

# CORRELATION

## Overview

- ① Univariate Data & Bivariate Data
- ② Meaning of Correlation & Its Method
- ③ Scatter Diagram
- ④ Karl Pearson's Method.
- ⑤ Spearman's Rank Correlation
- ⑥ Method of Concurrent Deviation
- ⑦ Miscellaneous Points.

## ① Univariate & Bivariate

When you collect data they are  
of Two Types



### Univariate Data

When you analyze  
a Single Data  
i.e. Only **One Variable**  
**at a time.**

M.O.C.T &

M.O.D

### Bivariate Data

When data is  
collected for  
**Two Variable &**  
Analyzed at  
**SAME Time**

**Correlation &**  
**Regression**

→ Bivariate Data.

① When FD is prepared with Bivariate  
Data it is called as Bivariate FD/  
Joint FD / Two Way Distribution

② No of Cells =  $m \times n$   
where  $m = \text{Rows}$   
 $n = \text{Columns}$

### ③ Marginal Distribution

① Marginal Means Total

② No of Marginal Distributions = 2  
→ MD of Stats  
↗

→ MD of Maths

MD  
of Stats

4  
12  
16  
11  
9  
—  
50

Class

0-4  
4-8  
8-12  
12-16  
16-20

MD  
of Maths

50  
50  
—  
50

### ④ Conditional Distribution

Marks in Stats when student scores  
(4-8) Maths

| <u>Class</u> | <u>f</u> |
|--------------|----------|
| 0-4          | 1        |
| 4-8          | 4        |
| 8-12         | 2        |
| 12-16        | 1        |
| 16-20        | 0        |
| Total        | <u>8</u> |

Marks of Maths when students  
score (12-16) in Statistics

|       | <u>f</u>  |
|-------|-----------|
| 0-4   | 0         |
| 4-8   | 1         |
| 8-12  | 3         |
| 12-16 | 2         |
| 16-20 | 5         |
|       | <u>11</u> |



No of Conditional  
Dist<sup>n</sup> =  $m + n$

$m$  = No of Row CI

$n$  = No of Column CI

(2) Meaning of Correlation & Its  
Method

Demand  
 $x$   
Qty

Price  
 $y$  (₹)

80

20

90

18

100

16

100

20

110

18

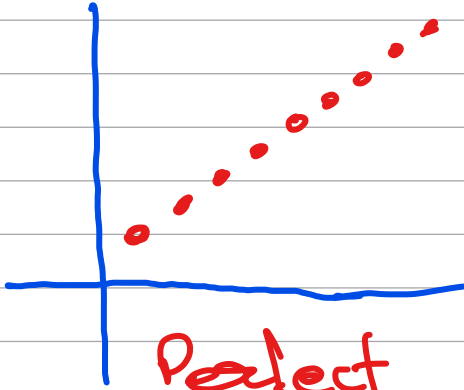
- (i) If change in **One Variable** causes changes in **Another Variable** **than** Both the variables are said to be **CORRELATED**.
- (ii) Correlation is the **Degree of Association** B/w Two/More Variables
- (iii) B/w two variables, they can be either **Positively Correlated** / **Negatively Correlated** / **No Correlation**
- (iv) Spurious / Nonsense Correlation.  
Ex. Shoe Size & Intelligence
- 
- The diagram illustrates a spurious correlation. It shows three entities: 'Tu', 'Babu', and 'Babu Ka Bestie'. 'Tu' and 'Babu' are connected by a double-headed arrow, indicating a correlation. 'Babu' is connected to 'Babu Ka Bestie' by a single arrow, and 'Babu Ka Bestie' is connected back to 'Tu' by a large red arrow. A red arrow points from 'Babu' to the text 'Lucky Variable'. A large red scribble underlines the entire diagram with the word 'Nonsense' written in the center, indicating that the correlation between 'Tu' and 'Babu' is spurious.

- ✓ Methods
- (a) Scatter Diagram
  - (b) Karl Pearson
  - (c) Spearman's Rank
  - (d) Concurrent Deviation Method

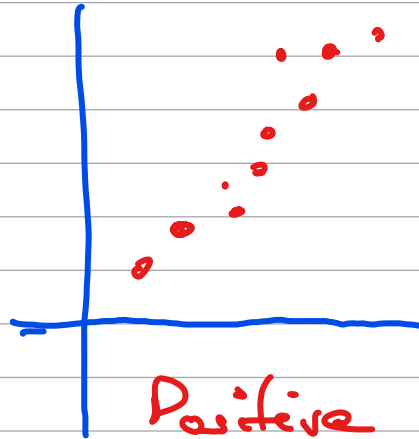
### (3) SCATTER DIAGRAM

- (i) It is Diagrammatic Representation of Bivariate Data
- (ii) It gives Nature of Relation  
But fails to Measure Extent of Association
- (iii) Suitable for linear as well as Curvilinear Relationship

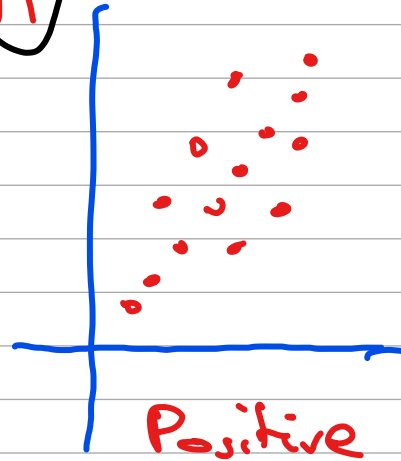
- (iv) Types of - Scatter Diagram
- $r = +ve$



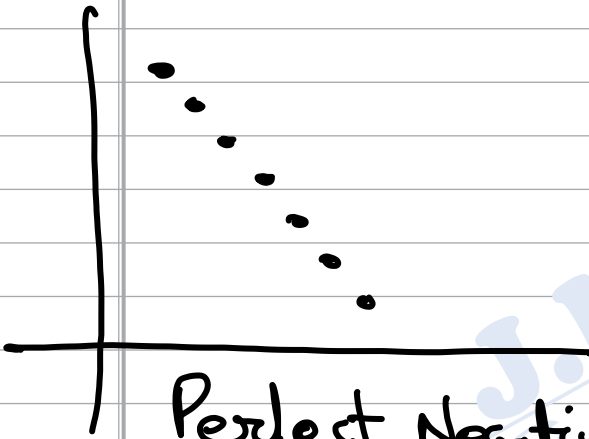
Perfect  
Positive  
 $r = +1$



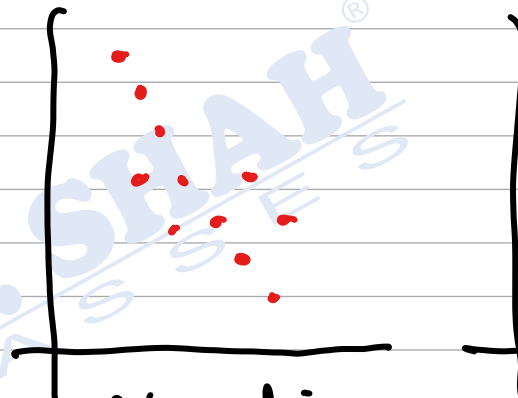
Positive  
(High Degree)



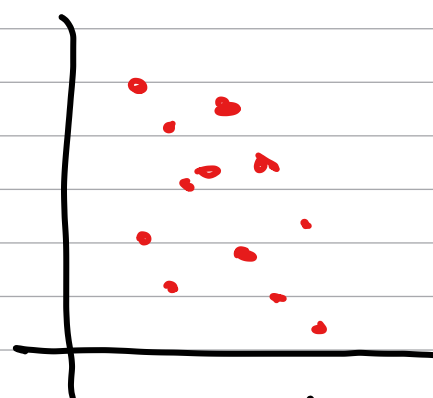
Positive  
(Low Degree)



Perfect Negative  
 $r = -1$

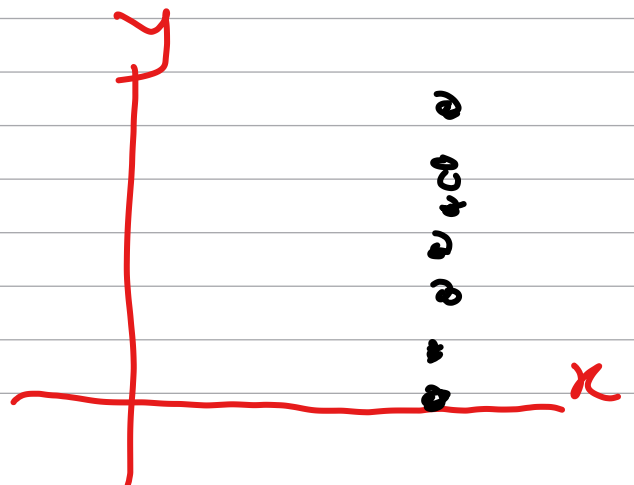
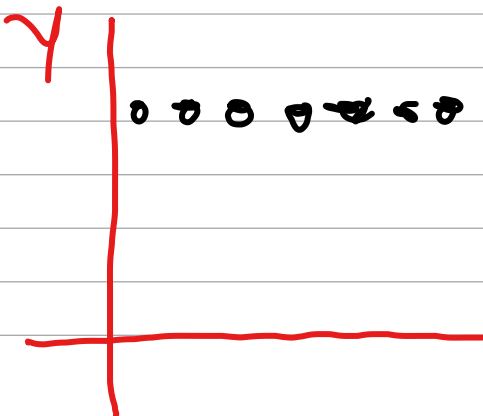


Negative  
(High Degree)



Negative  
(Low Degree)

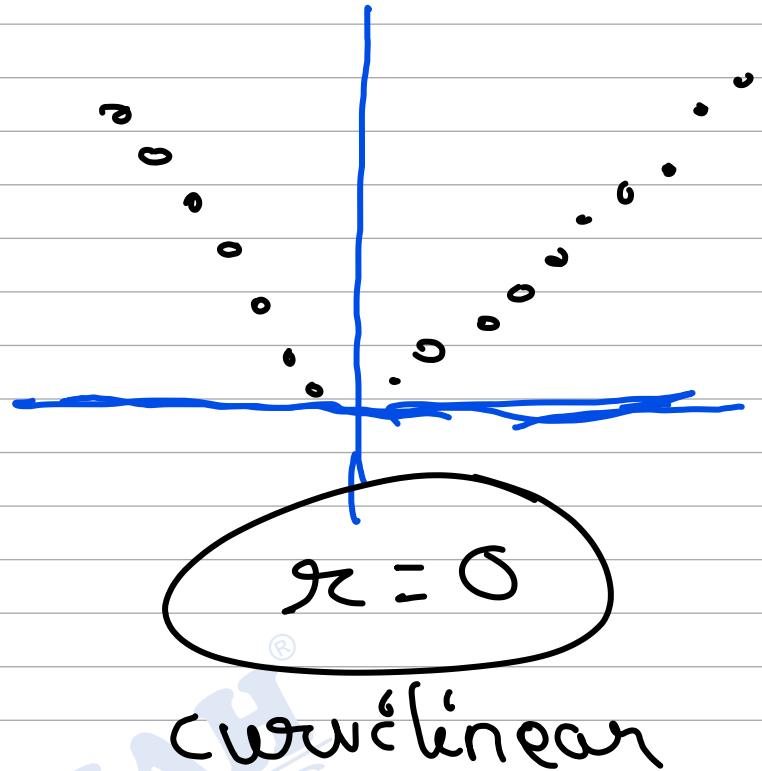
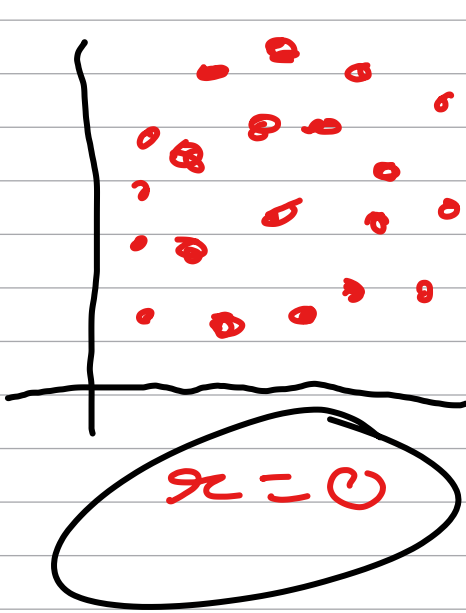
$r = 0$



$x \uparrow y$  constant

**J.K. SHAH**  
CLASSES

$y \uparrow x$  constant



## ④ KARL PEARSON'S CORRELATION'S COEFFICIENT

① Also known as Product  
Moment Correlation

② Also known as Correlation  
Coefficient By Covariance  
Method

- ③ Suitable for Quantitative Measures
- ④ Suitable for Linear Relationships Only.
- ⑤ It is Ratio of Covariance B/w  $x$  &  $y$  to product of their Standard deviation
- ⑥ It is denoted by " $r$ ".

⑦ Formula:-

⑧ 
$$r = \frac{\text{Cov}(x, y)}{\sigma_x \times \sigma_y} = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var } x} \cdot \sqrt{\text{Var } y}}$$

where

$$\text{Cov}(x, y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{n}$$

(or)

$$\text{Cor}(x, y) = \frac{\sum xy}{n} - \left( \frac{\sum x}{n} \right) \left( \frac{\sum y}{n} \right)$$

$\bar{x} \cdot \bar{y}$

$$\sigma_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} \quad \text{(or)} \quad \sigma_x = \sqrt{\frac{\sum x^2}{n} - (\bar{x})^2}$$

$$\sigma_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n}} \quad \text{(or)} \quad \sigma_y = \sqrt{\frac{\sum y^2}{n} - (\bar{y})^2}$$

ii

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

or

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

## Properties of 'r'

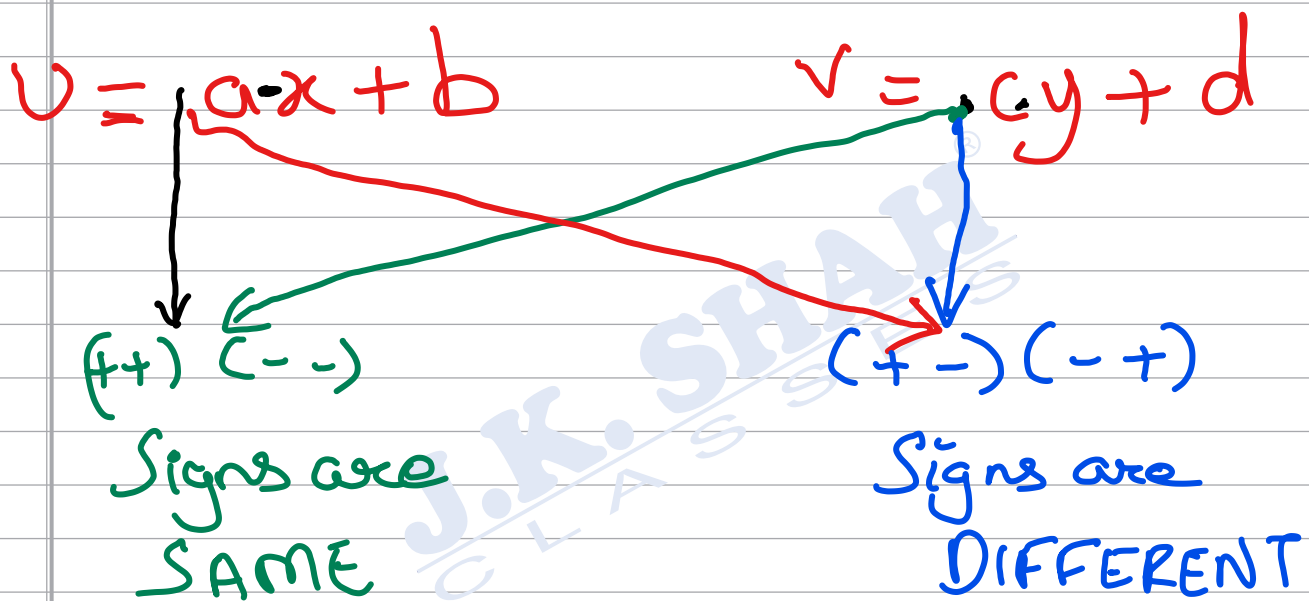
- ①  $r$  is commonly known as correlation coefficient.  
It is a Relative Measure  
Hence it is unit free.

②  $-1 \leq r \leq 1$   
 $0 \leq r^2 \leq 1$

- ③ 'r' is unaffected By  
Both Shift of Origin &  
Change of Scale.



④ " $r$ " is AFFECTED By  
"SIGN" of change of Scale.



$$+ r_{xy} = + r_{uv}$$

OR

$$- r_{xy} = - r_{uv}$$

$$r_{xy} = - r_{uv}$$

$$- r_{xy} = r_{uv}$$

Ex.

$$r_{xy} = 0.65$$

$$V = 5 - 2x$$

$$v = 10 + 3y$$

ve +ve

Different

$$r_{xy} = -r_{uv} = \underline{\underline{-0.65}}$$

Q.1

$$r = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y}$$

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

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

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

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{(\cancel{\sqrt{n} \cdot \sqrt{n}}) \frac{\sqrt{\sum (x - \bar{x})^2} \cdot \sigma_y}{\cancel{\sqrt{n}}}}$$

$$n = \sqrt{n} \times \sqrt{n}$$

$$0.25 = \frac{60}{\sqrt{n} \sqrt{90 \cdot 4}}$$

$$\sqrt{n} = \frac{60}{\sqrt{90 \cdot 4 \times 0.25}}$$

$$\sqrt{n} = \frac{60}{\sqrt{90}} \quad \text{Squaring on b.t.s}$$

$$n = \frac{3600}{90} = \textcircled{40}$$

Q.2  $\text{cov}(x, y) = \frac{-235}{64} = \textcircled{c}$   
-3.67

$$\underline{Q.3} \quad \text{Cov}(x, y) = \frac{\sum xy}{n} - \left( \frac{\sum x}{n} \right) \left( \frac{\sum y}{n} \right)$$

$$= \frac{411}{11} - \left( \frac{55}{11} \right) \left( \frac{74}{11} \right) = 3.727 \quad \text{(d)}$$

$$\underline{Q.4} \quad r = \frac{\text{Cov}(x, y)}{\sqrt{\text{var } x} \sqrt{\text{var } y}}$$

$$r = \frac{10}{\sqrt{6.25} \sqrt{31.36}} = 0.714 \quad \text{(a)}$$

$$\underline{Q.5} \quad r = \frac{25}{\sqrt{36} \sqrt{25}} = 0.8333 \quad \text{(c)}$$

Q.6

| <u>x</u>   | <u>y</u>   | <u>x<sup>2</sup></u> | <u>y<sup>2</sup></u> | <u>xy</u>   |
|------------|------------|----------------------|----------------------|-------------|
| 1          | 6          | 1                    | 36                   | 6           |
| 2          | 8          | 4                    | 64                   | 16          |
| 3          | 11         | 9                    | 121                  | 33          |
| 4          | 8          | 16                   | 64                   | 32          |
| 5          | 12         | 25                   | 144                  | 60          |
| <u>15</u>  | <u>45</u>  | <u>55</u>            | <u>429</u>           | <u>147</u>  |
| $\Sigma x$ | $\Sigma y$ | $\Sigma x^2$         | $\Sigma y^2$         | $\Sigma xy$ |

$n = 5$

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

$$r = \frac{5(147) - (15)(45)}{\sqrt{5(55) - (15)^2} \cdot \sqrt{5(429) - (45)^2}}$$

$$r = \frac{60}{(7.071)(10.954)}$$

$$= 0.7746 \approx 0.775$$

(a)

Q.8

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

$$r = \frac{225}{\sqrt{360} \sqrt{250}} = 0.75$$

Q.9

$$r = \frac{122}{\sqrt{136} \sqrt{138}} = 0.8905 \quad \textcircled{b}$$

|              |              |       |       |      |
|--------------|--------------|-------|-------|------|
| $x$          | $y$          | $x^2$ | $y^2$ | $xy$ |
| $\downarrow$ | $\downarrow$ |       |       |      |
| Bad          |              |       |       |      |
| No's         |              |       |       |      |

UV method.

$$u = x - 10 \quad \text{or} \quad v = y - 15$$

$$x \rightarrow u \quad y \rightarrow v$$

$x = ?$

$$n = 12 \quad \Sigma x = 120 \quad \Sigma y = 130$$

$$\Sigma (x - 8)^2 = 150 \quad \Sigma u^2$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{120}{12} = 10$$

$$\Sigma (y - 10)^2 = 200 \quad \Sigma v^2$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{130}{12} = 10.833$$

$$\Sigma (x - 8)(y - 10) = 50 \quad \Sigma uv$$

$$\text{let } u = x - 8 \quad \& \quad v = y - 10$$

$$\bar{u} = \bar{x} - 8$$

$$\frac{\Sigma u}{n} = \frac{\Sigma x}{n} - 8$$

$$\bar{v} = \bar{y} - 10$$

$$\frac{\Sigma v}{n} = \frac{\Sigma y}{n} - 10$$

Multiplying by  $n$  on all sides

$$\Sigma u = \Sigma x - 8 \cdot n$$

$$\Sigma v = \Sigma y - 10 \cdot n$$

$$\Sigma u = 120 - 8(12)$$

$$\Sigma v = 130 - 10(12)$$

$$\Sigma u = 24$$

$$\Sigma v = 10$$

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

$$r = \frac{12(50) - (24)(10)}{\sqrt{12(150) - (24)^2} \cdot \sqrt{12(200) - (10)^2}}$$

$$\sqrt{12(150) - (24)^2} \cdot \sqrt{12(200) - (10)^2}$$

$$r = \frac{360}{\sqrt{1224} \sqrt{2300}} = \underline{\underline{0.215}}$$

## 5) Spearman's Rank Correlation.

i) It is denoted by " $R$ "

ii)  $-1 \leq R \leq +1$

iii) It is used for Qualitative Data

iii)

### Formula.

When Ranks are  
NOT Repeated

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

where  $d$  : difference

When Ranks are  
Repeated

$$R = 1 - \frac{6 [\sum d^2 + \sum CF]}{n(n^2 - 1)}$$



$$\sum d = 0$$

where

CF is Correction Factor

$$CF = \frac{m(m^2 - 1)}{12}$$

where  $m$  = No of Times ranks get Repeated

Shortcut -

$$m = 2, CF = 0.5$$

$$m = 3, CF = 2$$

$$m = 4, CF = 5$$

Q.17.

| <u>x</u> | <u>y</u> | $R_1$ | $R_2$ | $d = R_1 - R_2$ | $d^2$ |
|----------|----------|-------|-------|-----------------|-------|
| 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     |

$$\sum d = 0$$

$$\sum d^2 = 6$$

∴ Ranks are not repeated

$$R = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6(6)}{8(8^2 - 1)} = 0.9285$$

(a)

Q-18

| Rank | Stats           | $R_1$ | $R_2$ | $d = R_1 - R_2$ | $\sum d^2$ |
|------|-----------------|-------|-------|-----------------|------------|
| 30   | 15 <sup>8</sup> | 5     | 8.5   |                 | 12.25      |
| 20   | 40 <sup>3</sup> | 7.5   | 3.5   |                 | 16         |
| 40   | 40 <sup>4</sup> | 3     | 3.5   |                 | 0.25       |
| 50   | 45              | 1.5   | 2     |                 | 0.25       |
| 30   | 20 <sup>6</sup> | 5     | 6.5   |                 | 2.25       |
| 20   | 30              | 7.5   | 5     |                 | 6.25       |
| 20   | 15 <sup>9</sup> | 5     | 8.5   |                 | 12.25      |
| 50   | 50 <sup>7</sup> | 1.5   | 1     |                 | 0.25       |
| 10   | 20 <sup>2</sup> | 9     | 6.5   |                 | 6.25       |
| 0    | 10              | 10    | 10    |                 | 0          |
|      |                 |       |       |                 | <u>56</u>  |

$\therefore$  Ranks are Repeated

$$R = 1 - \frac{6 [\sum d^2 + \sum CF]}{n(n^2 - 1)}$$

$$R = 1 - \frac{6 [56 + 4.5]}{10(10^2 - 1)} = 1 - \frac{363}{990} = 0.633 \text{ (a)}$$

WN For  $R_1$ ,

50 is Getting Repeated Twice

$\therefore m = 2$   $CF = 0.5$

30 is getting repeated Thrice  
 $\therefore m = 3$   $CF = 2$

CF

0.5

2

20 is getting repeated Twice

**J.K. SHAH**<sup>®</sup> m=2 CF=0.5  
CLASSES

For R<sub>2</sub>

40 ..... Twice  
m=2 CF=0.5

20 ..... Twice  
m=2 CF=0.5

15 ..... Twice  
m=2 CF=0.5

|            |            |
|------------|------------|
|            | 0.5        |
|            | 0.5        |
|            | 0.5        |
|            | 0.5        |
| <u>Σ f</u> | <u>4.5</u> |

4.15

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

$$\frac{2}{3} = 1 - \frac{6(55)}{(n^2-1)n}$$

$$\frac{6(55)}{n(n^2-1)} = 1 - \frac{2}{3}$$

$$\frac{330}{n(n^2-1)} \neq \frac{1}{3}$$

$$990 = \frac{n(n^2-1)}{10(10^2-1)}$$

a) 10

Q.14

$$n=6$$

wrong  $d=3$

right  $d=4$

wrong  $d^2=9$

right  $d^2=16$

$$\text{wrong } R = 1 - \frac{6 \sum d^2 \rightarrow \text{wrong}}{n(n^2-1)}$$

$$0.4 = 1 - \frac{6 \sum d^2}{6(6^2-1)}$$

$$\frac{\sum d^2}{35} = 1 - 0.4$$

$$\frac{\sum d^2}{35} = 0.6$$

$$\text{wrong } \sum d^2 = 21$$

Correct  $\sum d^2$

$$= \text{wrong } \sum d^2 \quad 21$$

$$(-) \text{ wrong } d^2 \quad (9)$$

$$(+1) \text{ correct } d^2 \quad +16$$

$$\underline{\underline{28}}$$

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

$$R = 1 - \frac{6(28)}{6(35)}$$

$$R = 0.2$$

(6)

## ⑥ CONCURRENT DEVIATIONS METHOD

- (i) It is denoted by  $r_c$   
&  $-1 \leq r_c \leq 1$
- (ii) It is Quickest Method of Finding Out Correlation

(iii)

$$r_c = \pm \sqrt{\pm \frac{2c - m}{m}}$$

where

- $r_c$  = Concurrent Correlation Coefficient
- $n$  = Total no of pairs
- $m = n - 1$   
Total no of signs (OR)  
Total no of deviations compared
- $c$  = Total no of Concurrent Deviations  
(OR) Total no of Positive Signs

4) If  $\frac{2c-m}{m} = +ve$ , then Bahar Bhi +ve

4) If  $\frac{2c-m}{m} = -ve$ , then Bahar Bhi -ve

5) Shortcuts:

(i) If  $c = m$ , then  $\pi_c = 1$

(ii) If  $c = 0$ , then  $\pi_c = -1$

(iii) If  $c = m/2$ , then  $\pi_c = 0$

|                 | 1             | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 |
|-----------------|---------------|----|----|----|----|----|----|----|----|----|----|
| Supply          | 68            | 43 | 38 | 78 | 66 | 83 | 38 | 23 | 83 | 53 | 48 |
| Demand          | 65            | 60 | 55 | 61 | 35 | 75 | 45 | 40 | 85 | 80 | 85 |
| $C_x$           | <del>NO</del> | -  | -  | +  | -  | +  | -  | -  | +  | -  | -  |
| $C_y$           | <del>NO</del> | -  | -  | +  | -  | +  | -  | -  | +  | -  | +  |
| $C_x \cdot C_y$ | <del>NO</del> | +  | +  | +  | +  | +  | +  | +  | +  | +  | -  |
|                 |               | ①  | ②  | ③  | ④  | ⑤  | ⑥  | ⑦  | ⑧  | ⑨  |    |

Both + / - → +ve

$$n = 11 \quad m = n - 1 = 11 - 1 = 10$$

$c = 9$  positive signs

$$r_c = \pm \sqrt{\pm \frac{2c - m}{m}}$$

$$r_c = + \sqrt{\frac{2(9) - 10}{10}}$$

$$r_c = + \underline{0.89} \quad \textcircled{C}$$

Q.21

$$n = 10 \quad c = 4$$
$$m = 10 - 1 = 9$$

$$r_c = \pm \sqrt{\pm \frac{2c - m}{m}}$$

$$r_c = - \sqrt{\frac{8 - 9}{9}} = - \sqrt{\frac{1}{9}} = -\frac{1}{3}$$

$\textcircled{C}$

q.23

$$r_c = \frac{1}{\sqrt{3}}$$

$$h = p = ??$$

$$c = 6$$

$$r_c = \sqrt{\frac{2c - m}{m}}$$

$$\frac{1}{\sqrt{3}} = \sqrt{\frac{2(6) - m}{m}}$$

Squaring on b.t.s

$$\frac{1}{3} = \frac{(12 - m)}{m}$$

$$m = 36 - 3m$$

$$4m = 36$$

$$m = 9$$

$$m = n - 1$$

$$n = m + 1 = 9 + 1$$

$$= 10$$

(a)

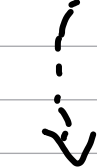
~~Ans~~



$$3y - 7v = 20$$

Q.45

$$v = 6 - 0.5x$$



-ve

+ve

$$7v = 3y - 20$$

$$v = \frac{3y}{7} - \frac{20}{7}$$



$$r_{xy} = -r_{uv} = -0.58 \text{ (b)}$$

## ⑦ Miscellaneous Properties

- ① Coefficient of Determination & Coefficient of Non Determination.

$$\text{Coeff of Determination} = r^2$$

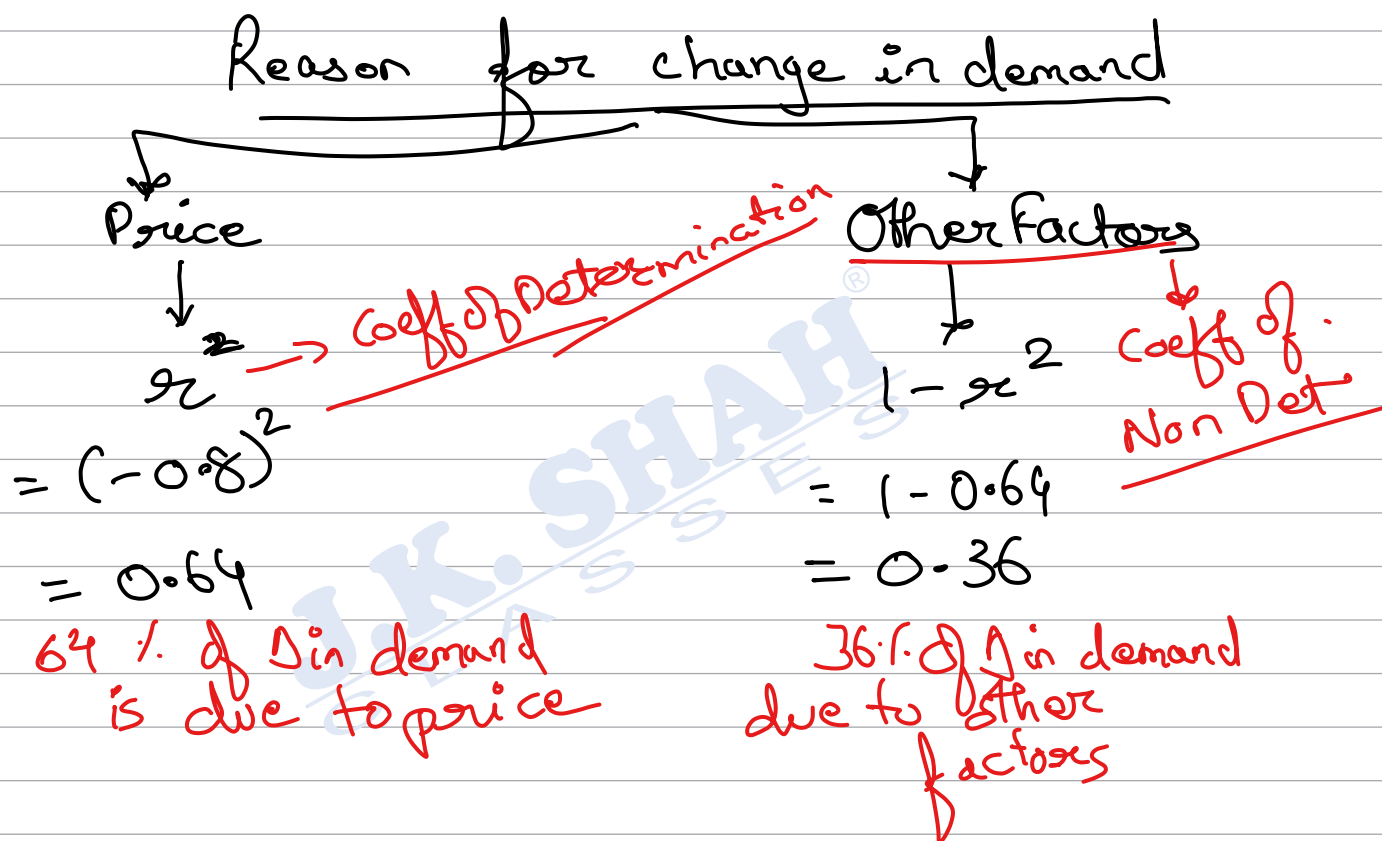
$$\text{where } r^2 = \frac{\text{Explained Variance}}{\text{Total Variance}}$$

$$\text{Coeff of Non Determination} = 1 - r^2$$

$$\text{where } 1 - r^2 = \frac{\text{Unexplained Variance}}{\text{Total Variance}}$$

Ex.

Correlation Coefficient Between  
Demand ( $x$ ) & Price ( $y$ ) is  $-0.8$   
i.e.  $r = -0.8$



(2) Coefficient of Alienation =  $\sqrt{1 - r^2}$   
(or) Square Root of Coeff of  
Non Determination

### ③ Probable Error

① PE is used for Determining Reliability of 'r'.

②  $PE = 0.6745 \times \text{Std Error (SE)}$   
where  $SE = \frac{1-r^2}{\sqrt{n}}$

③ Kaise Pata Chalega ki 'r' Kitna Reliable Hai????

→ If  $r < \underline{6 \cdot PE}$ , then 'r' is NOT SIGNIFICANT

→ If  $r = 6 \cdot PE$ , then 'r' has correlation

→ If  $r > \underline{6 \cdot PE}$ , then 'r' is SIGNIFICANT & there is CORRELATION

Popl<sup>n</sup>

$x$

=

Sample  
 $x$

④ If  $x$  &  $y$  are Correlated,

$$V(x \pm y) = V(x) + V(y) \pm 2 \cdot \text{Cor}(x, y)$$

If  $x$  &  $y$  are uncorrelated "

then  $\text{Cor}(x, y) = 0$

$$\therefore V(x \pm y) = V(x) + V(y)$$

⑤

$$y = ax + b$$

↓  
Slope

↖ Intercept

Slope  $> 0$

$$r = 1$$

Slope  $< 0$

$$r = -1$$

Q. 29  $PE = 0.6745 \times \left[ \frac{1 - z^2}{\sqrt{n}} \right]$

$\downarrow$   
S.E

$$= 0.6745 \times \left[ \frac{1 - (0.05)^2}{\sqrt{25}} \right]$$

$$PE = 0.1345$$

Q. 30

$$z = 0.526$$

$$n = 50$$

$$PE = 0.6745 \times \frac{\left[ 1 - (0.526)^2 \right]}{\sqrt{50}}$$

$$PE = 0.0689 \approx 0.069 \text{ (a)}$$

$$\begin{array}{l} 2 < 3 \\ 3 > 2 \end{array}$$

$$15 < \sigma_x \cdot \sigma_y$$

$$\sigma_x \cdot \sigma_y > 15$$

# REGRESSION ANALYSIS

In or  
Advt Exp  
(x)

In or  
Sales  
(y)

10

110

12

125

5

60

8

85

Dep vari

I

15

???

Dep

200

I

Cosae lation

Regression

## Overview:

- ① Meaning of Regression
- ② Calculation of Regression lines
- ③ Calculation of Regression Coefficient
- ④ Properties of Regression Coefficient
- ⑤ Miscellaneous Points.

## ① Meaning of Regression

- ① Regression Means Estimation/Prediction or Forecasting. Kiska???

Estimation of Dependent Variable with the given value of Independent Variable.  
means

- Estimate x when y is given (②)  
Dep. Var. Indep. Var.
- Estimate y when x is given

i) Regression is the **AVERAGE LINEAR RELATIONSHIP BETWEEN TWO VARIABLES**

ii) It establishes a **FUNCTIONAL** Relationship Between Two Variables

iii) Dep Variable is called as Regressed / Explained Variable  
& Indep Variable is called as Regressor / Unexplained Variable

iv) Correlation can only **MEASURE** Degree of Association

**BUT**

Regression **Quantifies** such mathematical relationship

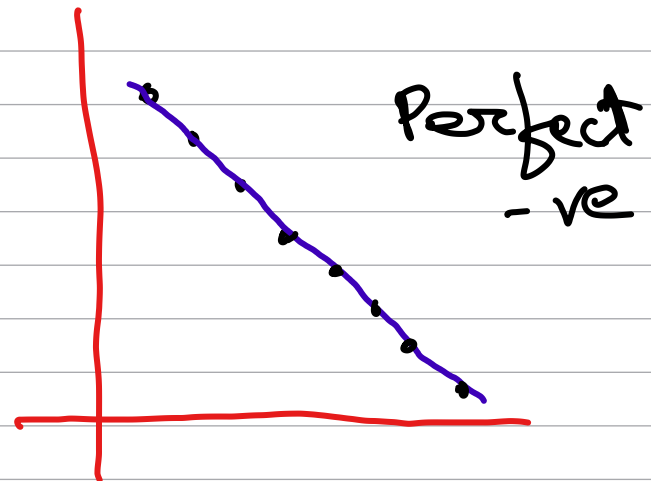
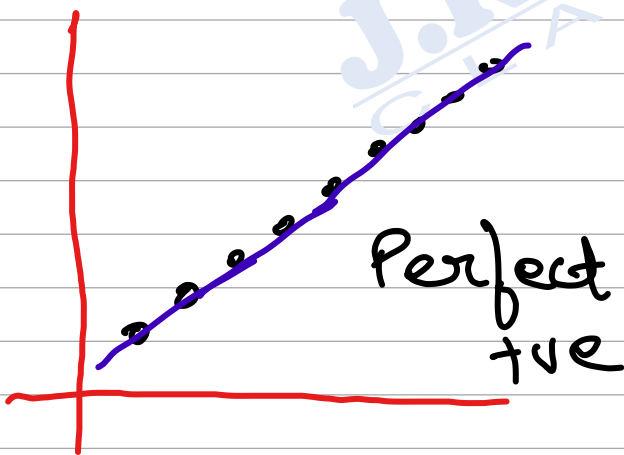


(vi) Reg. Analysis is **ABSOLUTE**  
**measure** whereas  
Corr. Analysis is **RELATIVE**  
**measure**

— x — x —

(2) Calculation of Regression Lines

Now let's see for  $x$  &  $y$  How we  
will draw a line for Perfect  
Correlation.



Perfect means a STRAIGHT line  
∴ There is a LINEAR  
RELATIONSHIP

If  $x$  &  $y$  are perfectly correlated  
then a linear equation  
will be formed

Ex.  $y = ax + b$

$a$  = Slope of equation  
 $b$  = Intercept

$a > 0$   
 $x = +1$

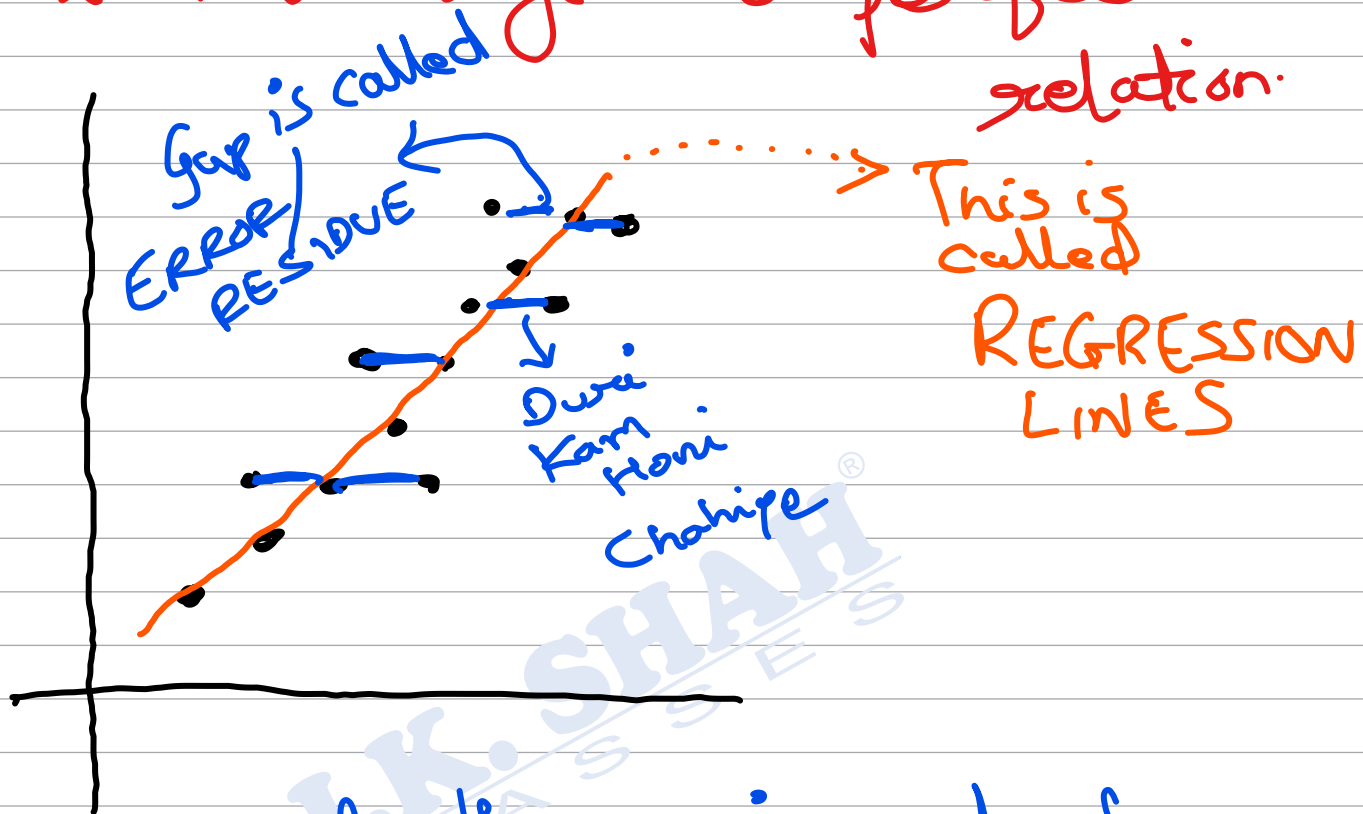
$a < 0$   
 $x = -1$

If I substitute value of  $x$   
then I will get answer of  $y$  &  
vice versa.

But this only holds True if  
Relations are PERFECT.

But what for Imperfect  
Relations

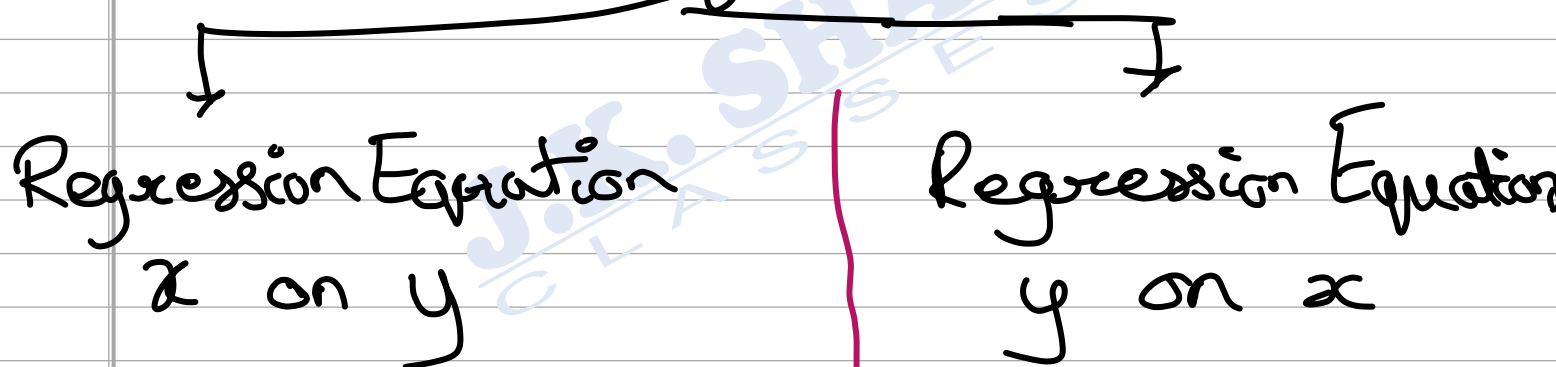
Now Let's Talk for variables  
which won't give a perfect  
relation.



In case of these imperfect  
correlation we have to make  
such a equation/line  
that it covers maximum  
variables & the remaining  
which are left should have  
the least distance from  
the equation/line so  
formed.

Hence the line which we prepare for the purpose of ESTIMATION is called REGRESSION LINES.

### Types of Regression Lines



- |                                                                                                                                 |                                                                                                                             |
|---------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|
| <p>(i) Used for Estimation of <math>x</math> when <math>y</math> is given<br/>i.e. <math>x</math> depends on <math>y</math></p> | <p>Used for Estimation of <math>y</math> when <math>x</math> is given<br/>i.e. <math>y</math> depends on <math>x</math></p> |
| <p>(ii) <math>b_{xy}</math> is regression coefficient</p>                                                                       | <p><math>b_{yx}</math> is regression coefficient</p>                                                                        |



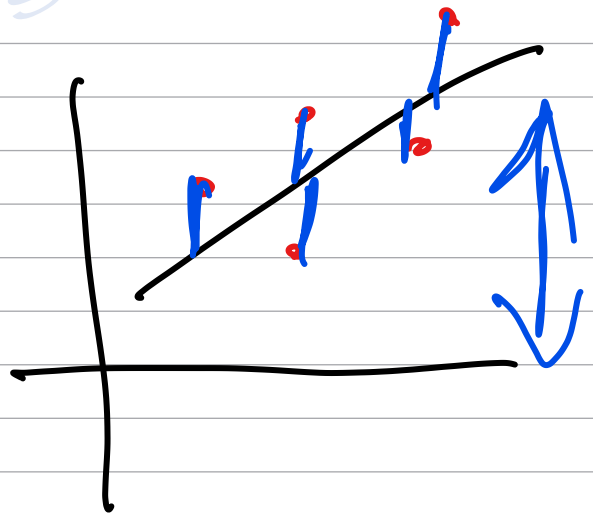
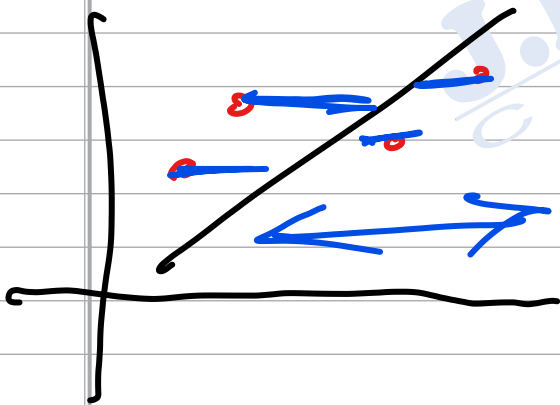
Equation is given  
as

$$x = a + by$$

Equation is given  
as

$$y = a + bx$$

- (iv) Reg line is derived by MINIMIZING HORIZONTAL DISTANCE Using Least Square Method
- Reg line is derived by MINIMIZING VERTICAL DISTANCE Using Least Square Method



**J.K. SHAH<sup>®</sup>**  
CLASSES

**J.K. SHAH<sup>®</sup>**  
CLASSES

**J.K. SHAH<sup>®</sup>**  
CLASSES



**J.K. SHAH<sup>®</sup>**  
CLASSES

**J.K. SHAH<sup>®</sup>**  
CLASSES

**J.K. SHAH<sup>®</sup>**  
CLASSES

**J.K. SHAH<sup>®</sup>**  
CLASSES

**J.K. SHAH<sup>®</sup>**  
CLASSES

**J.K. SHAH<sup>®</sup>**  
CLASSES

**J.K. SHAH<sup>®</sup>**  
**CLASSES**

**J.K. SHAH<sup>®</sup>**  
CLASSES



**J.K. SHAH<sup>®</sup>**  
CLASSES

**J.K. SHAH<sup>®</sup>**  
**CLASSES**

**J.K. SHAH<sup>®</sup>**  
**CLASSES**

**J.K. SHAH<sup>®</sup>**  
CLASSES

**J.K. SHAH<sup>®</sup>**  
CLASSES

**J.K. SHAH<sup>®</sup>**  
CLASSES

**J.K. SHAH<sup>®</sup>**  
CLASSES

**J.K. SHAH<sup>®</sup>**  
CLASSES