

Correlation & Regression

correlation

→ Degree of association between two variable

→ correlation can be negative positive or zero

→ $-1 \leq r \leq 1$

→ for Direction Relation b/w x & y

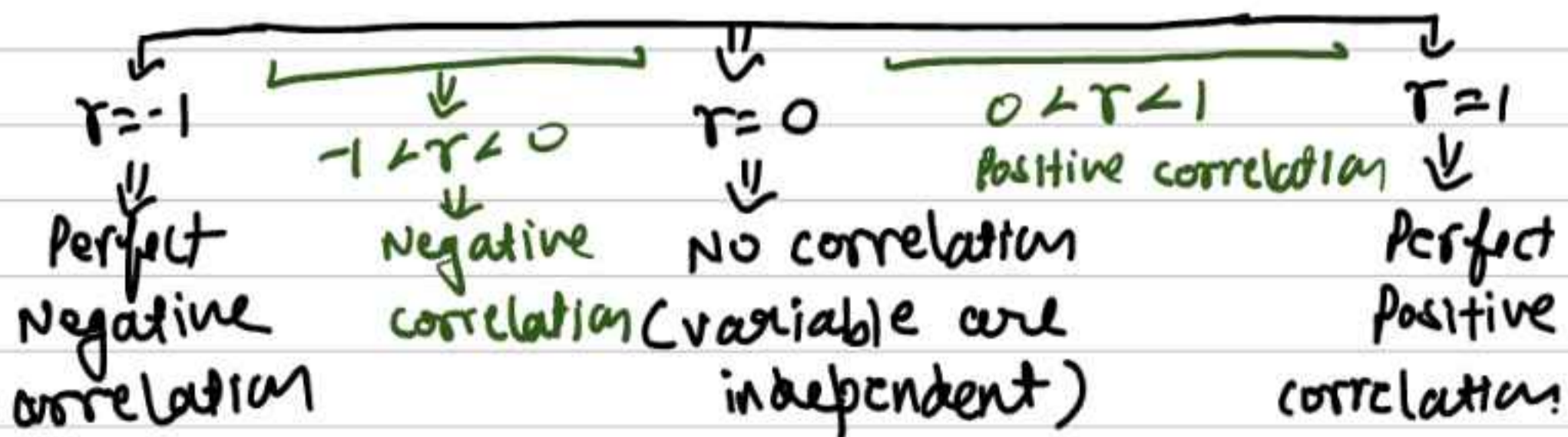
$x \uparrow y \uparrow$ or $x \downarrow y \downarrow$

r is positive

→ for inverse Relation b/w x & y

$x \downarrow y \uparrow$ or $x \uparrow y \downarrow$

r is negative



Methods of Calculating Correlation

Graphical method

↓
Scattered Diagram method

non graphical method

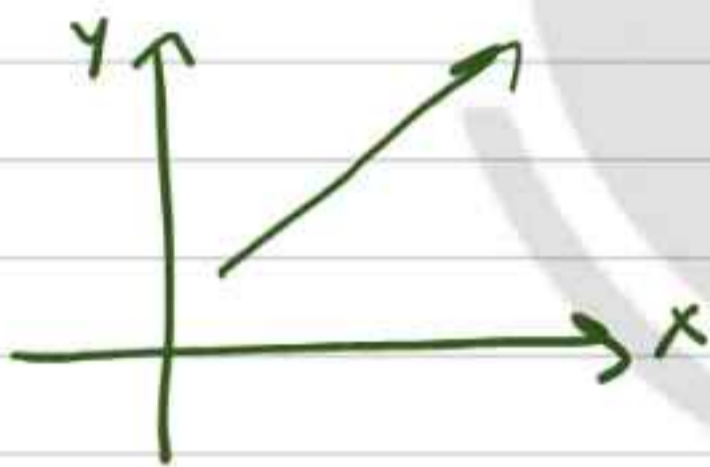
Karl Pearson method

Spearman method

Concurrent Deviation

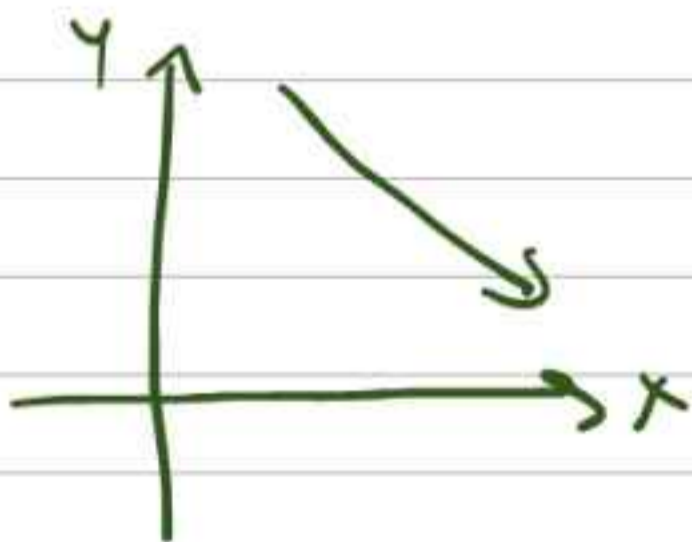
Scattered Diagram method

Plot the values of two variable on graph



For straight line
(Lower left to upper right)

$$r = 1$$



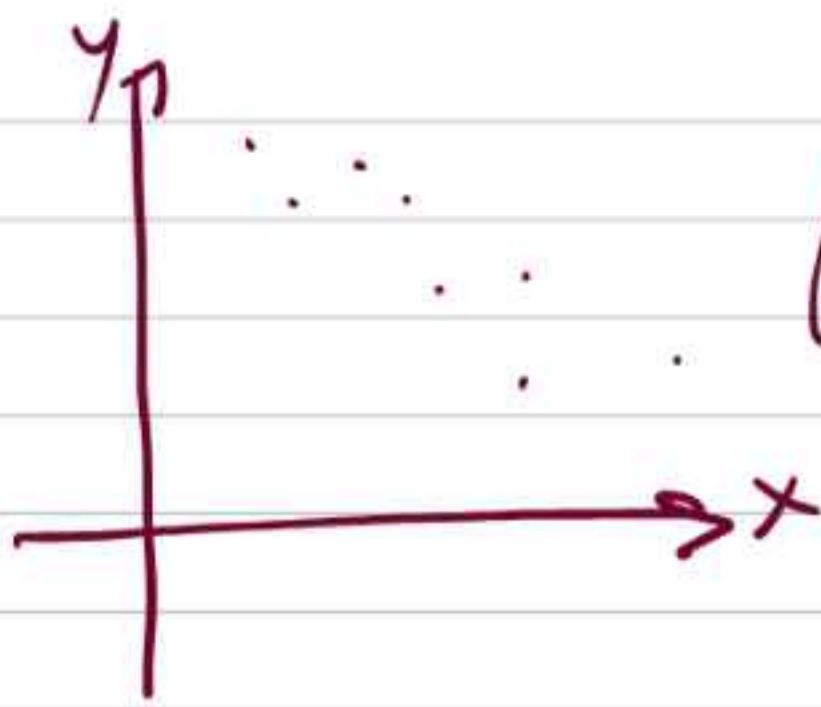
For straight line
(Upper left to Lower Right)

$$r = -1$$



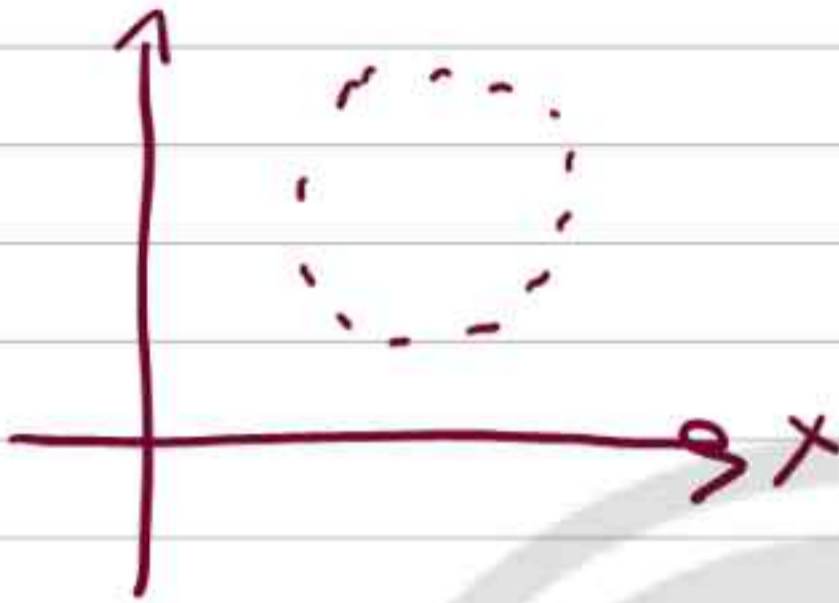
$$0 < r < 1$$

(No straight line)



$$-1 < r < 0$$

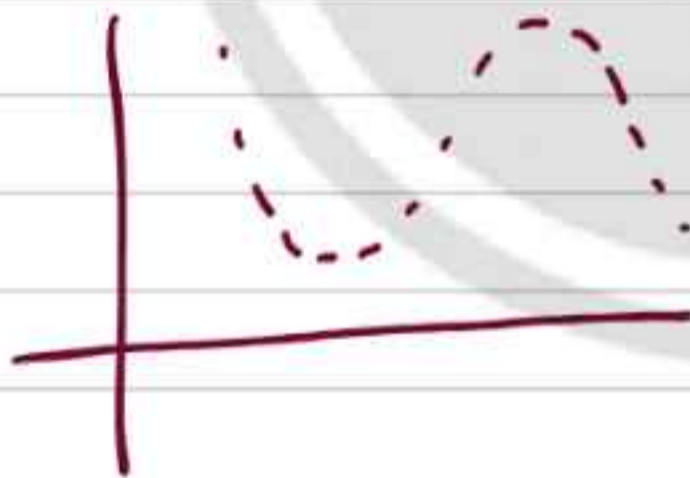
No straight
line



$$r = 0$$



$$r = 0$$



$$r = 0$$

→ Exact magnitude can not be calculated
in scattered diagram method

Karl Pearson's coefficient correlation

$$r = \frac{\text{cov.}(X, Y)}{\sigma_x \sigma_y}$$

or

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

$$r = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \times \sum (y_i - \bar{y})^2}}$$

or

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

→ covariance
= $\text{cov}(X, Y)$
= $\frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{N}$

→ $\text{cov}(X, Y)$ can be any real number (negative, positive & zero)

→ $\text{cov}(X, Y)$ does not change with origin but it changes with change of scale

→ Correlation does not change with change of origin

i.e. $u_i = x_i - A$ & $v_i = y_i - B$

then

$$r = \frac{\sum UV - \frac{\sum U \times \sum V}{N}}{\sqrt{\sum U^2 - \frac{(\sum U)^2}{N}} \sqrt{\sum V^2 - \frac{(\sum V)^2}{N}}}$$

→ Correlation does not change with change of scale (provided scale is positive)

→ when scale is negative, the sign of correlation may change but magnitude does not change

g
if $r(x, y) = 0.5$
if $u_i = 2x_i$ & $v_i = 4y_i$

then $r(u, v) = (+)(+)0.5 = 0.5$

g
if $r(x, y) = 0.6$
if $u_i = -2x_i$ & $v_i = 3y_i$

then $r(u, v) = (-)(+)0.6 = -0.6$

g
if $r(x, y) = 0.7$
if $u_i = -2x_i$ & $v_i = -4y_i$

then $r(u, v) = (-)(-)0.7 = (+)0.7$

Spearman's Rank Correlation

→ This method used for qualitative characters & level of agreements & disagreements b/w opinions of Judges

→ when no numbers repeat

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

→ when some numbers repeat

$$r = 1 - \frac{6 \left[\sum D^2 + \frac{1}{12} (m_1^3 - m_1) + \frac{1}{12} (m_2^3 - m_2) \right]}{N^3 - N}$$

→ $\sum D = 0$

X	Y	R ₁	R ₂	D = R ₁ - R ₂	D ²
				0	

Concurrent Deviation method

Concurrent Deviation



when x & y both increase
or both decrease

$$r = \pm \sqrt{\pm \left(\frac{2C - m}{m} \right)}$$

where $C =$ Total no of concurrent deviations
 $m = n - 1$

	X	Y		Concurrent Deviations
+	10	12	+	Yes (+)
+	12	15	-	No (-)
-	15	14	+	No (-)
+	14	15	+	Yes (+)
+	16	18	+	No (-)
-	12	20	+	Yes (+)

$$C = 2$$

$$m = 6 - 1 = 5$$

$$r = \sqrt{\frac{2(2) - 5}{5}}$$

→ when $\frac{2C - m}{m}$ is negative

then (-) sign will be taken
out from $\sqrt{\quad}$ sign

$$r = \sqrt{\frac{-1}{5}}$$

$$r = -\sqrt{\frac{1}{5}}$$

Bivariate Frequency Distribution Table

marks \ stats	0-5	5-10	10-15	15-20	Total
0-10	2	5	1	0	8
10-20	3	2	3	4	12
20-30	4	1	2	6	13
Total	9	8	6	10	33

Frequency Distribution of marks of maths

Marks	No of Students
0-10	8
10-20	12
20-30	13
	<u>33</u>

This is called marginal distribution

Frequency Distribution of marks in maths when score in stats is '5-10'

marks	No of Students
0-10	5
10-20	2
20-30	1
	<u>8</u>

This is called conditional distribution

For Bivariate distribution table of "m x n"

Total cells = $m \times n$

Total marginal distribution = 2

Total conditional distribution = $m + n$

Coefficient of Determination
 $= r^2 = \frac{\text{Explained variance}}{\text{Total variance}}$

Coefficient of Non Determination
 $= 1 - r^2$

Regressions

- Establishing mathematical relation b/w two variables (Independent + Dependent)
- Prediction of dependent variable
- For Linear Regression Least Square method is used

There two Linear Regression lines

when y depends on x

when x depends on y

$$y = a + bx$$

$$x = a + by$$

Standard form

How to find = ?

use formula

use formula

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

slope of line

$$\text{slope} = b_{yx}$$

$$\text{slope} = \frac{1}{b_{xy}}$$

b_{yx} & b_{xy} are known as regression coefficients

Calculation of Regression coefficients

$$b_{yx} = \frac{\text{cov}(x, y)}{\sigma_x^2}$$

$$b_{yx} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$b_{yx} = \frac{\sum xy - \frac{\sum x \sum y}{N}}{\sum x^2 - \frac{(\sum x)^2}{N}}$$

$$b_{yx} = r \frac{\sigma_y}{\sigma_x}$$

$$b_{xy} = \frac{\text{cov}(x, y)}{\sigma_y^2}$$

$$b_{xy} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum (y_i - \bar{y})^2}$$

$$b_{xy} = \frac{\sum xy - \frac{\sum x \sum y}{N}}{\sum y^2 - \frac{(\sum y)^2}{N}}$$

$$b_{xy} = r \frac{\sigma_x}{\sigma_y}$$

#

$$r = \pm \sqrt{b_{yx} \times b_{xy}}$$

⇓

→ 'r' will be positive if both b_{yx} & b_{xy} are positive

→ 'r' will be negative if both b_{yx} & b_{xy} are negative.

→

$$b_{yx} \times b_{xy} \leq 1$$

#

$$r \leq \frac{b_{yx} + b_{xy}}{2}$$

If Regression line

y on x is: $ax + by + c = 0$

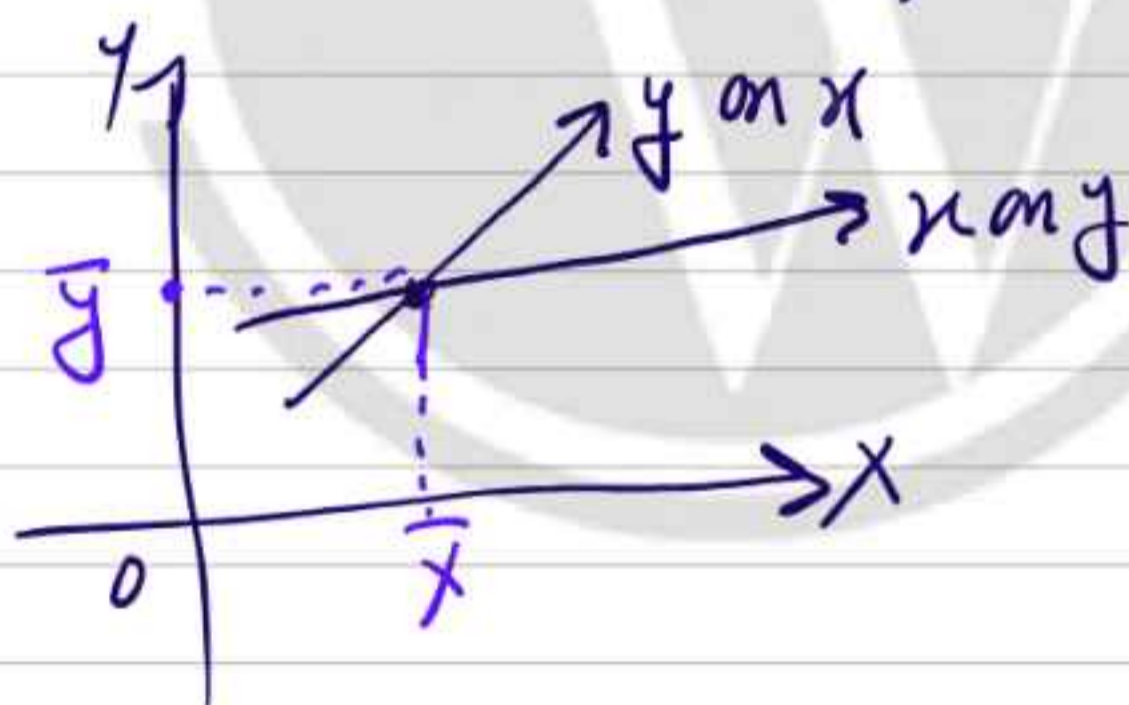
then $b_{yx} = -\frac{a}{b}$

If Regression line

x on y is $Ax + By + C = 0$

then $b_{xy} = -\frac{B}{A}$

Two Regression lines intersect each other at (\bar{x}, \bar{y})



Regression coefficients does not change with change of origin

$$u_i = x_i - A \quad \&$$

$$v_i = y_i - B$$

then $b_{yx} = b_{vu} \quad \& \quad b_{xy} = b_{uv}$

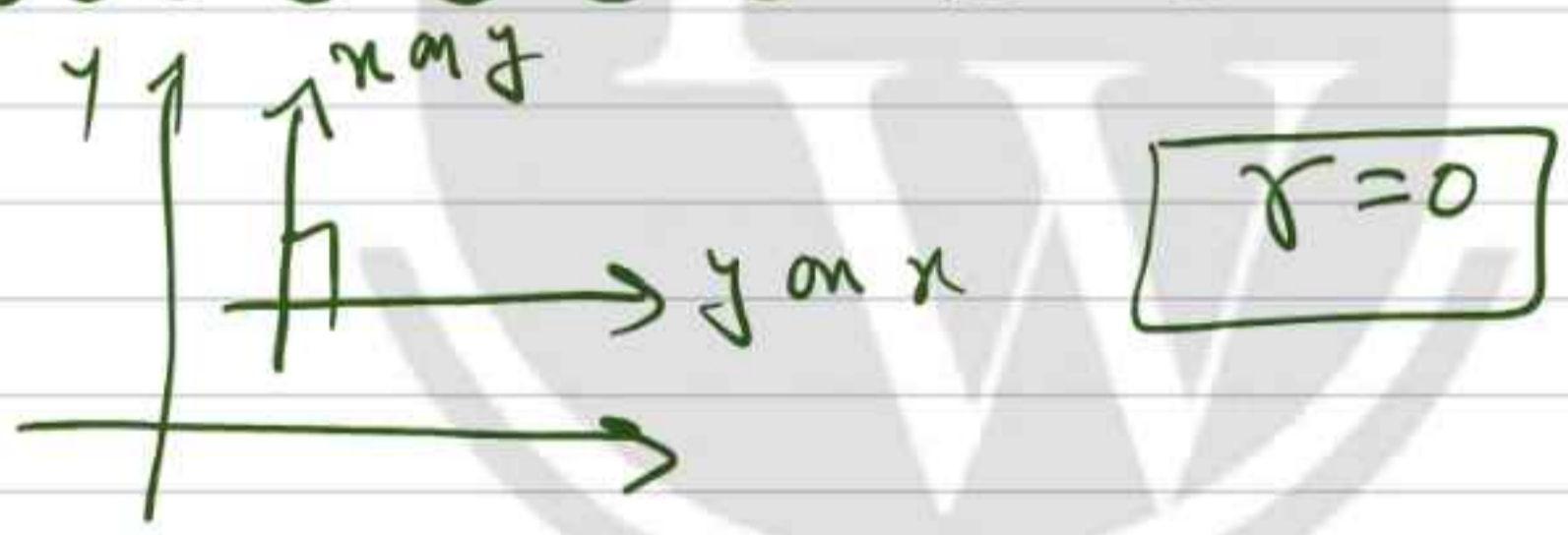
Regression coefficients changes with change of scale

$$U_i = a x_i + b \quad \& \quad v_i = c y_i + d$$

So
$$b_{vu} = \frac{c}{a} \times b_{yx}$$

$$b_{vu} = \frac{\text{Scale of } y}{\text{Scale of } x} \times b_{yx}$$

When two lines are perpendicular



When two lines are coincident

