Illustration 48. Compute Karl Pearson's coefficient of correlation from the following data and comment on its value :

X :	17	19	23	35	40	42	48	54
Y :	02	13	11	24	13	50	18	37

(MBA, Univ. of Lucknow, 2009)

Solution. CALCULATION OF KARL PEARSON'S COEFFICIENT OF CORRELATION

X	$(X-34)$ d_x	d_x^2	Y	$(Y-21)$ d_y	d_y^2	d_xd_y
17	−17	289	2	−19	361	+323
19	−15	225	13	− 8	64	+120
23	−11	121	11	−10	100	+110
35	+ 1	1	24	+ 3	9	+ 3
40	+ 6	36	13	− 8	64	− 48
42	+ 8	64	50	+29	841	+232
48	+14	196	18	− 3	9	− 42
54	+20	400	37	+16	256	+320
$\Sigma X = 278$	$\Sigma d_x = 6$	$\Sigma d_x^2 = 1332$	$\Sigma Y = 168$	$\Sigma d_y = 0$	$\Sigma d_y^2 = 1704$	$\Sigma d_xd_y = 1018$

$$r = \frac{\Sigma d_x d_y - \frac{(\Sigma d_x)(\Sigma d_y)}{N}}{\sqrt{\Sigma d_x^2 - \frac{(\Sigma d_x)^2}{N}}\sqrt{\Sigma d_y^2 - \frac{(\Sigma d_y)^2}{N}}}$$

$$\Sigma d_x d_y = 1018,\ \Sigma d_x = 6,\ \Sigma d_y = 0$$

$$\Sigma d_x^2 = 1332,\ \Sigma d_y^2 = 1704,\ N = 8$$

Substituting the values
$$r = \frac{1018 - \frac{(6)(0)}{8}}{\sqrt{1332 - \frac{(6)^2}{8}}\sqrt{1704 - \frac{(0)^2}{8}}}$$

$$= \frac{1018 - 0}{\sqrt{1332 - 4.5}\sqrt{1704}} = \frac{1018}{\sqrt{1327.5} \times \sqrt{1704}}$$

$$= \frac{1018}{36.435 \times 41.28} = \frac{1018}{1504.037} = 0.677$$

There is a positive degree of moderate correlation between the variables X and Y.

Illustration 49. 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. *(MA Econ., Madras Univ., 2012)*

Solution. From the given values of Y we can find out missing value.

$$\bar{Y} = \frac{\Sigma Y}{N};\ 8 = \frac{\Sigma Y}{5} \text{ or } \Sigma Y = 40.$$

The given values are 9 + 11 + 8 + 7 = 35. Hence the missing value is 40 − 35 = 5

CALCULATION OF CORRELATION COEFFICIENT

X	$(X-6)$ x	x^2	Y	$(Y-8)$ 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
$\Sigma X = 30$	$\Sigma x = 0$	$\Sigma x^2 = 40$	$\Sigma Y = 40$	$\Sigma y = 0$	$\Sigma y^2 = 20$	$\Sigma xy = -26$

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

$$= \frac{-26}{\sqrt{40 \times 20}} = \frac{-26}{28.28} = -0.919$$

Illustration 50. You are given the following information relating to a frequency distribution comprising 10 observations :

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

Solution.

$$\Sigma (X+Y)^2 = 947$$
$$\Sigma X^2 + \Sigma Y^2 + 2\Sigma XY = 947$$
$$385 + 192 + 2\Sigma XY = 947$$
$$2\Sigma XY = 947 - 577 = 370$$
$$\Sigma XY = 185$$

$$\sigma_x = \sqrt{\frac{\Sigma X^2}{N} - (\overline{X})^2}$$

$$= \sqrt{\frac{385}{10} - (5.5)^2} = \sqrt{8.25} = 2.872$$

$$\sigma_y = \sqrt{\frac{\Sigma Y^2}{N} - (\overline{Y})^2}$$

$$= \sqrt{\frac{192}{10} - (4)^2} = \sqrt{3.2} = 1.789$$

$$r_{xy} = \frac{\frac{\Sigma XY}{N} - \overline{X}\,\overline{Y}}{\sigma_x \sigma_y}$$

$$= \frac{\frac{185}{10} - (5.5)(4)}{2.872 \times 1.789}$$

$$= \frac{18.5 - 22}{5.138} = \frac{-3.5}{5.138} = -0.681.$$

Illustration 51. Find Karl Pearson's coefficient of correlation from the following series of marks secured by 10 students in a class test in Mathematics and Statistics :

Marks in Mathematics	:	45	70	65	30	90	40	50	75	85	60
Marks in Statistics	:	35	90	70	40	95	40	60	80	80	50

Also calculate the probable error. *(MBA, Univ. of Lucknow, 2008)*

Solution. Let 'X' denote marks in mathematics and 'Y' marks in statistics.

CALCULATION OF PEARSON'S COEFFICIENT OF CORRELATION

X	$(X-\bar{X})$ x	x^2	Y	$(Y-\bar{Y})$ y	y^2	xy
45	−16	256	35	−29	841	464
70	+ 9	81	90	+26	676	234
65	+ 4	16	70	+ 6	36	24
30	−31	961	40	−24	576	744
90	+29	841	95	+31	961	899
40	−21	441	40	−24	576	504
50	−11	121	60	− 4	16	44
75	+14	196	80	+16	256	224
85	+24	576	80	+16	256	384
60	− 1	1	50	−14	196	14
$\Sigma X = 610$	$\Sigma x = 0$	$\Sigma x^2 = 3490$	$\Sigma Y = 640$	$\Sigma y = 0$	$\Sigma y^2 = 4390$	$\Sigma xy = 3535$

Since deviations are taken from actual mean of 'X' and 'Y', we apply the following formula for calculating correlation coefficient :

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

$$\Sigma xy = 3535,\ \Sigma x^2 = 3490,\ \Sigma y^2 = 4390$$

$$r = \frac{3535}{\sqrt{3490 \times 4390}}$$

Using Logarithms

$$\log r = \log 3535 - \frac{1}{2}[\log 3490 + \log 4390]$$

$$= 3.5484 - \frac{1}{2}[3.5428 + 3.6425]$$

$$= 3.5484 - \frac{1}{2}[7.1853]$$

$$= 3.5484 - 3.5926 = \bar{1}.9558$$

$$r = \text{AL } \bar{1}.9558 = 0.903$$

$$\text{P.E.}_r = 0.6745\,\frac{1-r^2}{\sqrt{N}}$$

$$= 0.6745\,\frac{1-(0.903)^2}{\sqrt{10}} = 0.6745\,\frac{0.185}{3.1623}$$

$$= \frac{0.1248}{3.1623} = 0.04.$$

Illustration 53. Calculate Karl Pearson's ...

LIST OF FORMULAE

1. Karl Pearson's Correlation Coefficient (when deviations are taken from actual means) :

$$r = \frac{\Sigma\, x\, y}{N\, \sigma_x\, \sigma_y} \text{ or } \frac{\Sigma\, x\, y}{\sqrt{\Sigma\, x^2 \times \Sigma\, y^2}}$$

where $x = (X - \bar{X})$: $y = (Y - \bar{Y})$

2. When deviations are taken from assumed mean :

$$r = \frac{N\, \Sigma\, d_x\, d_y - (\Sigma\, d_x)\, (\Sigma\, d_y)}{\sqrt{N\, \Sigma\, d_x^2 - (\Sigma\, d_x)^2}\, \sqrt{N\, \Sigma\, d_y^2 - (\Sigma\, d_y)^2}}$$

where $d_x = (X - A)$ and $d_y = (Y - A)$

3. In a bivariate frequency distribution :

$$r = \frac{N\, \Sigma\, f\, d_x\, d_y - (\Sigma\, f\, d_x)\, (\Sigma\, f\, d_y)}{\sqrt{N\, \Sigma\, f\, d_x^2 - (\Sigma\, f\, d_x)^2}\, \sqrt{N\, \Sigma\, f\, d_y^2 - (\Sigma\, f\, d_y)^2}}$$

4. When we use actual values of X and Y :

$$= \frac{\Sigma\, X\, Y - (\Sigma\, X)\, (\Sigma\, Y)}{\sqrt{N\, \Sigma\, X^2 - (\Sigma\, X)^2}\, \sqrt{N\, \Sigma\, Y^2 - (\Sigma\, Y)^2}}$$

5. Spearman's Rank Correlation Coefficient (When ranks are not repeated) :

$$R = 1 - \frac{6\, \Sigma\, D^2}{N^3 - N}$$

In case ranks are repeated

$$R = 1 - \frac{6\left(\Sigma\, D^2 + \frac{1}{12}\,(m_1^{\,3} - m_1) + \frac{1}{12}\,(m_2^{\,3} - m_2)\right)}{N^3 - N}$$