PSOT : Module 2

Page No.
Date / /

- **Correlation** : Study of relations / variations b/w 2 or more variables.

Types :-

1 Nature of graph : i) Linear ii) Non linear.

2 Direction of Change : i) Positive ii) Negative

- **Regression** : Prediction of one category of data from other categories by using values of 1 or more variables, when co-related.

- **Formulas**

1 Mean $= \bar{x} = \frac{\sum x}{N} = u$

2) RMS $= \sqrt{\frac{(x-\bar{y})^2}{N}}$

3 $Cov(x,y) = \frac{\sum (x-\bar{x})(y-\bar{y})}{N}$ = Simultaneous variable

4 Karl Pearson's Coeff of Correlation :-

$(r) = \frac{Cov(x,y)}{\sigma_x \sigma_y}$

$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}$

$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{N \sigma_x \sigma_y}$

$r = \frac{\sum xy - N\bar{x}\bar{y}}{\sqrt{(\sum x^2 - N\bar{x}^2)(\sum y^2 - N\bar{y}^2)}}$ another way to express the formula

- **Properties** :-

1 $-1 \leq r \leq 1$

2 If x and y are independent variable, they aren't correlated.

3 Correlation coeff is independent of change of origin and change of scale :-

if $r \geq 0.95$: highly correlated

$\pm 0.75 < r < 0.95$: Correlated

$r < 0.35$: Not related.

- **Questions**

• From the following values of demand and price, find degree of correlation by using Karl Pearson's coeff of correl^n

1 Demand 65 66 67 67 68 69 70 72

Price 67 68 65 68 72 72 69 71

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
65	-3	9	67	-2	4	6
66	-2	4	68	-1	1	2
67	-1	1	65	-4	16	4
67	-1	1	68	-1	1	1
68	0	0	72	3	9	0
69	1	1	72	3	9	3
70	2	4	69	0	0	0
72	4	16	71	2	4	8

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}$$

$$\bar{x} = \frac{\sum x}{N} = \frac{544}{8} = 68$$

$$\bar{y} = \frac{\sum y}{N} = \frac{552}{8} = 69$$

$$\sum (y-\bar{y})^2 = 44 \qquad \sum (x-\bar{x})^2 = 36$$

$$\sum (x-\bar{x})(y-\bar{y}) = 24$$

$$r = 24 / \sqrt{44 \times 36}$$

2 x: 100 200 300 400 500

y: 30 40 50 60 70

x	$\hat{x}$	$\hat{x}-\bar{\hat{x}}$	$(\hat{x}-\bar{\hat{x}})^2$	y	$\hat{y}$	$(\hat{y}-\bar{\hat{y}})$	$(\hat{y}-\bar{\hat{y}})^2$	$(\hat{x}-\bar{\hat{x}})(y-\bar{\hat{y}})$
100	-2	-2	4	30	-2	-2	4	4
200	-1	-1	1	40	-1	-1	1	1
300	0	0	0	50	0	0	0	0
400	1	1	1	60	1	1	1	1
500	2	2	4	70	2	2	4	4
$\sum$			10				10	10

$$r = \frac{\sum (\hat{x}-\bar{x})(\hat{y}-\bar{y})}{\sqrt{\sum (\hat{x}-\bar{x})^2 \sum (\hat{y}-\bar{y})^2}}$$

$$\hat{x} = \frac{x-\bar{x}}{100} \qquad \hat{y} = \frac{y-\bar{y}}{10}$$

$\bar{x} = 300 \quad \bar{y} = 50$

$$r = \frac{10}{\sqrt{10 \times 10}} = \frac{10}{10} = 1$$

- Formula :

$$r = \frac{\Sigma d_x d_y - N \bar{d}_x \bar{d}_y}{\sqrt{(\Sigma d_x^2 - N(\bar{d}_x)^2)(\Sigma d_y^2 - N(\bar{d}_y)^2)}} \qquad d_x = \frac{x_i - a}{h} \quad d_y = \frac{y_i - b}{k}$$

h, k : common diff $h = k = 1$ when no common inputs

a : Nearest int. to $\bar{x}$ b : Nearest int to $\bar{y}$

- Questions

1 x : 10 12 14 15 16 17 18 10 14 15

y : 17 16 15 12 10 9 8 15 13 12

x	y	d_x	d_y	d_x^2	d_y^2	$d_x d_y$
10	17	-4	4	16	16	-16
12	16	-2	3	4	9	-6
14	15	0	2	0	4	0
15	12	1	-1	1	1	-1
16	10	2	-3	4	9	-6
17	9	3	-4	9	16	-12
18	8	4	-5	16	25	-20
10	15	-4	2	16	4	-8
14	13	0	0	0	0	0
15	12	1	-1	1	1	-1
Σ		1	-3	67	85	-70

$$\bar{x} = \frac{\Sigma x}{N} = \frac{141}{10} = 14.1 \qquad \bar{y} = \frac{\Sigma y}{N} = \frac{127}{10} = 12.7$$

$a = 14$ $b = 13$ $d_x = x - a$ $d_y = y - b$

$$\bar{d}_x = \frac{\Sigma d_x}{N} = \frac{1}{10} = 0.1 \qquad \bar{d}_y = \frac{-3}{10} = -0.3$$

$$r = \frac{(-70) - (10)(0.1)(-0.3)}{\sqrt{[(67) - (10)(0.1)^2][85 - (10)(-0.3)^2]}} = 0.9293$$

2 x = 28 45 40 38 35 33 40 32 36 33

y = 23 34 33 34 30 26 28 31 36 35

	x	y	d_x	d_y	d_x^2	d_y^2	$d_x d_y$
	28	23	-8	-8	64	64	64
	45	34	9	3	81	9	27
	40	33	4	2	16	4	8
	38	34	2	3	4	9	6
	35	30	-1	-1	1	1	1
	33	26	-3	-5	9	25	15
	40	28	4	-3	16	9	-12
	32	31	-4	0	16	0	0
	36	36	0	5	0	25	0
	33	35	-3	4	9	16	-12
Σ	360	310	0	0	216	162	97

$\bar{x} = \frac{\Sigma x}{N} = \frac{360}{10} = 36$ $\quad a = 36$ $\quad \bar{d}_x = 0$

$\bar{y} = \frac{\Sigma y}{N} = \frac{310}{10} = 31$ $\quad b = 31$ $\quad \bar{d}_y = 0$

$$r = \frac{97 - (10)(0)(0)}{\sqrt{216 \times 162}} = \frac{97}{187.061} = 0.5185$$

- Spearman Rank Correlation [can be used if method isn't specified in question]
  denote by R. given by $R = 1 - \left(\frac{6\Sigma d_i^2}{N^3 - N}\right)$

N = no. of data Find rank of x $\longrightarrow R_1$,
rank of y $\longrightarrow R_2$
difference b/w ranks : $d_i = R_1 - R_2$

1 X : 18 20 34 52 12
Y : 39 23 35 18 16

X	Y	R_1	R_2	$d_i = R_1 - R_2$	d_i^2
18	39	2	4	-2	4
20	23	3	2	1	1
34	35	4	3	1	1
52	18	5	1	4	16
12	46	1	5	-4	16

$$R = 1 - \left(\frac{6\sum d_i^2}{N^3 - N}\right) \qquad \sum d_i^2 = 38$$

$$R = 1 - \frac{6 \times 38}{125 - 5} = -0.9$$

- Equal values in Spearman Rank Correlation :
  Give average as per position and adjust $\frac{1}{12}(m_i^3 - m_i)$ term in formula.

$$R = 1 - 6\left[\frac{\sum d_i^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) \cdots}{N^3 - N}\right]$$

- Questions

1 X : 10 12 18 18 15 40
Y : 12 18 25 25 50 25

X	Y	R_1	R_2	$d_i = R_1 - R_2$	d_i^2
10	12	1	1	0	0
12	18	2	2	0	0
18	25	4.5	4	0.5	0.25
18	25	4.5	4	0.5	0.25
15	50	3	6	-3	9
40	25	6	4	2	4

$\sum d_i^2 = 13.5$ $\quad N = 6$

$m_1 = 2$ $\quad m_2 = 3$
(m_1 → no of similar terms)

X has 18 - 2 times, Y has 25 - 3 times.

$$R = 1 - 6\left[\frac{\sum d_i^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2)}{N^3 - N}\right]$$

$$= 1 - \frac{6}{210}\left[13.5 + \frac{1}{12}(6) + \frac{1}{12}(27-3)\right] = 0.5428$$

2 In a musical competition, 10 competitors were ranked by 3 judges X, Y and Z in the following order by using the rank correlation method. Discuss which pair of judges has nearest approach for evaluation in music

X rank (R_1): 1 6 5 10 3 2 4 9 7 8

Y (R_2): 3 5 8 4 7 10 2 1 6 9

Z (R_3): 6 4 9 8 1 2 3 10 5 7

X(R_1)	(R_2)Y	(R_3)Z	$d_{xy} = R_1 - R_2$	d_{xy}^2	$d_{yz} = R_2 - R_3$	d_{yz}^2	$d_{xz} = R_1 - R_3$	d_{xz}^2
1	3	6	-2	4	-3	9	-5	25
6	5	4	1	1	1	1	2	4
5	8	9	-3	9	-1	1	-4	16
10	4	8	6	36	-4	16	2	4
3	7	1	-4	16	-6	36	2	4
2	10	2	-8	64	8	64	0	0
4	2	3	2	4	-1	1	1	1
9	1	10	8	64	-9	81	-1	1
7	6	5	1	1	1	1	2	4
8	9	7	-1	1	2	4	1	1
Σ				200		214		60

$$R_{xy} = 1 - \frac{6 \times 200}{10^3 - 10} = \frac{990 - 1200}{990} = -0.2121$$

$$R_{yz} = 1 - \frac{6 \times 214}{990} = -\frac{49}{165} = -0.2969$$

$$R_{xz} = \frac{990 - 6 \times 60}{990} = \frac{21}{33} = 0.6363$$

$$R = 1 - \left[\frac{6 \times \sum d_i^2}{N^3 - N}\right]$$

∴ Judge X and Y have nearest approach in the evaluation of music

Page No.
Date

3. If Karl Pearson coefficient of the data is 0.4. Covariance b/w x and y is 1.6. Variance of y is given to be 25. Find variance of x.

→ $r = 0.4$ $\quad Cov(x,y) = 1.6$ $\quad \sigma_y^2 = 25$ $\quad \sigma_y = 5$

$$r = \frac{Cov(x,y)}{\sigma_x \times \sigma_y} \longrightarrow 0.4 = \frac{1.6}{5 \times \sigma_x}$$

$\sigma_x = 1.6 / 2 = 0.8$ $\quad$ Variance of $x = \sigma_x^2 = 0.64$

4. Rank correlation b/w x and y is given to be 0.143. Sum of square of diff b/w rank = 48. Find N

→ $\Sigma d_i^2 = 48$ $\quad R = 0.143$

$$R = 1 - \frac{6 \times \Sigma d_i^2}{N^3 - N} = 1 - \frac{288}{N^3 - N}$$

$-0.143 + 1 = 288 / (N^3 - N) = 0.857$

$N^3 - N = 288 / 0.857 = 336.05$

$N = 7.000$

5. The coefficient of rank correl^n of marks obtained by 10 students in Physics and Chem was found to be 0.5. After it is discovered diff in ranks for 1 student was wrongly taken as 3 instead of 1. Then find correct coeff. of rank of correl^n.

→ $N = 10$ $\quad$ wrong $R = 0.5$ $\quad$ wrong 3 → correct 7

$$R = 1 - \frac{6 \times \Sigma d_i^2}{N^3 - N}$$

wrong $\Sigma d_i^2 = \frac{0.5 \times 990}{6} = 82.5$

Correct Σd_i^2 = wrong $\Sigma d_i^2 - 3^2 + 7^2 = 82.5 + 49 - 9 = 122.5$

Correct $R = 1 - \left(\frac{6 \times 122.5}{990}\right) = 0.2575$

6. Sample of 25 pairs has the following result

$\Sigma x = 127$ $\quad \Sigma y = 100$ $\quad \Sigma x^2 = 760$ $\quad \Sigma y^2 = 449$

$\Sigma xy = 500$

Later it was found that 2 pairs :

(8, 14) was used instead of (8, 12)

and (8, 6) " " " " (6, 8)

↳ wrong $\quad$ correct ↙

→ correct $\Sigma x = 127 - 8 + 6 = 125$

correct $\Sigma y = 100 - 20 + 20 = 100$

correct $\Sigma x^2 = 760 - 64 + 36 = 732$

correct $\Sigma y^2 = 425$

correct $\Sigma xy = 500 - 8 \times 14 - 8 \times 6 + 8 \times 12 + 6 \times 8 = 484$

$$r = \frac{\Sigma xy - N\bar{x}\bar{y}}{\sqrt{[\Sigma x^2 - N(\bar{x})^2][\Sigma y^2 - N(\bar{y})^2]}} \qquad N = 25$$

correct $\bar{x} = \frac{125}{25} = 5$ correct $\bar{y} = \frac{100}{25} = 4$ $r = -0.309$

$$r = \frac{484 - 25 \times 5 \times 4}{\sqrt{(732 - 25 \times 25)(425 - 25 \times 16)}} = -0.309$$