

Problems on Correlation

Example (1) The table below gives the marks of 8 students in a test in two

subjects A and B.									
Student		1	2	3	4	5	6	7	8
Marks in A		10	6	9	10	12	13	11	9
Marks in B		9	4	6	9	11	13	8	4

Calculate the coefficient of correlation between the marks obtained by students in these two subjects.

Ans:→ Let we represent marks in the subject A by x and in B by y . Then

x	y	$x - \bar{x} = X$	$y - \bar{y} = Y$	XY	X^2	Y^2
10	9	0	1	0	0	1
6	4	-4	-4	16	16	16
9	6	-1	-2	2	1	4
10	9	0	1	0	0	1
12	11	2	3	6	4	9
13	13	3	5	15	9	25
11	8	1	0	0	1	0
9	4	-1	-4	4	1	16
$\Sigma x = 80$ $\Sigma y = 64$		$\Sigma XY = 43$		$\Sigma X^2 = 32$	$\Sigma Y^2 = 72$	

$$\therefore \bar{x} = \frac{80}{8} = 10, \text{ and } \bar{y} = \frac{64}{8} = 8$$

$$\begin{aligned} \therefore \text{Coefficient of Correlation} &= \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}} \\ &= \frac{43}{\sqrt{32 \times 72}} \\ &= 0.896 \end{aligned}$$

Ans

Example (2) The ages of husbands and their wives are given below in the table:

Husband's age (x)	23	27	28	29	30
Wife's age (y)	18	22	23	24	25

Find the coefficient of correlation between x and y.

Ans: →

x	y	$d_x = x - 28$	$d_y = y - 23$	d_x^2	d_y^2	$d_x d_y$
23	18	-5	-5	25	25	25
27	22	-1	-1	1	1	1
28	23	0	0	0	0	0
29	24	1	1	1	1	1
30	25	2	2	4	4	4
		-3	-3	9	9	9

$$\therefore r = \frac{n \sum d_x d_y - \sum d_x \sum d_y}{\sqrt{\{n \sum d_x^2 - (\sum d_x)^2\} \{n \sum d_y^2 - (\sum d_y)^2\}}}$$

$$= \frac{5 \times 31 - (-3)(-3)}{\sqrt{\{5 \times 31 - (-3)^2\} \{5 \times 31 - (-3)^2\}}}$$

$$= \frac{155 - 9}{155 - 9}$$

$$= 1$$

Hence the ages of the husbands and wives are perfectly correlated. Ans

Example (3) Following table gives a bivariate distribution showing frequency of marks according to age by a group of 100 students in an intelligence test.

Marks	Age (in years)				Total
	18	19	20	21	
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

Ans:-

		u	-1	0	1	2				
$\sqrt{N}$	$\frac{1}{\sqrt{N}}$	Marks(x)	18	19	20	21	Total(f)	fu	fu ²	fu ³
			(8)	(0)	(4)					
-2	15	10-20	4	2	2		8	-16	32	4
			(5)	(0)	(6)	(-8)				
-1	25	20-30	5	4	6	4	19	-19	19	-9
			(0)	(0)	(0)	(0)				
0	35	30-40	6	8	10	11	35	0	0	0
			(4)	(0)	(6)	(16)				
1	45	40-50	4	4	6	8	22	22	22	18
				(0)	(8)	(16)				
2	55	50-60		2	4	4	10	20	40	24
				(0)	(9)	(6)				
3	65	60-70		2	3	1	6	18	54	15
		Total	19	22	31	28	100	25	167	52
		$\sum u$	-19	0	31	56	68	$\sum fu$	$\sum fu^2$	$\sum fu^3$
		$\sum u^2$	19	0	31	112	162	$\sum fu^2$		
		$\sum u^3$	9	0	13	30	52	$\sum fu^3$		

Here $u = x - 19$

and $v = \frac{y - 35}{10}$

Students, don't get Confused here

u and v are just d_x and d_y respectively.

Some of the writers use u & v where the some others use d_x & d_y .

You see here marks is x and the mid values of the intervals are y .

Here u 's are in top row and v 's are in first column.

The numbers in the boxes in the corners can be obtained by the product of u and v correspondingly then multiply it with the number given.

Ex: $\left. \begin{aligned} 8 &= (-2)(-1)4 \\ 0 &= (-2)0.2 \\ -4 &= (-2)1.2 \end{aligned} \right\} \text{ in the first row of boxes.}$

Similarly,

$\left. \begin{aligned} 5 &= (-1)(-1)5 \\ 0 &= (-1)0.4 \\ -6 &= (-1)1.6 \end{aligned} \right\} \text{ in the second row of boxes etc.}$

Similarly we fill all the boxes.

Now, the Coefficient of Correlation is

$$\begin{aligned} r &= \frac{n \sum f d_x d_y - \sum f d_x \sum f d_y}{\sqrt{\{n \sum f d_x^2 - (\sum f d_x)^2\} \{n \sum f d_y^2 - (\sum f d_y)^2\}}} \\ &= \frac{100 \times 52 - 68 \times 25}{\sqrt{(100 \times 162 - 68^2)(100 \times 167 - 25^2)}} \\ &= 0.25 \text{ Ans} \end{aligned}$$