Assuming that $r_{XY} \neq -1$, we get

$$\sigma_X + k\sigma_Y = 0$$

$$\therefore \quad k = -\frac{\sigma_X}{\sigma_Y}$$

Example 9

If (X, Y) is a two-dimensional RV uniformly distributed over the triangular region R bounded by $y = 0$, $x = 3$ and $y = 4/3\ x$. Find $f_X(x)$, $f_Y(y)$, $E(X)$, $\mathrm{Var}(X)$, $E(Y)$, $\mathrm{Var}(Y)$ and ρ_{XY}.

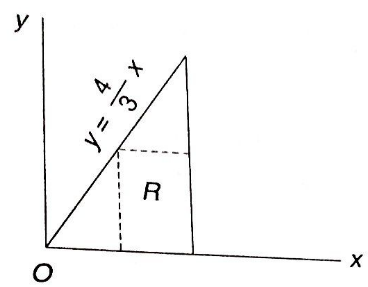

Fig. 4.3

Since (X, Y) is uniformly distributed, $f(x, y) =$ a constant $= k$

Now $$\iint f(x, y)\, dx\, dy = 1$$

i.e., $$\int_0^4 \int_{3y/4}^3 k\, dx\, dy = 1$$

i.e., $$k \int_0^4 \left(3 - \frac{3y}{4}\right) dy = 1$$

i.e., $$6k = 1$$

$$\therefore \quad k = \frac{1}{6}$$

$$f_Y(y) = \int_{3y/4}^{3} \frac{1}{6} dx = \frac{1}{8}(4 - y),\ 0 < y < 4$$

$$f_X(x) = \int_0^{4x/3} \frac{1}{6}\, dy = \frac{2}{9}x,\ 0 < x < 3$$

$$E(X) = \int x f_X(x)\, dx = \int_0^3 \frac{2}{9}x^2\, dx = 2$$

$$E(Y) = \int y f_Y(y)\, dy = \int_0^4 \frac{y}{8} \times (4 - y)\, dy = \frac{4}{3}$$

$$E(X^2) = \int_0^3 \frac{2}{9} \times x^3\, dx = \frac{9}{2}$$

$$E(Y^2) = \int_0^4 \frac{y^2}{8} \times (4 - y)\, dy = \frac{8}{3}$$

$$\text{Var}(X) = E(X^2) - \{E(X)\}^2 = \frac{9}{2} - 4 = \frac{1}{2}$$

$$\text{Var}(Y) = E(Y^2) - \{E(Y)\}^2 = \frac{8}{3} - \frac{16}{9} = \frac{8}{9}$$

$$E(XY) = \int_0^4 \int_{3y/4}^{3} \frac{1}{6}\, xy\, dx\, dy$$

$$= \frac{3}{64} \int_0^4 (16 - y^2) y\, dy = 3$$

$$\rho_{XY} = \frac{E(XY) - E(X) \times E(Y)}{\sigma_x \sigma_y} = \frac{3 - 2 \times \frac{4}{3}}{\frac{1}{\sqrt{2}} \times 2\frac{\sqrt{2}}{3}} = \frac{1}{2}$$

Example 10

Find the correlation co-efficient between X and Y, which are jointly normal distributed with

$$f(x, y) = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1 - r^2}} \exp\left\{-\frac{1}{2(1 - r^2)}\left(\frac{x^2}{\sigma_x^2} - \frac{2rxy}{\sigma_x \sigma_y} + \frac{y^2}{\sigma_y^2}\right)\right\}$$

$$\frac{x^2}{\sigma_x^2} - \frac{2rxy}{\sigma_x \sigma_y} + \frac{y^2}{\sigma_y^2} = \left(\frac{x}{\sigma_x} - \frac{ry}{\sigma_y}\right)^2 + (1 - r^2)\frac{y^2}{\sigma_y^2}$$

$$= \frac{1}{\sigma_x^2}\left(x - \frac{ry\sigma x}{\sigma_y}\right)^2 + (1 - r^2)\frac{y^2}{\sigma_y^2}$$

$$E(XY) = \frac{1}{2\pi\sigma_x \sigma_y \sqrt{1 - r^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp$$

$$\left\{-\frac{1}{2(1 - r^2)}\left[\frac{1}{\sigma_x^2}\left(x - \frac{ry\sigma_x}{\sigma_y}\right)^2 + \frac{(1 - r)^2 y^2}{\sigma_y^2}\right]\right\} xy\, dx\, dy$$

$$= \frac{1}{\sigma_y \sqrt{2\pi}} \int_{-\infty}^{\infty} y \exp\left(\frac{-y^2}{2\sigma_y^2}\right) \int_{-\infty}^{\infty} \frac{x}{\left(\sigma_x \sqrt{1 - r^2}\right)\sqrt{2\pi}}$$

48. If X is a RV for which $E(X) = 10$ and $\text{Var}(X) = 25$, for what positive values of a and b does $Y = aX - b$ have expectation 0 and variance 1?
49. If X is uniformly distributed in $(1, 2)$ and $Y = X^3$, find the mean and variance of Y.
50. If the continuous RV X has the density function $f(x) = 2xe^{-x^2}$, $x \geq 0$, and if $Y = X^2$, find the mean and variance Y.
51. If X and Y are independent random variables with density functions $f_X(x) = \dfrac{8}{x^3}$, $x > 2$, and $f_Y(y) = 2y$, $0 < y < 1$, respectively, and $Z = XY$, find $E(Z)$.
52. If each of the independent RVs X and Y follows $N(0, \sigma)$ and $Z = |X - Y|$, prove that $E(Z) = 2\sigma/\sqrt{\pi}$ and $E(Z^2) = 2\sigma^2$.
53. If the joint pdf of (X, Y) is given by $f(x, y) = 2$, $0 \leq x < y \leq 1$, find the conditional mean and conditional variance of X, given that $Y = y$.
54. If the joint pdf of (X, Y) is given by $f(x, y) = 21x^2y^3$, $0 \leq x < y \leq 1$, find the conditional mean and variance of X, given that $Y = y$, $0 < y < 1$.
55. If the joint pdf of (X, Y) is given by $f(x, y) = 3xy(x + y)$, $0 \leq x, y \leq 1$, verify that $E\{E(Y/X)\} = E(Y) = \dfrac{17}{24}$.

LINEAR CORRELATION

In many situations, the outcome of a random experiment will have two measurable characteristics, viz., will result in two random variables X and Y. Often we will be interested in finding whether the two different R.V.'s are related to each other. If they are related, we will try to determine the nature of relationship and degree of relationship (correlation). Assuming that there is some correlation between X and Y, we will then try to find a formula expressing the relationship and use this formula to predict the most likely value of one R.V. corresponding to any given value of the other R.V.

To examine whether the two R.V.'s are inter-related, we collect n pairs of values of X and Y corresponding to n repetitions of the random experiment. Let them be $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$. Then we plot the points with co-ordinates $(x_1, y_1), \cdots, (x_n, y_n)$ on a graph paper. The simple figure consisting of the plotted points is called a *scatter diagram*. From the scatter diagram, we can form a fairly good, though vague, idea of the relationship between X and Y. If the points are dense or closely packed, we may conclude that X and Y are correlated. On the other hand if the points are widely scattered throughout the graph paper, we may conclude that X and Y are either not correlated or poorly correlated.

Further if the points in the scatter diagram appear to lie near a straight line, we assume that the R.V.'s have *linear correlation*. If they cluster round a well defined curve other than a straight line, the R.V.'s are assumed to be *non-linear*. In this section we will assume linear correlation between the concerned R.V.'s and discuss how to measure the degree of linear correlation.

CORRELATION COEFFICIENT

As the variance $E\{X - E(X)\}^2$ measures the variations of the R.V. X from its mean value $E(X)$, the quantity $E\{[X - E(X)][Y - E(Y)]\}$ measures the simultaneous variation of two R.V.'s X and Y from their respective means and hence it is called *the covariance of* X, Y and denoted as Cov (X, Y).

Cov $(X, Y) = E\{[X - E(X)][Y - E(Y)]\}$ is also called the *product moment of* X and Y and is also denoted as $p(X, Y)$.

Though $p(X, Y)$ is a useful measure of the degree of correlation between X and Y, it is to be expressed in mixed units of X and Y. To avoid this difficulty and to express the degree of correlation in absolute units, we divide $p(X, Y)$ by $\sigma_x \cdot \sigma_y$, so that $\dfrac{p(x, y)}{\sigma_x \sigma_y}$ is a mere number, free from the units of X and Y.

$\dfrac{p(x, y)}{\sigma_x \sigma_y}$ is a measure of intensity of linear relationship between X and Y and is called *Karl Pearson's Product Moment Correlation Coefficient* or simply *correlation coefficient* between X and Y. It is denoted by $r(X, Y)$ or r_{XY} or simply r.

Thus
$$r_{XY} = \frac{E\{[X - E(X)][Y - E(Y)]\}}{\sqrt{E\{X - E(X)\}^2 E\{Y - E(Y)\}^2}} \tag{1}$$

since σ_x, the standard deviation of X is the positive square root of the variance of X.

Now
$$E[\{X - E(X)\}\{Y - E(Y)\}]$$
$$= E[XY - E(Y) \cdot X - E(X) \cdot Y + E(X) \cdot E(Y)]$$
$$= E(XY) - E(Y) \cdot E(X) - E(X) \cdot E(Y) + E(X) \cdot E(Y)$$
[Θ $E(X)$ and $E(Y)$ are non-random constants]
$$= E\{XY) - E(X) \cdot E(Y) \tag{2}$$

Also we know that $E\{X - E(X)\}^2 = E(X^2) - \{E(X)\}^2$ (3)

and $E\{Y - E(Y)\}^2 = E\{Y^2\} - \{E(Y)\}^2$ (4)

Using (2), (3) and (4) in (1), we get

$$r_{XY} = \frac{E(XY) - E(X) \cdot E(Y)}{\sqrt{\{E(X^2) - E^2(X)\}\{E(Y^2) - E^2(Y)\}}} \tag{5}$$

where $E^2(X)$ means $\{E(X)\}^2$.

We will mainly deal with linear correlation of discrete R.V.'s X and Y. X will take the values $x_1, x_2, \cdots, x_n$ with frequency 1 each and Y will simultaneously take the values $y_1, y_2, \cdots, y_n$ with frequency 1 each. Hence $E(X) = \frac{1}{n}\sum x_i$; $E(X^2) = \frac{1}{n}\sum x_i^2$, $E(XY) = \frac{1}{n}\sum x_i y_i$ etc. Using these values in (5), the working formula for the compution of r_{XY} is got as

$$r_{XY} = \frac{\frac{1}{n}\Sigma x_i y_i - \frac{1}{n}\Sigma x_i \frac{1}{n}\Sigma y_i}{\sqrt{\left\{\frac{1}{n}\Sigma x_i^2 - \left(\frac{1}{n}\Sigma x_i\right)^2\right\}\left\{\frac{1}{n}\Sigma y_i^2 - \left(\frac{1}{n}\Sigma y_i\right)^2\right\}}} \quad (6)$$

or

$$r_{XY} = \frac{n\Sigma xy - \Sigma x \cdot \Sigma y}{\sqrt{\{n\Sigma x^2 - (\Sigma x)^2\}\{n\Sigma y^2 - (\Sigma y)^2\}}} \quad (7)$$

Properties of Correlation Coefficient

1. $-1 \le r_{XY} \le 1$ or $|\text{Cov}(X, Y)| \le \sigma_X \cdot \sigma_Y$.

Let us consider

$$E[a\{X - E(X)\} + \{Y - E(Y)\}]^2 = a^2\,\sigma_x^2 + 2a\,C_{XY} + \sigma_Y^2 \quad (1)$$

The R.H.S. expression is a quadratic expression in a, that is a real quantity. It is positive, as it is the expected value of a perfect square.

Hence, by the property of quadratic expressions, the discriminant of the R.H.S. ≤ 0

i.e., $$4\,C_{XY}^2 - 4\,\sigma_X^2\,\sigma_Y^2 \le 0$$

i.e., $$C_{XY}^2 \le \sigma_X^2 \cdot \sigma_Y^2 \quad (2)$$

i.e., $$\frac{C_{XY}^2}{\sigma_X^2 \cdot \sigma_Y^2} \le 1$$

i.e., $$r_{XY}^2 \le 1$$

$\therefore$ $$|r_{XY}| \le 1 \text{ or } -1 \le r_{XY} \le 1$$

From step (2), it is clear that $|C_{XY}| \le \sigma_X \cdot \sigma_Y$

Note: When $0 < r_{XY} \le 1$, the correlation between X and Y is said to be *positive* or direct.

When $-1 \le r_{XY} \le 0$, the correlation is said to be *negative* or *inverse*.

When $-1 \le r_{XY} \le -0.5$ or $0.5 \le r_{XY} \le 1$, the correlation is assumed to be high, otherwise the correlation is assumed to be poor.

2. Correlation coefficient is independent of change of origin and scale.

i.e., If $U = \frac{X-a}{h}$ and $V = \frac{Y-b}{k}$, where $h, k > 0$, then $r_{XY} = r_{UV}$.

By the transformations, $X = a + hU$ and $Y = b + kV$

$\therefore \quad E(X) = a + hE(U)$ and $E(Y) = b + kE(V)$

$\therefore \quad X - E(X) = h\{U - E(U)\}$ and $Y - E(Y) = k\{V - E(V)\}$

Then
$$C_{XY} = E[h\{U - E(U)\} \cdot k\{V - E(V)\}] = hk\, C_{UV}$$
$$\sigma_X^2 = E[h^2\{U - E(U)\}^2] = h^2 \sigma_U^2$$
$$\sigma_Y^2 = E[k^2\{V - E(V)\}^2] = k^2 \sigma_V^2$$

$$\therefore \quad r_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\sigma_X^2 \cdot \sigma_Y^2}} = \frac{hk\, C_{UV}}{\sqrt{h^2 \cdot \sigma_U^2 \cdot k^2\, \sigma_V^2}}$$
$$= \frac{C_{UV}}{\sigma_U \cdot \sigma_V} = r_{UV}$$

Note *[If X and Y take considerably large values, computation of r_{XY} will become difficult. In such problems, we may introduce change of origin and scale and compute r using the above property.]*

3. Two independent R.V.'s X and Y are uncorrelated, but two uncorrelated R.V.'s need not be independent.

When X and Y are independent, $E(XY) = E(X) \cdot E(Y)$.

$\therefore C_{XY} = 0$ and hence $r_{XY} = 0$

viz., X and Y are uncorrelated.

The converse is not true, since $E(XY) = E(X) \cdot E(Y)$, when $r_{XY} = 0$. This does not imply that X and Y are independent, as X and Y are independent only when $f(x, y) = f_X(x) \cdot f_Y(y)$.

Note *Note When* $E(xy) = 0$, *X and Y are said to be orthogonal* R.V.'s.

4. $$r_{XY} = \frac{\sigma_X^2 + \sigma_Y^2 - \sigma_{(X-Y)}^2}{2\sigma_X \sigma_Y}$$

Let $\quad Z = X - Y$. Then $E(Z) = E(X) - E(Y)$

$\therefore Z - E(Z) = [X - E(X)] - [Y - E(Y)]$

$$\therefore \quad \sigma_Z^2 = E[Z - E(Z)]^2 = E[\{X - E(X)\} - \{Y - E(Y)\}]^2$$
$$= E\{X - E(X)\}^2 + E\{Y - E(Y)\}^2 - 2E[\{X - E(X)\}\{Y - E(Y)\}]$$
$$= \sigma_X^2 + \sigma_Y^2 - 2C_{XY}$$

$$\therefore \quad C_{XY} = \frac{\sigma_X^2 + \sigma_Y^2 - \sigma_Z^2}{2}$$

$$\therefore \quad r_{XY} = \frac{C_{XY}}{\sigma_X \sigma_Y} = \frac{\sigma_X^2 + \sigma_Y^2 - \sigma_{(X-Y)}^2}{2\sigma_X \sigma_Y}$$

Similarly we can prove that

$$\sigma^2_{(X+Y)} = \sigma^2_X + \sigma^2_Y + 2C_{XY}$$

$$\text{and hence } r_{XY} = \frac{\sigma^2_{(X+Y)} - \sigma_X^2 - \sigma_Y^2}{2\sigma_X \sigma_Y}$$

Rank Correlation Coefficient

Sometimes the actual numerical values of X and Y may not be available, but the positions of the actual values arranged in order of merit (ranks) only may be available. The ranks of X and Y will in general, be different and hence may be considered as random variables. Let them be denoted by U and V. The correlation coefficient between U and V is called *the rank correlation coefficient* between (the ranks of) X, Y and denoted by ρ_{XY}.

Let us now derive a formula for ρ_{XY} or r_{UV}. Since U represents ranks of n values of X, U takes the values 1, 2, 3, $\cdots$, n.

Similarly V takes the same values 1, 2, 3, $\cdots$, n in a different order.

$$E(U) = E(V) = \frac{1}{n}(1 + 2 + \cdots + n) = \frac{n+1}{2}$$

$$E(U^2) = E(V^2) = \frac{1}{n}(1^2 + 2^2 + \cdots + n^2) = \frac{(n+1)(2n+1)}{6}$$

$$\therefore \quad \sigma_U^2 = \sigma_V^2 = E(U^2) - E^2(U)$$

$$= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}$$

$$= \frac{(n+1)}{12}\{2(2n+1) - 3(n+1)\}$$

$$= \frac{n^2 - 1}{12}$$

Let $\quad D = U - V \quad \therefore E(D) = 0$

and $\quad \sigma_D^2 = E(D^2)$

By property (4) given above,

$$\rho_{XY} = r_{UV} = \frac{\sigma_U^2 + \sigma_V^2 - \sigma_D^2}{2\sigma_U \sigma_V}, \text{ where } D = U - V$$

$$= \frac{\left(\frac{n^2-1}{6}\right) - \sigma_D^2}{2\left(\frac{n^2-1}{12}\right)}$$

$$= 1 - \frac{6}{n^2-1}\sigma_D^2 \text{ or } 1 - \frac{6E(D^2)}{n^2-1}$$

$$= 1 - \frac{6\,\Sigma d^2}{n(n^2-1)} \quad \left[\because E(D^2) = \frac{1}{n}\Sigma d^2\right]$$

[**Note:** The formula for the rank correlation coefficient is known as *spearman's formula*. The values of r_{XY} and ρ_{XY} (or r_{UV}) will be, in general, different.

2

Worked Example 4(B)

Example 1

Compute the coefficient of correlation between X and Y, using the following data:

X:	1	3	5	7	8	10
Y:	8	12	15	17	18	20

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	8	1	64	8
3	12	9	144	36
5	15	25	225	75
7	17	49	289	119
8	18	64	324	144
10	20	100	400	200
34	90	248	1446	582

Thus $n = 6$

$$\Sigma x_i = 34, \Sigma y_i = 90$$

$$\Sigma x_i^2 = 248, \Sigma y_i^2 = 1446$$

$$\Sigma x_i y_i = 582$$

$$r_{XY} = \frac{n\,\Sigma xy - \Sigma x \cdot \Sigma y}{\sqrt{\left\{n\Sigma x^2 - (\Sigma x)^2\right\}\left\{n\Sigma y^2 - (\Sigma y)^2\right\}}}$$

$$= \frac{6 \times 582 - 34 \times 90}{\sqrt{\{6 \times 248 - (34)^2\}\{6 \times 1446 - (90)^2\}}}$$

$$= \frac{432}{\sqrt{332 \times 576}} = 0.9879$$

Y A

Example 2

Compute the coefficients of correlation between X and Y using the following data:

X:	65	67	66	71	67	70	68	69
Y:	67	68	68	70	64	67	72	70

We effect change of origin in respect of both X and Y. The new origins are chosen at or near the average of extreme values. Thus we take $\frac{65+71}{2} = 68$ as the new origin for X and $\frac{64+72}{2} = 68$ as the new origin for Y. viz., we put $u_i = (x_i - 68)$ and $v_i = y_i - 68$ and find r_{UV}.

$X = x_i$	$Y = y_i$	$u_i = x_i - 68$	$v_i = y_i - 68$	u_i^2	v_i^2	$u_i v_i$
65	67	−3	−1	9	1	3
67	68	−1	0	1	0	0
66	68	−2	0	4	0	0
71	70	3	2	9	4	6
67	64	−1	−4	1	16	4
70	67	2	−1	4	1	−2
68	72	0	4	0	16	0
69	70	1	2	1	1	2
	Total	−1	2	29	39	13

$$r_{XY} = r_{UV} = \frac{n\Sigma uv - \Sigma u \cdot \Sigma v}{\sqrt{\{n\Sigma u^2 - (\Sigma u)^2\}\{n\Sigma v^2 - (\Sigma v)^2\}}}$$

$$= \frac{8 \times 13 - (-1) \times 2}{\sqrt{(8 \times 29 - 1)(8 \times 39 - 4)}} = \frac{106}{\sqrt{231 \times 308}} = 0.3974$$

Example 3

Find the coefficient of correlation between X and Y using the following data:

X:	5	10	15	20	25
Y:	16	19	23	26	30

As the values of X are in arithmetic progression, we make the change of origin and scale, by choosing the middle most value 15 as the new origin and the common difference 5 as the new scale.

i.e., we put $U = \dfrac{X - 15}{5}$

As the values of Y are not in A.P., we are content with effecting a change of origin only i.e., we put $V = Y - \left(\dfrac{30 + 16}{2}\right) = Y - 23$.

x	y	$u = \dfrac{x - 15}{5}$	$v = y - 23$	u^2	v^2	uv
5	16	−2	−7	4	49	14
10	19	−1	−4	1	16	4
15	23	0	0	0	0	0
20	26	1	3	1	9	3
25	30	2	7	4	49	14
	Total	0	−1	10	123	35

$$r_{XY} = r_{UV} = \frac{n\Sigma uv - \Sigma u \cdot \Sigma v}{\sqrt{\left\{n\Sigma u^2 - (\Sigma u)^2\right\}\left\{n\Sigma v^2 - (\Sigma v)^2\right\}}}$$

$$= \frac{5 \times 35 - 0 \times (-1)}{\sqrt{(5 \times 10 - 0)\,(5 \times 125 - 1)}}$$

$$= \frac{175}{\sqrt{50 \times 624}} = 0.9907$$

Example 4

The following table gives the bivariate frequency distribution of marks in an intelligence test obtained by 100 students according to their age:

Age (x) in yrs / *Marks (y)*	*18*	*19*	*20*	*21*	*Total*
10–20	4	2	2	–	8
20–30	5	4	6	4	19
30–40	6	8	10	11	35
40–50	4	4	6	8	22
50–60	–	2	4	4	10
60–70	–	2	3	1	6
Total	19	22	31	28	100

Calculate the coefficient of correlation between age and intelligence.

Since the frequencies of various values of x and y are not equal to 1 each the formula for the computation of r_{XY} is taken with a slight modification as given below:

$$\exp\left\{\frac{-\left(x-\dfrac{ry\sigma_x}{\sigma_y}\right)^2}{2\left(1-r^2\right)\sigma_x^2}\right\} dx\,dy \qquad (1)$$

The inner integral is the mean of the normal distribution with mean $\dfrac{ry\sigma_x}{\sigma_y}$ and variance $(1-r^2)\,\sigma_X^2$.

$\therefore$ the inner integral $= \dfrac{ry\sigma_x}{\sigma_y}$

Using this value in (1),

$$E(XY) = \left(\frac{r\sigma_x}{\sigma_y}\right) \times \frac{1}{\sigma_y\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 \exp\left(-\frac{y^2}{2\sigma_y^2}\right) dy$$

$$= \frac{r\sigma_x}{\sigma_y} E(Y^2) \text{ for } N(0, \sigma_Y)$$

$$= \frac{r\sigma_x}{\sigma_y} \times \sigma_Y^2$$

$$= r\sigma_X\sigma_Y$$

$$\therefore \qquad \rho_{XY} = \frac{E(XY) - E(X)E(y)}{\sigma_x\,\sigma_y} = r$$

Example 11

Ten students got the following percentage of marks in Mathematics and Physical sciences:

Students:	1	2	3	4	5	6	7	8	9	10
Marks in Mathematics:	78	36	98	25	75	82	90	62	65	39
Marks in Phy. Sciences:	84	51	91	60	68	62	86	58	63	47

Calculate the rank correlation coefficient.

Denoting the ranks in Mathematics and in Phy. Sciences by U and V, we have the following values of U and V:

U:	4	9	1	10	5	3	2	7	6	8	
V:	3	9	1	7	4	6	2	8	5	10	
D:	1	0	0	3	1	-3	0	-1	1	-2	
D^2:	1	0	0	9	1	9	0	1	1	4	: $\Sigma d^2 = 26$

$$r_{XY} = r_{UV} = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 26}{10 \times 99} = 0.8424$$

Example 12

Ten competitors in a beauty contest were ranked by three judges as follows:

Judges	Competitors 1	2	3	4	5	6	7	8	9	10
A:	6	5	3	10	2	4	9	7	8	1
B:	5	8	4	7	10	2	1	6	9	3
C:	4	9	8	1	2	3	10	5	7	6

Discuss which pair of judges have the nearest approach to common taste of beauty.

Rank by A (U)	Rank by B (V)	Rank by C (W)	$d_1 = U - V$	$d_2 = V - W$	$d_3 = U - W$	d_1^2	d_2^2	d_3^2
6	5	4	1	1	2	1	1	4
5	8	9	−3	−1	−4	9	1	16
3	4	8	−1	−4	−5	1	16	25
10	7	1	3	6	9	9	36	81
2	10	2	−8	8	0	64	64	0
4	2	3	2	−1	1	4	1	1
9	1	10	8	−9	−1	64	81	1
7	6	5	1	1	2	1	1	4
8	9	7	−1	2	1	1	4	1
1	3	6	−2	−3	−5	4	9	25
					Total:	157	214	158

$$r_{UV} = 1 - \frac{6\Sigma d_1^2}{n(n^2-1)} = 1 - \frac{6 \times 157}{10 \times 99} = 0.0485$$

$$r_{VW} = 1 - \frac{6\Sigma d_2^2}{n(n^2-1)} = 1 - \frac{6 \times 214}{10 \times 99} = -0.2970$$

$$r_{UW} = 1 - \frac{6\Sigma d_3^2}{n(n^2-1)} = 1 - \frac{6 \times 158}{10 \times 99} = 0.0424$$

Since r_{UV} is maximum, the judges A and B may be considered to have common taste of beauty to some extent compared to other pairs of judges.

Exercise 4(B)

Part A

(Short Answer Questions)

1. What is a scatter diagram? What is its role in correlation analysis?
2. What do you mean by correlation between two random variables?
3. What is linear correlation? How will you find that two R.V.'s are linearly correlated?
4. Define covariance of X, Y and coefficient of correlation between X and Y.
5. Why is r_{XY} preferred for measuring the degree of linear correlation to Cov (X, Y)?
6. State the properties of correlation coefficient.
7. State two different formulas used to compute r_{XY}.
8. Define rank correlation coefficient and write down the formula for computing it.
9. Prove that $-1 \le r_{XY} \le 1$.
10. Prove that $\sigma^2_{(X+Y)} - \sigma^2_{(X-Y)} = 4 \text{ Cov } (X, Y)$
11. If C_{XY} is the covariance of X and Y, prove that $C_{XY} = E(XY) - E(X) \cdot E(Y)$.
12. If X and Y are indejpendent R.V.'s prove that $r_{XY} = 0$. Is the converse true?
13. If X and Y are uncorrelated, prove that $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$
14. When are two R.V.s said to be orthogonal?

Part B

15. Ten students got the following marks in Mathematics and Basic Engineering:

Marks in Mathematics	78	36	98	25	75	82	90	62	65	39
Marks in Basic Engg.	84	51	91	60	68	62	86	58	53	47

Calculate the coefficient of correlation.

16. Calculate the correlation coefficient between X and Y from the following data:

X:	65	66	67	67	68	69	70	72
Y:	67	68	65	68	72	72	69	71

17. Find the coefficient of correlation between X and Y using the following data:

X:	5.5	3.6	2.6	3.4	3.1	2.7	3.0	3.1	3.2	3.8
Y:	27	36	39	39	32	35.	40	36	44	36

18. Compute the coefficient of correlation between X and Y from the following data:

X:	80	45	55	56	58	60	65	68	70	75	85
Y:	82	56	50	48	60	62	64	65	70	74	90