(5) # Covariance & Correlation :

Covariance :

- It is relationship between a pair of ~~freedom~~ random variables where one variable causes change in another variable.

- values can be from $-\alpha$ to $+\alpha$

- covariance is actually find proportionality between two variable.
  if covariance is negative - Inversely proportional
  if covariance is positive - Directly Proportional

- It is used for the linear relationship between variables.
- It gives direction of relationship between variables.

Population

$$cov(x,y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Sample

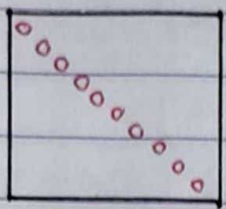
$$cov(x,y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

$\bar{x}, \bar{y} \rightarrow$ mean of given sample set
$n \rightarrow$ total no of sample

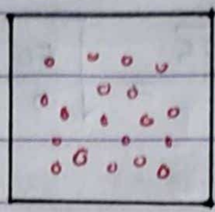
★ $cov(x,x) = E\,\dfrac{(x_i-\bar{x})(x_i-\bar{x})}{N-1} = \dfrac{\Sigma (x_i-\bar{x})^2}{N-1}$

$\therefore \boxed{cov(x,x) = var(x)}$

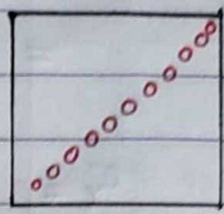
Covariance



Large Negative covariance



Nearly Zero Covariance



Large positive Covariance

Correlation:

- It shows whether and how strongly pair of variables are related to each other.

- Correlation takes value between -1 to +1, where in values close to +1 represents strong positive.
  -1 represents strong negative.
- correlation value will tell us how strongly variable are related.

Pearson Correlation Coefficient:

$$\boxed{\rho_{x,y} = \frac{cov(x,y)}{\sigma_x \sigma_y}}$$

x	y
5	10
10	15
15	20

$\bar{x} = (5+10+15)/3 = 10$

$\bar{y} = (10+15+20)/3 = 15$

$$cov(x,y) = \frac{(5-10)\times(10-15) + (10-10)\times(15-15) + (15-10)\times(20-15)}{} = \frac{25+0+25}{3} = \frac{50}{3}$$

$corr(x,y) = \frac{50}{3} \times \sigma_x \times \sigma_y =$

$$\sigma_x = \sqrt{\frac{(5-10)^2 + (10-10)^2 + (15-10)^2}{3}}$$

$$= \sqrt{\frac{25+0+25}{3}} = \sqrt{\frac{50}{3}}$$

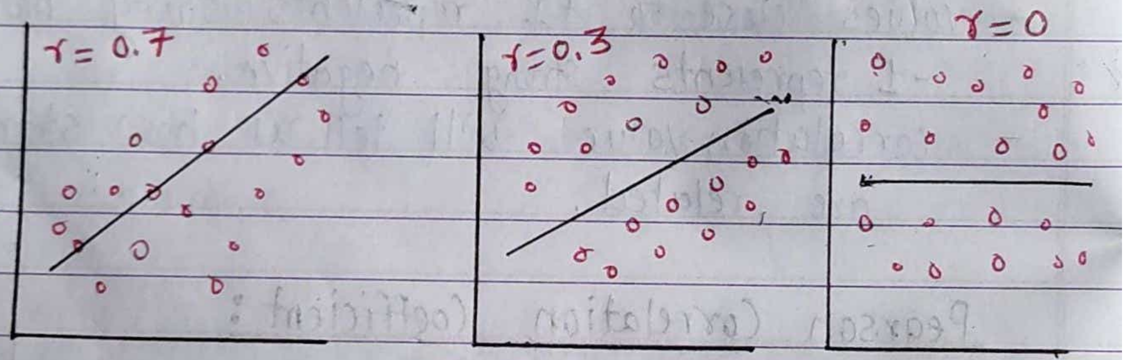
$\sigma_x = 4.0824$

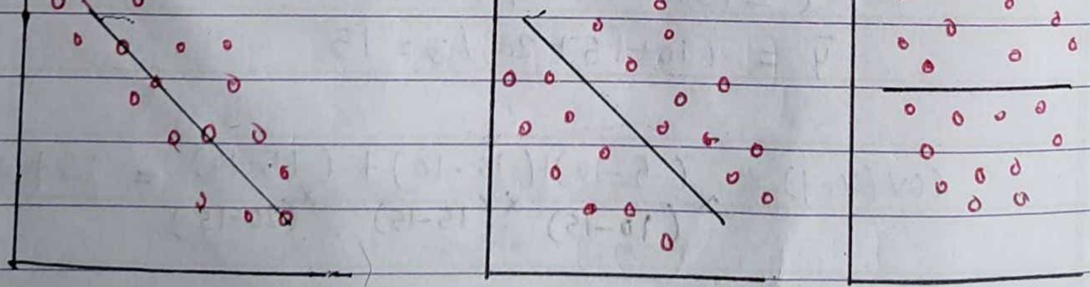
similarly,

$\sigma_y = 4.0824$

$\boxed{Corr(x,y) \approx 1}$ Highly correlated.

- Pearson correlation can evaluate only a Linear Relationship between two continuas variables





- Pearson Correlation use only when data follows Linear type distribution, it can be negative slope or positive.

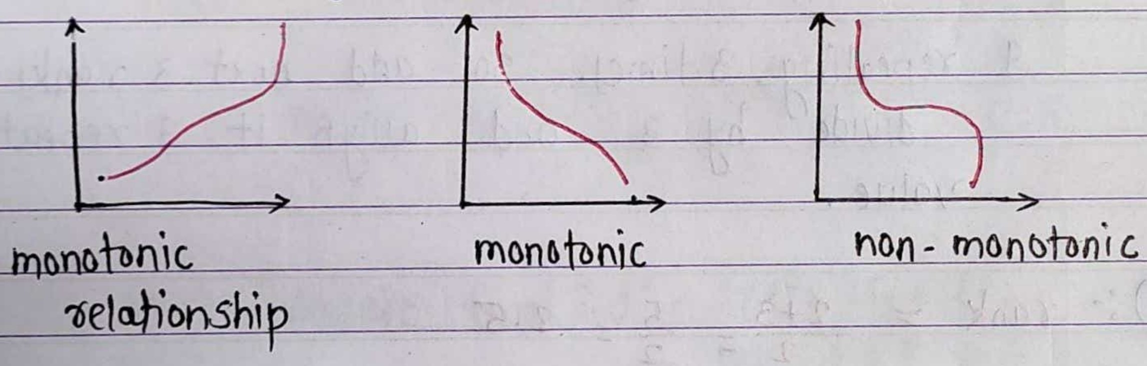
Spearman Correlation Coefficient : (Range :- -1 to +1)

- It is Non-parametric measure of rank correlation.

- Spearman correlation can evaluate a monotonic relationship between two variables.

monotonic relationship:

- proportionality should be either directly proportional or Inversly proportional to each other.



Comparison of Pearson & Spearman coefficients :

- Pearson can only used when linear relationship between two variables.

- Spearman works on any type of data but it should have monotonic relation ship.

$$\boxed{\rho = 1 - \frac{6 \sum d^2}{n(n^2-1)}}$$

Spearman correlation calculation:-

A	B	rank of A	Rank (B)	d	d^2
35	2	1	8	-7	49
23	17	4	5	-1	1
14	23	5	4	1	1
25	45	3	1	2	4
34	12	2	7	-5	25
9	30	7	2.5	4.5	20.25
9	30	7	2.5	4.5	20.25
9	14	7	6	1	1

$\sum d^2 = 121.5$

~~rank = $\frac{9+9+9}{3}$ $\frac{6+6+6}{3} = \frac{18}{3} =$~~

column (A):- rank = $\frac{6+7+8}{3} = \frac{21}{3} = 7$

9 repeating 3 times. so add next 3 ranks and divide by 3 and assign it 3 repeating value.

column (B):- rank = $\frac{2+3}{2} = \frac{5}{2} = 2.5$

value 30 repeating twice so adding next 2 ranks and divide by 2 and assign as rank to those two values.

step (1):- Give ranks (High value → 1) in Descending order.

note:- Suppose, there is tie (repeated values) then. calculate rank as above given in rectangular block.

Step ②:- Calculate d :- Find difference between ranks given.

Step ③: Calculate d^2:- Find Squares of d values. and Find Σd^2

Step ④:- note :- only if there are repeated values in column

In col Ⓐ :- 9 repeats thrice. $\therefore m = 3$

correction factor (CF):- $\frac{m(m^2-1)}{12}$

$$Cf_1 = \frac{3(3^2-1)}{12} = \frac{3(9-1)}{12} = \frac{24}{12} = 2$$

In col Ⓑ :- 30 repeats twice $\therefore m = 2$

$$Cf_2 = \frac{2(2^2-1)}{12} = \frac{2(3)}{12} = \frac{6}{12} = 0.5$$

Step ⑤ :- Calculate final $\Sigma d^2 = \Sigma d^2 + Cf_1 + Cf_2$

$$= 121.5 + 2 + 0.5$$

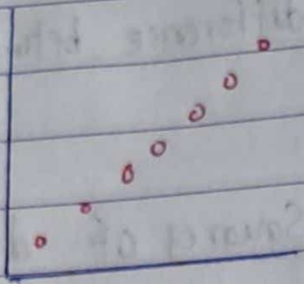
$$\Sigma d^2 = 124$$

Step ⑥ :- put Σd^2_{final} in Spearman Corr. Coeff. $= 1 - \frac{6 \Sigma d^2_{final}}{n(n^2-1)}$

$$\rho = 1 - \frac{6 \times 124}{8 \times (8^2-1)} = 1 - \frac{744}{504}$$

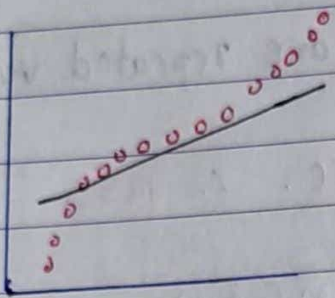
$$= 1 - 1.476$$

$$\boxed{\rho = -0.476}$$



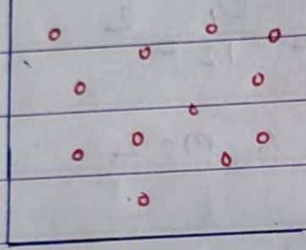
pearson +1

Spearman +1



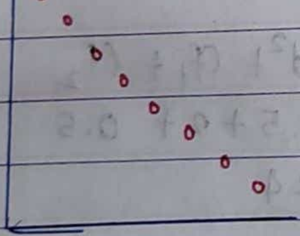
pearson + 0.85

Spearman + 1



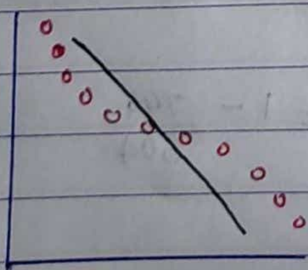
Pearson − 0.093

Spearman − 0.093



Pearson − 1

Spearman − 1



Pearson − 0.85

Spearson − 1