Recall

1) Measures of Central tendency

    a) Mean

    b) Median

    c) Mode

    d) Geometric mean

    e) Harmonic mean.

2) Measures of dispersion.

    a) Range   b) Quartile deviation   c) Mean deviation about Mean.

    d) Standard deviation   e) Moments.

Correlation (Linear).

Amount of relationship between two Variables.

X	⇑	↓	↑	↓
Y	↑	↓	↓	↑

⟶ Karl Pearson's Coefficient of Correlation.

γ or P.

$$\gamma = \frac{N \sum' xy - \sum' x \sum' y}{\sqrt{N \sum' x^2 - (\sum' x)^2} \sqrt{N \sum' y^2 - (\sum' y)^2}}$$

⟶ Covariance b/w X & Y.

$$\gamma = \frac{Cov(X,Y)}{\sigma_x \; \sigma_y}$$

Multiple correlation & Partial Correlation

$\gamma_{12.3}$

↓

Relationship b/w x_1 & x_2 keeping x_3 constant.

$R_{12.3}$ ⟶ Multiple correlation

Properties of correlation coefficient

1) r lies b/w -1 & 1. ie $-1 \leq r \leq 1$.

2) Correlation is unaffected by the change of scale & origin.

Individual observations.

$$r = \frac{N \Sigma' xy - \Sigma' x \, \Sigma' y}{\sqrt{N \Sigma' x^2 - (\Sigma' x)^2} \sqrt{N \Sigma' y^2 - (\Sigma' y)^2}}$$

Discrete / Continuous

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

$$N = \Sigma' f ; \quad dx = \frac{x - A}{h} ; \quad dy = \frac{y - B}{k} .$$

Calculate correlation coeff:

X: 65   66   67   67   68   69   70   72

Y: 67   68   65   68   72   72   69   71.

Soln

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

$N \to$ no. of observation

(or)

$$r = \frac{N \sum dx \, dy - \sum dx \sum dy}{\sqrt{N \sum dx^2 - (\sum dx)^2} \; \sqrt{N \sum dy^2 - (\sum dy)^2}}$$

$$; \quad dx = \frac{x - A}{h}$$

$$dy = \frac{y - B}{k}.$$

X	y	$dx = X - 68$	$dy = y - 69$	dx^2	dy^2	$dx\,dy$
65	67	-3	-2	9	4	6
66	68	-2	-1	4	1	2
67	65	-1	-4	1	16	4
67	68	-1	-1	1	1	1
68	72	0	3	0	9	0
69	72	1	3	1	9	3
70	69	2	0	4	0	0
72	71	4	2	16	4	8
		0	0	36	44	24

$N = 8$

$\sum dx = 0 = \sum dy$

$\sum dx^2 = 36$

$\sum dy^2 = 44$

$\sum dx\,dy = 24.$

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

$$= \frac{8(24) - 0 \times 0}{\sqrt{8(36) - 0^2}\, \sqrt{8(44) - 0^2}} = \frac{192}{6\sqrt{8} \times 352}$$

$$\frac{\frac{3}{44 \times 8}}{\frac{352 \times 8}{2816}}$$

$$= \frac{\cancel{192}\,32}{\cancel{6}\sqrt{2816}} = 0.603$$

The following table gives the number of blind per lakh of population in different age groups. Find out the correlation between age & blindness.

X → age:

	0-10	10-20	20-30	30-40	40-50
Y → No. of blind per lakh:	55	67	100	111	150

50-60	60-70	70-80
200	300	500

Age X	m	y	$dx = \dfrac{m-45}{10}$	dx^2	$dy = y - 150$	dy^2	$dx\,dy$
0-10	5	55	-4	16	-95	9025	380
10-20	15	67	-3	9	-83	6889	249
20-30	25	100	-2	4	-50	2500	100
30-40	35	111	-1	1	-49	2401	49
40-50	(45)	150	0	0	0	0	0
50-60	55	200	1	1	50	2500	50
60-70	65	300	2	4	150	22500	300
70-80	75	500	3	9	350	122500	1050
			-4	44	273	168315	2178

$N = 8$

$\sum dx = -4$

$\sum dx^2 = 44$

$\sum dy = 273$

$\sum dy^2 = 168315$

$\sum dx\,dy = 2178.$

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

$$= \frac{8(2178) - (-4)(273)}{\sqrt{8(44) - (-4)^2}\ \sqrt{8(168315) - (273^2)}}$$

$$= \frac{18516}{\sqrt{336}\ \sqrt{1271991}} = \frac{18516}{18.33 \times 1127.82}$$

$$= 0.8956$$