

### ***UNIT - 3 : STATISTICAL TOOLS AND INTERPRETATION***

#### **Correlation**

##### **Points to Remember**

- \* Meaning of correlation :  
Correlation is a statistical tool which studies the relationship between two variables. For e.g. change in price leads to change in quantity demanded.
- \* Correlation studies and measures the direction and intensity of relationship among variables. It measures covariation not causation.

- \* Types of Correlation

Correlation is classified into positive and negative correlation.

The correlation is said to be positive when the variables move together in the same direction. For e.g. sale of Ice cream and temperature move in same direction.

The correlation is said to be negative when the variables move in opposite direction. For e.g. When you spend more time in studying chances of your failing decline.

- \* Examples of positive correlation are :
  1. Price of commodity and amount of supply
  2. Increase in height and weight.
  3. Age of husband and age of wife.
  4. The family income and expenditure on luxury items.

\* Examples of negative correlation are :

1. Sale of woollen garments and day temperature.
2. Demand of a commodity may go down as a result of rise in prices.
3. Yield of crops and price.

\* 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 & -0.75
Low	Between 0 and + 0.25	Between 0 and - 0.25
Zero	0	0

\* Methods of estimating correlation.

- (a) Scatter diagram
- (b) Karl pearson's coefficient of correlation.
- (c) Spearman's rank correlation.

- Scatter diagram offers a graphic expression of the direction and degree of correlation.
- Karl pearson's coefficient of correlation is a quantitative method of calculating correlation. It gives a precise numerical value of the degree of linear relationship between two variables.
- Karl pearson's coefficient of correlation is also known as product moment correlation.

Formula :

$$r = \frac{\sum x y}{N \sigma x \sigma y}$$

Here,

$r$  = Coefficient of correlation

$$x = (X - \bar{X})$$

$$y = (Y - \bar{Y})$$

$\sigma_x$  = Standard deviation of  
X - series

$\sigma_y$  = Standard deviation of  
Y- Series

$N$  = Number of observations

Karl pearson's coefficient of correlation is calculated by following methods :

(a) Actual mean method : 
$$r = \frac{\sum xy}{\sqrt{\sum x^2 \cdot \sum y^2}}$$

Here,

$r$  = Coeff. of correlation

$$x = (X - \bar{X})$$

$$y = (Y - \bar{Y})$$

(b) Assumed mean method :

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

(or)

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

Here,

$dx$  = Deviations of x- series from assumed mean =  $(X-A)$

$dy$  = Deviation of Y- series from assumed mean =  $(Y-A)$

$\sum dx \cdot dy$  = Sum of the multiple of  $dx$  and  $dy$

$\sum dx^2$  = Sum of the square of  $dx$

$\sum dy^2$  = Sum of the square of  $dy$

$\sum dx$  = Sum of the deviation of x- series

$\sum dy$  = Sum of the deviation of Y-series

$N$  = Number of pairs of observations

(c) Step deviation method :

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

$$\text{Here, } dx' = \frac{dx}{C_1}$$

$$dy' = \frac{dy}{C_2}$$

$C_1$  is common factor for series - x

$C_2$  is common factor for series - y

\* Properties of correction coefficient (r)

(i) Correlation coefficient (r) has no unit

(ii) A negative value of r indicates an inverse relation.

(iii) If r is positive then two variables move in the same direction.

(iv) The value of r lies between minus one (1) and plus one,

$$-1 \leq r \leq 1$$

(v) If r is zero, the two variables are uncorrelated.

(vi) If  $r = +1$  or  $r = -1$ , the correlation is perfect.

(viii) A high value of r indicates strong linear relationship and a low

value or  $r$  indicates a weak linear relationship.

- (viii) The value of  $r$  is unaffected by the change of origin and change of scale.

Given two variables  $x$  and  $y$  let us define two new variables

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

Here  $A$  and  $C$  are assumed means of  $X$  and  $Y$  respectively.  $B$  and  $D$  are common factors. They  $r_{xy} = r_{uv}$

- Spearman's rank correlation method is used to calculate coefficient of correlation of qualitative variables such as beauty, bravery, wisdom, ability, virtue etc.

**Formula** 
$$r_s = 1 - \frac{6 \sum D^2}{N^3 - N}$$

Here,  $r_s$  = Coefficient of rank correlation

$D$  = Rank differences

$N$  = Numbers of pairs

When ranks are repeated the formula is

$$r_s = \frac{6[\sum d^2 + \frac{(m_1^3 - m_1)}{12} + \frac{(m_2^3 - m_2)}{12} + \dots]}{N^3 - N}$$

Where  $m_1, m_2, \dots$  are number of repetitions of ranks.

**VERY SHORT ANSWER TYPE QUESTIONS  
(ONE 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 of 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?

### 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  $x$  and  $y$  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  $x$  and  $y$  and comment on their relationship :

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

(Ans.  $r = +1$ )

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 Karl Pearson's coefficient of correlation on the following data :

X	15	18	21	24	27	30	36	39	42	48
Y	25	25	27	27	31	33	35	41	41	45

(Ans.  $r = 0.98$ )

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

	<u>X series</u>	<u>Y series</u>
(i) No. of itmes	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 their 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 spearman's rank correlation from the following data :

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:

X	1	2	3	4	5	6	7	8	9	10	11	12
Y	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 of 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 MARKS 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	10	3	2	4	9	7	8
Judge II	3	5	8	4	7	10	2	1	6	9
Judge III	6	4	9	8	1	2	3	10	5	7

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

(Ans.  $r_s \text{ I \& II} = -0.21$ ;  $r_s \text{ II \& III} = -0.29$ ;  $r_s \text{ I \& III} = +0.64$ )

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 the marks obtained by 12 students in 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 :  $r = + 0.78$ )

**ANSWER OF ONE 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 variables move in opposite direction.
4. The correlation is said to be positive when the variables 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. Scatter 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) Scatter diagram  
(ii) Karl pearson's coefficient of correlation.  
(iii) Spearman's Rank correlation.
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.