CORRELATION IN A BIVARIATE DATA

$$r = \frac{N\Sigma fdxdy - \Sigma fdx \cdot \Sigma fdy}{\sqrt{N\Sigma fdx^2 - (\Sigma fdx)^2}\sqrt{N\Sigma fdy^2 - (\Sigma fdy)^2}}$$

Example :- Calculate Karl Pearson's Correlation coeff. for the following data on family income (X) and percentage expenditure on food (Y) for 100 families.

Y \ X	200-300	300-400	400-500	500-600	600-700	Total
10-15	—	—	-	3	7	10
15-20	—	4	9	4	3	20
20-25	7	6	12	5	—	30
25-30	3	10	19	8	—	40
Total	10	20	40	20	10	100

In this table the variable X (Annual Income) and variable Y (% expenditure on food) is given in class intervals. Notice that the class Intervals should be equal for each variable. The no. of classes of X and that of Y need not be equal.

The row totals when added gives us 100 which is also equal to the column totals

The following table shows the solution. First take midpoints of each class.

COMPUTATION OF CORRELATION COEFFICIENT

← Annual Income (Rs'000) →

% Exp. on Food ↓ / C.I. →	Y	dX → / dY ↓	200 to 300	300 to 400	400 to 500	500 to 600	600 to 700	f	f·dy	f·dy²	f·dx·dy
% ↓	Y	dY ↓	−2	−1	0	+1	+2	↓	↓	↓	↓
10-15	12.5	−1	·			3 [-3]	7 [-14]	10	−10	10	−17
15-20	17.5	0	—	4 [0]	9 [0]	4 [0]	3 [0]	20	0	0	0
20-25	22.5	+1	7 [-14]	6 [-6]	12 [0]	5 [5]	—	30	30	30	−15
25-30	27.5	+2	3 [-12]	10 [-20]	19 [0]	8 [16]	—	40	80	160	−16
		f	10	20	40	20	10	N = 100	Σfdy = 100	Σfdy² = 200	Σfdxdy = −48
		$f d_x$	−20	−20	0	20	20	Σfdx = 0			
		$f d_x^2$	40	20	0	20	40	Σfdx^2 = 120			
		fdxdy	−26	−26	0	18	−14	Σfdxdy = −48			

Σfdxdy = Σfdxdy = −48

$$\rho = \frac{N\Sigma fdxdy - \Sigma fdx\,\Sigma fdy}{\sqrt{N\Sigma fdx^2 - (\Sigma fdx)^2}\sqrt{N\Sigma fdy^2 - (\Sigma fdy)^2}}$$

Substituting Values we get,

$$\rho = \frac{(100)(-48) - (0)(100)}{\sqrt{100(120) - (0)^2}\sqrt{100(200) - (100)^2}} = -0.438$$

Purple row → dX Purple column → dY Y Assumed Mean = 17.5

Frequency is given in each cell. X Assumed Mean = 450

Blue subcells in each cell represent f·dx·dy

for example In first row 1st blue cell & 2nd blue cell

dx is −1 and dy is +1 ∴ (3)(−1)(+1) = −3

dx is −1 and dy is +2 ∴ (7)(−1)(+2) = −14

Sum total of all blue cells = Σfdxdy = −48

Practice Questions :-

Q.1 Calculate Karl Pearson's coeff. of correlation :

Age → / Marks ↓	18	19	20	21	22	Total
20-25	3	2	–	–	–	5
15-20	–	5	4	–	–	9
10-15	–	–	7	10	–	17
5-10	–	–	–	3	2	5
0-5	–	–	–	–	4	4
Total	3	7	11	13	6	40

Answer $r = -0.903$

Q.2 Calculate r between X and Y

Y ↓ / X →	30-50	50-70	70-90	Total
0-5	10	6	2	18
5-10	3	5	4	12
10-15	4	7	9	20
Total	17	18	15	50

Answer $r = 0.376$

Tip to minimize calculations : Take assumed means of X and Y at different positions.
e.g. if Assumed mean of X is 3rd value, Taken assumed mean of Y as 2nd value [Refer above Solved example]