STATS

CH: Measures of Correlation

→ The term correlation indicates the relationship between two such variables in which change in the value of one variable, also changes the value of other variable.

Price	Demand
₹ 10	100
₹ 12	80
₹ 13	70
ज्यादा	कम

inverse

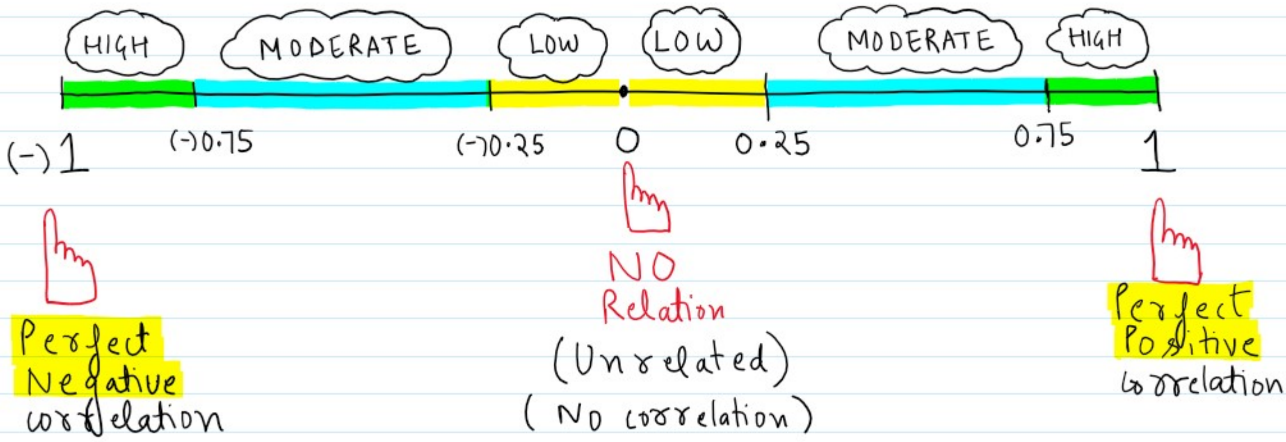
Price	Supply
₹10	100
₹12	120
₹13	140

Direct

→ Kinds of Correlation

- Positive + Negative correlation
- linear & Curvilinear (Non-Linear) Correlation
- Simple, Partial and Multiple correlation

→* **Degree of Correlation**



$$\boxed{(-)1 \leq r \leq 1}$$

* (-)1.25 , 2.4 , 1.8 Slap yourself

× ——— × ——— × ——— × ——— × ——— × ——— ×

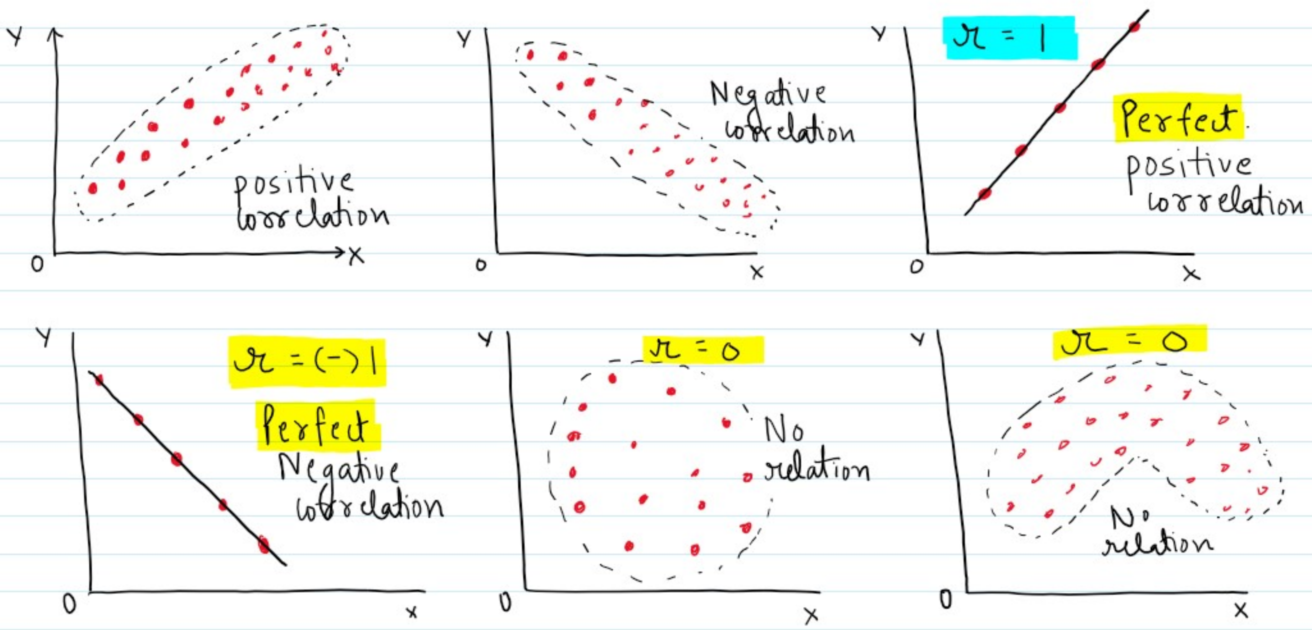
* Methods of Calculating Correlation *

Scatter diagram
↓
This method gives only general idea, not exact value

Karl Pearson's coefficient of correlation **
↓
Best method which is applicable in linear relationship

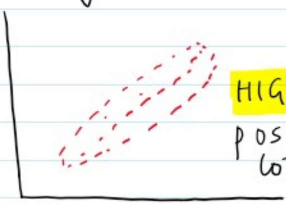
Spearman's **Rank** correlation
↓
It is used only when correlation is to be calculated Based on **Ranks**

I. SCATTER DIAGRAM



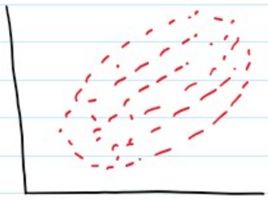
Q- Identify correlation

(i)



HIGH positive correlation

(ii)



LOW / MODERATE positive correlation

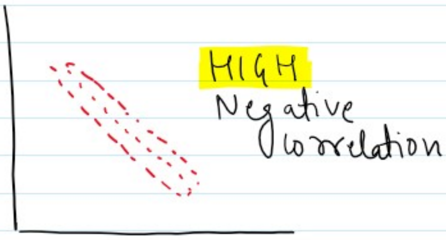
(iii)

HIGH

(iv)

LOW

(iii)



(iv)

II. Karl Pearson's Coefficient of Correlation

- Actual Mean method
- Direct method
- Assumed Mean method
- Step deviation method

(A) Actual Mean Method

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

Here $x = (X - \bar{X})$

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

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

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

eg :-

Year	Birth rate (X)	Death Rate (Y)	$X - \bar{X}$	x^2	$Y - \bar{Y}$	y^2	xy
2011	24	15	(−) 6	36	(−) 7	49	42
2012	26	20	(−) 4	16	(−) 2	4	8
2013	32	22	2	4	0	0	0
2014	33	24	3	9	2	4	6
2015	35	27	5	25	5	25	25
2016	30	24	0	0	2	4	0
	$\Sigma X = 180$	$\Sigma Y = 132$	$\Sigma x = 0$	$\Sigma x^2 = 90$	$\Sigma x = 0$	$\Sigma y^2 = 86$	81

Find r

Sol :-

Step 1: $\bar{X} = \frac{\Sigma X}{N} = \frac{180}{6} = 30$

Step 2: $\bar{Y} = \frac{\Sigma Y}{N} = \frac{132}{6} = 22$

Step 3: $(X - \bar{X})$ ☞ x

Step 4: x^2

Step 4 x^2

Step 5 $(Y - \bar{Y})$ ∃ y

Step 6 y^2

Step 7 xy (Step 3 × Step 5)

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}} = \frac{81}{\sqrt{90 \times 86}}$$

$$= \frac{81}{\sqrt{7740}} = \frac{81}{87.97} = 0.92$$

High Positive Correlation

II DIRECT Method

When $\bar{X}$ & $\bar{Y}$ are in fraction	IMP When $\bar{X}$ & $\bar{Y}$ are Not in fraction
$r = \dfrac{\Sigma XY - \dfrac{\Sigma X . \Sigma Y}{N}}{\sqrt{\Sigma X^2 - \dfrac{(\Sigma X)^2}{N}} \times \sqrt{\Sigma Y^2 - \dfrac{(\Sigma Y)^2}{N}}}$	$r = \dfrac{\Sigma XY - N.(\bar{X}).(\bar{Y})}{\sqrt{\Sigma X^2 - N(\bar{X})^2} \times \sqrt{\Sigma Y^2 - N(\bar{Y})^2}}$

eg:-

X	Y	X^2	Y^2	X Y
10	5	100	25	50
10	6	100	36	60
11	4	121	16	44
12	3	144	9	36
12	2	144	4	24
$\Sigma X = 55$	$\Sigma Y = 20$	$\Sigma X^2 = 609$	$\Sigma Y^2 = 90$	$\Sigma XY = 214$

Find r

Sol:- Step 1 :- $\bar{X} = \dfrac{\Sigma X}{N} = \dfrac{55}{5} = 11$

$$\overline{X} = \frac{\Sigma X}{N} = \frac{\quad}{5} = 11$$

Step 2 :- $\overline{Y} = \frac{\Sigma Y}{N} = \frac{20}{5} = 4$

Step 3 :- X^2

Step 4 :- Y^2

Step 5 :- $X.Y$

$$r = \frac{\Sigma XY - N.\overline{X}.\overline{Y}}{\sqrt{\Sigma X^2 - N(\overline{X})^2} \times \sqrt{\Sigma Y^2 - N(\overline{Y})^2}}$$

$$= \frac{214 - 5 \times 11 \times 4}{\sqrt{609 - 5(11)^2} \times \sqrt{90 - 5(4)^2}}$$

$$= \frac{214 - 220}{\sqrt{609 - 605} \times \sqrt{90 - 80}}$$

$$= \frac{(-)\,6}{\sqrt{4} \times \sqrt{10}} \Rightarrow \frac{(-)\,\cancel{6}^{\,3}}{\cancel{2} \times 3.162}$$

$$= \frac{(-)\,3}{3.162} = (-)\,0.949$$

High Negative correlation

III $\boxed{\text{ASSUMED MEAN method}}$

$$r = \frac{\Sigma d_x d_y - \frac{\Sigma d_x . d_y}{N}}{\sqrt{\Sigma d_x^2 - \frac{(\Sigma d_x)^2}{N}} \times \sqrt{\Sigma d_y^2 - \frac{(\Sigma d_y)^2}{N}}}$$

eg :-

X	Y	$d_x = X - A$	$d_y = Y - A$	d_x^2	d_y^2	$d_x . d_y$
65	67	−3	2	9	4	−6
66	56	−2	−9	4	81	18
57	65 A	−11	0	121	0	0
67	68	−1	3	1	9	−3
A 68	72	0	7	0	49	0
69	72	1	7	1	49	7
70	69	2	4	4	16	8
72	71	4	6	16	36	24
		$\Sigma d_x = (-)10$	$\Sigma d_y = 20$	$\Sigma d_x^2 = 156$	$\Sigma d_y^2 = 244$	$\Sigma d_x d_y = 48$

Find r

Sol:-

Step	1	:-	d_x
Step	2	:-	d_y
Step	3	:-	d_x^2
Step	4	:-	d_y^2
Step	5	:-	$d_x . d_y$

$$r = \frac{\Sigma d_x . d_y - \dfrac{\Sigma d_x \times \Sigma d_y}{N}}{\sqrt{\Sigma d_x^2 - \dfrac{(\Sigma d_x)^2}{N}} \times \sqrt{\Sigma d_y^2 - \dfrac{(\Sigma d_y)^2}{N}}}$$

$$= \frac{48 - \dfrac{(-10) \times (20)}{8}}{\sqrt{156 - \dfrac{(-10)^2}{8}} \times \sqrt{244 - \dfrac{(20)^2}{8}}}$$

$$= \frac{48 - (-)25}{\sqrt{156 - 12.5} \times \sqrt{244 - 50}}$$

$$= \frac{48+25}{\sqrt{143.5} \times \sqrt{194}} = \frac{73}{11.97 \times 13.92}$$

$$= \frac{73}{166.62} \Rightarrow 0.438$$

Moderate positive correlation

IV $\boxed{\text{Step Deviation Method}}$

$$r = \frac{\Sigma d'_x d'_y - \frac{\Sigma d'_x \cdot \Sigma d'_y}{N}}{\sqrt{\Sigma {d'_x}^2 - \frac{(\Sigma d'_x)^2}{N}} \times \sqrt{\Sigma {d'_y}^2 - \frac{(\Sigma d'_y)^2}{N}}}$$

here $d'_x = \frac{X - A}{C}$ and $d'_y = \frac{Y - A}{C}$

Application & Steps same as Assumed Mean method

× —— × —— × —— × —— × —— × —— × —— ×

III Spearman's Rank correlation

→ Under Spearman's Rank correlation method, correlation value is based on Ranks

$$r = 1 - \frac{6 \Sigma D^2}{N(N^2-1)}$$

here D is deviation in Ranks

eg ① $\boxed{\text{When Ranks are given}}$

Contestants	Judge 1	Judge 2	D	D^2

Step 1 & only

Contestants	Judge 1	Judge 2	D	D^2
A	1	2	-1	1
B	2	3	-1	1
C	3	1	2	4
D	4	6	-2	4
E	5	4	1	1
F	6	5	1	1
G	7	8	-1	1
H	8	7	1	1
I	9	10	-1	1
J	10	11	-1	1
K	11	9	2	4
N = 11				$\Sigma d^2 = 20$

Step 1 only

$$r = 1 - \frac{6 \cdot \Sigma d^2}{N(N^2-1)}$$

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

$$= 1 - \frac{6 \times 20}{11^3 - 11}$$

$$= 1 - \frac{120}{1331 - 11}$$

$$= 1 - \frac{120}{1320}$$

$$= 1 - 0.091$$

$$= 0.909$$

eg 2 When Ranks are not given

Student	Marks (Accounts)	Marks (Economics)	R_1	R_2	D	D^2

Student	Marks (Accounts)	Marks (Economics)	R_1	R_2	D	D^2
Manas	21	02	4	4	0	0
Rohit	05	99	5	2	3	9
Gopal	31	100	3	1	2	4
Sonu	99	47	1	3	−2	4
Moksh	63	00	2	5	−3	9
						26

Step 1 :- Calculate R_1 and R_2

Step 2 :- D and D^2

Sol :-

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

$$= 1 - \frac{6 \times 26}{5^3 - 5}$$

$$= 1 - \frac{156}{125 - 5}$$

$$= 1 - \frac{156}{120} = 1 - 1.3$$

$$= \boxed{(-)\ 0.3}$$

× ——— × ——— × ——— × ——— × ——— × ——— ×

Important Questions

Q1 :- Is correlation affected by change in origin ?

All items (+) / (−)

+7

Ans $\boxed{\text{NO}}$, Correlation is not affected by change in origin

X	Y
2	10
3	12
4	13
5	15

$r = 0.83$

X	Y
9	17
10	19
11	20

9	17
10	19
11	20
12	22

$r = 0.83$

$r = 0.83$

Q2 - Is correlation affeted by change in scale ??

All items $\times$ or $\div$ ($\times 5$)

Ans - **NO**, it is not affected by change in Scale

X	Y
2	10
5	9
7	8
10	7

$r = (-)0.73$

X	Y
10	50
25	45
35	40
50	35

$r = (-)0.73$

Q3* If covariance between X & Y is $\boxed{12.3}$ and Variance of X and Y are 13.8 and 16.4 Find coefficient of correlation.

Sol:- Covariance between X & Y (COV(x, y)) $= \dfrac{\Sigma xy}{N} = \dfrac{\Sigma(X-\bar{X})(Y-\bar{Y})}{N}$

$$\text{Variance of } X = (S.D_x)^2 = 13.8$$

$$\text{Variance of } Y = (S.D_y)^2 = 16.4$$

$$r = \frac{\Sigma xy}{N} \times \frac{1}{S.D_x} \times \frac{1}{SD_y}$$

$$= 12.3 \times \frac{1}{\sqrt{13.8}} \times \frac{1}{\sqrt{16.4}}$$

$$= 12.3 \times \frac{1}{3.71} \times \frac{1}{4.05}$$

$$\ldots \times \frac{1}{3.71} \times \frac{1}{4.05}$$

$$= 12.3 \times 0.27 \times 0.25 = 0.83$$

Q4- $r = 0.8$, $\Sigma xy = 200$

$S.D_y = 5$, $\Sigma x^2 = 100$

Find (i) Σy^2

(ii) N

Sol:- (i)

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

$$0.8 = \frac{200}{\sqrt{100 \times \Sigma y^2}}$$

squaring both sides

$$0.64 = \frac{400\cancel{00}}{1\cancel{00} \times \Sigma y^2}$$

$$\Sigma y^2 = \frac{400}{0.64} = 625$$

(ii)

$$SD_y = \sqrt{\frac{\Sigma y^2}{N}}$$

$$5 = \sqrt{\frac{625}{N}}$$

squaring both sides

$$25 = \frac{625}{N}$$

$$\therefore N = \frac{\cancel{625}}{\cancel{25}} = 25$$

Q5 - Rank correlation of marks of 10 students was 0.5.

Q5 - Rank correlation of marks of 10 students was 0.5. Later it was discovered that difference in ranks for one item was wrongly taken as 3 instead of 7. Find correct rank correlation.

Sol :-

$$N = 10 \qquad r = 0.5$$

$$r = 1 - \frac{6\sum d^2}{N^3 - N}$$

$$0.5 = 1 - \frac{6\sum d^2}{10^3 - 10}$$

$$0.5 - 1 = \frac{-6\sum d^2}{1000 - 10}$$

$$\cancel{(-)}\,0.5 = \frac{\cancel{(-)}\,6\sum d^2}{990}$$

$$495 = 6\sum d^2$$

$$\sum d^2 = \frac{495}{6} = 82.5$$

$$\text{Correct } \sum d^2 = 82.5 - (3)^2 + (7)^2$$

$$= 82.5 - 9 + 49$$

$$= 122.5$$

$$\text{Correct } r = 1 - \frac{6\sum d^2}{N^3 - N}$$

$$= 1 - \frac{6 \times 122.5}{10^3 - 10}$$

$$= 1 - \frac{735}{990}$$

$$= 1 - 0.742 = \boxed{0.258}$$