CORRELATION

Definition:- The relationship between two variables such that a change in one variable results in a positive or negative change in the other variable and also a greater change in one variable results in corresponding greater ~~and~~ or smaller change in other variable is known as correlation.

$$X \xrightarrow[\text{Correlated}]{} Y$$

(degree of relationship)

Example:-

1. Advertisement and sales — Positive correlation
2. Price and Demand — Negative correlation
3. Income and Expenditure — positive
4. competition and sales — Negative
5. Price and supply — positive

Types of correlation

1. Positive Correlation

X	Y
↑	↑
↓	↓

2. Negative correlation

X	Y
↑	↓
↓	↑

Coefficient of correlation

The degree to which the two Variables are interrelated is measured by a coefficient which is called the coefficient of correlation.

It summarizes in one figure the DIRECTION and DEGREE of CORRELATION.

The coefficient of correlation between two Variables x and y is gene[illegible]lly denoted by r.

Properties of coeffi[illegible]t of correlation

The degree to which the two Variables are interrelated is measured by a coefficient which is called the coefficient of correlation.
It summarizes in one figure the DIRECTION and DEGREE of CORRELATION.
The coefficient of correlation between two Variables x and y is generally denoted by r.

Properties of coefficient of correlation

① $-1 \leq r \leq 1$

② $r=-1$ Perfect -ive correlation
$r=+1$ Perfect +ive correlation
$r=0$ No correlation

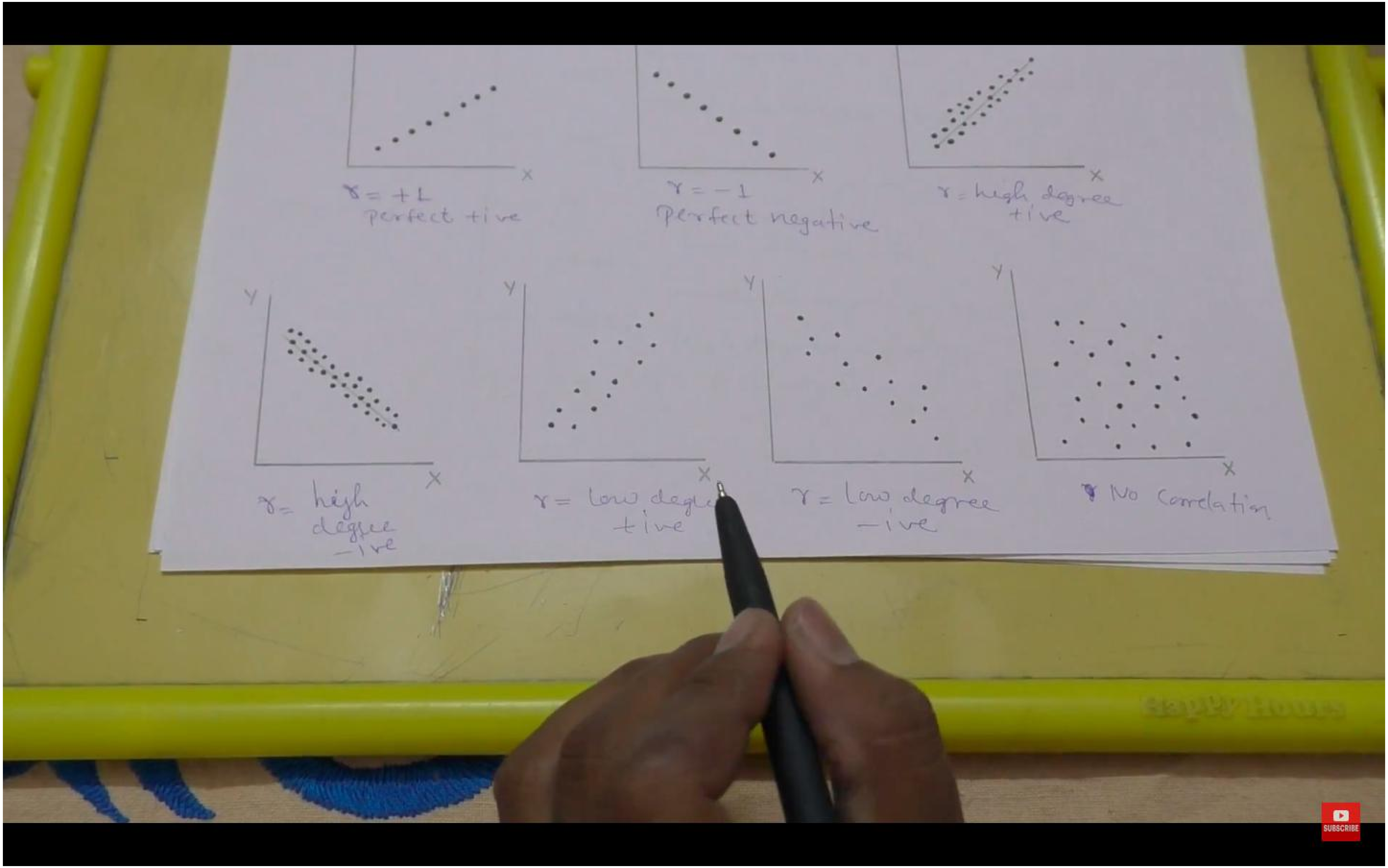
③ r is a pure no. (does not depends on units of variable)

④ r is Independent of change of origin and change of scale.

Interpretation of Value of coefficient of correlation (r)

Value of r

- +1 — Perfect positive
- +1 to 0.75 — High degree +ive
- 0.75 to 0.25 — moderate degree +ive
- 0.25 to 0 — Low degree +ive
- 0 — No correlation
- 0 to −0.25 — Low degree −ive
- −0.25 to −0.75 — moderate degree −ive
- −0.75 to −1 — High degree −ive
- −1 — Perfect negative



x
r = +1
perfect +ive
x
r = −1
Perfect negative
x
r = high degree
+ive
Y
x
r = high
degree
−ive
Y
x
r = low degree
+ive
Y
x
r = low degree
−ive
Y
x
r No Correlation
SUBSCRIBE

... their respective mean is known as Covariance.

Covariance, $cov(x,y) = \dfrac{\Sigma (x-\bar{x})(y-\bar{y})}{n}$

Var. = $(S.D.)^2$

Covariance

① Degree of relationship
② Cov. 0, -ive, +ive
③ absolute (depends on units of variable)

correlation coefficient

① ✓
② ✓
③ r is a pure no. (r is independent of units of variable)

Q. what will be we prefer? — — — — — — —

Coefficient of Correlation.

Karl Pearson's Coefficient of Correlation

Ist form :-

$$r = \frac{cov(x,y)}{\sqrt{var(x)}\sqrt{var(y)}}$$

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

$\bar{x}$ = mean of x- series
$\bar{y}$ = mean of y- series

IInd form:-

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

IIIrd form:-

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

$u = x - A_x$, A_x = Assumed mean in x-series
$v = y - A_y$, A_y = Assumed mean in y-series

BeingGourav.com ①

Karl Pearson's Coefficient of Correlation

Ist form :-

$$r = \frac{cov(X,Y)}{\sqrt{var(X)}\,\sqrt{var(Y)}}$$

$$r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2}\,\sqrt{\Sigma(Y-\bar{Y})^2}}$$

X	Y	$(X-\bar{X})$	$(Y-\bar{Y})$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
ΣX	ΣY	$\Sigma(X-\bar{X})$	$\Sigma(Y-\bar{Y})$	$\Sigma(X-\bar{X})$	$\Sigma(Y-\bar{Y})$	$\Sigma(X-\bar{X})(Y-\bar{Y})$

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

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

$\bar{X}$ = mean of X- series

$\bar{Y}$ = mean of Y- series

IInd form:-

Direct method

$$r = \frac{n\Sigma XY - \Sigma X \Sigma Y}{\sqrt{n\Sigma X^2 - (\Sigma X)^2}\,\sqrt{n\Sigma Y^2 - (\Sigma Y)^2}}$$

X	Y	X^2	Y^2	XY
ΣX	ΣY	ΣX^2	ΣY^2	ΣXY

IIIrd form:-

Shortcut method

$$r = \frac{n\Sigma uv - \Sigma u \Sigma v}{\sqrt{n\Sigma u^2 - (\Sigma u)^2}\,\sqrt{n\Sigma v^2 - (\Sigma v)^2}}$$

X	Y	$u = X - A_x$	$V = Y - A_y$	u^2	v^2	uv
		Σu	Σv	Σu^2	Σv^2	Σuv

$u = X - A_x$, A_x = Assumed mean in X-series

$V = Y - A_y$, A_y = Assumed mean in Y-series

Q1. Calculate Karl Pearson's coefficient of correlation from following data:

- X : 10 6 9 10 12 13 11 9
  Y : 9 4 6 9 11 13 8 4

$n = 8$

Solution:-

X	Y	X^2	Y^2	XY
10	9	100	81	90
6	4	36	16	24
9	6	81	36	54
10	9	100	81	90
12	11	144	121	132
13	13	169	169	169
11	8	121	64	88
9	4	81	16	36
$\Sigma X = 80$	$\Sigma Y = 64$	$\Sigma X^2 = 832$	$\Sigma Y^2 = 584$	$\Sigma XY = 683$

we know that

$$r = \frac{n\Sigma XY - \Sigma X \Sigma Y}{\sqrt{n\Sigma X^2 - (\Sigma X)^2}\sqrt{n\Sigma Y^2 - (\Sigma Y)^2}}$$

$$r = \frac{8(683) - (80)(64)}{\sqrt{8(832) - (80)^2}\sqrt{8(584) - (64)^2}}$$

$$r = \frac{5464 - 5120}{\sqrt{6656 - 6400}\sqrt{4672 - 4096}}$$

$$r = \frac{344}{\sqrt{256}\sqrt{576}}$$

$$r = \frac{\overset{86}{\cancel{344}}}{\underset{4}{\cancel{16}} \times 24} =$$

$$\boxed{r = 0.896}$$

High degree
tive correlation.

Spearman's Rank Correlation method

(I) when Ranks are Given

$-1 \le R \le 1$

Q. Calculate Rank Correlation from following data:

R_x	R_y	$D = R_x - R_y$	D^2
2	1	1	1
1	3	-2	4
4	2	2	4
3	4	-1	1
5	5	0	0
7	6	1	1
6	7	-1	1
			$\Sigma D^2 = 12$

Charles Edward Spearman - 1904

$$R = 1 - \frac{6 \Sigma D^2}{n^3 - n}$$

$$R = 1 - \frac{6 \times 12}{7^3 - 7}$$

$$R = 1 - \frac{72}{343 - 7}$$

$$R = 1 - \frac{72}{336} = 1 - 0.214$$

$$\boxed{R = 0.786}$$

SUBSCRIBE

Spearman's Rank Correlation method

(I) when Ranks are Given

Q. Calculate Rank Correlation from following data:

Rank by Judge A	Rank by Judge B	Rank by Judge C	$D_{AB} = R_A - R_B$	D_{AB}^2	$D_{BC} = R_B - R_C$	D_{BC}^2	$D_{AC} = R_A - R_C$	D_{AC}^2
1	3	6	−2	4	−3	9	−5	25
6	5	4	1	1	1	1	2	4
5	8	9	−3	9	−1	1	−4	16
10	4	8	6	36	−4	16	2	4
3	7	1	−4	16	6	36	2	4
2	10	2	−8	64	8	64	0	0
4	2	3	2	4	−1	1	1	1
9	1	10	8	64	−9	81	−1	1
7	6	5	1	1	1	1	2	4
8	9	7	−1	1	2	4	1	1
				$\Sigma D_{AB}^2 = 200$		$\Sigma D_{BC}^2 = 214$		$\Sigma D_{AC}^2 = 60$

Spearman's Rank Correlation method

(I) when Ranks are Given

$\boxed{n=10}$

(i) Rank Correlation

$$R_{AB} = 1 - \frac{6 \Sigma D_{AB}^2}{n^3 - n}$$

$$= 1 - \frac{6 \times 200}{10^3 - 10}$$

$$= 1 - \frac{1200}{990}$$

$$= 1 - \frac{120}{99}$$

$$= 1 - 1.212$$

$$\boxed{R_{AB} = -0.212} \quad \text{— (1)}$$

D_{AB}^2	$D_{BC} = R_B - R_C$	D_{BC}^2	$D_{AC} = R_A - R_C$	D_{AC}^2
4	−3	9	−5	25
1	1	1	2	4
9	−1	1	−4	16
36	−4	16	2	4
16	6	36	2	4
64	8	64	0	0
4	−1	1	1	1
64	−9	81	−1	1
1	1	1	2	4
1	2	4	1	1
$\Sigma D_{AB}^2 = 200$		$\Sigma D_{BC}^2 = 214$		$\Sigma D_{AC}^2 = 60$

Spearman's Rank Correlation method

(I) when Ranks are Given

$\boxed{n = 10}$

(ii) Rank Correlation

$$R_{BC} = 1 - \frac{6 \sum D_{BC}^2}{n^3 - n}$$

$$= 1 - \frac{6 \times 214}{10^3 - 10}$$

$$= 1 - \frac{1284}{990}$$

$$= 1 - 1.296$$

$$\boxed{R_{BC} = -0.296} \quad \text{— } 2$$

$$\boxed{R_{AB} = -0.212} \quad \text{— } 1$$

D_{AB}^2	$D_{BC} = R_B - R_C$	D_{BC}^2	$D_{AC} = R_A - R_C$	D_{AC}^2
4	−3	9	−5	25
1	1	1	2	4
9	−1	1	−4	16
36	−4	16	2	4
16	6	36	2	4
64	8	64	0	0
4	−1	1	1	1
64	−9	81	−1	1
1	1	1	2	4
1	2	4	1	1
$\sum D_{AB}^2 = 200$		$\sum D_{BC}^2 = 214$		$\sum D_{AC}^2 = 60$

Spearman's Rank Correlation method

(I) when Ranks are given

$\boxed{n=10}$

(iii) Rank Correlation

$$R_{AC} = 1 - \frac{6 \Sigma D_{AC}^2}{n^3 - n}$$

$$R_{AC} = 1 - \frac{6 \times 60}{10^3 - 10}$$

$$= 1 - \frac{360}{990}$$

$$\boxed{R_{AC} = 1 - 0.36 = 0.64} \quad - (3)$$

$$\boxed{R_{BC} = -0.296} \quad - (2)$$

$$\boxed{R_{AB} = -0.212} \quad - (1)$$

D_{AB}^2	$D_{BC} = R_B - R_C$	D_{BC}^2	$D_{AC} = R_A - R_C$	D_{AC}^2
4	−3	9	−5	25
1	1	1	2	4
9	−1	1	−4	16
36	−4	16	2	4
16	6	36	2	4
64	8	64	0	0
4	−1	1	1	1
64	−9	81	−1	1
1	1	1	2	4
1	2	4	1	1
$\Sigma D_{AB}^2 = 200$		$\Sigma D_{BC}^2 = 214$		$\Sigma D_{AC}^2 = 60$

Spearman's Rank Correlation method

(II) when Ranks are not given

Q. Calculate Spearman's Rank correlation coefficient from the following data:-

$\boxed{n = 8}$

Price of tea (in Rs.) (X)	Price of coffee (in Rs.) (Y)	R_X	R_Y	$D = R_X - R_Y$	D^2
75	120	5	5	0	0
88	134	2	4	−2	4
95	150	1	1	0	0
70	115	6	6	0	0
60	110	7	7	0	0
80	140	4	3	1	1
81	142	3	2	1	1
50	100	8	8	0	0
					$\Sigma D^2 = 6$

$$R = 1 - \frac{6 \Sigma D^2}{n^3 - n}$$

$$R = 1 - \frac{6 \times 6}{8^3 - 8}$$

$$R = 1 - \frac{36}{512 - 8}$$

$$R = 1 - \frac{36}{504}$$

$$R = 1 - 0.071$$

$$\boxed{R = 0.929}$$

Spearman's Rank Correlation method

(III) when Ranks are not given and Repeated

Q. Calculate Spearman's Rank correlation coefficient from the following data:-

X	Y	R_x	R_y	$D = R_x - R_y$	D^2
68	62	4			
64	58	6			
75	68	2.5			
50	45	9			
64	81	6			
80	60	1			
75	68	2.5			
40	48	10			
55	50	8			
64	70	6			

$$R = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \ldots\right]}{n^3-n}$$

In x-series, 75 repeated 2 times, $m=2$

In x-series, 64 repeated 3 times, $m=3$

$$\frac{2+3}{2} = \frac{5}{2} = 2.5$$

$$\frac{5+6+7}{3} = \frac{18}{3} = 6$$

Spearman's Rank Correlation method

$3 \times 3 \times 3 - 3$

(III) when Ranks are not given and Repeated

Q. Calculate Spearman's Rank correlation coefficient from the following data:-

X	Y	R_x	R_y	$D = R_x - R_y$	D^2
68	62	4	5	-1	1
64	58	6	7	-1	1
75	68	2.5	3.5	-1	1
50	45	9	10	-1	1
64	81	6	1	5	25
80	60	1	6	-5	25
75	68	2.5	3.5	-1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16
					$\Sigma D^2 = 72$

$$R = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \ldots\right]}{n^3-n}$$

In X-series, 75 repeated 2 times, $m=2$

In X-series, 64 repeated 3 times, $m=3$

In Y-series, 68 repeated 2 times, $m=2$

$$R = 1 - \frac{6\left[72 + \frac{1}{12}(2^3-2) + \frac{1}{12}(3^3-3) + \frac{1}{12}(2^3-2)\right]}{10^3-10}$$

$$R = 1 - \frac{6[72 + 0.5 + 2 + 0.5]}{990}$$

$$R = 1 - \frac{6 \times 75}{990}$$

$$R = 1 - \frac{450}{990}$$

$$R = 1 - 0.45$$

$$\boxed{R = 0.55}$$