Date 13/10/22

Module 5

Thursday

Correlation Analysis and Regression Lines

1] mean of $x = \bar{x} = \frac{\Sigma x}{N}$

2] mean of $y = \bar{y} = \frac{\Sigma y}{N}$

3] Variance of $x = V(x) = \sigma_x^2 = \frac{\Sigma(x-\bar{x})^2}{N}$

4] Variance of $y = V(y) = \sigma_y^2 = \frac{\Sigma(y-y)^2}{N}$

5] Standard deviation $= SD = \sqrt{V(x)} = \sigma x$

6] Standard deviation $= SD \sqrt{V(y)} = \sigma y$

7] Co-variance $= cov(x, y) = \frac{\Sigma[(x-\bar{x})(y-\bar{y})]}{N}$

Note : If x and y are independent of each other, then $cov(x, y) = 0$.

Correlation:

① Karl Pearson's Coefficient of Correlation (r)

r = coefficient of correlation $= \frac{cov(x, y)}{\sigma x \cdot \sigma y}$

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

$$\therefore \boxed{r = \frac{\Sigma[(x-\bar{x})(y-\bar{y})]}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}} \quad \boxed{-1 \le r \le 1}$$

(If $\bar{x}$ and $\bar{y}$ are integer form) or

Assumed Mean Method

Let A and B are the means of x and y respectively.

$$\therefore \boxed{r = \frac{\Sigma[(x-A)(y-B)] - \frac{\Sigma[(x-A)].\Sigma[(y-B)]}{N}}{\sqrt{\Sigma[(x-A)^2] - \frac{[\Sigma(x-A)]^2}{N}}\sqrt{\Sigma[(y-B)^2] - \frac{[\Sigma(y-B)]^2}{N}}}}$$

or

Actual method

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

Ex. 1] 10 students got following % of marks in EM III and DS.

Maths (x)	DS (y)	$x-\bar{x}$ $x-65$	$y-\bar{y}$ $y-66$	$(x-\bar{x})(y-\bar{y})$		
78	84	13	18	$(13\times18)=234$	169	324
36	51	-29	-15	$(-29\times-15)=435$	841	225
98	91	33	25	$(33\times25)=825$	1089	625
25	60	-40	-6	$(-40\times-6)=240$	1600	36
75	68	10	2	$(10\times2)=20$	100	4
82	62	17	-4	$(17\times-4)=-68$	289	16
90	86	25	20	$(25\times20)=500$	625	400
62	58	-3	-8	$(-3\times-8)=24$	9	64
65	53	0	-13	$(0\times-13)=0$	0	169
39	47	-26	-19	$(-26\times-19)=494$	676	361
$\bar{x}=650/10=65$	$\bar{y}=660/10=66$			~~2840~~ 2704	5398	2224

classmate

Date
Page

② Rank Correlation method : (Spearman's coefficient) (R)

$$\boxed{R = 1 - 6\left[\frac{\Sigma (R_1 - R_2)^2}{N^3 - N}\right]} \qquad \boxed{-1 \le R \le 1}$$

[Only for non-repeated data]

$R_1 \Rightarrow$ Rank of x

$R_2 \Rightarrow$ Rank of y

ex.1] Panel of 2 judges A and B given the marks for dramatic performance

R(x)	S(y)	R_1 = Rank for R	R_2 = Rank for S	$R_1 - R_2$	$(R_1 - R_2)^2$
35	51	9	8	1	1
38	37	8	10	-2	4
43	48	7	9	-2	4
30	62	10	6	4	16
54	93	5	1	4	16
68	73	3	3	0	0
76	56	2	7	-5	25
92	72	1	4	-3	9
44	70	6	5	1	1
56	92	4	2	2	4
					80

$$R = 1 - 6\left[\frac{80}{(10)^3 - 10}\right] \Rightarrow R = 1 - \frac{480}{990} = 1 - 0.4848$$

$$\therefore \boxed{R = 0.5151}$$

For repeated data :

$$R = 1 - 6\left[\frac{\Sigma(R_1-R_2)^2 + \frac{1}{12}(m_1^3-m_1) + \frac{1}{12}(m_2^3-m_2) + \dots \frac{1}{12}(m_3^3-m_3)}{N^3 - N}\right]$$

$m_1 \Rightarrow$ Data repeated how many times.

Fitting of First degree curve

① Regression line y on x
(x independent)

Method 1 :

$$(y-\bar{y}) = byx(x-\bar{x})$$

$$byx = r.\frac{\sigma y}{\sigma x} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

Method 2 :

$$y = ax+b$$

$\Sigma y = a\Sigma x + bN \dots ①$ $(\Sigma 1 = N)$

multiply by x

$\Rightarrow \Sigma xy = a\Sigma x^2 + b\Sigma x \dots ②$

② Regression line x on y
(y independent)

Method 1 :

$$(x-\bar{x}) = bxy(y-\bar{y})$$

$$bxy = r.\frac{\sigma x}{\sigma y} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2}$$

Method 2 :

$$x = ay+b$$

$\Sigma x = a\Sigma y + bN \dots ①$

multiply by y

$\Rightarrow \Sigma xy = a\Sigma y^2 + b\Sigma y \dots ②$

$$r = \sqrt{byx . bxy}$$

this method.

Second degree curve

$y = ax^2 + bx + c$

$\Sigma y = a\Sigma x^2 + b\Sigma x + cN \ldots$ (1)

Multiply by x

$\Rightarrow \Sigma xy = a\Sigma x^3 + b\Sigma x^2 c\Sigma x \ldots$ (2)

Multiply by x

$\Rightarrow \Sigma x^2 y = a\Sigma x^4 + b\Sigma x^3 + c\Sigma x^2 \ldots$ (3)

For large values,

$Y = aX^2 + bX + C$

$\Rightarrow \Sigma Y = a\Sigma X^2 + b\Sigma X + CN$

$\Rightarrow \Sigma XY = a\Sigma X^3 + b\Sigma X^2 + C\Sigma X$

$\Rightarrow \Sigma X^2 Y = a\Sigma X^4 + b\Sigma X^3 + c\Sigma X^2$

ex. 1] Obtain eqⁿs of line of regression for following data.

x	y	xy	x^2	y^2
65	67			
66	68			
67	65			
67	68			
68	72			
69	72			
70	69			
72	71			
		$\Sigma xy =$	$\Sigma x^2 =$	$\Sigma y^2 =$

1] Regression line y on x

Let $y = ax + b$ .... (*)

$\Sigma y = a\Sigma x + Nb$ ... (1)

Multiply by x and Σ.

$\Rightarrow$ $\Sigma xy = a\Sigma x^2 + b\Sigma x$ .... (2)

2] Regression line x one y

Let $x = ay + b$ ... (*)

$\Sigma x = a\Sigma y + Nb$ ... (1)

Multiply by y and Σ

$\Rightarrow$ $\Sigma xy = a\Sigma y^2 + b\Sigma y$ ... (2)

ex. 2] Fit second degree curve (Parabola)

Yrs (x)	Profit in crores (y) y=Y	X=(x-1968.5)2	X^2	X^3	X^4	XY	X^2Y
1965	125	-7	49	-343	2401	-875	
1966	140	-5	25	-125	625	-700	
1967	165	-3	9	-27	81	-495	
1968	195	-1	1	-1	1	-195	
1969	200	1	1	1	1	200	
1970	215	3	9	27	81	645	
1971	220	5	25	125	625	1100	
1972	230	7	49	343	2401	1610	
		$\Sigma X=0$	$\Sigma X^2=168$	$\Sigma X^3=0$	$\Sigma X^4=6216$		

Avg = 1968.5

Find a,bc,

$$y = a\left[(x-1968.5)2\right]^2 + b\left[(x-1968.5)2\right] + c$$