Wife's age*
30	22
32	25
34	27
35	28
37	29
38	30
40	31
42	32
44	33

Solution:

Calculation of Coefficient of Correlation

Husband's Age (X-Series)			*Wife's Age (Y-Series)*			
X	*dx = X – A* *A = 37*	*dx²*	*Y*	*dy = Y – A* *A = 28*	*dy²*	*dxdy*
30	– 7	49	22	– 6	36	42
32	– 5	25	25	– 3	9	15
34	– 3	9	27	– 1	1	3
35	– 2	4	**28 (A)**	0	0	0
37 (A)	0	0	29	+ 1	1	0

38	+ 1	1	30	+ 2	4	2
40	+ 3	9	31	+ 3	9	9
42	+ 5	25	32	+ 4	16	20
44	+ 7	49	33	+ 5	25	35
	Σdx = –1	**Σdx² = 171**		**Σdy = 5**	**Σdy² = 101**	**Σdxdy = 126**

$$r = \frac{N\Sigma dxdy - \Sigma dx \times \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2} \times \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

Here, $\Sigma dxdy = 126$; $\Sigma dx = -1$; $\Sigma dy = 5$; $N = 9$; $\Sigma dx^2 = 171$; $\Sigma dy^2 = 101$

$$= \frac{9 \times 126 - (-1)(5)}{\sqrt{9 \times 171 - (-1)^2} \times \sqrt{9 \times 101 - (5)^2}}$$

$$= \frac{1134 + 5}{\sqrt{1538} \times \sqrt{884}} = \frac{1139}{1166.015} = 0.976$$

Ans. Coefficient of Correlation = 0.976. It shows very high degree of positive correlation between the age of husband and age of wife.

11.10 ASSUMPTIONS OF COEFFICIENT OF CORRELATION

Karl Pearson's coefficient of correlation is based on the following assumptions:

1. Linear Relationship: In this method, it is assumed that there is a linear relationship between the variables, i.e., if the paired observations of both the variables are plotted on a scatter diagram, the plotted points will form a straight line.
2. Causal Relationship: There is no cause and effect relationship between the two variables under study. However, there exists a cause and effect relationship between the forces affecting the two variables. Correlation is meaningless if there is no such relationship.
3. Normal Distribution: The two variables under study are affected by a large number of independent causes of such a nature as to produce normal distribution. Variables such as height, weight, colour of skin, commodity prices, demand, supply, etc., are affected by a multiplicity of forces.
4. Error of Measurement: The coefficient of correlation is more reliable if the error of measurement is reduced to the minimum.

11.11 PROPERTIES OF COEFFICIENT OF CORRELATION

1. Coefficient of correlation lies between – 1 and + 1. This property of r will also serve as a useful check on the correctness of our calculations. If the calculated value of r lies outside these limits, then there is an error in the calculations.
2. The coefficient of correlation is independent of the change of origin and scale of measurements. We will observe that changes of origin and scale of both the variables will have no effect on correlation coefficient *(refer Examples 12 and 13)*.

3. The coefficient of correlation (r) is a measure of the linear relationship; r is positive when both the variables increase or decrease together; r is negative when values of one variable increase as values of other decrease and vice-versa.
4. If two variables X and Y are independent, coefficient of correlation between them will be zero.

11.12 MERITS AND DEMERITS OF COEFFICIENT OF CORRELATION

Merits

The merits of Karl Pearson's method are:

1. Popular Method: It is the most popular and most widely used mathematical method of studying correlation between two variables.
2. Degree and direction of correlation: The correlation coefficient summarises in one figure not only the degree of correlation but also the direction, i.e., whether correlation is positive or negative.

Demerits

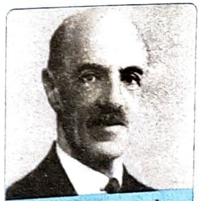
The main limitations of Karl Pearson's method are:

1. Affected by extreme values: The values of correlation are unduly affected by the value of extreme items.
2. Time Consuming Method: As compared to other methods, it is more time consuming.
3. Assumption of linear relationship: The correlation coefficient always assume linear relationship regardless of the fact whether that assumption is correct or not.
4. Possibility of wrong interpretation: One has to be very careful in interpreting the value of the coefficient of correlation as very often the coefficient is misinterpreted.

Coefficient of correlation by Karl Pearson's method can be ascertained only when quantitative measurements of various items of a series are available. However, in many cases, the direct measurement of phenomenon under study is not possible. For example, qualitative variables like efficiency, honesty, intelligence, ability, bravery, beauty, etc., cannot be measured in quantitative terms. ***If we want to study the correlation between two such qualitative variables, say intelligence and beauty, then Spearman's Rank Correlation is used.***

11.13 SPEARMAN'S RANK CORRELATION

The measure of rank correlation was developed by British Psychologist *'Charles Edward Spearman' in 1904. In this method, various items are* ***assigned ranks according to their characteristics and a correlation is computed between these ranks.***



Charles Edward Spearman (1863-1945)

Rank correlation method permits us to correlate two sets of qualitative observations which are subject to ranking. **The Spearman's Rank Correlation formula for computing correlation is:**

$$r_k = 1 - \frac{6\Sigma D^2}{N^3 - N}$$

Where: r_k = Coefficient of rank correlation; ΣD^2 = Sum of square of Rank Differences; N = Number of pairs of observations;

The value of the coefficient r_k is interpreted in the same way as Karl Pearson's coefficient of correlation. Its value ranges between + 1 and – 1. Rank correlation is equal to product moment correlation between the ranks.

11.14 COMPUTATION OF RANK CORRELATION

There are three types of problems in Rank method:

1. When Ranks are given.
2. When Ranks are not given.
3. When Ranks are equal or repeated.

When Ranks are Given

When order of ranks is given, the correlation coefficient can be determined by the following steps:

Steps in Calculation

Step 1. Represent the ranks of first variable by R_1 and second variable by R_2.

Step 2. Take the difference of the two ranks (i.e., $R_1 - R_2$) and denote these differences by D.

Step 3. Compute the squares of these differences and total them to get ΣD^2.

Step 4. Apply the following formula: $r_k = 1 - \dfrac{6\Sigma D^2}{N^3 - N}$

For Better Understanding

Perfect Positive Rank Correlation (r_k = +1)

r_k = +1, when there is complete agreement in the order of ranks and the direction of ranks is also same.

Calculation of Rank Correlation

Candidate	*Rank by 1st Judge (R_1)*	*Rank by 2nd Judge (R_2)*	*($R_1 - R_2$) (D)*	*D^2*
A	1	1	0	0
B	2	2	0	0
C	3	3	0	0
D	4	4	0	0
N = 4				**ΣD^2 = 0**

Rank Coefficient Correlation (r_k) $= 1 - \dfrac{6\Sigma D^2}{N^3 - N}$

Here, $\Sigma D^2 = 0$; N = 4

$$r_k = 1 - \frac{6(0)}{(4)^3 - 4} = 1 - \frac{0}{60} = +1$$

It shows a perfect positive correlation between 1st Judge and 2nd Judge.

Perfect Negative Rank Correlation (r_k = –1)

r_k = –1, when there is complete disagreement in order of ranks and they are in opposite direction.

Calculation of Rank Correlation

Candidate	*Rank by 1st Judge (R_1)*	*Rank by 2nd Judge (R_2)*	*($R_1 - R_2$) (D)*	*D^2*
A	1	4	–3	9
B	2	3	–1	1
C	3	2	1	1
D	4	1	3	9
N = 4				**ΣD^2 = 20**

Rank Coefficient Correlation (r_k) $= 1 - \dfrac{6\Sigma D^2}{N^3 - N}$

Here, $\Sigma D^2 = 20$; N = 4

$$r_k = 1 - \frac{6(20)}{(4)^3 - 4} = 1 - \frac{120}{60} = -1$$

It shows a perfect negative correlation between 1st Judge and 2nd Judge.

Examples 16, 17 and 18 will illustrate the computation of rank correlation coefficient.

Example 16. In a singing competition, two judges rank the 7 contestants as follows:

Judge 1	5	4	7	3	1	2	6
Judge 2	6	5	2	1	3	4	7

Calculate coefficient of rank correlation.

Solution:

Calculation of Rank Correlation

Rank by Judge 1 (R_1)	*Rank by Judge 2 (R_2)*	*($R_1 - R_2$) (D)*	*D^2*
5	6	– 1	1
4	5	– 1	1
7	2	+ 5	25
3	1	+ 2	4
1	3	– 2	4
2	4	– 2	4
6	7	– 1	1
N = 7			**ΣD^2 = 40**

Rank Coefficient Correlation (r_k) $= 1 - \dfrac{6\Sigma D^2}{N^3 - N}$

$\Sigma D^2 = 40$; N = 7

$$r_k = 1 - \frac{6(40)}{(7)^3 - 7} = 1 - \frac{240}{336} = 0.285$$

Ans. Rank Coefficient of correlation = 0.285. There is low degree of positive correlation.

Example 17. Calculate Spearman's Rank correlation of coefficient from the ranks given below:

X	2	1	4	3	5	7	6
Y	1	3	2	4	5	6	7

Solution:

Calculation of Rank Correlation

X (R_1)	*Y (R_2)*	*($R_1 - R_2$) (D)*	*D^2*
2	1	+ 1	1
1	3	– 2	4
4	2	+ 2	4
3	4	– 1	1
5	5	0	0
7	6	+ 1	1
6	7	– 1	1
N = 7			**ΣD^2 = 12**

Rank Coefficient of Correlation (r_k) $= 1 - \dfrac{6\Sigma D^2}{N^3 - N}$

$\Sigma D^2 = 12; N = 7$

$$r_k = 1 - \frac{6(12)}{(7)^3 - 7} = 1 - \frac{72}{336} = 0.786$$

Ans. Rank Coefficient of correlation = 0.786. There is a fairly high degree of positive correlation.

Example 18. Ten competitors in the singing competition are ranked by three judges in the following order:

Judge 1	1	5	4	8	9	6	10	7	3	2
Judge 2	4	8	7	6	5	9	10	3	2	1
Judge 3	6	7	8	1	5	10	9	2	3	4

Use rank correlation coefficient to discuss which pair of judges have the nearest approach to common tastes in singing competition.

Solution:

In order to determine which pair of judges has the nearest approach to common taste in beauty, we shall have to calculate the rank coefficient of correlation between the rankings of:

(i) First and Second Judge

(ii) Second and Third Judge

(iii) First and Third Judge

(i) First and Second Judge

Rank Correlation between First and Second Judge

Rank by Judge 1 (R_1)	*Rank by Judge 2 (R_2)*	*($R_1 - R_2$) (D)*	*D^2*
1	4	– 3	9
5	8	– 3	9
4	7	– 3	9
8	6	+ 2	4
9	5	+ 4	16
6	9	– 3	9
10	10	0	0
7	3	+ 4	16
3	2	+ 1	1
2	1	+ 1	1
N = 10			**$\Sigma D^2 = 74$**

Rank Coefficient of Correlation (r_k) $= 1 - \dfrac{6\Sigma D^2}{N^3 - N}$

$\Sigma D^2 = 74; N = 10$

$$r_k = 1 - \frac{6(74)}{(10)^3 - 10} = 1 - \frac{444}{990} = 0.552$$

(ii) Second and Third Judge

Rank Correlation between Second and Third Judge

Rank by Judge 2 (R_2)	*Rank by Judge 3 (R_3)*	*($R_2 - R_3$) (D)*	*D^2*
4	6	– 2	4
8	7	+ 1	1
7	8	– 1	1
6	1	+ 5	25
5	5	0	0
9	10	– 1	1
10	9	+ 1	1
3	2	+ 1	1
2	3	– 1	1
1	4	– 3	9
N = 10			**$\Sigma D^2 = 44$**

Rank Coefficient of Correlation (r_k) $= 1 - \dfrac{6\Sigma D^2}{N^3 - N}$

$\Sigma D^2 = 44; N = 10$

$$r_k = 1 - \frac{6(44)}{(10)^3 - 10} = 1 - \frac{264}{990} = 0.733$$

(iii) First and Third judge

Rank Correlation between First and Third Judge

Rank by Judge 1 (R_1)	*Rank by Judge 3 (R_3)*	*($R_1 - R_3$) (D)*	*D^2*
1	6	– 5	25
5	7	– 2	4
4	8	– 4	16
8	1	+ 7	49
9	5	+ 4	16
6	10	– 4	16
10	9	+ 1	1
7	2	+ 5	25
3	3	0	0
2	4	– 2	4
N = 10			**$\Sigma D^2 = 156$**

Rank Coefficient of Correlation $(r_k) = 1 - \dfrac{6\Sigma D^2}{N^3 - N}$

$\Sigma D^2 = 156;\ N = 10$

$$r_k = 1 - \frac{6(156)}{(10)^3 - 10} = 1 - \frac{936}{990} = 0.0545$$

Ans. The second judge and third judge have the nearest approach in common tastes in singing competition, because the rank coefficient of correlation is highest (0.733) between them.

When Ranks are not given

Spearman's method of computing coefficient of rank correlation is used to find correlation of qualitative variables. However, this method can be used for calculating rank correlation in case of quantitative data as well.

When we are given the actual data and not the ranks, it is necessary to assign the ranks. ***Once the actual data is converted into ranks, we use the same formula noted above for computing the coefficient of rank correlation.***

Steps for Calculating Rank Coefficient of Correlation (When Ranks not Given)

Step 1. Assign 1st rank to the highest (lowest) value, 2nd rank to the next highest (lowest) item and so on.

Note: *Ranks can be assigned by taking either the highest value as 1 or the lowest value as 1. But, whether we start with the lowest value or the highest value, the same method must be followed in case of both the variables.*

Step 2. Calculate difference of ranks $(R_1 - R_2)$;

Step 3. Denote these difference by D;

Step 4. Squares these differences to get ΣD^2;

Step 5. Apply the following formula: $r_k = 1 - \dfrac{6\Sigma D^2}{N^3 - N}$

The calculation of rank correlation in absence of ranks will be clear with the help of Examples 19 and 20.

Example 19. Calculate rank correlation of coefficient between marks assigned to ten students by judge A and B in a competitive test as shown below:

Marks by Judge A	52	53	42	60	45	41	37	38	25	27
Marks by Judge B	65	68	43	38	77	48	35	30	25	50

Solution:

Since we are given actual marks and not the ranks, it is necessary to assign ranks to different values. Assigning rank from the highest to the lowest (descending order), we get the following table:

R_1 = Marks by Judge A; R_2 = Marks by Judge B

Calculation of Rank Correlation

Marks by Judge A	*Marks by Judge B*	*R_1*	*R_2*	*$(R_1 - R_2)$ (D)*	*D^2*
52	65	3	3	0	0
53	68	2	2	0	0
42	43	5	6	– 1	1
60	38	1	7	– 6	36
45	77	4	1	+ 3	9
41	48	6	5	+ 1	1
37	35	8	8	0	0
38	30	7	9	– 2	4
25	25	10	10	0	0
27	50	9	4	+ 5	25
N = 10					$\Sigma D^2 = 76$

Rank Coefficient of Correlation $(r_k) = 1 - \dfrac{6\Sigma D^2}{N^3 - N}$

$\Sigma D^2 = 76;\ N = 10$

$$r_k = 1 - \frac{6(76)}{(10)^3 - 10} = 1 - \frac{456}{990} = 0.539$$

Ans. Rank Coefficient of correlation = 0.539. There is moderate degree of positive correlation.

Example 20. The following table gives the marks obtained by 10 students in Economics and Accounts. Calculate the rank correlation.

Marks in Economics	35	90	70	40	95	45	60	85	80	50
Marks in Accounts	45	70	65	30	90	40	50	75	85	60

Solution:

We are given actual marks and not the ranks. So, assigning ranks from the lowest to the highest, we get:

Calculation of Rank Correlation

Marks in Economics	Marks in Accounts	*Ranks (R_1) in Economics*	*Ranks (R_2) in Accounts*	*$(R_1 - R_2)$ (D)*	*D^2*
35	45	1	3	– 2	4
90	70	9	7	+ 2	4
70	65	6	6	0	0
40	30	2	1	+ 1	1
95	90	10	10	0	0
45	40	3	2	+ 1	1
60	50	5	4	+ 1	1
85	75	8	8	0	0
80	85	7	9	– 2	4
50	60	4	5	– 1	1
N = 10					$\Sigma D^2 = 16$

Rank Coefficient of Correlation (rk) $= 1 - \dfrac{6\Sigma D^2}{N_3 - N}$

$\Sigma D^2 = 16; N = 10$

$$r_k = 1 - \frac{6(16)}{(10)^3 - 10} = 1 - \frac{96}{990} = 0.903$$

Ans. Rank Coefficient of correlation = 0.903. There is very high degree of positive correlation.

When Ranks are Equal or Repeated

When two or more than two items of a group have same value, then they are assigned a common rank. **This common rank is the average of the ranks which these items would have got had they differed slightly from each other.**

For example:

- If *two* individuals are ranked equal at 3rd place, then each of them are given $\frac{3+4}{2} = 3.5^{th}$ Rank.
- If *three* individuals are ranked equal at 3rd place, then each of them are given $\frac{3+4+5}{3} = 4^{th}$ Rank.

Need for Correction in Formula

The Spearman's formula is based upon the assumption of different ranks to different individuals. ***So, in case of equal or tied ranks, the correction in formula becomes necessary.*** In the formula, we add the correction factor $\frac{1}{12}(m^3 - m)$ to ΣD^2, where 'm' is equal to the number of times an item is assigned equal rank. The correction factor is to be added for each and every repeated value. The formula in this case becomes:

$$r_k = 1 - \frac{6\left(\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \text{.........}\right)}{N^3 - N}$$

The case of equal ranks is explained in Examples 21 and 22.

Example 21. Calculate the coefficient of correlation of the following data by the Spearman's Rank Correlation method:

X	19	24	12	23	19	16
Y	9	22	20	14	22	18

Solution:

Calculation of Rank Correlation

X	Y	R_1	R_2	$(R_1 - R_2)$ (D)	D^2
19	9	3.5	1	+ 2.5	6.25
24	22	6	5.5	+ 0.5	0.25
12	20	1	4	− 3	9
23	14	5	2	+ 3	9
19	22	3.5	5.5	− 2	4
16	18	2	3	− 1	1
N = 6					$\Sigma D^2 = 29.50$

Here, number 19 is repeated twice in series X and number 22 is repeated twice in series Y. Therefore, in X, m = 2 and in Y, m = 2.

$$r_k = 1 - \frac{6\left(\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \text{.........}\right)}{N^3 - N}$$

$\Sigma D^2 = 29.50; N = 6$

$$= 1 - \frac{6\left(29.50 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)\right)}{6^3 - 6}$$

$$= 1 - \frac{6(29.50 + 0.5 + 0.5)}{216 - 6} = 1 - \frac{6 \times 30.5}{210} = 0.128$$

Ans. Rank Coefficient of correlation = 0.128. There is very low degree of positive correlation.

Example 22. Calculate coefficient rank correlation of the following data:

X	75	73	72	72	63	62	55	50
Y	10	11	13	13	13	20	16	28

Solution:

Calculation of Rank Correlation

X	Y	R_1	R_2	$D = (R_1 - R_2)$	D^2
75	10	1	8	− 7	49
73	11	2	7	− 5	25
72	13	3.5	5	− 1.5	2.25
72	13	3.5	5	− 1.5	2.25
63	13	5	5	0	0
62	20	6	2	+ 4	16
55	16	7	3	+ 4	16
50	28	8	1	+ 7	49
N = 8					$\Sigma D^2 = 159.50$

Here, number 72 is repeated twice in series X and number 13 is repeated thrice in series Y. Therefore, in X, m = 2 and in Y, m = 3

$$r_k = 1 - \frac{6\left(\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \text{........}\right)}{N^3 - N}$$

$\Sigma D^2 = 159.50;\ N = 8$

$$= 1 - \frac{6\left(159.50 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(3^3 - 3)\right)}{8^3 - 8}$$

$$= 1 - \frac{6(159.50 + 0.5 + 2)}{512 - 8} = 1 - \frac{972}{504} = 1 - 1.9285 = -0.9285$$

Ans. Rank Coefficient of correlation = – 0.9285. There is very high degree of negative correlation.

Incorrect Values

Example 23. The coefficient of rank correlation of the marks obtained by 10 students in Accounts and Maths 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 coefficient of rank correlation.

Solution:

We are given: $n = 10,\ r_k = 0.8$

$$\text{Rank Coefficient of Correlation } (r_k) = 1 - \frac{6\Sigma D^2}{N^3 - N}$$

$$0.8 = 1 - \frac{6\Sigma D^2}{(10)^3 - 10}$$

$$\frac{6\Sigma D^2}{990} = 1 - 0.8$$

$$\Sigma D^2 = \frac{990 \times 0.2}{6} = 33$$

Since one difference is wrongly taken as 7 instead of 9, the corrected value of ΣD² is given

Corrected $\Sigma D^2 = 33 - 7^2 + 9^2$

$= 33 - 49 + 81$

$= 65$

$$\text{Corrected } (r_k) = 1 - \frac{6 \times 65}{990} = 1 - \frac{390}{990} = 1 - 0.394 = 0.606$$

Ans. Correct coefficient of rank correlation = 0.606

Example 24. A person while calculating coefficient of correlation between two variables X and Y, obtained the following results.

$N = 8,\ \Sigma X = 120,\ \Sigma X^2 = 600,\ \Sigma Y = 90;\ \Sigma Y^2 = 250;$ and $\Sigma XY = 356$

However, at the time of checking the calculations, he discovered that two pairs of observations (8, 10) and (12, 7) were wrongly entered instead of (8, 12) and (10, 8). Determine the correct value of coefficient of correlation.

Solution:

Corrected ΣX $= 120 - 8 - 12 + 8 + 10 = 118$

Corrected ΣX² $= 600 - 8^2 - 12^2 + 8^2 + 10^2 = 556$

Corrected ΣY $= 90 - 10 - 7 + 12 + 8 = 93$

Corrected ΣY² $= 250 - 10^2 - 7^2 + 12^2 + 8^2 = 309$

Corrected ΣXY $= 356 - (8 \times 10) - (12 \times 7) + (8 \times 12) + (10 \times 8) = 368$

$$\text{Coefficient of Correlation (r)} = \frac{N\Sigma XY - \Sigma X.\Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2} \times \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

$$= \frac{30 \times 368 - 118 \times 93}{\sqrt{30 \times 556 - (118)^2} \times \sqrt{30 \times 309 - (93)^2}}$$

$$= \frac{11{,}040 - 10{,}974}{\sqrt{16{,}680 - 13{,}924} \times \sqrt{9{,}270 - 8{,}649}}$$

$$= \frac{66}{\sqrt{2{,}756} \times \sqrt{621}} = \frac{66}{52.50 \times 24.92} = \frac{66}{1{,}308.3} = 0.05$$

Ans. Coefficient of Correlation = 0.05. There is very low degree of positive correlation.

~~11.15~~ MERITS AND DEMERITS OF RANK CORRELATION

Merits

1. This method is easy to calculate and simple to understand as compared to Karl Pearson's method. It takes less time in computation.
2. Rank method is very useful when the data is qualitative in nature like honesty, beauty, intelligence, voice quality, etc. In such cases, ranks are assigned to different items under consideration.
3. This is the only method that can be used where we are given the ranks but not the actual data.
4. When actual values are given (instead of ranks), then this method can be used to get rough idea about the degree of correlation.

Demerits

1. Rank method cannot be used for finding out correlation in a bivariate (grouped) frequency distribution.

2. If the number of values is quite large, it becomes a difficult task to ascertain the ranks and their differences. That is why it is advisable to use this method only when the number of observations is less than 30.

3. This method lacks precision as compared to Karl Pearson's method. It uses ranks instead of the original values.

Karl Pearson's Method Vs Spearman's Rank Method

The coefficient of correlation by both the methods ranges between −1 and +1. Still, there exist the following differences:

1. Karl Pearson's method of Correlation measures correlation for *quantitative data*, whereas Spearman's method of rank correlation measures coefficient of correlation for *qualitative data*.
2. Karl Pearson's method calculates *deviations from actual or assumed mean*, whereas Spearman's method calculates the *rank differences*.
3. Rank correlation gives *less importance to the extreme values* because it gives them rank. However, Karl Pearson's method of correlation gives *more importance* to *extreme values* as it is based on actual values.

FORMULAE AT A GLANCE

Karl Pearson's Coefficient of Correlation

1. ***Actual Mean Method***	$r = \dfrac{\Sigma xy}{N \times \sigma_x \times \sigma_y} = \dfrac{\Sigma xy}{N \times \sqrt{\dfrac{\Sigma x^2}{N}} \times \sqrt{\dfrac{\Sigma y^2}{N}}} = \dfrac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$ ✓
2. ***Direct Method***	$r = \dfrac{N\Sigma XY - \Sigma X.\Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2} \times \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$
3. ***Short – Cut Method***	$r = \dfrac{N\Sigma dxdy - \Sigma dx \times \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2} \times \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$
4. ***Step Deviation Method***	$r = \dfrac{N\Sigma dx'dy' - \Sigma dx' \times \Sigma dy'}{\sqrt{N\Sigma dx'^2 - (\Sigma dx')^2} \times \sqrt{N\Sigma dy'^2 - (\Sigma dy')^2}}$

Spearman's Rank Correlation Coefficient

When Ranks are not Equal	$r_k = 1 - \dfrac{6\Sigma D^2}{N^3 - N}$ ✓
When Ranks are Equal	$r_k = 1 - \dfrac{6\left(\Sigma D^2 + \dfrac{1}{12}(m^3 - m) + \dfrac{1}{12}(m^3 - m) + \ldots\ldots\right)}{N^3 - N}$

Abbreviations Used

r = Karl Pearson's Coefficient of Correlation.

N = Number of pair of observations.

x = Deviation of X series from mean $(X - \overline{X})$.

y = Deviation of Y series from mean $(Y - \overline{Y})$.

σ_x = Standard deviation of X series, i.e., $\sqrt{\dfrac{\Sigma x^2}{N}}$.

σ_y = Standard deviation of Y series, i.e., $\sqrt{\dfrac{\Sigma y^2}{N}}$.

Σdx = Sum of deviations of X values from assumed mean.

Σdy = Sum of deviations of Y values from assumed mean.

Σdx^2 = Sum of squared deviations of X values from assumed mean.

Σdy^2 = Sum of squared deviations of Y values from assumed mean.

$\Sigma dxdy$ = Sum of the products of deviations dx and dy.

$\Sigma dx'$ = Sum of step deviations of X values from assumed mean .

$\Sigma dy'$ = Sum of step deviations of Y values from assumed mean.

$\Sigma dx'^2$ = Sum of squared step deviations of X values from assumed mean.

$\Sigma dy'^2$ = Sum of squared step deviations of Y values from assumed mean.

$\Sigma dx'dy'$ = Sum of the products of step deviations dx′ and dy′.

r_k = Coefficient of rank correlation.

ΣD^2 = Sum of square of rank differences.

m = Number of times an item is assigned equal rank.

SUMMARY OF KARL PEARSON'S COEFFICIENT OF CORRELATION

Example: Calculate the Coefficient of Correlation (r) from the following data by different methods.

X	2	4	6	8	10	12
Y	6	12	18	24	30	36

1st Method: Actual Mean Method

X-Series			Y-Series			
X	$x = X-\bar{X}$	x^2	Y	$y = Y-\bar{Y}$	y^2	xy
2	−5	25	6	−15	225	75
4	−3	9	12	−9	81	27
6	−1	1	18	−3	9	3
8	1	1	24	3	9	3
10	3	9	30	9	81	27
12	5	25	36	15	225	75
$\Sigma X = 42$		$\Sigma x^2 = 70$	$\Sigma Y = 126$		$\Sigma y^2 = 630$	$\Sigma xy = 210$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{42}{6} = 7 \qquad \bar{Y} = \frac{\Sigma Y}{N} = \frac{126}{6} = 21$$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}} = \frac{210}{\sqrt{70 \times 630}} = \frac{210}{210} = 1$$

2nd Method: Direct Method

X-Series		Y-Series		
X	X^2	Y	Y^2	XY
2	4	6	36	12
4	16	12	144	48
6	36	18	324	108
8	64	24	576	192
10	100	30	900	300
12	144	36	1296	432
$\Sigma X = 42$	$\Sigma X^2 = 364$	$\Sigma Y = 126$	$\Sigma Y^2 = 3,276$	$\Sigma XY = 1,092$

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

$$= \frac{6 \times 1092 - 42 \times 126}{\sqrt{6 \times 364 - (42)^2} \times \sqrt{6 \times 3,276 - (126)^2}}$$

$$= \frac{6,552 - 5292}{\sqrt{2,184 - 1,764} \times \sqrt{19,656 - 15,876}}$$

$$= \frac{1,260}{\sqrt{420 \times 3,780}} = \frac{1,260}{1,260} = 1$$

3rd Method: Short-Cut Method or Assumed Mean Method

X-Series			Y-Series			
X	dx = X − A A = 8	dx^2	Y	dy = Y − A A = 24	dy^2	dxdy
2	−6	36	6	−18	324	108
4	−4	16	12	−12	144	48
6	−2	4	18	−6	36	12
8	0	0	24	0	0	0
10	2	4	30	6	36	12
12	4	16	36	12	144	48
	$\Sigma dx = -6$	$\Sigma dx^2 = 76$		$\Sigma dy = -18$	$\Sigma dy^2 = 684$	$\Sigma dxdy = 228$

$$r = \frac{N\Sigma dxdy - \Sigma dx \times \Sigma dy}{\sqrt{N\Sigma dx^2 \times (\Sigma dx)^2} \times \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

$$= \frac{\{6 \times 228\} - \{-6 \times -18\}}{\sqrt{6 \times 76 - (-6)^2} \times \sqrt{6 \times 684 - (-18)^2}}$$

$$= \frac{1,368 - 108}{\sqrt{456 - 36} \times \sqrt{4,104 - 324}}$$

$$= \frac{1,260}{\sqrt{420 \times 3,780}} = \frac{1,260}{1,260} = 1$$

4th Method: Step Deviation Method

X-Series				Y-Series				
X	dx = X − A A = 8	$dx' = \frac{dx}{C}$ C = 2	dx'^2	Y	dy = Y − A A = 24	$dy' = \frac{dy}{C}$ C = 6	dy'^2	dx′dy′
2	−6	−3	9	6	−18	−3	9	9
4	−4	−2	4	12	−12	−2	4	4
6	−2	−1	1	18	−6	−1	1	1
8	0	0	0	24	0	0	0	0
10	2	1	1	30	6	1	1	1
12	4	2	4	36	12	2	4	4
		$\Sigma dx' = -3$	$\Sigma dx'^2 = 19$			$\Sigma dy' = -3$	$\Sigma dy'^2 = 19$	$\Sigma dx'dy' = 19$

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

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

$$= \frac{114 - 9}{\sqrt{114 - 9} \times \sqrt{114 - 9}}$$

$$= \frac{105}{\sqrt{105 \times 105}} = \frac{105}{105} = 1$$

SUMMARY OF SPEARMAN'S RANK CORRELATION

1st Case: When Ranks are Given

Example 1. In a competition, two judges rank the 5 contestants as follows:

Judge 1	1	2	3	4	5
Judge 2	4	2	1	3	5

Calculate coefficient of rank correlation.

Solution:

Ranks by Judge 1 (R_1)	Ranks by Judge 1 (R_2)	$D = R_1 - R_2$	D^2
1	4	−3	9
2	2	0	0
3	1	2	4
4	3	1	1
5	5	0	0
			$\Sigma D^2 = 14$

$$\text{Rank Correlation } (r_k) = 1 - \frac{6\Sigma D^2}{N^3 - N}$$

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

2nd Case: When Ranks are NOT Given

Example 2. Calculate Spearman's Rank correlation of coefficient from the following data:

X	87	22	33	75	37
Y	29	63	52	46	48

Solution: *It is necessary to assign ranks. Assigning rank from the highest to the lowest.*

X	Ranks (R_1)	Y	Ranks (R_2)	$D = R_1 - R_2$	D^2
87	1	29	5	−4	16
22	5	63	1	4	16
33	4	52	2	2	4
75	2	46	4	−2	4
37	3	48	3	0	0
					$\Sigma D^2 = 40$

$$\text{Rank Correlation } (r_k) = 1 - \frac{6\Sigma D^2}{N^3 - N}$$

$$r_k = 1 - \frac{6 \times 40}{(5)^3 - 5} = 1 - \frac{240}{120} = -1$$

3rd Case: When Ranks are EQUAL or REPEATED

Example 3. Calculate Spearman's Rank correlation of coefficient from the following data:

X	90	88	75	75	74	70	65	62
Y	18	25	34	34	34	42	38	47

Solution: *It is necessary to assign ranks. Assigning rank from the highest to the lowest.*

X	Ranks (R_1)	Y	Ranks (R_2)	$D = R_1 - R_2$	D^2
90	1	18	8	−7	49
88	2	25	7	−5	25
75	3.5	34	5	−1.5	2.25
75	3.5	34	5	−1.5	2.25
74	5	34	5	0	0
70	6	42	2	4	16
65	7	38	3	4	16
62	8	47	1	7	49
					$\Sigma D^2 = 159.5$

Number 75 is repeated twice in series X and 34 is repeated thrice in series Y. Therefore, in X, m = 2 and in Y, m = 3.

$$r_k = 1 - \frac{6\left(\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)\right)}{N^3 - N}$$

$$= 1 - \frac{6\left(159.5 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(3^3 - 3)\right)}{8^3 - 8}$$

$$= 1 - \frac{6(159.5 + 0.5 + 2)}{512 - 8} = 1 - \frac{6 \times 162}{504} = -0.93$$