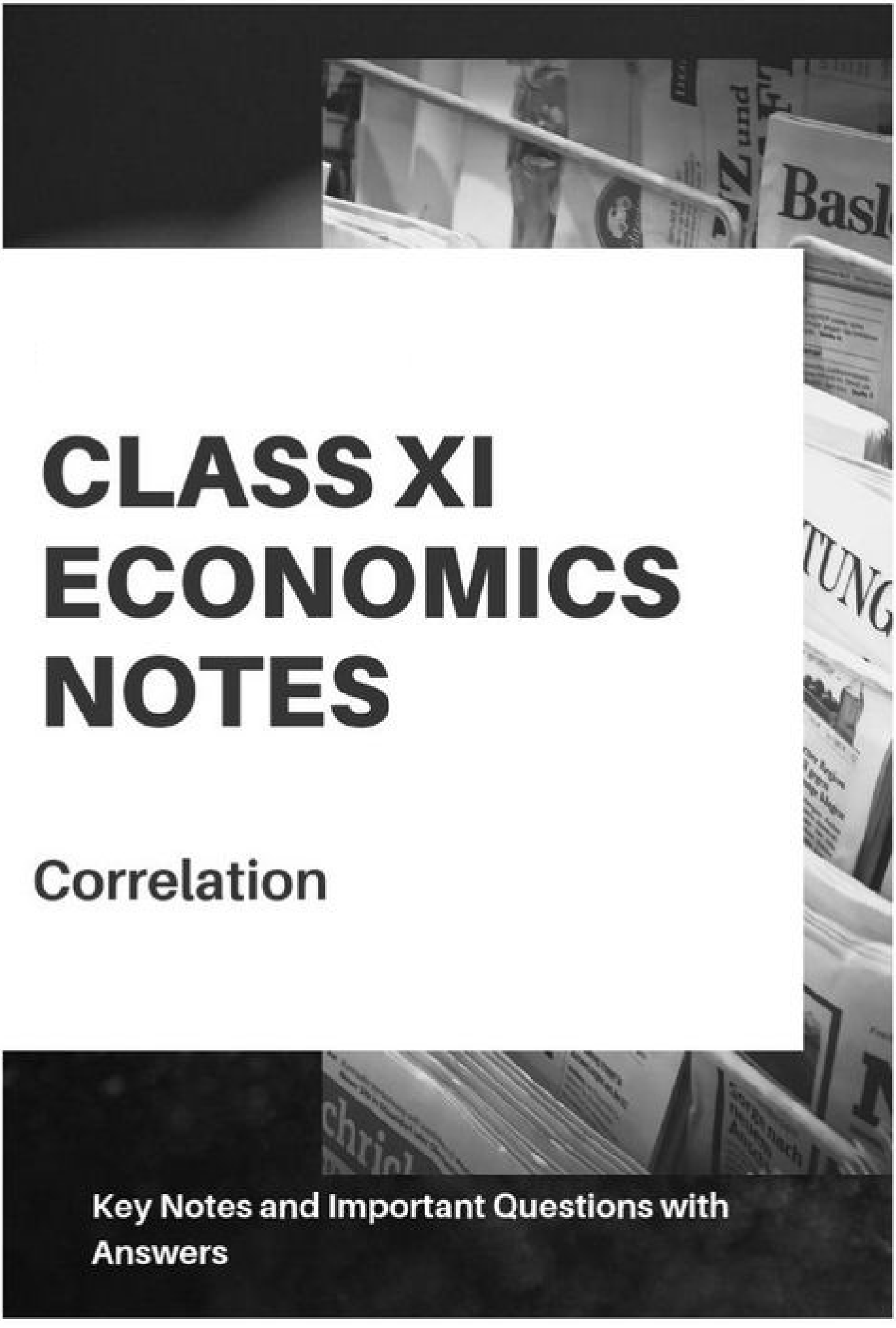


CLASS XI ECONOMICS NOTES

Correlation

Key Notes and Important Questions with Answers

(iii) CORRELATION

Correlation is a statistical device or tool which measures the quantitative relationship between two variables. It measures the direction and intensity of relationship among variables. Thus, correlation measures co-variation, not causation.

- **Types of Correlation**

1) **Positive and negative correlation :**

When both the variables (say X and Y) move in the same direction then it is called positive correlation. For eg. increase in temperature and increase in sale of AC.

When both the variable (say X and Y) move in the opposite direction then it is called negative correlation for eg. increase in temperature and decrease in sale of Heater.

2) **Linear and Non-linear correlation :**

When both the variables (say X and Y) change in the same proportion, it is called linear correlation. When both the variables (say X and Y) change in the different proportions, it is called non-linear correlation.

3) **Simple, Partial and Multiple Correlation :**

When we study the correlation between two variables then it is called simple correlation. In this correlation. There are two variables one is independent and another is dependent.

When we study the correlation between more than two variables is called multiple correlation. If we study correlation between two variables keeping the content of all other variables then it is called partial correlation.

DEGREE OR MAGNITUDE OF CORRELATION

Degree	Positive	Negative
Perfect	+1	–1
Higher	(+ 0.75) – (+ 1)	(– 0.75) – (– 1)
Medium	(+ 0.25) – (+ 0.75)	(– 0.25) – (– 0.75)
Low	0 – (+ 0.25)	(0) – (– 0.25)
Zero (Absence of correlation)	0	0

- **Methods of measurement of correlation :**

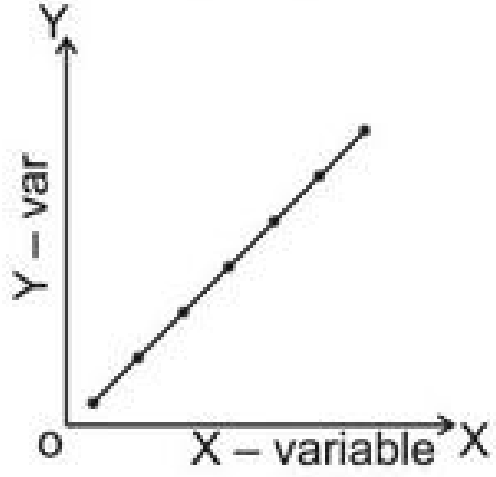
There are following methods.

1. Scattered Diagram Method
2. Karl Person's Coefficient
3. Spearman's Coefficient

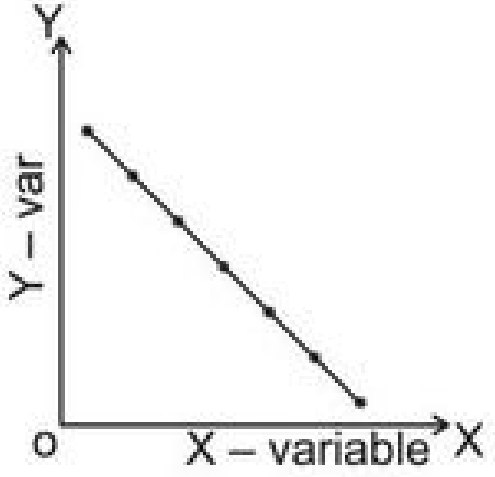
1) Scattered Diagram Method

It is a graphical method. In this method we use graph paper. We show X-variable on X-axis and Y-variable on the Y-axis. We plote the corresponding value of both the variable by dot (.) on the graph paper.

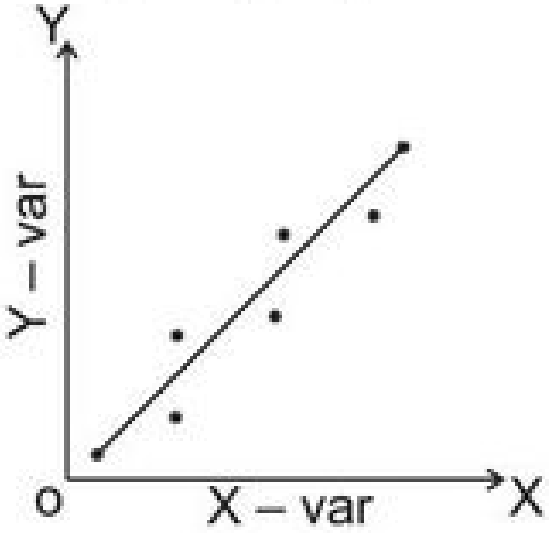
Perfect degree positive correlation



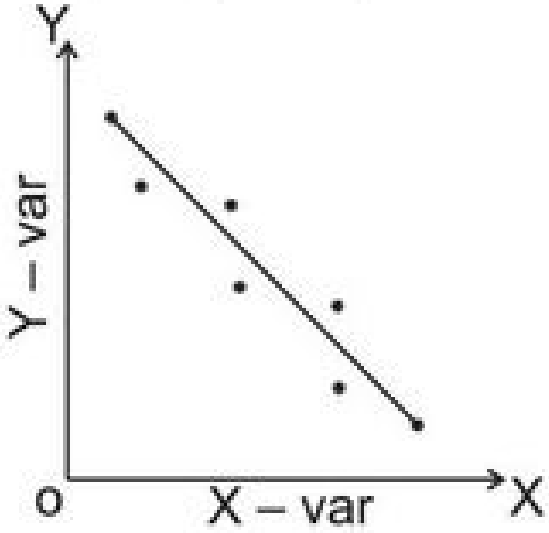
Perfect degree negative correlation



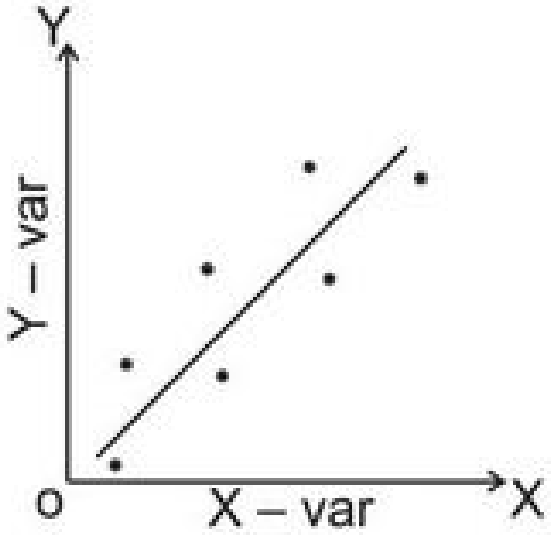
Higher degree positive correlation



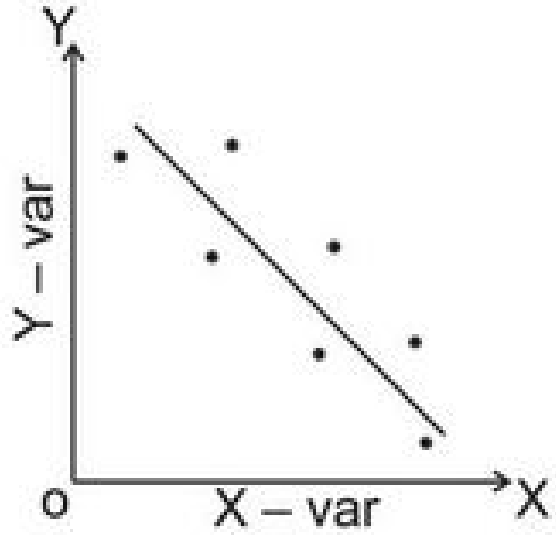
Higher degree negative correlation



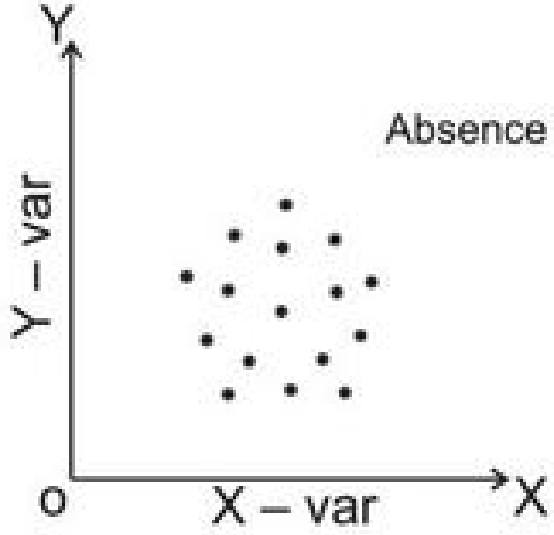
Lower degree positive correlation



Lower degree negative correlation



Absence of correlation



2. Karl Pearson's Method

It is also called product moment method of correlation coefficient. It is indicated by r. It is based on arithmetic mean and standard deviation.

Let there are two variables X and Y.

Mean of x-series is $\overline{X} = \frac{\Sigma X}{N}$ and mean of y-series is $\overline{Y} = \frac{\Sigma Y}{N}$.

Standard deviation of x-series is $\sigma x = \sqrt{\frac{\Sigma x^2}{N}}$ and standard deviation of y-series is $\sigma y = \sqrt{\frac{\Sigma y^2}{N}}$. Here $x = X - \overline{X}$ and $y = Y - \overline{Y}$

Covariance of variable x and y is

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

then we find Karl Pearson's coefficient of correlation

$$r = \frac{\text{Cov. } (X, Y)}{\sigma x \, . \, \sigma y}$$

OR

$$r = \frac{\Sigma xy}{N.\sigma x \, . \, \sigma y}$$

OR

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

OR

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2} \times \sqrt{\Sigma y^2}} = \frac{\Sigma (X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma (X-\overline{X})^2} \sqrt{\Sigma (Y-\overline{Y})^2}}$$

Therefore, the following methods are used to measure Karl Pearson's coefficient of correlation.

1) Actual Mean Method

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

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

$$\bar{X} = \frac{\sum X}{N} ; \quad \bar{Y} = \frac{\sum Y}{N}$$

N = No. of observations

2) Assumed Mean Method

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

OR

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

Where $dx = X - A$; $dy = Y - A$

A = Assumed mean from X and Y series.

3) Step-deviation Method

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

OR

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

Where $dx' = \frac{X - A}{i}$; $dy' \frac{Y - A}{i}$

If we assume that $dx^1 = U = \frac{X-A}{i}$ and

$$dy^1 = V = \frac{Y-A}{i}$$

then above formula can be written as

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

then correlation between X and Y (r_{xy}) is same as correlation between r_{uv}.

- Direct Method

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

OR

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

- **Properties of correlation**

i) It is independent from unit.

ii) Negative value of coefficient of correlation indicates negative correlation while positive value of coefficient of correlation indicates positive correlation.

iii) Coeff. of correlation lies between –1 and +1

i.e $-1 \leq r \leq +1$

iv) If $r = 0$, it means of absence of correlation.

v) If higher value of r shows higher degree linear correlation and a lower value of r shows lower degree of linear corr.

vi) If $r = +1$, it means perfect degree positive correlation between two variable and i f $r = -1$, it means perfect degree negative correlation between two variables.

vii) It is independent of change of origin and change of scale of the variables. It is proved by value of r which is calculated by step deviation method.

3. Spearman's coefficient of correlation.

It is also called rank order coefficient of correlation. It is useful for qualitative observations. When values of variables are not impressed in quantitative measures then it is used to measure correlation. For example honesty, morality, character, beautifulness, originality, leadership, quality, wisdom etc. It is better alternative to determine the ranks instead of quantification of qualitative information.

It is indicated by rk or Roh (P).

This method is useful in the following three situation :

1) When ranks are given :

If ranks are already given then

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

Where N = No. of observations

D = Deviation / Difference between ranks of two variables

2) When ranks are not given.

1. First of all rank the values of variables. Ranking can be done either in ascending order or in descending order. 1st rank to largest value 2nd rank to second largest value, 3rd rank to 3rd largest value and so on. and vice-versa.
2. Find deviation or difference between ranks of two series.
3. Use the following formula to find correlation

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

3. When values are repeated.

- When two or more than two values are equal then average rank is given.
- Next rank to next value and so on.
- Use the following formula to find correlation.

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

Where $m_1, m_2, \ldots\ldots$ indicate number of repeation of values and $\frac{1}{12}(m_1^3 - m_1)$, $\frac{1}{12}(m_2^3 - m_2)$ ..... indicate their corresponding correction coefficient.

- **Similarities between Karl Pearson's and Spearman's. Correlation.**

(i) The values of both correlation lie between ± 1.

(ii) When $r_R = -1$; it means perfectly disagree. In this case ranks are such that highest ranking X goes with the lowest ranking Y and so on, we have perfect negative correlation with coefficient of – 1.

(iii) When $r_R = +1$, it means perfectly agree. If each X and its paired Y have exactly the same rank, we have perfect positive correlation with coefficient if + 1.

- **Dissimilarities**

(i) Ranks correlation give less importance to the extreme values and it does not based on the numerical value of all the informations. So, result of this method is not accurate as compared to product moment method. It is because that product moment method gives more importance to extreme values because it is based on all actual values.

(ii) It is more useful when number of items are small, data are given as ranks, scores etc. and data are not numerically expressed than product moment method.

QUESTION BANK

Very Short Answer Type Questions (1 Mark Questions) :-

1) What is meant by correlation?
2) List some variables where accurate measurement is difficult.
3) What is negative correlation?
4) Give the meaning of positive correlation.

5) What is the range of simple correlation coefficient?

6) State the type of correlation when two variables change in the same ratio.

7) Give two examples of positive correlation.

8) Mention the principal short coming of scatter diagram as a method of estimating correlation.

9) Give two examples of negative correlation.

10) When is rank correlation method used?

11) Mention the names of different methods for measuring correlation.

12) What is the main demerit of spearman's rank method?

13) Mention the principal short coming of Karl Pearson's coefficient correlation.

14) If $r_{xy} = 0$, then the variables X and Y are :

    i) Linearly related

    ii) Not linearly related

    iii) Independent

15) The unit of correlation coefficient between height in feet and weight in kilograms is :

    i) kg / feet

    ii) percentage

    iii) non-existent

16) Which method of measuring correlation measures any type of relationship?

    a) Karl Pearson's Co-efficient of correlation.

    b) Spearman's rank correlation.

    c) Scattered Diagram.

17) If precisely measured data are available, the simple co-efficcient correlation is :

    a) more accurate than rank correlation co-efficient

    b) less accurate than rank correlation co-efficient

    c) as accurate as the rank correlation co-efficient

Short Answer Type Questions (3/4 Mark Questions) :-

1. What is meant by correlation? What are the properties of coefficient of correlations?
2. Interpret the values of r as 1, –1 and 0.
3. Calculate the correlation coefficient between X & Y and comment on their relationship.

X	–3	–2	–1	1	2	3
Y	9	4	1	1	4	9

(Ans. r = 0)

4. Calculate the correlation coefficient between X & Y and comment on their relationship :

X	1	2	3	4	5
Y	3	4	6	7	10

(Ans. r = +0.98)

5. Plot the following data as a scatter diagram and comment over the result :

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

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

X	20	25	30	35	40	45	50	55	60
Y	16	20	23	25	33	38	46	50	55

(Ans. r = +0.99)

7. From the following data, compute the product movement correlation between x and y.

		X series	**Y series**
i)	No. of items	15	15
ii)	Arithmetic mean	25	18
iii)	Square of deviations From arithmetic mean	136	138
iv)	Summation of products of deviations of X and Y series from respective means = 122		

(Ans. r = 0.89)

8. Number of pairs of observations of X and Y series = 10

X series Arithmetic average = 65

Standard deviation = 23.33

Y series Arithmetic average = 66

Standard deviation = 14.9

Summation of products of corresponding deviation of X and Y series = + 2704

Calculate product moment correlation of x and y series.

(Ans. r = + 0.78)

9. Calculate the Spearman's rank correlation from the following data X 10

X	10	12	8	15	20	25	40
Y	15	10	6	25	16	12	8

(Ans. r = + 0.14)

10. Two judges in a beauty competition rank the twelve entries as follows :

Without Make-up	1	2	3	4	5	6	7	8	9	10	11	12
With Make-up	12	9	6	10	3	5	4	7	8	2	11	1

(Ans. r = –0.45) Calculate rank correlation coefficient.

11. Calculate the rank coefficient correlation of the following data :

X	68	75	90	75	50	62	40	35
Y	10	12	14	10	10	13	9	8

(Ans. r = + 0.76)

12. Does correlation imply causation?

13. Does zero correlation mean independence ?

14. Why does rank correlation coefficient differ from Karl Pearson's coefficient of correlation?

15. When is rank correlation coefficient more precise than simple correlation coefficient?

Long Answer Type Questions (6 Mark Questions) :-

1) Discuss Karl Pearson's method of calculating coefficient of correlation. Give its merits and limitations.

2) In a beauty contest, three judges accorded following ranks to 10 participants :

Judge I	1	6	5	1	0	3	2	4	9	7	8
Judge II	3	5	8	4	7	1	0	2	1	6	9
Judge III	6	4	9	8	1	2	3	1	0	5	7

Find out by Spearman's rank difference method which pair of judges has a common taste in respect of beauty.

(Ans. rs I & II = – 0.21; rs II & III = – 0.29)

3. What are the advantages of spearman's rank correlation coefficient over Karl Pearson's correlation coefficient? Explain the method of calculating Spearman's rank correlation coefficient.

4. Following are the heights and weights of 10 students in a class. Draw a scatter diagram and indicate whether the correlation is positive or negative.

Height (in inches)	72	60	63	66	70	75	58	78	72	62
Weight (in Kg.)	65	54	55	61	60	54	50	63	65	50

5. Calculate the correlation coefficient of ten marks obtained by 12 students Mathematics and Statistics and interpret it.

Marks (in Maths)	50	54	56	59	60	62	61	65	67	71	71	74
Marks (in Statistics)	22	25	34	28	26	30	32	30	28	34	36	40

(Ans. r = 0.78)

ANSWERS OF ONE (1) MARK QUESTIONS

1. Correlation is a statistical tool which studies the relationship between two variables.
2. Beauty, bravery, wisdom, ability etc.
3. The correlation is said to be negative when the variable move in opposite direction.
4. The correlation is said to be positive when the variable move together in the same direction.
5. $-1 \leq r \leq 1$
6. Perfect correlation.
7. i) Age of husband and age of wife.
   ii) Increase in height and weight.
8. Scattered diagram does not indicate the exact numerical value of correlation.
9. i) Sale of wollen garments and day temperature.
   ii) Yield of crops and price.
10. When data are of qualitative nature like beauty, honesty etc.
11. i) Scatterred diagram
    ii) Karl Pearson's coefficient of correlation.
    iii) Spearman's Rank correlation coefficient.
12. This method can not be employed for finding out correlation in a grouped frequency distribution.
13. The value of the coefficient is affected by extreme items.
14. Independent
15. Non-existent
16. Karl Pearson's coefficient of correlation.

Frequently Asked Questions

Q1. Who gave the rank difference method of correlation? (1 Marks)

Hint : Prof. Charles Speaman

Q2. Define correlation. Give an example each of positive and negative correlation. (3 Marks)

Hints : Positive correlation :– Increase in Price and increase in supply.

Negative correlation :– Rise in price and fall in demand.

Q3. Compute the Karl Pearson's coefficient from following data : (6 Marks)

X	10	12	11	13	12	14	9	12	14	13
Y	7	9	12	9	13	8	10	2	7	13

Hints :

i) Compute the mean of both series $(\bar{x}\,\bar{y})$

ii) Take the deviation from the mean (xy)

iii) Square the deviation of (x^2y^2)

iv) Compute the product of (xy)

v) Use the following formula :

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

Ans. – 0.115 (Low degree of negative correlation)

4. Find out rank difference correlated of X and Y. (6 Marks)

X	80	78	75	75	58	67	60	59
Y	12	13	14	14	14	16	15	27

X	R1	Y	R2	D=R1–R2	D^2
80	1	12	8	–7	49
78	2	13	7	–5	25
75	3.5	14	5	–1.5	2.25
75	3.5	14	5	–1.5	2.25
58	8	14	5	3	9
67	5	16	2	3	9
60	6	15	3	3	9
59	7	17	1	6	36

$\sum D^2 = 141.5$

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

$$rk = 1 - \frac{6\left[141.5 + \frac{2^3 - 2}{12} + \frac{3^3 - 3}{12}\right]}{8^3 - 8}$$

$$= 1 - \frac{6(141.5 + 0.5 + 2)}{504}$$

$$= 1 - \frac{6 \times 144}{504}$$

$$= \frac{504 - 864}{504}$$

$$= \frac{-360}{504}$$

$$= -0.71$$