Module 3 Statistical Techniques

Karl Pearson's coefficient of correlation

$\sigma^2 x = \frac{1}{N}\Sigma(x-\bar{x})^2$ measure of variation in x, $\sigma^2 y = \frac{1}{N}\Sigma(y-\bar{y})^2$

variation $= \frac{1}{N}\Sigma(x-\bar{x})(y-\bar{y})$

Coefficient of correlation between x & y denoted by r

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{N\,\sigma x\,\sigma y} \quad -①$$

$$r = \frac{\mathrm{cov}(x,y)}{\sigma x\,\sigma y} \quad -②$$

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2\;\Sigma(y-\bar{y})^2}} \quad -③$$

$x-\bar{x} = x \quad y-\bar{y} = y$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}} \quad -④$$

$$r = \frac{\Sigma xy - N\bar{x}\bar{y}}{\sqrt{(\Sigma x^2 - N\bar{x}^2)(\Sigma y^2 - N\bar{y}^2)}} \quad -⑤$$

for direct values ⑤ use

$-1 \leq r \leq 1$

(i) If dx & dy denote the deviations of x & y from assumed mean A & B

$dx = x - A$, $dy = y - B$

$$r = \frac{\Sigma dx dy - \frac{(\Sigma dx)(\Sigma dy)}{N}}{\sqrt{\left(\Sigma d^2x - \frac{(\Sigma dx)^2}{N}\right)} \sqrt{\Sigma d^2y - \frac{(\Sigma dy)^2}{N}}} \quad - (6)$$

(1) Actual mean method

$$r = \frac{\Sigma xy}{\Sigma x^2 \Sigma y^2} \quad - (6)$$

(i) mean $\bar{X}$ and then ~~tak~~ take deviation of ~~from~~ ~~x~~ x from $\bar{X}$. $x = X - \bar{X}$

(ii) $y = \bar{Y} - \bar{Y}$

(iii) xy, x^2, y^2

Ex (1) Calculate Karl Pearson's coefficient of correlation betn x & y

x :	28	45	40	38	35	33	40	32	36	33
Y :	23	34	33	34	30	26	28	31	36	35

15

Soln $\bar{X} = \frac{\Sigma X}{n} = \frac{360}{10} = 36$

$\bar{Y} = \frac{\Sigma Y}{n} = \frac{310}{10} = 31$

Calculation r betn X & Y

Sr No.	X	$x = X - \bar{X}$	x^2	Y	$y = Y - \bar{Y}$	y^2	xy
1	28	-8	64	23	-8	64	64
2	45	9	81	34	3	9	81
3	40	4	16	33	2	4	16
4	38	2	4	34	3	9	18
5	35	-1	1	30	-1	1	+1
6	33	-3	9	26	-5	25	15
7	40	4	16	28	-3	9	-12
8	32	-4	16	31	0	0	0
9	36	0	0	36	3	9	~~6~~ 0
10	33	-3	9	35	4	16	-12
N = 10	$\Sigma X = 360$	~~Σx~~ =	$\Sigma x^2 = 216$	$\Sigma Y = 310$		$\Sigma y^2 = 162$	$\Sigma xy = 97$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \, \Sigma y^2}} = \frac{97}{\sqrt{(216)(162)}}$$

$$= \frac{97}{187.1} = 0.5186$$

Ex: X: 65 66 67 67 68 69 70 72

Y: 67 68 65 68 72 72 69 71

Find r

H.W.

(2) Step-deviation method

Ex (1) Calculate the coefficient of correlation

X:	100	200	300	400	500	
Y:	30	40	50	60	70	15

Soln: Calculation of r betn x & y

Sr No.	X	$x = X - \bar{x}$	$x = \frac{X-\bar{x}}{100}$	x^2	Y	$y = Y - \bar{y}$, $Y - 50$	$y = \frac{Y-\bar{y}}{10}$	y^2	xy
1	100	-200	-2	4	30	-20	-2	4	4
2	200	-100	-1	1	40	-10	-1	1	1
3	300	0	0	0	50	0	0	0	0
4	400	100	1	1	60	10	1	1	1
5	500	200	2	4	70	20	2	4	4
	$\Sigma X = 1500$			$\Sigma x^2 = 10$	$\Sigma Y = 250$			$\Sigma y^2 = 10$	$\Sigma xy = 10$

$\Sigma X = 1500$

$\bar{x} = \frac{\Sigma X}{N} = \frac{1500}{5} = 300$

$\Sigma Y = 150$

$\bar{y} = \frac{\Sigma Y}{N} = \frac{250}{5} = 50$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} = \frac{10}{\sqrt{10 \cdot 10}} = 1$$

(3) Assumed mean method

$$r = \frac{\Sigma dxdy - \frac{\Sigma dx dy}{N}}{\sqrt{\Sigma dx^2 - \frac{(\Sigma dx)^2}{N}}\sqrt{\Sigma dy^2 - \frac{(\Sigma dy)^2}{N}}}$$

Ex (5) Calculate the correlation coefficient

x : 23 27 28 29 30 31 33 35 36 39

Y : 18 22 23 24 25 26 28 29 30 32

15

Soln r between x & Y

Sr No	x	$dx = x-30$	d^2x	Y	$dY = Y-25$	d^2Y	$dxdY$
1	23	-7	49	18	-7	49	49
2	27	-3	9	22	-3	9	9
3	28	-2	4	23	-2	4	4
4	29	-1	1	24	-1	1	1
5	30	0	0	25	0	0	0
6	31	1	1	26	1	1	1
7	33	3	9	28	3	9	9
8	35	5	25	24	4	16	20
9	36	6	36	30	5	25	30
10	39	9	81	32	7	49	63
N=10		$\Sigma dx = 11$	$\Sigma d^2x = 215$		$\Sigma dY = 7$	$\Sigma d^2Y = 163$	$\Sigma dxdY = 186$

~~$\bar{x} = \frac{\Sigma x}{N}$~~ A = 30, B = 25 be assumed mean

$$r = \frac{\Sigma dxdY - \frac{\Sigma dx \cdot \Sigma dY}{N}}{\sqrt{\left(\Sigma d^2x - \frac{(\Sigma dx)^2}{N}\right)}\sqrt{\Sigma d^2Y - \frac{(\Sigma dY)^2}{N}}}$$

$$= \frac{186 - \frac{11 \times 7}{10}}{\sqrt{\left(215 - \frac{(11)^2}{10}\right)\left(163 - \frac{(7)^2}{10}\right)}} = \frac{186 - 7.7}{\sqrt{215 - 12.1}\sqrt{163 - 4.9}}$$

$$= \frac{178.3}{\sqrt{202.9}\sqrt{158.1}} = 0.9948$$