| | | | | | |
|---|---|---|---|---|---|
| 65 | 53 | 0 | | | |
| 62 | 58 | – 3 | – 8 | 9 | 64 |
| 90 | 86 | 25 | 20 | 625 | 400 |
| 82 | 62 | 17 | – 4 | 289 | 16 |
| 75 | 68 | 10 | 2 | 100 | 4 |
| 25 | 60 | – 40 | – 6 | 1600 | 36 |
| 98 | 91 | 33 | 25 | 1089 | 625 |
| 36 | 51 | – 29 | – 15 | 841 | 225 |
| 78 | 84 | 13 | 18 | 169 | 324 |
| $\Sigma x = 650$ | $\Sigma y = 660$ | $\Sigma dx = 0$ | $\Sigma dy = 0$ | $\Sigma dx^2 = 5398$ | $\Sigma dy^2 = 2224$ |

$$\bar{x} = \frac{\Sigma x}{n} = \frac{650}{10} = 65 \;;\; \bar{y} = \frac{\Sigma y}{n} = \frac{660}{10} = 66 \;; \quad \therefore \quad dx = x - \bar{x} = x - 65 \;;\; dy = y - \bar{y} =$$

$$r = \frac{\Sigma dxdy}{\sqrt{\Sigma dx^2 . \Sigma dy^2}} = \frac{2704}{\sqrt{5398 \times 2224}} = \frac{2704}{\sqrt{12005152}} = \frac{2704}{3464 \cdot 8451} = 0 \cdot 7804$$

**Aliter.**

$$\Rightarrow \qquad \log r = \log 2704 - \tfrac{1}{2}[\log 5398 + \log 2224]$$

$$= 3 \cdot 4320 - \tfrac{1}{2}(3 \cdot 7325 + 3 \cdot 3472) = 3 \cdot 4320 - 3 \cdot 53985 = -0 \cdot 10785 = \bar{1} \cdot 89215$$

$$\Rightarrow \qquad r = \text{Antilog}(\bar{1} \cdot 89215) = 0 \cdot 7802$$

Hence, there is a fairly high degree of positive correlation between expenditure on advertisement and sales. We may, therefore, conclude that in general, sales have increased with an increase in the advertisement expenses.

**Example 8·3.** *From the following table calculate the coefficient of correlation by Karl Pearson's method.*

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

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

**Solution.** First of all we shall find the missing value of $Y$. Let the missing value in $Y$-series be $a$. Then mean $Y$ is given by :

$$\overline{Y} = \frac{\sum Y}{n} = \frac{9 + 11 + a + 8 + 7}{5} = \frac{35 + a}{5} = 8 \text{ (given)}$$

$$35 + a = 5 \times 8 = 40 \quad \Rightarrow \quad a = 40 - 35 = 5$$

**CALCULATION OF CORRELATION COEFFICIENT**

| $X$ | $Y$ | $X - \overline{X} = X - 6$ | $(Y - \overline{Y}) = Y - 8$ | $(X - \overline{X})^2$ | $(Y - \overline{Y})^2$ | $(X - \overline{X})(Y - \overline{Y})$ |
|---|---|---|---|---|---|---|
| [illegible] | 9 | 0 | 1 | 0 | 1 | 0 |
| [illegible] | 11 | $-4$ | 3 | 16 | 9 | $-12$ |
| [illegible] | 5 | 4 | $-3$ | 16 | 9 | $-12$ |
| [illegible] | 8 | $-2$ | 0 | 4 | 0 | 0 |
| [illegible] | 7 | 2 | $-1$ | 4 | 1 | $-2$ |
| $\Sigma X = 30$ | $\Sigma Y = 40$ | 0 | 0 | $\Sigma(X - \overline{X})^2 = 40$ | $\Sigma(Y - \overline{Y})^2 = 20$ | $\Sigma(X - \overline{X})(Y - \overline{Y}) = -26$ |

$$\overline{X} = \frac{\Sigma X}{5} = \frac{30}{5} = 6, \qquad \overline{Y} = \frac{\Sigma Y}{5} = \frac{40}{5} = 8$$

Karl Pearson's correlation coefficient is given by :

$$r = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y} = \frac{\Sigma(X - \overline{X})(Y - \overline{Y})}{\sqrt{\Sigma(X - \overline{X})^2 \, \Sigma(Y - \overline{Y})^2}} = \frac{-26}{\sqrt{40 \times 20}} = \frac{-26}{\sqrt{800}} = \frac{-26}{28{\cdot}2843} = -0{\cdot}9192 \simeq -0{\cdot}92$$

**Example 8·4.** *Calculate the coefficient of correlation between X and Y series from the following data :*

| | *Series X* | *Series Y* |
|---|---|---|
| *No. of pairs of observations* | *15* | *15* |
| *Arithmetic mean* | *25* | *18* |
| *Standard deviation* | *3·01* | *3·03* |
| *Sum of squares of deviations from mean* | *136* | *138* |

*Summation of product deviations of X and Y series from their respective arithmetic means = 122.*

**Solution.** In the usual notations, we are given :

$$n = 15, \quad \bar{x} = 25, \quad \bar{y} = 18, \quad \sigma_x = 3{\cdot}01, \quad \sigma_y = 3{\cdot}03$$

$$\Sigma(x - \bar{x})^2 = 136, \quad \Sigma(y - \bar{y})^2 = 138, \quad \text{and} \quad \Sigma(x - \bar{x})(y - \bar{y}) = 122.$$

Karl Pearson's correlation coefficient between $X$-series and $Y$-series is given by

$$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{n \, \sigma_x \sigma_y} = \frac{122}{15 \times 3{\cdot}01 \times 3{\cdot}03} = \frac{122}{136{\cdot}8049} = 0{\cdot}8918$$

**Remark.** Here some of the given data are superfluous *viz.*, $\bar{x}$, $\bar{y}$, $\Sigma(x - \bar{x})^2$, $\Sigma(y - \bar{y})^2$.

**Aliter.** We may also compute the correlation coefficient using the formula

$$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{\Sigma(x - \bar{x})^2 \, \Sigma(y - \bar{y})^2}} = \frac{122}{\sqrt{136 \times 138}} = \frac{122}{\sqrt{18768}} = \frac{122}{136{\cdot}9964} = 0{\cdot}8905$$

If we use this formula, then the data relating to $n$, $\bar{x}$, $\bar{y}$, $\sigma_x$ and $\sigma_y$ are superfluous.

**Example 8·5.** *The coefficient of correlation between two variables X and Y is 0·48. The covariance is 36. The variance of X is 16. Find the standard deviation of Y.*

**Solution.** We are given :

$$r_{xy} = 0{\cdot}48, \ \text{Cov}(X, Y) = 36, \ \sigma_x^2 = 16 \quad \Rightarrow \quad \sigma_x = 4$$

We have :

$$r_{xy} = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y} \quad \Rightarrow \quad \sigma_y = \frac{\text{Cov}(X, Y)}{\sigma_x \, r_{xy}} = \frac{36}{4 \times 0{\cdot}48} = \frac{9}{0{\cdot}48} = 18{\cdot}75.$$

other.

However, if in any ... correlation is termed as *chance correlation* ...

[Also see § 8·1·2.]

**8·4·3. Interpretation of $r$.** The following general points may be borne in mind while interpreting an observed value of correlation coefficient $r$ :

(*i*) $r = +1$ implies that there is perfect positive correlation between the variables. In other words, scatter diagram will be a straight line starting from left bottom and rising upwards to the right top as in Fig. 8·1, § 8·3.

(*ii*) If $r = -1$, there is perfect negative correlation between the variables. In this case scatter diagram will again be a straight line as shown in Fig. 8·1, § 8·3.

(*iii*) If $r = 0$, the variables are uncorrelated. In other words, there is no linear (straight line) relationship between the variables. However, $r = 0$ does not imply that the variables are independent [*c.f.* Property ... page 8·11].

(*iv*) For other values of $r$ lying between $+1$ and $-1$, there are no set guidelines for its interpretation. The maximum we can conclude is that nearer is the value of $r$ to 1, the closer is the relation between the variables and nearer is the value of $r$ to 0, the less close is the relationship between them. One should be very careful in interpreting the value of $r$ as it is often mis-interpreted.

(*v*) The reliability or the significance of the value of the correlation coefficient depends on a number of factors. One of the ways of testing the significance of $r$ is finding its probable error [*c.f.* § 8·5], which in addition to the value of $r$ takes into account the size of the sample also.

(*vi*) Another more useful measure for interpreting the value of $r$ is the coefficient of determination [*c.f.* § 8·9]. It is observed there that *the closeness of the relationship between two variables is not proportional to r.*

## 8·5. PROBABLE ERROR

After computing the value of the correlation coefficient, the next step is to find the extent to which it is dependable. Probable error of correlation coefficient, usually denoted by $P.E.(r)$ is an old measure of testing the reliability of an observed value of correlation coefficient in so far as it depends upon the conditions of random sampling.

If $r$ is the observed correlation coefficient in a sample of $n$ pairs of observations then its standard error, usually denoted by $S.E.\ (r)$ is given by :

$$S.E.(r) = \frac{1 - r^2}{\sqrt{n}}$$

Probable error of the correlation coefficient is given by :

$$P.E.\ (r) = 0{\cdot}6745 \times S.E.\ (r) = 0{\cdot}6745\ \frac{(1 - r^2)}{\sqrt{n}}$$

The reason for taking the factor 0·6745 is that in a normal distribution 50% of the observations lie in the range $\mu \pm 0{\cdot}6745\ \sigma$, where $\mu$ is the mean and $\sigma$ is the *s.d.*

*According to Secrist, "The probable error of the correlation coefficient is an amount which if added to and subtracted from the mean correlation coefficient, produces amounts within which the chances are even that a coefficient of correlation from a series selected at random will fall."*

... in that year

8. (a) ... correlation coe... (i) $r = +1$, (ii) $r = -1$ and (iii) $r = 0$, where $r$ is the coefficient of correlation.
[I.C.W.A. (Intermediate), Dec. 2001]

(b) What is a scatter diagram ? Give the procedure of drawing a scatter diagram. Draw scatter diagrams when the coefficient of correlation $r = +1$ and $r = -1$.
[C.A. (Foundation), May 2000]

9. The production manager of a company maintains that the flow time in days ($y$), depends on the number of operations ($x$) to be performed. The following data give the necessary information :

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $x$ : | 2 | 2 | 3 | 4 | 4 | 5 | 6 | 6 | 7 | 7 |
| $y$ : | 8 | 13 | 14 | 11 | 20 | 10 | 22 | 26 | 22 | 25 |

Plot a scatter diagram. Calculate the value of the Karl Pearson's Product Moment Correlation Coefficient .
[I.C.W.A. (Intermediate), Dec. 1995]

**Ans.** $r(x, y) = 0{\cdot}78$.

10. Making use of the data given below, calculate the coefficient of correlation $r_{12}$

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Case* : | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* |
| $X_1$ : | 10 | 6 | 9 | 10 | 12 | 13 | 11 | 9 |
| $X_2$ : | 9 | 4 | 6 | 9 | 11 | 13 | 8 | 4 |

**Ans.** $r_{12} = 0{\cdot}8958$.

11. Calculate Karl Pearson's coefficient of correlation from the following data, using 20 as the working mean for price and 70 as the working mean for demand :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Price* : | 14 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| *Demand* : | 84 | 78 | 70 | 75 | 66 | 67 | 62 | 58 | 60 |

[Delhi Univ. B.Com. (Pass), 1999]

**Ans.** $r = -0{\cdot}954$.

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

| *No.* | *Subject* | *Percentage of Marks* First Term | Second Term | *No.* | *Subject* | *Percentage of Marks* First Term | Second Term |
|---|---|---|---|---|---|---|---|
| 1. | Hindi | 75 | 62 | 5. | Commerce | 77 | 69 |
| 2. | English | 81 | 68 | 6. | Mathematics | 81 | 72 |
| 3. | Economics | 70 | 65 | 7. | Statistics | 84 | 76 |
| 4. | Accounts | 76 | 60 | 8. | Costing | 75 | 72 |

[Delhi Univ. B.Com. (Pass), 2000]

**Ans.** $r = 0{\cdot}623$.

**13.** Calculate the Karl Pearson's coefficient of correlation for the following ages of husbands and wives at the time of their marriage

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Age of husband (in years)* | 23 | 27 | 28 | 28 | 28 | 30 | 30 | 33 | 35 |
| *Age of wife (in years)* | 18 | 20 | 22 | 27 | 21 | 29 | 27 | 29 | 28 |

**Ans.** $r = 0.8013$.

**14.** Calculate the Pearson's coefficient of correlation from the following data using 44 and 26 respectively as the origin of $X$ and $Y$:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $X$: | 43 | 44 | 46 | 40 | 44 | 42 | 45 | 42 | 38 | 40 | 42 | 57 |
| $Y$: | 29 | 31 | 19 | 18 | 19 | 27 | 27 | 29 | 41 | 30 | 26 | 10 |

[Osmania Univ. B.Com.]

**Ans.** $r_{xy} = -0.7326$.

**15.** The following table gives the distribution of the total population and those who are totally or partially blind among them. Find out if there is any relation between age and blindness.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Age (Years)* | : | 0—10 | 10—20 | 20—30 | 30—40 | 40—50 | 50—60 | 60—70 |
| *No. of Persons ('000)* | : | 100 | 60 | 40 | 36 | 24 | 11 | 6 |
| *Blind* | : | 55 | 40 | 40 | 40 | 36 | 22 | 18 |

**Hint.** Here we shall find the correlation coefficient between age ($X$) and the number of blinds per lakh given in the following table.

| $X$ | 5 | 15 | 25 | 35 | 45 | 55 | 65 | 75 |
|---|---|---|---|---|---|---|---|---|
| $Y$ | 55 | 67 | 100 | 111 | 150 | 200 | 300 | 500 |

**Ans.** $r = 0.8982$.

**16.** With the following data in 6 cities, calculate the coefficient of correlation by Pearson's method between density of population and the death rate.

| *Cities* | *Area in square miles* | *Population (in '000)* | *No. of deaths* |
|---|---|---|---|
| A | 150 | 30 | 300 |
| B | 180 | 90 | 1440 |
| C | 100 | 40 | 560 |
| D | 60 | 42 | 840 |
| E | 120 | 72 | 1224 |
| F | 80 | 24 | 312 |

[C.A. (Intermediate), May [illegible]]

**Hint.** Find $r$ between, Density $= \dfrac{\text{Population}}{\text{Area}}$ ; and Death Rate $= \dfrac{\text{No. of deaths}}{\text{Population}} \times 1000$.

**Ans.** $r = 0.9876$.

**17.** Calculate the correlation coefficient from the following data :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $X$: | 12 | 9 | 8 | 10 | 11 | 13 | 7 |
| $Y$: | 14 | 8 | 6 | 9 | 11 | 12 | 3 |

Let now each value of $X$ be multiplied by 2 and then 6 be added to it. Similarly multiply each value of $Y$ by 3 and subtract 2 from it. What will be the correlation coefficient between the new series of $X$ and $Y$.

[C.A. (Foundation), May [illegible]]

**Ans.** Let $U = 2X + 6$, $V = 3Y - 2$. Since correlation coefficient is independent of change of origin and scale.

$$r(U, V) = r(X, Y) = 0.9485.$$

**18.** (*a*) Given : $\Sigma X = 125$, $\Sigma Y = 100$, $\Sigma X^2 = 650$, $\Sigma Y^2 = 436$, $\Sigma XY = 520$ and $n = 25$, obtain the value of Karl Pearson's correlation coefficient $r(X, Y)$.

**Ans.** 0·67.

**19.** You are given the following information relating to a frequency distribution comprising of 10 observations

$\overline{X} = 5.5$, $\overline{Y} = 4.0$, $\Sigma X^2 = 385$, $\Sigma Y^2 = 192$; $\Sigma(X + Y)^2 = 947$.

Find $r_{xy}$. [Punjab Univ. B.Com.]

**Hint.** Use $\Sigma(X + Y)^2 = \Sigma X^2 + \Sigma Y^2 + 2\,\Sigma XY$ and find $\Sigma XY = 185$.

**Ans.** $r(X, Y) = -0.681$.

| | |
|---|---|
| 8 | 10 |
| 12 | 7 |

while the correct values were,

| | |
|---|---|
| 8 | 12 |
| 10 | 8 |

Obtain the correct value of the correlation coefficient between $X$ and $Y$. [I.C.W.A. Dec., 2003]

**Ans.** $r = 0{\cdot}0504$

**21.** Coefficient of correlation between $X$ and $Y$ for 20 items is 0·3; mean of $X$ is 15 and that of $Y$ 20, standard deviations are 4 and 5 respectively. At the time of calculations one pair ($x = 27$, $y = 30$) was wrongly taken as ($x = 17$, $y = 35$). Find the correct coefficient of correlation. [*Delhi Univ. B.Com. (Hons.), (External), 2007*]

**Ans.** Correct value of correlation coefficient = 0·5153.

**22.** In order to find the correlation coefficient between two variables $X$ and $Y$ from 12 pairs of observations, the following calculations were made :

$$\Sigma X = 30,\quad \Sigma Y = 5,\quad \Sigma X^2 = 670,\quad \Sigma Y^2 = 285,\quad \Sigma XY = 334$$

On subsequent verification it was found that the pair ($X = 11$, $Y = 4$) was copied wrongly, the correct value being ($X = 10$, $Y = 14$). Find the correct value of correlation coefficient.

**Ans.** 0·78.

**23.** What do you understand by the probable error of correlation coefficient ? Explain how it can be used to :

(*i*) Interpret the significance of an observed value of sample correlation coefficient.

(*ii*) Determine the limits for the population correlation coefficient.

**24.** Calculate the coefficient of correlation and find its probable error from the following data :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $X$ : | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
| $Y$ : | 18 | 16 | 14 | 12 | 10 | 6 | 8 |

**Ans.** $r_{xy} = 0{\cdot}9643$; $P.E.\ (r) = 0{\cdot}0179$.

**25.** Find Karl Pearson's correlation coefficient between age and playing habit of the following students :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *Age (years)* | : | 15 | 16 | 17 | 18 | 19 | 20 |
| *No. of students* | : | 250 | 200 | 150 | 120 | 100 | 80 |
| *Regular players* | : | 200 | 150 | 90 | 48 | 30 | 12 |

Also calculate the probable error and point out if coefficient of correlation is significant.

[*Himachal Pradesh Univ. M.B.A. 1998; Delhi Univ. B.Com. (Hons.), 1996*]

**Hint.** Find $r$ between age ($X$) and percentage of regular players ($Y$).

**Ans.** $r_{xy} = -0{\cdot}9912$ ; $P.E.(r) = 0{\cdot}0048$; $r$ is highly significant.

**26.** Calculate Karl Pearson's coefficient of correlation for the following series.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *Price (in Rs.)* | : | 110—111 | 111—112 | 112—113 | 113—114 | 114—115 | 115—116 |
| *Demand (in kg.)* | : | 600 | 640 | 640 | 680 | 700 | 780 |
| *Price (in Rs.)* | : | 116—117 | 117—118 | 118—119 | | | |
| *Demand (in kg.)* | : | 830 | 900 | 1,000 | | | |

Also calculate the probable error of the correlation coefficient. From your result can you assert that the demand is correlated with price ?

**Hint.** Find correlation coefficient between $x$ : Mid-value of price (in Rs.) and $y$ demand (in kg.)

Ans. ... 0·9651 ; $P.E.\ (r) = 0{\cdot}0154$.