Economics

(Statistics)

Chapter 7: Correlation

Correlation

Important terms and concepts –

Correlation refers to the associations between variables. When an association exists between two variables, it means that the average value of one variable changes as there is a change in the value of the other variable.

Kinds of correlation –

- Positive and Negative correlation.
- Linear and non – linear correlation.
- Simple and multiple correlations.

Types of Correlation –

- **Positive correlation:** If two variables change in the same direction (i.e. if one increases the other also increases, or if one decreases, the other also decreases), then this is called a positive correlation. For example: Advertising and sales.

  Some other examples of series of positive correlation are:

  - Heights and weights.
  - Household income and expenditure.
  - Price and supply of commodities.
  - Amount of rainfall and yield of crops.

- **Negative correlation:** If two variables change in the opposite direction (i.e. if one increases, the other decreases and vice versa), then the correlation is called a negative correlation. For example: T.V. registrations and cinema attendance.
- **Linear Correlation:** When two variables change in a constant proportion.
- **Non- linear correlation:** When two variables do not change in the same proportion.
- **Simple correlation:** Relationship between two variables are studied.
- **Multiple Correction:** Relationship between three or more than three variables are studied.

Degrees of Correlation –

- **Perfect Correlation:** When values of both variables changes at a constant rate.

Types:

- **Perfect positive correlation:** When values of both variables changes at a constant ratio in the same direction correlation coefficient value (r) is +1.
- **Perfect negative correlation:** When values of both the variables change at a constant ratio in opposite direction. Value of coefficient of correlation is -1.

❖ **Absence of correlation:** If two series of two variables exhibit no relations between them or change in one variable does not lead to a change in the other variable, then we can firmly say that there is no correlation or absurd correlation between the two variables. In such a case the coefficient of correlation is 0. 3. Limited degrees of correlation.

❖ **Limited degree correlation:** If two variables are not perfectly correlated or there is a perfect absence of correlation, then we term the correlation as Limited correlation.

Types:

- High: r his between ± 0.7 & 0.999
- Moderate = r lies between ± 0.5 and + 0.699
- Low: r < ± 0.5

Different methods of finding correlation –

1. Karl Pearson's coefficient method.
2. Rank method/ Spearman's coefficient method.
3. Scatter Diagram.

❖ **Karl Pearson's Method:** It gives the precise numerical expression for the measure of correlation. It is denoted by 'r'. The value of 'r' gives the magnitude of correlation and its sign denotes its direction.

Merits of Karl Pearson's Method:

- Helps to find direction of correlation.
- Most widely used method.

Demerits of Karl Pearson's method:

- Based on large number of assumptions.
- Affected by extreme values.

❖ **Spearman's Rank Correlation Method:** This method is based on the ranks of the items rather than on their actual values. The advantage of this method over the others in that it can be used even when the actual values of items are unknown. For example if you want to know the correlation between honesty and wisdom of the boys of your class, you can use this method by giving ranks to the boys. It can also be used to find the degree of agreements between the judgments of two examiners or two judges.

Merits of Spearman's Rank Correlation:

- Simple and easy to calculate.
- Not affected by extreme values.

Demerits of Spearman's Rank Correlation:

- Not Suitable for grouped data.
- Not based on original values of observations.

❖ **Scatter Diagram:** Scatter Plots (also called scatter diagrams) are used to graphically investigate the possible relationship between two variables without calculating any numerical value. In this method, the values of the two variables are plotted on a graph paper. One is taken along the horizontal (X-axis) and the other along the vertical (Y-axis). By plotting the data, we get points (dots) on the graph which are generally scattered and hence the name 'Scatter Plot'.

Merits of Scatter Diagram:

- Most simplest method.
- Not affected by size of extreme values.

Demerits of Scatter Diagram:

- It gives only an approximate idea of the relationship.
- It is only an qualitative expression of the quantitative change.

Class : 11th Economics (Statistics)
Chapter-7 : Correlation

Spearman's Rank Correlation

Spearman's rank correlation measures the strength and direction of association between two ranked variables.

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

ρ = Spearman's rank correlation coefficient
d_i = difference between the two ranks of each observation
n = number of observations

Karl Pearson's Method

Karl Pearson's coefficient of correlation is an extensively used mathematical method in which the numerical representation is applied to measure the level of relation between linearly related variables.

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

Where, $\bar{X}$ = mean of X variable
$\bar{Y}$ = mean of Y variable

Correlation

Scatter Diagram

Meaning

A type of diagram used to show the relationship between data items that have two numeric properties along both the axes.

Application of Scatter Diagram Examples

- To show the amount of sleep needed per day by age
- To show the average income for adults based on the number of years of education completed.
- To show the earnings and savings relation of different people.

Important Questions

Multiple Choice questions-

1. When the relation of three or more variables is studied simultaneously, it is called:
   (a) simple correlation
   (b) partial correlation
   (c) multiple correlation
   (d) none of these
2. When coefficient of correlation lies between +0.25 and + 0.75, it is called:
   (a) perfect degree of correlation
   (b) high degree of correlation
   (c) moderate degree of correlation
   (d) low degree of correlation
3. Coefficient of correlation lies always between:
   (a) 0 and +1
   (b) -1 and 0
   (c)-1 and+1
   (d) none of these
4. Rank correlation is a superior method of analysis in case of ____ distribution.
   (a) qualitative
   (b) quantitative
   (c) frequency
   (d) none of these
5. When two variables change in a constant proportion, it is called:
   (a) linear correlation
   (b) non-linear correlation
   (c) partial correlation
   (d) none of these
6. A scatter diagram:
   (a; Is a statistical test
   (b) Must be linear
   (c) Must be curvilinear
   (d) Is a graph of x and y values
7. Maximum value of rank correlation coefficient is:
   (a) 0
   (b) +1
   (c) -1
   (d) None of these
8. The correlation coefficient will be -1 if the slope of the straight line in a scatter diagram is:

(a) Positive
(b) Negative
(c) Zero
(d) None of these

9. In a 'negative' relationship
(a) As x increases, y increases
(b) As x decreases, y decreases
(c) As x increases, y decreases
(d) Both (a) and (b)

10. Scatter diagram helps us to:
(a) Find the nature of correlation between two variables
(b) Obtain the mathematical relationship between two variables
(c) Compute the extent of correlation between two variables
(d) Both (a) and (c)

11. The lowest strength of association is reflected by which of the following correlation coefficients?
(a) 0.95
(b) - 0.60
(c) -0.35
(d) 0.29

12. Relation between price and demand is:
(a) positive
(b) negative
(c) one to one
(d) no relationship

13. If the relationship between x and y is positive, as variable y decreases, variable x:
(a) Increases
(b) Decreases
(c) Remains same
(d) Changes linearly

14. Kari Pearson's coefficient is defined from:
(a) Ungrouped data
(b) Grouped data
(c) Both (a) and (b)
(d) None of these

15. When r = 1, all the points in a scatter diagram would lie:
(a) On a straight line directed from lower left to upper right
(b) On a straight line
(c) On a straight line directed from upper left to lower right
(d) Both (a) and (b)

Very Short Questions:

1. Define correlation.
2. Define partial correlation.
3. Define the line of best fit.
4. Explain the principal methods for calculating the coefficient of correlation?
5. What is the difference between negative and positive correlation?
6. What is the nature of correlation of two variables, when they move in the same direction?
7. Coefficient of correlation is between -1 and +. How would you express it arithmetically?
8. When is the rank correlation method used?
9. What is a simple correlation?
10. What are the multiple correlations?

Short & Long Questions:

1. Concept and definition of correlation?
2. What does Correlation Measure?
3. Linear Correlation?
4. Non-Linear Correlation?
5. Degree of correlation?
6. Importance or significance of correlation?

Assertion Reason Question:

1. Read the following statements given below and choose the correct alternative.

(a) Both assertion and reason are true. The reason is the correct explanation of the assertion

(b) Both assertion and reason are not true. The reason is not the correct explanation of the assertion

(c) Assertion is true but the reason is not

(d) Reason is true but the assertion is not

Assertion- Quantitative method of calculating correlation is given by Karl Pearson.

Reason- Karl Pearson's coefficient of correlation is generally written as 'y'.

2. Read the following statements given below and choose the correct alternative.

   (a) Both assertion and reason are true. The reason is the correct explanation of the assertion

   (b) Both assertion and reason are not true. The reason is not the correct explanation of the assertion

   (c) Assertion is true but the reason is not

   (d) Reason is true but the assertion is not

Assertion- Karl Pearson formula applies only to those series that comes out from the actual average.

Reason- the value of the coefficient of correlation can vary between + 1 and -1

ANSWER KEY

Multiple Choice Answers-

1. C
2. C
3. C
4. A
5. A
6. D
7. B
8. B
9. C
10. A
11. D
12. B
13. B
14. A
15. A

Very Short Answers:

1. Correlation is a statistical method or a technique that measures, a quantitative relationship between different variables, such as demand and price.

2. When more than two variables are involved, and out of these, the relationship

between only two variables is suited to treating other variables as constant, then the correlation is partial.

3. The line of the best fit is one of that passes through the scattered points such that it represents most of these points. Roughly, half of the scattered points should be on either side.
4. The principal methods for calculating the coefficient of correlation are.
   - Scattered Diagram Method
   - Karl Person's Coefficient of Correlation, and
   - Spearman's Rank Correlation Coefficient.
5. The difference between negative and positive correlation is that in positive correlation, variables move in the same direction, whereas, in the negative correlation, they move in the opposite direction.
6. When two variables move in the same direction, such correlation is positive.
7. When the coefficient of correlation is -1, correlation is perfect negative, and when the coefficient of correlation is +1, correlation is perfect positive.
8. The rank correlation method is used when variables are qualitative, such as bravery, beauty, virtue, and wisdom, etc.
9. A simple correlation implies the study or a relationship between two variables only.
10. When the relation between three or more variables are studied simultaneously, it is known as multiple correlations.

Short & Long Answers:

1. **CONCEPT AND DEFINITION OF CORRELATION**

The statistical methods so far studied in this book focus on the analysis of one variable or one statistical series only. In real life however, two or more than two statistical series may be found to be mutually related. For instance, change in price leads to change in quantity demanded. Increase in supply of money causes increase in price level. Increase in level of employment results in increase in output. Such situations necessitate simultaneous study of two or more statistical series. The focus of study in such situations is on the degree of relationship between different statistical series. The statistical technique that studies the degree of such relationships is called the technique of correlation.

Definition

According to 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." In the words of Boddington, "Whenever some definite connection exists between the two or more groups, classes or series or data there is said to be correlation. "

Relationship between Two Variables may just be a Coincidence One may find a relationship between two variables which is just a coincidence. Example: When there is a departure of migratory birds from a sanctuary, you may find a fall in wedding ceremonies in the country. Such relationships are meaningless. These are in other words, spurious relationships which are devoid of any meaningful conclusion. Such relationships are not to be treated as correlations, Only those relationships are to be treated as correlations which offer some meaningful conclusions.

Example: Increase in rainfall and increase in rice production is a relationship that makes sense: increase in per capita income and decrease in death rate is a meaningful relationship; Good percentage of marks in physics may be related to good percentage of marks in mathematics; and so on.

Positive and Negative Correlation Correlation between different variables may either be positive or negative. Here is a brief description of the two:

(1) Positive Correlation

When two variables move in the same direction, that is, when one increases the other also increases and when one decreases the other also decreases, such a relation is called positive correlation.

Relationship between price and supply may be cited as an example. Check the following table as an illustration:

(A) Simultaneous Increase in the Values of both Variables		(B) Simultaneous Decrease in the Values of both Variables	
X	Y	X	Y
10	100	50	200
20	150	40	150
30	200	30	100
40	250	20	50

(2) Negative Correlation

When two variables change in different directions, it is called negative correlation. Relationship between price and demand, may be cited as an example. The following table demonstrates this relation:

Rise in the Value of One Variable is X	Accompanied with a fall in the other Y
1	5
2	4
4	2
5	1

2. Often the students tend to believe that correlation suggests a relationship between two variables where one is the cause of the other. Example: This is correlation between price and quantity demanded of a commodity. Clearly, an increase in price causes a decrease in quantity demanded, and vice versa. Change in price causes changes in quantity demanded. But to be emphatically noted is the point that cause and effect relationship between the variables is not at all any pre-condition in the theory of correlation.

   Correlation just measures the degree and intensity of relationship between the two variables, with or without any cause and effect relationship. Of course, the established relationship between the variables should be capable of offering us some meaningful conclusion. Example: Students who are good in academics may be good in sports also. Certainly, it is a meaningful relationship (or correlation) if one finds it. But surely there is no cause and effect relationship between the two variables. Linear and Non-Linear Correlation.

3. **Linear Correlation**

When two variables change in a constant proportion, it is called linear correlation. If the two sets of data bearing fixed proportion to each other are shown on a graph paper, their relationship will be indicated by a straight line. Thus, linear correlation implies a straightline relationship.

Illustration.

Linear Correlation

(a)	2	4	6	8	10	12	14
(b)	5	10	15	20	25	30	35
(a)	2	4	6	8	10	12	14
(b)	3	7	12	18	25	35	45

Thus, for every change in variable (a) by 2 units there is a change in variable (b) by 5 units.

4. Non-linear Correlation

Here there is no specific relationship between the two variables, though both tend to change in the same direction. That is, both are increasing, but not in any constant proportion.

Simple and Multiple Correlation

(1) Simple Correlation

Simple correlation implies the study of relationship between two variables only. Like the relationship between price and demand or the relationship between money supply and price level.

(2) Multiple Correlation

When the relationship among three or more than three variables is studied simultaneously, it is called multiple correlations. In case of such correlation, the entire set of independent and dependent variables is simultaneously studied. For instance, effects of rainfall, manure, water, etc., on per hectare productivity of wheat are simultaneously studied.

5. Degree of correlation refers to the Coefficient of Correlation. There can be the following degrees of positive and negative correlation.
(1) Perfect Correlation: When two variables change in the same proportion it is called perfect correlation. It may be of two kinds:.

(i) Perfect Positive: Correlation is perfectly positive when proportional change in two variables is in the same direction. In this case, coefficient of correlation is positive (+1).

(ii) Perfect Negative: Correlation is perfectly negative when proportional change in two variables is in the opposite direction. In this case, coefficient of correlation is negative (-1).

(2) Absence of Correlation: If there is no relation between two series or variables, that is, change in one has no effect on the change in other, then those series or variables lack any correlation between them.

(3) Limited Degree of Correlation: Between perfect correlation and absence of correlation there is a situation of limited degree of correlation. In real life, one mostly finds limited degree of correlation. Its coefficient (r) is more than zero and less than one ($r > 0$ but < 1). The degree of correlation between 0 and 1 may be rated as:

(i) High: When correlation of two series is close to one, it is called high degree of correlation. Its coefficient lies between 0.75 and 1,

(ii) Moderate: When correlation of two series is neither large nor small, it is called moderate degree of correlation. Its coefficient lies between 0.25 and 0.75.

(iii) Low: When the degree of correlation of two series is very small, it is called low degree of correlation. Its coefficient lies between 0 and 0.25. All these degrees of correlations may be positive or negative.

Degree of Correlation

Degree	Positive	Negative
Perfect	+ 1	-1
High	Between + 0.75 and +1	Between - 0.75 and -1
Moderate	Between + 0.25 and + 0.75	Between - 0.25 and - 0.75
Low	Between 0 and + 0.25	Between 0 and - 0.25
Zero	o	0

6. Following observations highlight the importance or significance of correlation as a statistical method:

   (1) Formation of Laws and Concepts: The study of correlation shows the direction and degree of relationship between the variables. This has helped the formation of various laws and concepts in economic theory, such as, the law of demand and the concept of elasticity of demand.

   (2) Cause and Effect Relationship: Correlation coefficient sometimes suggests cause and effect relationship between different variables. This helps in understanding why certain variables behave the way they behave.

   (3) Business Decisions: Correlation analysis facilitates business decisions because the trend path of one variable may suggest the expected changes in the other. Accordingly, the businessman may plan his business decisions for the future.

   (4) Policy Formulation: Correlation analysis also helps policy formulation. If the Government finds a negative correlation between tax rate and tax collection, it should pursue the policy of low tax rate. Because, low tax rate would lead to high tax collection.

Assertion Reason Answer:

1. (c) Assertion is true but the reason is not.

2. (a) Both Assertion and Reason are true. The reason is the correct explanation of the assertion.