Correlation: It measures the degree of relationship between variables

for eg demand & supply , rainfall and amount of crop,
work effort & result etc

1. **Positive correlation**: one vari ↑ then second vari ↑
   one vari ↓ then 2nd vari ↓

   eg demand & sale

2. **Negative correlation**: one vari ↑ then second vari ↓
   one vari ↓ then other vari ↑

   eg speed of car (↑) & time to cover distance (↓)

3. **Simple correlation**: study betn two variables only

4. **Multiple correlation**: study betn more than two variable
   for eg: crop yield = f (rain, temp, humidity, soil)

5. **Linear correlation**: ratio of change in one series is same as another.

   for eg

x	2	4	6	8	10	12
y	3	6 9	12	15	18	

   Here x is ↑ by amount of 2 in each obse, y is ↑ by amount of 3

6. **Nonlinear correlation**: Opposite to linear

   for eg. If double the amount of rainfall give not gurantee exactly double the crop.

7. **Partial correlation**: When output depends on (as such) more than one factor but Just restrict the study betn output & only one factor keeping other factor constant is known as partial correlation

   for eg Total crop yield being study only based on rainfall (not paying much attention to temperature).

8. **Total correlation**: like nonlinear correlation whose output variable is being studied by

Note: Correlation coefficient is denoted by r, the value of r is between -1 and 1.

Formulas: 1. $\cancel{cov(x,y) = \sqrt{\Sigma(x-\bar{x})}}$

$$cov(x,y) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n} = \text{covariance of } x \,\&\, y$$

$$\sigma_x = \text{std. deviation of } x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

$$\sigma_y = \text{std. deviation of } y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n}}$$

Karl Pearson's coefficient of correlation: r

$$\boxed{r = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y}} \quad ; \quad -1 \le r \le 1$$

or

$$\boxed{r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2} \cdot \sqrt{\Sigma(y-\bar{y})^2}}}$$

$\bar{x}, \bar{y}$:- actual Mean

or

$$\boxed{r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}}$$

where $x = x - \bar{x}$

$y = y - \bar{y}$

∴ Above all three versions are for correlation formula using actual mean

Correlation without deviation: [Direct method actual data]

Derivation:

$$r = \frac{\sum(xy - x\bar{y} - \bar{x}y + \bar{x}\bar{y})/n}{\sqrt{\frac{\sum(x^2 - 2x\bar{x} + \bar{x}^2)}{n}} \cdot \sqrt{\frac{\sum(y^2 - 2y\bar{y} + \bar{y}^2)}{n}}}$$

$$= \frac{(\sum xy - \bar{y}\sum x - \bar{x}\sum y + \bar{x}\bar{y}\sum 1)/n}{\left(\sqrt{\sum x^2 - 2\bar{x}\sum x + \bar{x}^2\sum_{i=1}^{n} 1}\right)/n \left(\sqrt{\sum y^2 - 2\bar{y}\sum y + \bar{y}^2\sum 1}\right)/n}$$

$$= \frac{\frac{\sum xy}{n} - \frac{\sum y}{n}\frac{\sum x}{n} - \frac{\sum x}{n}\frac{\sum y}{n} + \bar{x}\bar{y}\frac{n}{n}}{\sqrt{\frac{\sum x^2}{n} - 2\frac{\sum x}{n}\frac{\sum x}{n} + \left(\frac{\sum x}{n}\right)^2 \cdot \frac{n}{n}} \cdot \sqrt{\frac{\sum y^2}{n} - 2\frac{\sum y}{n}\frac{\sum y}{n} + \left(\frac{\sum y}{n}\right)^2 \cdot \frac{n}{n}}}$$

$$= \frac{\frac{\sum xy}{n} - 2\frac{\sum y \sum x}{n^2} + \frac{\sum x}{n}\frac{\sum y}{n}}{\sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} \cdot \sqrt{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2}}$$

$$\boxed{r = \frac{n\sum xy - \sum x\sum y}{\sqrt{n\sum x^2 - (\sum x)^2} \cdot \sqrt{n\sum y^2 - (\sum y)^2}}}$$

Remember This formula to calculate r without deviation

Using above formula, find r without changing the data

Correlation using assumed mean:

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

where $dx = x - A_1$

$dy = y - A_2$

A_1 is assumed mean of x-series

A_2 " " y-series

Probable error (P.E.) $= 0.6745\left(\frac{1-r^2}{n}\right)$

This error helps in interpreting value of coeffi. of correlation..

- If value of r is $<$ P.E. : No evidence of correlation
- $r > 6$(P.E.) : r is significant

Spearman's Rank Correlation Coefficient: (R)

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

d = difference of rank between paired item

N = Total no. of data in a series

When rank is not given (when actual data are given)

Steps to assign Rank:

Step 1: Assign rank 1 to the highest or lowest value

Step 2: Assign rank 2 to next lowest or next highest value, continue in same way

Note: If any data is repeat then assign rank as average of no. of repeat data.

eg If ~~three/data~~ are same then

When ranks are repeated

$$R = 1 - \frac{6\sum d^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \cdots}{n(n^2-1)}$$

Here d : difference of rank

m : Total no. of items whose ranks are common,

$\frac{1}{12}(m^3 - m)$ to be added for every such group of items

eg

Price of Tea	75	88	95	70	60	80	81	50
Price of Coffee	120	134	150	115	110	140	142	100

Assign rank:

Note: rank is normally in one digit.

next lower value assign 2

→ highest no - assign rank 1

Price of tea	75	$\boxed{88}$	$\boxed{95}$	70	60	80	81	50
R_1	5	2	1	6	7	4	3	8

||ly do for price of coffee

→ showing this eg starting with low value as rank 1

eg

Judge X:	$\boxed{20}$	22	28	23 ✓	30	30	23 ✓	24
Rank R_1:	1	2	6	3.5	7.5	7.5	3.5	5

23 Repeat 2 times : 1 & 2 already assigned so to assign rank to 23 Take $\frac{3+4}{2} = 3.5$

Now next rank is 5, assign to next high value

to calculate R → Here $m_1 = 2$ (for group (23))
$m_2 = 2$ (for group 30, as 30 also repeat 2 times)

Important Note: Higher value of R (betn two vari.) means nearest approach or not much difference