# Measures of Correlation

PART 1

## Objective Questions

### • Multiple Choice Questions

**1.** Which of the following techniques deals with the association between two or more variables?

(a) Index number (b) Correlation
(c) Dispersion (d) None of these

*Ans.* (b) Correlation

**2.** When two variables move together in the same direction, it is said to be

(a) no correlation (b) negative correlation
(c) positive correlation (d) zero correlation

*Ans.* (c) positive correlation

**3.** Correlation is an analysis of ...... between two or more variables.

(a) relationship (b) covariation
(c) determination (d) calculation

*Ans.* (b) covariation

**4.** If all the points lie on the same downward sloping line, the correlation is said to be

(a) perfect correlation
(b) perfect positive correlation
(c) perfect negative correlation
(d) negative correlation

*Ans.* (c) Downward sloping line indicates inverse relationship between variable X and Y and as they lie in a straight line, it indicates perfect negative correlation.

**5.** Scatter diagram can be used to indicate which of the following degrees of correlation?

(a) Perfect positive correlation
(b) Perfect negative correlation
(c) No correlation (d) All of the above

*Ans.* (d) All of the above

**6.** A modified version of Karl Pearson's formula is

(a) $r = \dfrac{\Sigma xy}{\Sigma x^2 \cdot \Sigma y^2}$ (b) $r = \dfrac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$

(c) $r = \dfrac{\Sigma xy}{n\,\Sigma X^2 \cdot \Sigma Y^2}$ (d) $r = \dfrac{\Sigma xy}{n\,\delta_x \cdot \delta_y}$

*Ans.* (b) $r = \dfrac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$

**7.** Karl Pearson's coefficient of correlation indicates the ...... and also the degree of relationship between the two variables.

(a) direction (b) relation
(c) interpretation (d) None of these

*Ans.* (a) Karl Pearson's method is the most useful method of correlation, which can be used to indicate direction as well as magnitude of change.

**8.** The coefficient of correlation is independent of

(a) change of scale only
(b) change of origin only
(c) both change of scale and origin
(d) None of the above

*Ans.* (c) Coefficient of correlation remains unaffected due to change in either scale or origin.

**9.** When the mean of series is a decimal number, then which method should be used for computing Karl Pearson's coefficient of correlation?

(a) Direct Method (b) Short-cut Method
(c) Step Deviation Method (d) None of these

*Ans.* (b) Under short-cut method of computing correlation, assumed mean is used in place of actual mean. Thus, it can be used when actual mean comes in decimal points.

**10.** Which of the following pair is correctly matched?

| Column I (Method) | Column II (Formula) |
|---|---|
| A. Direct Method | (i) $r = \dfrac{\Sigma dxdy \cdot n - (\Sigma dx)(\Sigma dy)}{\sqrt{\Sigma dx^2 \cdot n - (\Sigma dx)^2} \times (\sqrt{\Sigma dy^2 . n - (\Sigma dy)^2}}$ |
| B. Short-cut Method | (ii) $r = \dfrac{\Sigma xy}{n \cdot \sigma_x \cdot \sigma_y}$ |
| C. Step Deviation Method | (iii) $r = \dfrac{\Sigma dx'dy' \cdot n - (\Sigma dx')(\Sigma dy')}{\sqrt{\Sigma dx'^2 \cdot n - (\Sigma dx')^2} \times \sqrt{\Sigma dy'^2 . n - (\Sigma dy')^2}}$ |

**Codes**

(a) A–(i) (b) B–(ii)
(c) C–(iii) (d) None of these

*Ans.* (c) C–(iii)

**11.** The minimum limit of correlation under Karl Pearson's method is

(a) –1 (b) 0
(c) 1 (d) None of these

***Ans.*** (a) Minimum limit of correlation is –1 as coefficient of correlation ranges from –1 to 1.

**12. Statement I** Correlation is a multivariate analysis.

**Statement II** Partial correlation considers all other variables to be constant.

**Alternatives**

(a) Statement I is correct and Statement II is incorrect
(b) Statement II is correct and Statement I is incorrect
(c) Both the statements are correct
(d) Both the statements are incorrect

***Ans.*** (c) Both the statements are correct

**13. Statement I** Non-linear correlation is also called curvy linear correlation.

**Statement II** Numerical measure of correlation is called coefficient of determination.

**Alternatives**

(a) Statement I is correct and Statement II is incorrect
(b) Statement II is correct and Statement I is incorrect
(c) Both the statements are correct
(d) Both the statements are incorrect

***Ans.*** (a) Statement I is correct and Statement II is incorrect

**14.** If the dots in a scatter diagram fall on a narrow band, it indicates a ..........degree of correlation.

(a) zero (b) high
(c) low (d) None of these

***Ans.*** (c) low

**15.** Coefficient of correlation lies between

(a) 0 and +1 (b) 0 and –1
(c) –1 and +1 (d) – 3 and +3

***Ans.*** (c) –1 and +1

## • Assertion–Reasoning MCQs

**Direction** *(Q. Nos. 1 to 4) There are two statements marked as Assertion (A) and Reason (R). Read the statements and choose the appropriate option from the options given below.*

(a) Both Assertion (A) and Reason (R) are true and Reason (R) is the correct explanation of Assertion (A)
(b) Both Assertion (A) and Reason (R) are true, but Reason (R) is not the correct explanation of Assertion (A)
(c) Assertion (A) is false, but Reason (R) is true
(d) Both Assertion (A) and Reason (R) are false

**1. Assertion** (A) Sale of ice-cream increases during summer, shows positive correlation.

**Reason** (R) When two variables move in same direction, it shows positive correlation.

***Ans.*** (b) Sale of ice cream and season indicates cause and effect relationship and this relation as per question is positive.

**2. Assertion** (A) Correlation analysis is a means for examining inter relationships systematically.

**Reason** (R) Causation explain the cause and effect relationship between variables.

***Ans.*** (b) Correlation is a measure of interrelation between two variables and does not indicate the cause and effect relationship.

**3. Assertion** (A) Karl Pearson's method is non-mathematical in nature.

**Reason** (R) Degree of correlation helps in identifying the nature of correlation between variables.

***Ans.*** (c) Karl Pearson's method is purely a mathematical method which is used to measure the magnitude of change in the two or more variables.

**4. Assertion** (A) Broadly, there only two types of correlation i.e., positive and negative.

**Reason** (R) The correlation is said to be positive when the variables move together in the same direction, the correlation is negative when they move in opposite directions.

***Ans.*** (b) There are only two types of correlation i.e., positive and negative correlation, rest of the degrees are the sub-types of positive and negative correlation only.

## • Case Based MCQs

**1. Direction** *Read the following case study and answer the question no. (i) to (vi) on the basis of the same.*

Coefficient of correlation is an important statistical tool which is used to measure the relationship between two variables. This is not only useful in the field of statistics but also used in other disciplines like Economics, Geography, Psychology.

In the present time due to the outbreak of Covid-19 corona virus, demand has gradually come down in almost all areas. As per the latest estimates, demand for car has decreased after the nationwide lockdown is lifted in phased manner.

(i) What will be the coefficient of correlation between demand for car and varied level of income due to nationwide lockdown?

(a) Positive correlation (b) Negative correlation
(c) Perfect positive correlation (d) No correlation

***Ans.*** (a) Due to lockdown, income level declined which further led to fall in demand for cars and thus, indicates positive correlation between the two.

(ii) If one variable change exactly in the reverse direction of the other variable, .......... should be the degree of correlation.

(a) positive correlation
(b) negative correlation
(c) perfect positive correlation
(d) perfect negative correlation

***Ans.*** (d) perfect negative correlation

(iii) Which of the following tools can be used to know the pattern of demand during lockdown?

(a) Correlation
(b) Causation
(c) Both (a) and (b)
(d) Neither (a) nor (b)

Ans. (b) Causation

(iv) Which method of calculating correlation uses actual mean?

(a) Karl Pearson's coefficient of correlation
(b) Scatter diagram method
(c) Spearman's rank correlation method
(d) Both (a) and (c)

***Ans.*** (a) Karl Pearson's coefficient of correlation

(v) In perfect positive correlation, the value of '$r$' is ...... .

(a) –1 (b) 0
(c) 1 (d) infinity

***Ans.*** (c) 1

(vi) The correlation between using mask and being injected by the virus will be

(a) positive (b) negative
(c) Not correlated (d) Either (a) or (b)

***Ans.*** (b) Wearing mask reduces the chance of being infected by the virus, thus shows an inverse relationship between the two.

# PART 2
# Subjective Questions

## • Short Answer (SA) Type Questions

**1.** What do you understand by 'spurious' or 'non-sense' correlation?

***Ans.*** If there is no evident or sensible connection between two variables, then the correlation between these variables is said to be spurious, non-sense or chance correlation. For example, correlation between rainfall recorded and production of steel.

These two variables are not connected by any way. So, the correlation between these variables is said to be spurious.

**2.** Consider the examples given below

(i) As price falls, demand for product 'A' increases.
(ii) Effect of adequate irrigation facilities, fertilisers and pesticides on per hectare productivity of wheat.

On the basis of above examples explain the main difference between simple correlation and multiple correlation.

***Ans.*** The first example involves only two variables, viz. price and demand. Therefore, it relates to simple correlation.

The second example involves more than two variables, i.e., how the productivity of wheat is affected by use of irrigation facilities, fertilisers and pesticides. Therefore, it relates to multiple correlation.

The main difference between simple correlation and multiple correlation is

| Simple Correlation | Multiple Correlation |
|---|---|
| When the relationship between only two variables is studied, it is called simple correlation. | When the relationship among three or more than three variables is studied simultaneously, it is called multiple correlation. |

**3.** Why is '**r**' preferred to covariance as a measure of association? **(NCERT)**

***Ans.*** Both, correlation coefficient and covariance measure the degree of linear relationship between two variables but correlation coefficient is generally preferred to covariance.

It is due to the following reasons

(i) The correlation coefficient (**r**) has no unit.
(ii) The correlation coefficient is independent of origin as well as scale.

**4.** Can $r$ lie outside the **−1** and 1 range depending on the type of data? **(NCERT)**

***Ans.*** No, the value of the correlation coefficient lies between minus one and plus one i.e., $-1 \le r \le 1$. If the value of $r$ is outside this range in any type of data, it indicates error in calculation as in between two or more variables, there can be either a perfect or an imperfect relationship. A perfect relationship is indicated by –1 or 1 and imperfect relationships are indicated by a value between –1 and 1, excluding 0.

**5.** List some variables where accurate measurement is difficult. **(NCERT)**

***Ans.*** Accurate measurement is difficult in case of

(i) Qualitative variables such as beauty, intelligence, honesty, etc.
(ii) It is also difficult to measure subjective variables such as poverty, development, etc, which are interpreted differently by different people.
(iii) Where the cause and effect relationship is not known.

**6.** Interpret the values of $r$ as 1, –1 and 0. (**NCERT**)

***Ans.*** (i) If $r = 0$, the two variables are uncorrelated. There is no linear relation between them. However, other types of relation may be there and hence the variables may not be independent.

(ii) If $r = 1$, the correlation is perfectly positive. The relation between them is exact in the sense that if one increases,

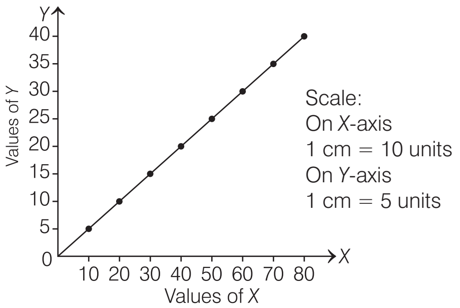
the other also increases in the same proportion and if one decreases, the other also decreases in the same proportion.

(iii) If $r = -1$, the correlation is perfectly negative. The relation between them is exact in the sense that if one increases, the other decreases in the same proportion and if one decreases, the other increases in the same proportion.

**7.** Draw a scatter diagram and indicate the nature of correlation.

| **X** | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|---|---|---|---|---|---|---|---|---|
| **Y** | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 |

***Ans.*** Now, we plot the points on a graph paper which is shown below



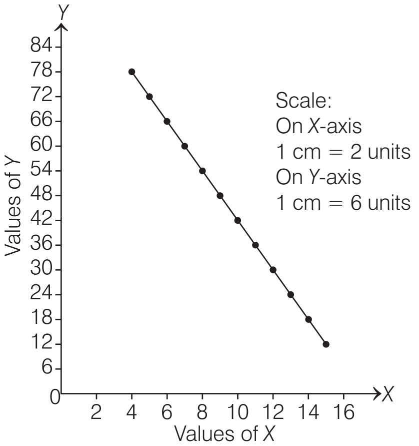

The diagram indicates that there is perfect positive correlation between the values of the two variables **X** and **Y**.

**8.** Draw a scatter diagram and interpret whether the correlation is positive or negative.

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

***Ans.*** The pair of points are (4, 78), (5, 72), (6, 66), (7, 60), (8, 54), (9, 48), (10, 42), (11, 36), (12, 30), (13, 24), (14, 18) and (15, 12).

Now we plot the points on a graph paper, which is shown below




The diagram indicates that there is perfect negative correlation between the values of the two variables **X** and **Y**.

**9.** Calculate Karl Pearson's coefficient of correlation between **X** and **Y** from the following data

$\mathbf{n = 8, \overline{X} = 11, \overline{Y} = 10, \Sigma x^2 = 184, \Sigma y^2 = 148,}$

$\mathbf{\Sigma xy = 164}$

***Ans.*** Given that, $\mathbf{n = 8, \overline{X} = 11, \overline{Y} = 10, \Sigma x^2 = 184, \Sigma y^2 = 148}$ and $\mathbf{\Sigma xy = 164}$. Applying the formula,

$$\mathbf{r} = \frac{\mathbf{\Sigma xy}}{\sqrt{\mathbf{\Sigma x^2 \times \Sigma y^2}}} = \frac{\mathbf{164}}{\sqrt{\mathbf{184 \times 148}}}$$

$$= \frac{\mathbf{164}}{\sqrt{\mathbf{27,232}}}$$

$$= \frac{\mathbf{164}}{\mathbf{164.02}} = \mathbf{0.99}$$

**10.** The following results are obtained regarding two series. Compute coefficient of correlation, when sum of products of deviations of **X** and **Y** series from their respective mean is 122.

| | **X Series** | **Y Series** |
|---|---|---|
| Number of Items | 15 | 15 |
| Arithmetic Mean | 25 | 18 |
| Standard Deviation | 3.01 | 3.03 |

***Ans.*** We are given that, $\mathbf{n = 15, \overline{X} = 25, \sigma_x = 3.01,}$

$\mathbf{\overline{Y} = 18, \sigma_y = 3.03}$

and $\mathbf{\Sigma xy = 122}$

$$\mathbf{r} = \frac{\mathbf{\Sigma xy}}{\mathbf{n \cdot \sigma_x \cdot \sigma_y}} = \frac{\mathbf{122}}{\mathbf{15 \times 3.01 \times 3.03}} = \frac{\mathbf{122}}{\mathbf{136.80}} = 0.89$$

**11.** Give the advantages of Karl Pearson's coefficient of correlation.

***Ans.*** The advantages of Karl Pearson's coefficient of correlation are

(i) Karl Pearson's coefficient of correlation indicates the relationship as positive or negative and thus direction of the relationship can be ascertained.

(ii) This measure gives summarised and precise quantitative figure of correlation which can be interpreted easily and can provide meaningful results.

(iii) This coefficient of correlation indicates the direction and also the degree of relationship between the two variables. It shows whether the relationship is high, moderate or low.

**12.** Give the disadvantages of Karl Pearson's coefficient of correlation.

***Ans.*** The disadvantages of Karl Pearson's coefficient of correlation are

(i) The value of coefficient is affected by extreme items.

(ii) The calculation process consumes a lot of time.

(iii) Correlation coefficient needs very careful interpretation, otherwise it may be misinterpreted.

**13.** From the following data, calculate Karl Pearson's coefficient of correlation.

| **X** | 6 | 2 | 10 | 4 | 8 |
|---|---|---|---|---|---|
| **Y** | 9 | 11 | ? | 8 | 7 |

Arithmetic mean of **X** and **Y** series are 6 and 8, respectively.

***Ans.*** Let the missing value be a

$$\bar{Y} = \frac{\Sigma Y}{n} = \frac{9+11+a+8+7}{5} = \frac{35+a}{5}$$

$$\Rightarrow \quad 8 = \frac{35+a}{5} \Rightarrow 40 = 35 + a \Rightarrow a = 5$$

Thus, the completed series is

| **X** | 6 | 2 | 10 | 4 | 8 |
|---|---|---|---|---|---|
| **Y** | 9 | 11 | 5 | 8 | 7 |

Now, we find coefficient of correlation.

**Calculation of Coefficient of Correlation**

| X | $x\,(X-\bar{X})$ | $x^2$ | Y | $y\,(Y-\bar{Y})$ | $y^2$ | xy |
|---|---|---|---|---|---|---|
| 6 | 0 | 0 | 9 | 1 | 1 | 0 |
| 2 | –4 | 16 | 11 | 3 | 9 | –12 |
| 10 | 4 | 16 | 5 | –3 | 9 | –12 |
| 4 | –2 | 4 | 8 | 0 | 0 | 0 |
| 8 | 2 | 4 | 7 | –1 | 1 | –2 |
| $\Sigma X = 30$ | | $\Sigma x^2 = 40$ | $\Sigma Y = 40$ | | $\Sigma y^2 = 20$ | $\Sigma xy = -26$ |

Here, $n = 5, \Sigma X = 30, \Sigma x^2 = 40, \Sigma Y = 40, \Sigma y^2 = 20$ and $\Sigma xy = -26$

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

$$= \frac{-26}{\sqrt{40 \times 20}} = \frac{-26}{\sqrt{800}} = \frac{-26}{28.28} = -0.9193$$

It indicates that there is high degree of negative correlation between **X** and **Y**.

**14.** Calculate the correlation coefficient between *X* and *Y* and comment on the relationship. **(NCERT)**

| X | –3 | –2 | –1 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|
| Y | 9 | 4 | 1 | 1 | 4 | 9 |

*Ans.* **Calculation of Coefficient of Correlation**

| X | $X^2$ | Y | $Y^2$ | XY |
|---|---|---|---|---|
| –3 | 9 | 9 | 81 | –27 |
| –2 | 4 | 4 | 16 | –8 |
| –1 | 1 | 1 | 1 | –1 |
| 1 | 1 | 1 | 1 | 1 |
| 2 | 4 | 4 | 16 | 8 |
| 3 | 9 | 9 | 81 | 27 |
| $\Sigma X = 0$ | $\Sigma X^2 = 28$ | $\Sigma Y = 28$ | $\Sigma y^2 = 196$ | $\Sigma xy = 0$ |

Here, $n = 6,\ \Sigma XY = 0,\ \Sigma X^2 = 28$ and $\Sigma Y^2 = 196$

$$\therefore \quad r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \times \Sigma Y^2}} = \frac{0}{\sqrt{28 \times 196}} = 0$$

**r = 0** shows that there is absence of correlation between the variables **X** and **Y** but we observe that it remains a non-linear correlation between the two variables as $y = x^2$. So, in this question, the correlation coefficients fails to indicate the correct relationship between these two variables.

**15.** Calculate the correlation coefficient between *X* and *Y* and comment on their relationship. **(NCERT)**

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

*Ans.* **Calculation of Coefficient of Correlation**

| X | Y | XY | $X^2$ | $Y^2$ |
|---|---|---|---|---|
| 1 | 2 | 2 | 1 | 4 |
| 3 | 6 | 18 | 9 | 36 |
| 4 | 8 | 32 | 16 | 64 |
| 5 | 10 | 50 | 25 | 100 |
| 7 | 14 | 98 | 49 | 196 |
| 8 | 16 | 128 | 64 | 256 |
| $\Sigma X = 28$ | $\Sigma Y = 56$ | $\Sigma XY = 328$ | $\Sigma X^2 = 164$ | $\Sigma Y^2 = 656$ |

Here, $n = 6, \Sigma XY = 328,\ \Sigma X^2 = 164$ and $\Sigma Y^2 = 656$

$$\therefore \quad r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \times \Sigma Y^2}}$$

$$= \frac{328}{\sqrt{164 \times 656}} = \frac{328}{328} = 1$$

As the correlation coefficient between the two variables is + 1, so the two variables are perfectly positively correlated.

**16.** Compute Karl Pearson's coefficient of correlation by direct method and interpret the result.

| Marks in Mathematics | 15 | 18 | 21 | 24 | 27 |
|---|---|---|---|---|---|
| **Marks in Accountancy** | 25 | 25 | 27 | 31 | 32 |

***Ans.*** Let **X** and **Y** denote marks in mathematics and accountancy, respectively.

**Calculation of Coefficient of Correlation**

| X | x $(X-\bar{X})$, $\bar{X}=21$ | $x^2$ | Y | y $(Y-\bar{Y})$ $\bar{Y}=28$ | $y^2$ | xy |
|---|---|---|---|---|---|---|
| 15 | –6 | 36 | 25 | –3 | 9 | 18 |
| 18 | –3 | 9 | 25 | –3 | 9 | 9 |
| 21 | **0** | 0 | 27 | –1 | 1 | 0 |
| 24 | 3 | 9 | 31 | 3 | 9 | 9 |
| 27 | 6 | 36 | 32 | 4 | 16 | 24 |
| $\Sigma X = 105$ | | $\Sigma x^2 = 90$ | $\Sigma Y = 140$ | | $\Sigma y^2 = 44$ | $\Sigma xy = 60$ |

Here, $\Sigma X = 105$, $\Sigma Y = 140$, $\Sigma xy = 60$, $\Sigma x^2 = 90$ and $\Sigma y^2 = 44$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{105}{5} = 21;\ \bar{Y} = \frac{\Sigma Y}{n} = \frac{140}{5} = 28;\ r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}} = \frac{60}{\sqrt{90 \times 44}} = \frac{60}{\sqrt{3{,}960}} = \frac{60}{62.928} = 0.95$$

It indicates that there is high degree of positive correlation between marks in mathematics and accountancy.

## • Long Answers (LA) Type Questions

**1.** Calculate the correlation coefficient between the height of fathers in inches (**X**) and their sons (**Y**). **(NCERT)**

| X | 65 | 66 | 57 | 67 | 68 | 69 | 70 | 72 |
|---|---|---|---|---|---|---|---|---|
| **Y** | 67 | 56 | 65 | 68 | 72 | 72 | 69 | 71 |

***Ans.***

**Calculation of Coefficient of Correlation**

| X | $x(X-\bar{X})\ \bar{X} = 66.75$ | $x^2$ | Y | $y\,(Y-\bar{Y})$ $\bar{Y} = 67.5$ | $y^2$ | xy |
|---|---|---|---|---|---|---|
| 65 | – 1.75 | 3.0625 | 67 | – 0.5 | 0.25 | 0.875 |
| 66 | – 0.75 | 0.5625 | 56 | – 11.5 | 132.25 | 8.625 |
| 57 | – 9.75 | 95.0625 | 65 | – 2.5 | 6.25 | 24.375 |
| 67 | 0.25 | 0.0625 | 68 | 0.5 | 0.25 | 0.125 |
| 68 | 1.25 | 1.5625 | 72 | 4.5 | 20.25 | 5.625 |
| 69 | 2.25 | 5.0625 | 72 | 4.5 | 20.25 | 10.125 |
| 70 | 3.25 | 10.5625 | 69 | 1.5 | 2.25 | 4.875 |
| 72 | 5.25 | 27.5625 | 71 | 3.5 | 12.25 | 18.375 |
| $\Sigma X = 534$ | | $\Sigma x^2 = 1{,}43.5$ | $\Sigma Y = 540$ | | $\Sigma y^2 = 194$ | $\Sigma xy = 73$ |

Here, $n = 8$, $\Sigma X = 534$, $\Sigma x^2 = 143.5$, $\Sigma Y = 540$, $\Sigma y^2 = 194$ and $\Sigma xy = 73$

Now, $\bar{X} = \frac{\Sigma X}{n} = \frac{534}{8} = 66.75$, and $\bar{Y} = \frac{\Sigma Y}{n} = \frac{540}{8} = 67.5$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}} = \frac{73}{\sqrt{143.5 \times 194}} = \frac{73}{\sqrt{27{,}839}} = \frac{73}{166.85} = 0.438$$

It indicates that there is low degree of positive correlation between heights of fathers and sons.

**2.** Calculate coefficient of correlation between age group and rate of mortality from the following data.

| **Age Group** | 0–20 | 20–40 | 40–60 | 60–80 | 80–100 |
|---|---|---|---|---|---|
| **Rate of Mortality** | 350 | 280 | 540 | 760 | 900 |

***Ans.*** Since, class interval are given for age, so mid value should be used for the calculation of **r**.

**Calculation of Coefficient of Correlation**

| Age Group | Mid-value (X) | dx(X – A), A = 50 | $dx'\left(\frac{dx}{c_1}\right)$, $c_1 = 20$ | $dx'^2$ | Rate of Mortality (Y) | dy(Y – A), A = 540 | $dy'\left(\frac{dx}{c_2}\right)$, $c_2 = 10$ | $dy'^2$ | $dx'dy'$ |
|---|---|---|---|---|---|---|---|---|---|
| 0–20 | 10 | –40 | –2 | 4 | 350 | –190 | –19 | 361 | 38 |
| 20–40 | 30 | –20 | –1 | 1 | 280 | –260 | –26 | 676 | 26 |
| 40–60 | 50 | 0 | 0 | 0 | 540 | 0 | 0 | 0 | 0 |
| 60–80 | 70 | 20 | 1 | 1 | 760 | 220 | +22 | 484 | 22 |
| 80–100 | 90 | 40 | 2 | 4 | 900 | 360 | +36 | 1,296 | 72 |
| | | | $\Sigma dx' = 0$ | $\Sigma dx'^2 = 10$ | | | $\Sigma dy' = 13$ | $\Sigma dy'^2 = 2,817$ | $\Sigma dx'dy' = 158$ |

Here, $n=5, \Sigma dx' = 0, \Sigma dx'^2 = 10, \Sigma dy' = 13, \Sigma dy'^2 = 2,817$ and $\Sigma dx'dy' = 158$

Now,
$$r = \frac{\Sigma dx'dy' - \frac{\Sigma dx' \times \Sigma dy'}{n}}{\sqrt{\Sigma dx'^2 - \frac{(\Sigma dx')^2}{n}} \times \sqrt{\Sigma dy'^2 - \frac{(\Sigma dy')^2}{n}}} = \frac{158 - \frac{0 \times 13}{5}}{\sqrt{10 - \frac{(0)^2}{5}} \times \sqrt{2,817 - \frac{(13)^2}{5}}} = \frac{158}{\sqrt{10 - 0} \times \sqrt{2,817 - \frac{169}{5}}}$$

$$= \frac{158}{\sqrt{10} \times \sqrt{2,817 - 33.8}} = \frac{158}{\sqrt{10} \times \sqrt{2,783.2}} = \frac{158}{3.16 \times 52.8} = \frac{158}{166.8} = +\,0.95$$

There is high degree of positive correlation between age group and rate of mortality.

**3.** From the following data, calculate coefficient of correlation between age and playing habits.

| **Age Group** | 20–30 | 30–40 | 40–50 | 50–60 | 60–70 |
|---|---|---|---|---|---|
| **Number of Students** | 25 | 60 | 40 | 20 | 20 |
| **Number of Regular Players** | 10 | 30 | 12 | 2 | 1 |

***Ans.*** First, we shall find the percentage of regular players in the following way

**Calculation of Percentage of Regular Players**

| Number of Students | Number of Regular Players | Percentage of Regular Players |
|---|---|---|
| 25 | 10 | $\frac{10}{25} \times 100 = 40$ |
| 60 | 30 | $\frac{30}{60} \times 100 = 50$ |
| 40 | 12 | $\frac{12}{40} \times 100 = 30$ |
| 20 | 2 | $\frac{2}{20} \times 100 = 10$ |
| 20 | 1 | $\frac{1}{20} \times 100 = 5$ |

Denoting mid-value of age as **X** and percentage of regular players as **Y**.

**Calculation of Coefficient of Correlation**

| Age Group | Mid-value (X) | dx (X – A), A = 45 | $dx'\left(\frac{dx}{c_1}\right)$, $c_1 = 10$ | $dx'^2$ | Percentage of Regular Players (Y) | dy (Y – A), A = 30 | $dy'\left(\frac{dy}{c_2}\right)$, $c_2 = 5$ | $dy'^2$ | $dx'dy'$ |
|---|---|---|---|---|---|---|---|---|---|
| 20–30 | 25 | –20 | –2 | 4 | 40 | 10 | 2 | 4 | –4 |
| 30–40 | 35 | –10 | –1 | 1 | 50 | 20 | 4 | 16 | –4 |
| 40–50 | 45 | 0 | 0 | 0 | 30 | 0 | **0** | 0 | **0** |
| 50–60 | 55 | 10 | 1 | 1 | 10 | –20 | –4 | 16 | –4 |
| 60–70 | 65 | 20 | 2 | 4 | 5 | –25 | –5 | 25 | –10 |
| | | | $\Sigma dx' = 0$ | $\Sigma dx'^2 = 10$ | | | $\Sigma dy' = -3$ | $\Sigma dy'^2 = 61$ | $\Sigma dx'dy' = -22$ |

Here, $n = 5, \Sigma dx' = 0, \Sigma dx'^2 = 10, \Sigma dy' = -3, \Sigma dy'^2 = 61$

and $\Sigma dx'dy' = -22$

$$\therefore \qquad r = \frac{\Sigma dx'dy' - \frac{\Sigma dx' \times \Sigma dy'}{n}}{\sqrt{\Sigma dx'^2 - \frac{(\Sigma dx')^2}{n}} \times \sqrt{\Sigma dy'^2 - \frac{(\Sigma dy')^2}{n}}} = \frac{-22 - \frac{0 \times -3}{5}}{\sqrt{10 - \frac{(0)^2}{5}} \times \sqrt{61 - \frac{(-3)^2}{5}}}$$

$$= \frac{-22}{\sqrt{10} \times \sqrt{61 - 1.8}} = \frac{-22}{\sqrt{10} \times \sqrt{59.2}} = \frac{-22}{3.16 \times 7.69} = \frac{-22}{24.3} = -0.90$$

It indicates that there is a high degree of negative correlation between age and playing habits. It shows that as age increases, the tendency to play decreases.

**4.** From the data given below, calculate Karl Pearson's coefficient of correlation between density of population and death rate by step deviation method.

| Region | **Area** (in sq km) | Population | Death |
|---|---|---|---|
| A | 200 | 40,000 | 480 |
| B | 150 | 75,000 | 1,200 |
| C | 120 | 72,000 | 1,080 |
| D | 80 | 20,000 | 280 |

***Ans.*** First of all, we shall compute density of population i.e., population per sq km and death rate per 1,000.

Density of Population $= \frac{\textbf{Population}}{\textbf{Area}}$, Death Rate $= \frac{\textbf{Number of Deaths}}{\textbf{Population}} \times \mathbf{1{,}000}$

**Calculation of Coefficient of Correlation**

| Region | Density (X) | dx (X – A), A = 500 | $dx'\left(\frac{dx}{c_1}\right)$, $c_1 = 50$ | $dx'^2$ | Death Rate (Y) | dy (Y – A), A = 16 | $dy'\left(\frac{dy}{c_2}\right)$, $c_2 = 1$ | $dy'^2$ | $dx'dy'$ |
|---|---|---|---|---|---|---|---|---|---|
| A | 200 | –300 | –6 | 36 | 12 | –4 | –4 | 16 | 24 |
| B | 500 | 0 | 0 | 0 | 16 | 0 | 0 | 0 | 0 |
| C | 600 | 100 | 2 | 4 | 15 | –1 | –1 | 1 | –2 |
| D | 250 | –250 | –5 | 25 | 14 | –2 | –2 | 4 | 10 |
| | | | $\Sigma dx' = -9$ | $\Sigma dx'^2 = 65$ | | | $\Sigma dy' = -7$ | $\Sigma dy'^2 = 21$ | $\Sigma dx'dy' = 32$ |

Here, $dx' = -9, \Sigma dx'^2 = 65, \Sigma dy' = -7, \Sigma dy'^2 = 21$ and $\Sigma dx'dy' = 32$

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

$$= \frac{32 - \frac{(-9 \times -7)}{4}}{\sqrt{65 - \frac{(-9)^2}{4}} \times \sqrt{21 - \frac{(-7)^2}{4}}} = \frac{32 - 15.75}{\sqrt{65 - 20.25} \times \sqrt{21 - 12.25}}$$

$$= \frac{16.25}{\sqrt{44.75} \times \sqrt{8.75}} = \frac{16.25}{6.69 \times 2.96} = \frac{16.25}{19.80} = 0.82$$

There is high degree of positive correlation between density of population and death rate.

**5.** Calculate coefficient of correlation between the price and quantity supplied (using short-cut method)

| **Price** (₹) | 4 | 6 | 7 | 12 | 20 |
|---|---|---|---|---|---|
| **Supply** (kg) | 6 | 12 | 18 | 20 | 24 |

*Ans.*

**Calculation of Coefficient of Correlation**

| Price (X) | Deviation ($dx = X - A$) $A = 12$ | Square of Deviation ($dx^2$) | Supply (Y) | Deviation ($dy = Y - A$) $A = 18$ | Square of Deviation ($dy^2$) | Multiple of deviations (dxdy) |
|---|---|---|---|---|---|---|
| 4 | –8 | 64 | 8 | –10 | 100 | 80 |
| 6 | –6 | 36 | 12 | –6 | 36 | 36 |
| 7 | –5 | 25 | 18 | 0 | 0 | 0 |
| 12 | 0 | 0 | 20 | 2 | 4 | 0 |
| 20 | 8 | 64 | 24 | 6 | 36 | 48 |
| $N = 5$ | $\Sigma dx = -11$ | $\Sigma dx^2 = 189$ | $N = 5$ | $\Sigma dx = -8$ | $\Sigma dy^2 = 176$ | $\Sigma dxdy = 164$ |

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

$$= \frac{164 - \frac{(-11) \times (-8)}{5}}{\sqrt{189 - \frac{121}{5}} \times \sqrt{176 - \frac{64}{5}}}$$

$$= \frac{164 - \frac{88}{5}}{\sqrt{189 - 24.2} \times \sqrt{176 - 12.8}}$$

$$= \frac{164 - 17.6}{\sqrt{164.8} \times \sqrt{163.2}}$$

$$= \frac{146.4}{12.84 \times 12.77} = \frac{146.4}{163.97} = 0.89$$

# Chapter Test

## Multiple Choice Questions

**1.** Which of the following represents cyclic causation?

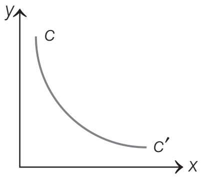
(i) $L \text{ Causes} \begin{matrix} \nearrow M \\ \searrow N \end{matrix}$ (ii) $M \xrightarrow{\text{Causes}} L \xrightarrow{\text{Causes}} N$ (iii) $M \xrightarrow{\text{Causes}} N \xrightarrow{\text{Causes}} M$

**Alternatives**

(a) Only (i) (b) Only (ii) (c) Only (iii) (d) None of these

**2.** The curve *cc′* depicts

(a) linear positive correlation
(b) non-linear positive correlation
(c) linear negative positive correlation
(d) non-linear negative correlation




**3.** The value of correlation coefficient of two variables 'alpha' and 'beta' has been computed as 0.39. What does this value convey?

(a) High degree of positive correlation
(b) Moderate degree of positive correlation
(c) Low degree of positive correlation
(d) Very low degree of positive correlation

**4.** The techniques which provide the decision maker a systematic and powerful means of analysis to explore policies for achieving predetermined goals are called......... .

(a) Correlation techniques
(b) Mathematical techniques
(c) Quantitative techniques
(d) None of these

## Short Answers (SA) Type Questions

**1.** "Correlation is preferred to covariance as a measure of association." Explain.

**2.** Calculate the coefficient of correlation from the following data $\Sigma xy = 4{,}880$, $\sigma_x = 28.70$, $\sigma_Y = 18.02$, $n = 10$

**3.** Compute coefficient of correlation from the following data

| | X Series | Y Series |
|---|---|---|
| Mean | 15 | 28 |
| Sum of Squares of Deviation from Mean | 144 | 225 |

Sum of products of deviation of *X* and *Y* series from their respective mean is 20. Number of pairs of observations is 10.

**4.** Does correlation imply causation?

**5.** What are the properties of Karl Pearson's coefficient of correlation?

## Long Answers (LA) Type Questions

**1.** Calculate coefficient of correlation between the *X* and *Y* variables.

| *X* | 43 | 48 | 56 | 64 | 67 | 70 |
|---|---|---|---|---|---|---|
| *Y* | 128 | 120 | 138 | 143 | 141 | 152 |

**2.** From the following data, calculate coefficient of correlation between age and playing habits.

| Age Group | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|
| Number of Students | 25 | 60 | 40 | 20 | 20 |
| Number of Regular Players | 10 | 30 | 12 | 2 | 1 |

## Answers

### Multiple Choice Questions

***1.* (c)** ***2.* (d)** ***3.* (c)** ***4.* (c)**