Correlation

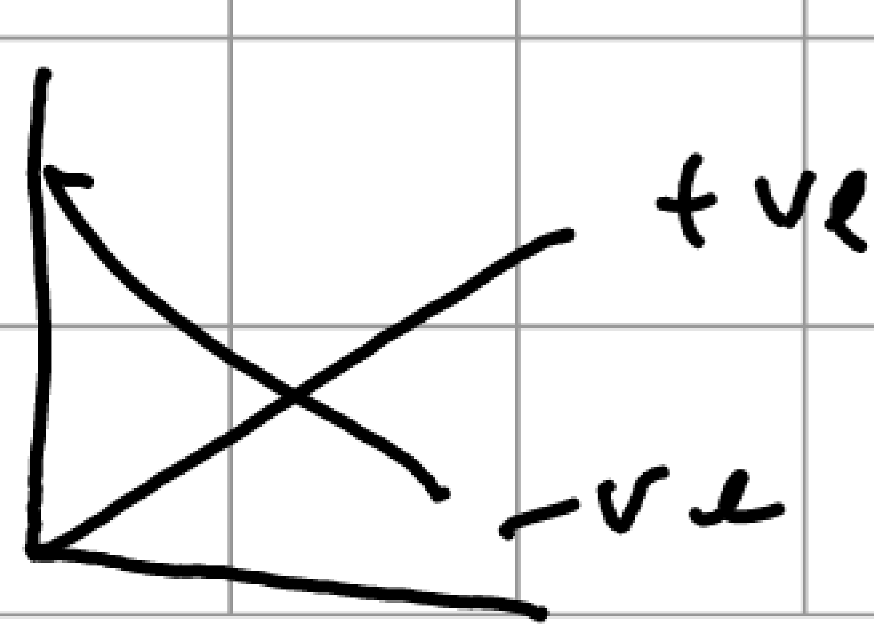
① Positive & Negative Correlation

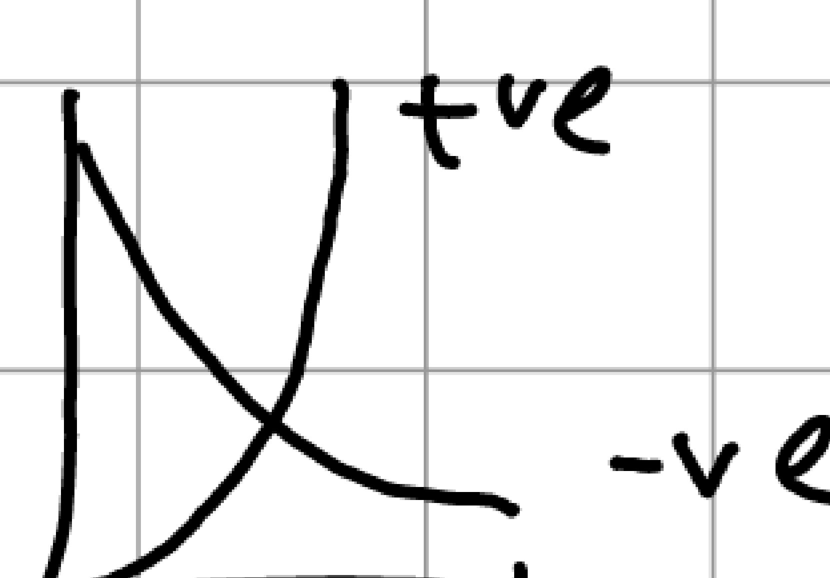
If both variables change in same direction
Then +ve correlation else -ve correlation

Example

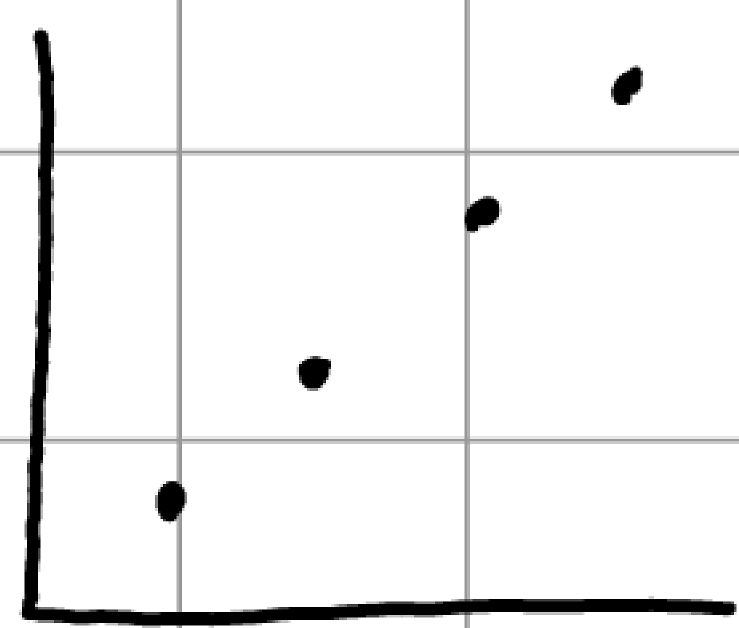
+ve : effort & marks
-ve : difficulty & marks

② Linear & non Linear Correlation
If graph between two variables is straight line then Linear correlation else it is non linear correlation





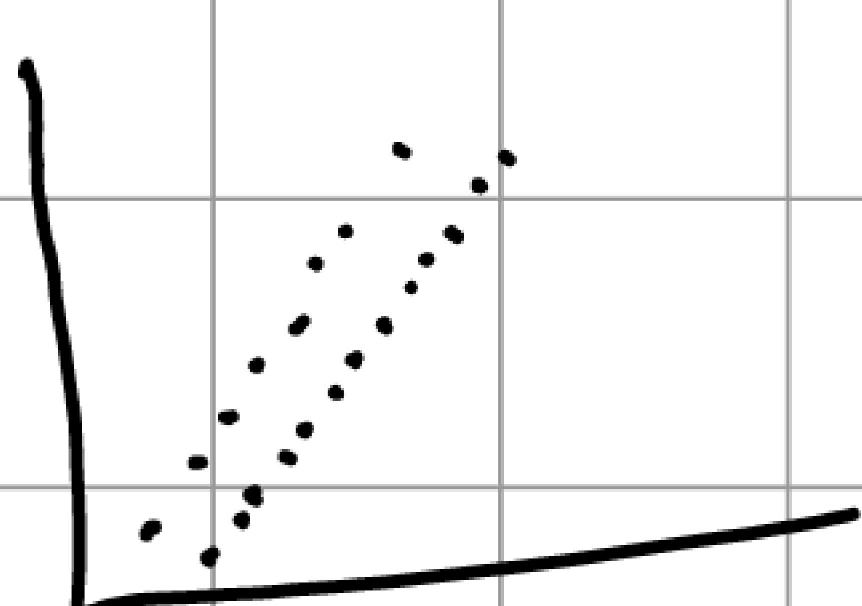
Scatter diagram



Perfect positive Correlation

All are in One st. line & upward

$\gamma = 1$



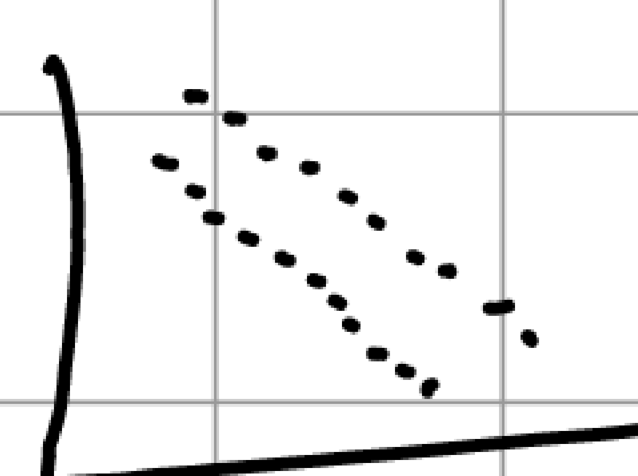
High degree +ve Correlation

$\gamma \to 1$

Perfect Negative Correlation

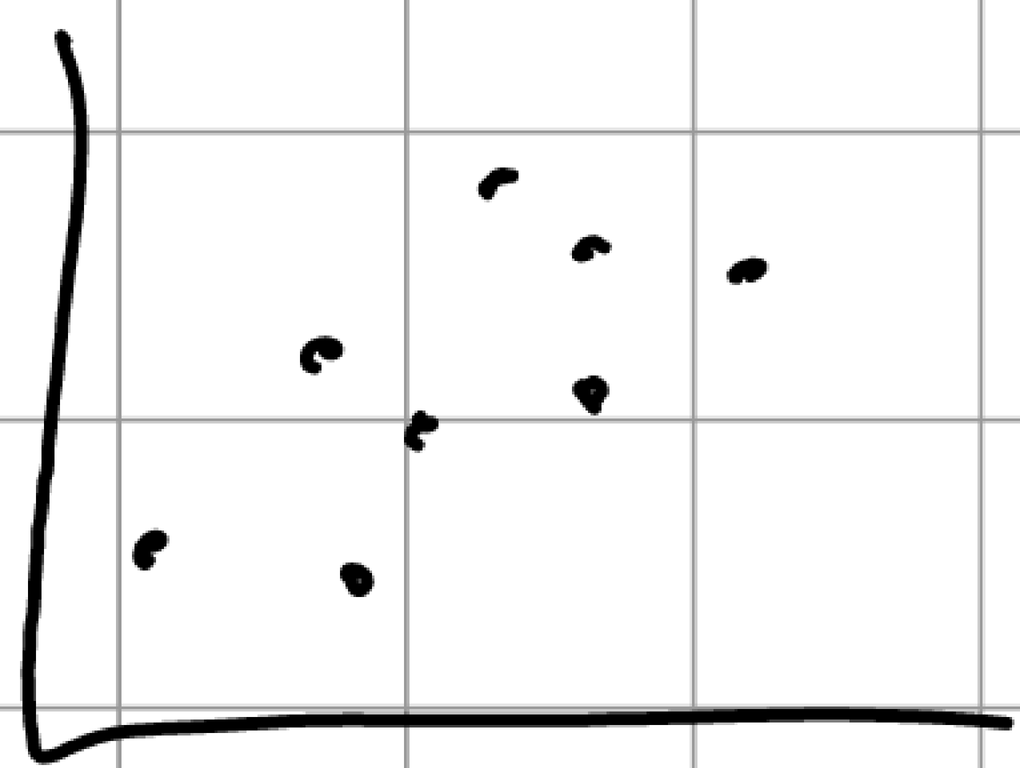
All are in One st. line & downward

$\gamma = -1$



High degree -ve Correlation

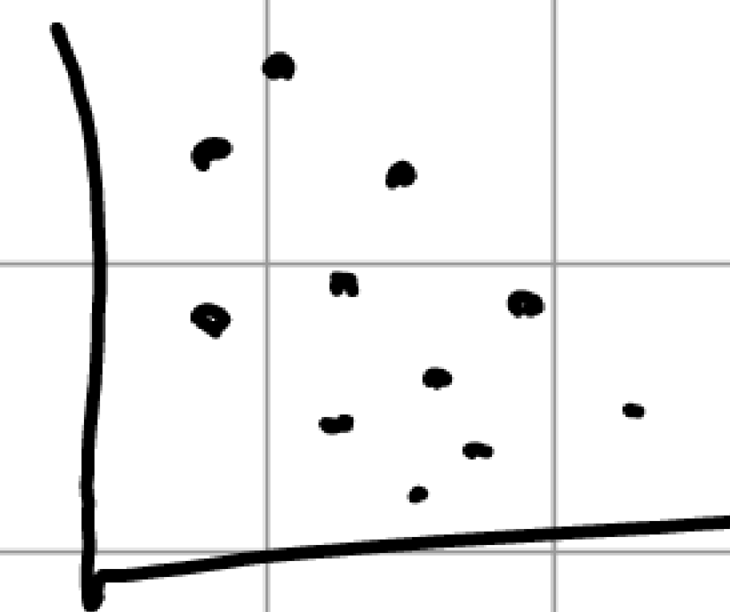
$\gamma \to -1$



Low degree
Positive Correlation

$r < 1$

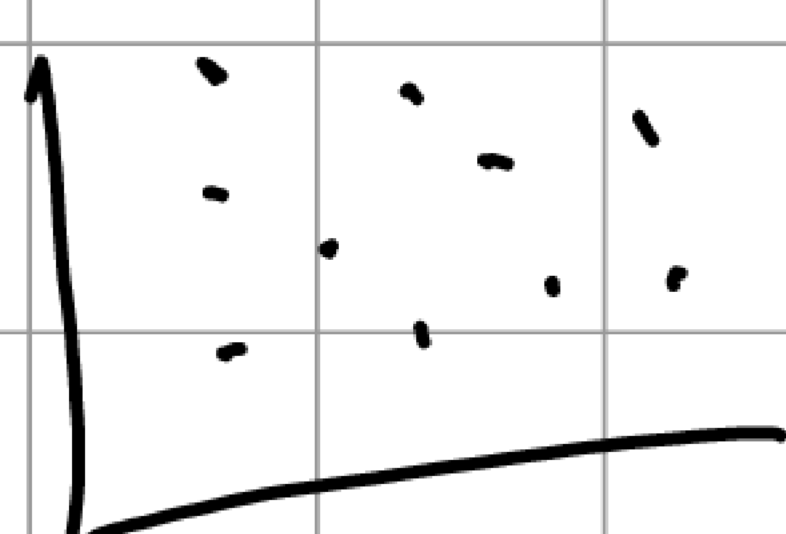
Distance between
points more



Low degree
Negative Correlation

$r < 1$

Distance between
points more



No correlation
$r = 0$

Karl-pearson's coorelation coefficient

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

$\sigma \rightarrow$ standered deviation

$$\text{Covariance} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N}$$

$$r = \frac{\text{Covariance}}{\sigma_x \cdot \sigma_y}$$

$N \rightarrow$ Number of Samples

$$\sigma_x = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{N}}$$

$$\sigma_y = \sqrt{\frac{\Sigma (y_i - \bar{y})^2}{N}}$$

On Expanding

$$r = \frac{(\Sigma x_i y_i - N\bar{x}\bar{y})}{\sqrt{(\Sigma x_i^2 - N\bar{x}^2)(\Sigma y_i^2 - N\bar{y}^2)}}$$

Properties :-

$-1 \le r \le 1$

If x & y are independant $r = 0$

r is independant of change of origin & change of scale

① Calculate r for

x	y
23	18
21	22
28	23
29	24
30	25
31	26
33	28
35	29
36	30
39	32

Ans $r = 0.995$

$\therefore$ +ve correlation

High degree

$u = x_i - 30$	$v = y_i - 23$	u^2	v^2	uv
-7	-7	49	49	49
-3	-3	9	9	6
-2	-2	4	4	4
-1	-1	1	1	1
0	0	0	0	0
1	1	1	1	1
3	3	9	9	9
5	4	25	16	20
6	5	36	25	30
9	7	81	49	63
$\Sigma u = 11$	7	215	163	186

$$\bar{u} = \frac{\Sigma u}{N}$$

$$= \frac{11}{10} = 1.1$$

$$\bar{v} = \frac{\Sigma v}{N} = \frac{7}{10} = 0.7$$

$$r = \frac{\Sigma uv - N\bar{u}\bar{v}}{\sqrt{(\Sigma u_i^2 - N\bar{u}^2)(\Sigma v^2 - N\bar{v}^2)}}$$

$$= \frac{186 - 10 \times 1.1 \times 0.7}{\sqrt{(215 - 10(1.1)^2)}\sqrt{163 - 10 \cdot 0.7^2}}$$

$$= 0.995$$

Spearmans rank correlation coefficient

Used for non measurable (graded) variables

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

$$-1 \leq R \leq 1$$

$$d_i = x - y$$

① Calculate Spearman's rank corelation for the data

SR No	R_1 Rank in english	R_2 Rank in maths	d_i	d_i^2
1	1	3	2	4
2	3	1	2	4
3	7	4	3	9
4	5	5	0	0
5	4	6	2	4
6	6	9	3	9
7	2	7	5	25
8	10	8	2	4
9	9	10	1	1
10	8	2	6	36

$\Sigma d_i^2 = 96$

$N = 10$

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

$$= 1 - \frac{96 \times 6}{1000 - 10}$$

$$= 0.4182$$

If ranks are repeated then

$$R = 1 - \frac{6 \Sigma d_i^2 + \frac{1}{12} \Sigma m_i^3 - m_i}{N^3 - N}$$

m_i = Number of items having equal ranks

③ find R from data

x	y	Rank x	Rank y	d_i^2
32	40	3	5	1
55	30	9	3.5	9.15
49	70	7.5	8	0.25
60	20	10	1	81
43	30	5.5	3.5	4
57	50	4	7	4
43	72	5.5	10	12.25
49	60	7.5	8	0.25
10	45	1	6	16
20	25	2	2	0

$\Sigma d_i = 176$

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

$$= 1 - \frac{6\left(176 + \frac{1}{12}(6+6+6)\right)}{10^3 - 10}$$

$$= 1 - \frac{6(176 + 1.5)}{990}$$

$$= -0.076$$

$M_1 = 2$

$M_2 = 2$

$M_3 = 2$

(4) Let $r_{xy} = 0.4$

$cov(x, y) = 1.6$

$\sigma_y^2 = 25$ find σ_x

$\sigma_y = 5$

$\rightarrow$

$$r_{xy} = \frac{cov(x, y)}{\sigma_x \sigma_y}$$

$$\frac{0.4}{1.6} \times 5 = \frac{1}{\sigma_x}$$

$\sigma_x = 0.8$

(5) $R_{xy} = 0.143$

$\Sigma d_i^2 = 48$

$\rightarrow$

$$R_{xy} = 1 - \frac{6 \Sigma d_i^2}{N^3 - N}$$

On sub

$N^3 - N = 336 \cdot 0.5$

$N(N^2-1) = 336$

$16 \quad 7 \quad 3$

④ Calculate corelation coef from data

$N = 10$

$\Sigma x = 140 \qquad \bar{x} = 14 \qquad \bar{u} = \bar{x} - 10 = 4$

$\Sigma y = 150 \qquad \bar{y} = 15 \qquad \bar{v} = \bar{y} - 15 = 0$

$\Sigma (x-10)^2 = 180 = u$

$\Sigma (y-15)^2 = 215 = v$

$\Sigma (x-10)(y-15) = 60$

$$r = \frac{\Sigma uv - N\bar{u}\bar{v}}{\sqrt{\Sigma u^2 - N\bar{u}^2}\sqrt{\Sigma v^2 - N\bar{v}^2}} =$$

Substituting

$$r = \frac{60}{\sqrt{180-10\times 16}\ \sqrt{213}}$$

$$= 0.914$$

(3) for 25 pairs of values

$\sum x = 127 \quad \bar{x} = 127/25$

$\sum y = 100 \quad \bar{y} = 100/25$

$\sum x^2 = 760$

$\sum y^2 = 449$

$\sum xy = 500$

Later on it was found 2 pairs were (8, 14) (8, 6) instead of (8, 12) & (6, 8) find correct r

$\rightarrow N = 25$

Correct value

$\Sigma x = 127 - 8 - 8 + 8 + 6$

$= 125$

$\Sigma y = 100 - 14 - 6 + 12 + 8$

$= 100$

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

$\Sigma y^2 = 449 - 228 - 36 + 144 + 64 = 425$

$\Sigma xy = 500 - 14 \times 8 - 8 \times 6$

$+ 8 \times 12 + 6 \times 8 = 484$

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

On substitution

Answer = -0.309

For 10 pairs of values of x & y the following are determined. Later on it was found that one pair of values was $(34, 47)$ instead of $(43, 74)$. Determine correct value of coeficiant of correlation if

Mean $(x) = 30.1$ Mean $(y) = 47.8$

$S.D.(x) = 6.2$ $S.D.(y) = 9.5$

$r = 0.72$

$\longrightarrow$

old $\bar{x} = 30.1$

$\bar{y} = 47.8$

$\sigma x = 6.2$

$\sigma y = 9.5$

$r = 0.72$

$$\cdot\ \Sigma x_2 = \Sigma x_1 - x_7 + x_2$$

$$= 10 \times 30.1 - 34 + 43$$

$$\Sigma x_2 = 310$$

$$\bar{x} = 31$$

$$\Sigma y_2 = \Sigma y_1 - 47 + 74$$

$$= 505$$

$$\bar{y} = 50.5$$

$$\sigma x_1^2 = \Sigma x_1^2 - N(\bar{x})^2$$

$$38.44 = \Sigma x_1^2 - 10(30.1)^2$$

$$\Sigma x_1^2 = 9098.54$$

$$\Sigma x_2^2 = \Sigma x_1^2 - (34)^2 + (43)^2$$

$$= 9098.54 - (34)^2 + (43)^2$$

$$= 9791.54$$

$$\sigma x_2^2 = 9791.54 - 10(31.)^2$$

$\therefore \sigma_{x_2}^2 = 181.34$

$\sigma_x = 13.47367804$

old $\sigma_y^2 = \Sigma(y)^2 - N(\bar{y})^2$

$(9.5)^2 = \Sigma(y)^2 - 10 \times (47.8)^2$

$\Sigma(y)^2 = 22938.65$

New

$\Sigma y^2 = 22938.65 - (47)^2 + (75)^2$

$= 26205.65$

New $\sigma_y^2 = 26205.65 - 10 \times (50.5)^2$

$= 703.15$

$= 26.51697569$

old

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

$\therefore \quad 0.72 \times 6.2 \times 9.5 = \Sigma xy - 10\,(30.1)(57.8)$

$\therefore \quad \Sigma xy = 14430.208$

New $\Sigma xy = 14430.208 - 34 \times 47 + 43 \times 75$

$= 16014.208$

New $r = \dfrac{\Sigma xy - N(\bar{x})(\bar{y})}{\sigma x \, \sigma y}$

$= \dfrac{359.208}{} = 1.005$

$$r = \frac{\sum x_i y_i - N\bar{x}\bar{y}}{\sqrt{\left(\sum x_i^2 - N\bar{x}^2\right)\left(\sum y_i^2 - \bar{y}^2 N\right)}}$$

$$R = 1 - \frac{\frac{1}{6}\sum d_i^2 + \frac{1}{12}\sum M_i^3 - M_i}{N^3 - N}$$