

CORRELATION PART-3

BIVARIATE FREQUENCY DISTRIBUTION

Question: Calculate Correlation between Marks in Statistics and Marks in Mathematics scored by 100 students:

Marks in Mathematics	Marks in Statistics				Total
	40-50	50-60	60-70	70-80	
50-55	4	7	5	2	18
55-60	6	10	7	4	27
60-65	6	12	10	7	35
65-70	3	8	6	3	20
Total	19	37	28	16	100

- Now here you can clearly see instead of 1 continuous frequency distribution, we have 2 continuous frequency distribution. To solve these type of questions we need to follow some steps. Let us Start:

- **Step 1**

- The variable indicated in rows is denoted as **X** and variable in columns is denoted as **Y**.
- In our example Marks in Statistics is in rows, so it is **X**. Similarly Marks in Mathematics is in columns, so it is **Y**.

- **Step 2**

- Find Mid points of both variables.
- Let us look at the table till here

Marks in Mathe Matics (y)			Marks in Statistics(x)				
			40-50	50-60	60-70	70-80	Total
	Mid-Point (Y)	Mid-Point (X) →	45	55	65	75	
50-55	52.5		4	7	5	2	18
55-60	57.5		6	10	7	4	27
60-65	62.5		6	12	10	7	35
65-70	67.5		3	8	6	3	20
Total			19	37	28	16	100

- **Step 3**

- Transform variables **X** and **Y** to new variables **u** and **v** as follows:

- $U = \frac{X - A}{h}$ and $V = \frac{Y - B}{k}$

- Where **A** is Assumed Mean of **X** = 55

- **H** = Size of class interval of **X** = 10

- **B** is assumed mean of **Y** = 57.5

- **K** = Size of class interval of **Y** = 5

- **Step 4**

- Find total of all cell frequencies of **X** and **Y**. In this question totals are already given to us as highlighted in next table.

Marks in Mathe Matics (y)			Marks in Statistics(x)				Total
			40-50	50-60	60-70	70-80	
		Mid-Point (X)	45	55	65	75	
	Mid-Point (Y)	$U = \frac{X-55}{10}$ $V = \frac{Y-57.5}{5}$	-1	0	1	2	f
50-55	52.5	-1	4	7	5	2	18
55-60	57.5	0	6	10	7	4	27
60-65	62.5	1	6	12	10	7	35
65-70	67.5	2	3	8	6	3	20
Total		f	19	37	28	16	100



Total of cell frequencies of all classes of Y (Marginal freq)



Total of cell frequencies of all classes of X (Marginal frequency)

- **Step 5**
- **Now in next step we multiply U with V and respective cell frequency and write the number in a circle(f_{uv}).**
- **Number in circle (f_{uv})= $U \times V \times$ Respective cell Frequency**
- **In the next slide red boxes indicate how the circles are obtained.**
- **Add together all these values of f_{uv} across each rows and each columns to get respective $\sum f_{uv}$ for each row and column.**

Marks in Mathe Matics (y)	Mid-Point (Y)	U = $\frac{X-55}{10}$ V = $\frac{Y-57.5}{5}$	Marks in Statistics(x)				Total	fv	fv ²	fuv
			40-50	50-60	60-70	70-80				
	52.5	-1	4	7	5	2	18			4+0-5-4=-5
	57.5	0	6	10	7	4	27			0+0+0+0=0
	62.5	1	6	12	10	7	35			-6+0+10+14=18
	67.5	2	3	8	6	3	20			-6+0+12+12=18
Total		f	19	37	28	16	100	Total		-5+0+18+18=31
		fu								
		fu ²								
		fuv	4+0- 6-6= -8	0+0+ 0+0= 0	-5+0 +10 +12=17	-4+0 +14 +12=22	-8+0 +17+ 22=31			Should be same $\sum fuv$

- **Step 6**
- Calculate F_u values by multiplying respective marginal frequency(f) with respective u values.
- This is shown with help of black arrow in the next slide.
- Add all these values to compute $\sum f_u$.

- Similarly Calculate F_v values by multiplying respective marginal frequency(f) with respective v values.
- This is shown with help of red arrow in the next slide.
- Add all these values to compute $\sum f_v$.

Marks in Mathematics (y)	Mid-Point (X)	Marks in Statistics(x)					f	fv	fv ²	fuv			
		40-50	50-60	60-70	70-80	Total							
	Mid-Point (Y)												
		$U = \frac{X-55}{10}$											
		$V = \frac{Y-57.5}{5}$											
50-55	52.5	-1	4	4	7	0	5	-5	2	-4	18	-18	-5
55-60	57.5	0	6	0	10	0	7	0	4	0	27	0	0
60-65	62.5	1	6	-6	12	0	10	10	7	14	35	35	18
65-70	67.5	2	3	-6	8	0	6	12	3	12	20	40	18
Total		f	19	37	28	16	N=100	$\sum fv = 57$					31
		fu	-19	0	28	32	$\sum fu = 41$						
		fu ²											
		fuv	-8	0	17	36	31						

Step 7

- Calculate Fu^2 values by multiplying respective marginal frequency(f) with respective u^2 values.
- Add all these values to compute $\sum fu^2$.
- $Fu^2 = f \times u \times u$

- Calculate Fv^2 values by multiplying respective marginal frequency(f) with respective v^2 values.
- Add all these values to compute $\sum fv^2$.
- $Fv^2 = f \times v \times v$

Marks in Mathe Matics (y)	Mid-Point (X)	Marks in Statistics(x)					Total			
		40-50	50-60	60-70	70-80					
	Mid-Point (Y)	$U = \frac{X-55}{10}$ $V = \frac{Y-57.5}{5}$	-1	0	1	2	f	fv	fv ²	fuv
50-55	52.5	-1	4	7	5	2	18	-18	18	-5
55-60	57.5	0	6	10	7	4	27	0	0	0
60-65	62.5	1	6	12	10	7	35	35	35	18
65-70	67.5	2	3	8	6	3	20	40	80	18
Total		f	19	37	28	16	N=100	$\sum fv = 57$	$\sum fv^2 = 111$	$\sum fuv = 31$
		fu	-19	0	28	32	$\sum fu = 41$			
		fu ²	19	0	28	64	$\sum fu^2 = 111$			
		fuv	-8	0	17	36	$\sum fuv = 31$			

Step 8: To compute correlation coefficient r we apply the following formula

$$r = \frac{N \sum f_{uv} - \sum f_u \cdot \sum f_v}{\sqrt{N \sum f_u^2 - (\sum f_u)^2} \cdot \sqrt{N \sum f_v^2 - (\sum f_v)^2}}$$

$$r = \frac{100 \times 31 - 41 \times 57}{\sqrt{100 \times 111 - (41)^2} \cdot \sqrt{100 \times 133 - (57)^2}}$$

$$r = \frac{3100 - 2337}{\sqrt{11100 - 1681} \cdot \sqrt{13300 - 3249}}$$

$$r = \frac{763}{\sqrt{9419} \cdot \sqrt{10051}}$$

$$r = 0.0784$$

- A video explaining the same has been shared on whats app group.
- Practise similar questions from book.