CORRELATION PART-2

NOW LET US START WITH NEXT METHOD CORRELATION FROM ACTUAL MEAN:

- The formula is :

$$r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}}$$

where

$$X = x - \bar{x}$$

$$Y = y - \bar{y}$$

- r = correlation coefficient
- $\bar{X}$ = mean of X
- $\bar{Y}$ = mean of Y

LET US TAKE A EXAMPLE TO UNDERSTAND THIS

Question:

Find the following data X & Y positively or negatively correlated?

X	81	85	96	75	65	90	82	75
Y	95	96	99	82	85	60	57	75

- step 1 : Add values of X and Y
  data set x = { 81 + 85 + 96 + 75 + 65 + 90 + 82 + 75}= 649
  data set y = {95 + 96 + 99 + 82 + 85 + 60 + 57 + 75} = 649
- Total number of elements (n) = 8

  step 2 : Find $\bar{x}$ & $\bar{y}$
  $\bar{x}$ = 649/8
  $\bar{x}$ = 81.125
  $\bar{y}$ = 649/8
  $\bar{y}$ = 81.125

- Step 3: To find coefficient correlation follow the table below:

x	y	$X = x - \bar{x}$	$Y = y - \bar{y}$	X^2	Y^2	XY
81	95	-0.125	13.875	0.0156	192.5156	-1.7344
85	96	3.875	14.875	15.0156	221.2656	57.6406
96	99	14.875	17.875	221.2656	319.5156	265.8906
75	82	-6.125	0.875	37.5156	0.7656	-5.3594
65	85	-16.125	3.875	260.0156	15.0156	-62.4844
90	60	8.875	-21.125	78.7656	446.2656	-187.4844
82	57	0.875	-24.125	0.7656	582.0156	-21.1094
75	75	-6.125	-6.125	37.5156	37.5156	37.5156
$\sum x = 649$	$\sum y = 649$	$\sum X = 0$	$\sum Y = 0$	$\sum X^2 = 650.875$	$\sum Y^2 = 1814.875$	$\sum XY = 82.875$

- **Step 4: Substitute $\sum x$, $\sum y$, $\sum xy$, $\sum x^2$ & $\sum y^2$ value in the below correlation coefficient formula**

$$r = \frac{\sum XY}{\sum X^2 . \sum Y^2}$$

$\sum XY = 82.875$, $\sum X^2 = 650.875$, $\sum Y^2 = 1814.875$

$$r = \frac{82.875}{650.875 \times 1814.875}$$

step 5: Simplify above expression

$r = 0.0763$

Exercise: Practise similar questions in Book.

ASSUMED MEAN METHOD

- Formula is:

Assumed Mean Method:

$$d_x = X - A$$

$$d_y = Y - A$$

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

LET US TAKE AN EXAMPLE TO ELABORATE FURTHER:

Calculate Karl Pearson's correlation co-efficient bv the assumed mean method.

X	14	15	18	20	25	30
Y	40	45	65	28	30	40

Step 1: Let us assume mean of X series as 20 and of Y series as 45.

Step 2 : Calculate deviations of X series and denote as dX using formula
$dX = X - 20$
Similarly calculate deviations of Y series and denote as dY using formula
$dY = Y - 45$

X	Y	DX= X - 20	DY = Y - 45
14	40	14 - 20 = -6	40 -45 = -5
15	45	15 - 20 = -5	45 – 45 = 0
18	65	18 – 20 = -2	65 – 45 = 20
20	28	20 – 20 = 0	28 – 45 = -17
25	30	25 – 20 = 5	30 – 45 = -15
30	40	30 – 20 = 10	40 – 45 = -5
	TOTAL:	∑DX = 2	∑DY = - 22

Step 3: Multiply Dx with Dy and calculate ∑DXDY.

X	Y	DX= X - 20	DY =Y - 45	DX * DY
14	40	-6	-5	(-6) * (-5) = 30
15	45	-5	0	(-5) * (0) = 0
18	65	-2	20	(-2) * (20) = -40
20	28	0	-17	(0) * (-17) = 0
25	30	5	-15	(5) * (-15) = -75
30	40	10	-5	(10) * (-5) = -50
	TOTAL:	∑DX = 2	∑DY = -22	∑DXDY= -135

Step 4: Calculate dx^2 and dy^2 by squaring dx and dy and find total of each.

X	Y	DX= X - 20	DY =Y - 45	DX * DY	DX^2	DY^2
14	40	-6	-5	30	$(-6)^2 = 36$	$(-5)^2 = 25$
15	45	-5	0	0	$(-5)^2 = 25$	$(0)^2 = 0$
18	65	-2	20	-40	$(-2)^2 = 4$	$(20)^2 = 400$
20	28	0	-17	0	$(0)^2 = 0$	$(-17)^2 = 289$
25	30	5	-15	-75	$(5)^2 = 25$	$(15)^2 = 225$
30	40	10	-5	-50	$(10)^2 = 100$	$(-5)^2 = 25$
	TOTAL:	$\sum DX = 2$	$\sum DY = -22$	$\sum DXDY = -135$	$\sum DX^2 = 190$	$\sum DY^2 = 964$

- Step 5: Apply the formula

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

- So putting values from step 4 we get: (n=6)
- r =. $\frac{(6) \times (-135) - (2) \times (-22)}{\sqrt{6} \times 190 - (2)^2 \quad \sqrt{6} \times 964 - (-22)^2}$

$$r = \frac{-810 + 44.}{\sqrt{1136}\ \sqrt{5300}} =. \frac{-766}{33.70 \times 72.80} =. \frac{-766}{2453.36} = -0.312$$

- **Example 24 (Page 388)** : Find correlation between age and blindness using foll:

Age(yrs)	50-55	55-60	60-65	65-70	70-75	75-80
No of persons	25000	20000	15000	12000	10000	8000
No of blinds	200	150	90	48	30	12

- In this problem we have to find correlation between age n blindness. In such type of questions we have to convert no of blinds in ratio form. Let us learn how to do it.
- Out of 25000 persons, no of blinds = 200
- Out of 1 person number of blind =200/25000
- Out of 100 persons, number of blind= $\frac{200}{25000} * 100 = 0.80$

Unitary method

Similarly for other also.

Age	Mid point (X)	No of persons	Blind	No of blinds in ratio(Y)
50-55	52.5	25000	200	0.80
55-60	57.5	20000	150	0.75
60-65	62.5	15000	90	0.60
65-70	67.5	12000	48	0.40
70-75	72.5	10000	30	0.30
75-80	77.5	8000	12	0.15

- Now we have to work with only 2 variables i.e. age which we denote X and no of blinds in ratio which is Y.
- To find correlation we use the formula:

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

- So DX= X – 67.5 and DY = Y – 0.40. Also N=6

Mid point (X)	Dx	Dx^2	No of blinds in ratio(Y)	Dy	Dy^2	Dx Dy
52.5	-15	225	0.80	0.40	0.16	-6
57.5	-10	100	0.75	0.35	0.1225	-3.5
62.5	-5	25	0.60	0.20	0.04	-1
67.5	0	0	0.40	0	0	0
72.5	5	25	0.30	-0.10	0.01	-0.5
77.5	10	100	0.15	-0.25	0.0625	-2.5
	$\sum DX = -15$	$\sum DX^2 = 475$		$\sum DY = 0.60$	$\sum DY^2 = 0.395$	$\sum DXDY = -13.5$

- Substituting the values in above formula we get:
- $r = \dfrac{(6) \times (-13.5) - (-15)(0.60)}{\sqrt{6}\ X\ 475 - (-15)^2\ \sqrt{6}\ X\ 0.395 - (0.60)^2}$
- $r = \dfrac{-81 + 9}{\sqrt{2625}\ \sqrt{2.01}}$
- $r = \dfrac{-72}{51.23 \times 1.417}$
- $r = \dfrac{-72}{72.5929} = 0.9918$

- This shows high degree of positive correlation between age and blindness.