UNIT-5

Correlation:

* 2 Variables x, y are considering is called correlation.

* The degree of relationship between two (or) more variables is known as correlation.

degree:

* Positive (or) Negative (or) there is no relationship is called degree

x	y		
↑	↑	= positive relationship	
↓	↓	= positive relationship	
↑	↓	= Negative relationship	
↓	↑	= Negative relationship	
↓	-	= NO relationship (correlation)	0 value
-	↑	= NO relationship (correlation)	0 value

Remark:

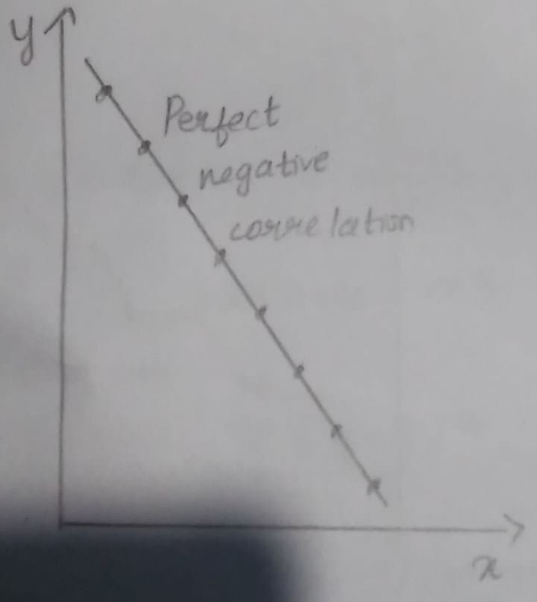
* The correlation value always lies between -1 to 1.

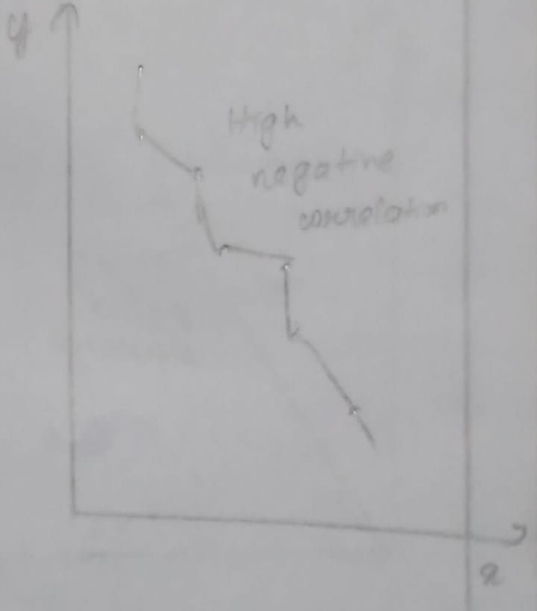
Types of correlation:

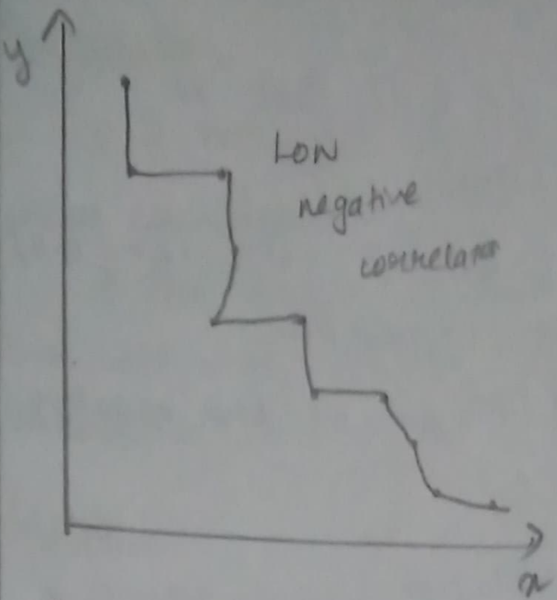
* If the value is (-1) that is perfect negative correlation.
* If value is $(-0.5 \le r < -1)$ $(-1 < r \le 0.5)$ that is high negative correlation.
* If value is $(-0.5 < r < 0)$ low negative correlation
* $r = 0$ No correlation
* $(+1)$ perfect positive correlation
* $(0.5 < r < 1)$ high positive correlation
* $(0 < r < 0.5)$ low positive correlation.

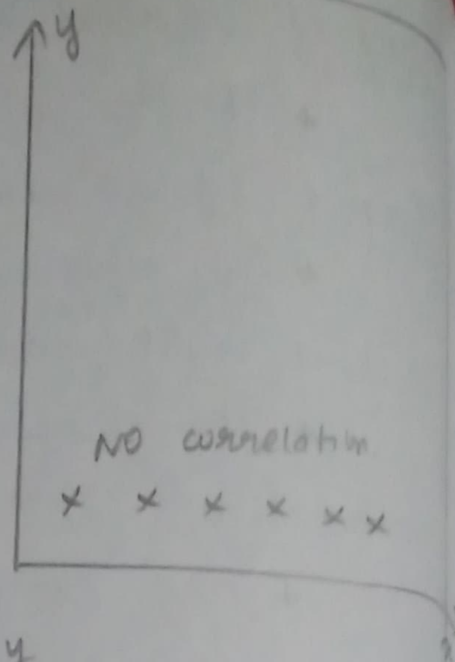
Scatter Diagram:

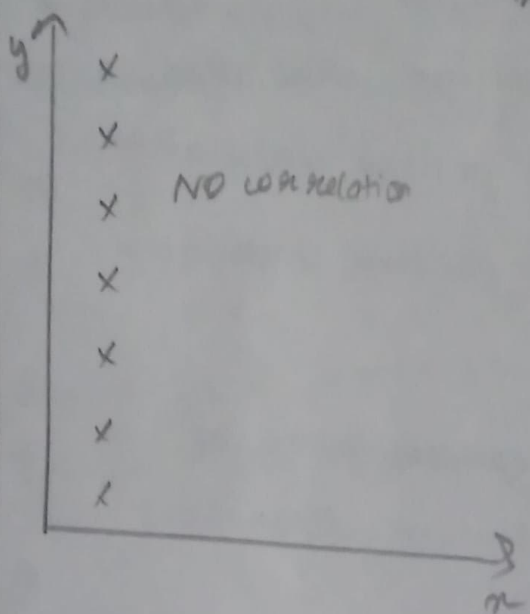
* A graphical representation of correlation is called scatter diagram

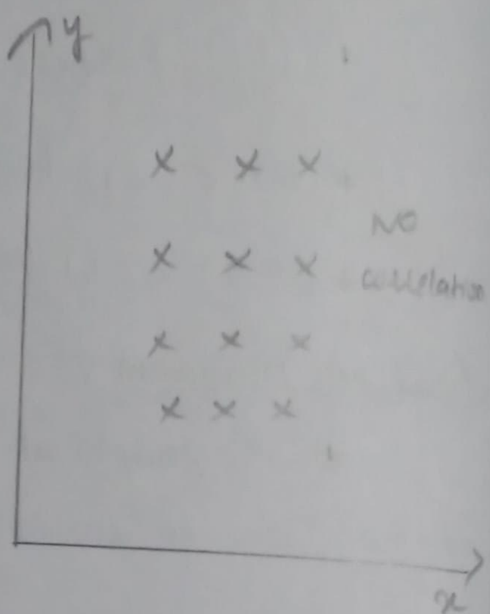


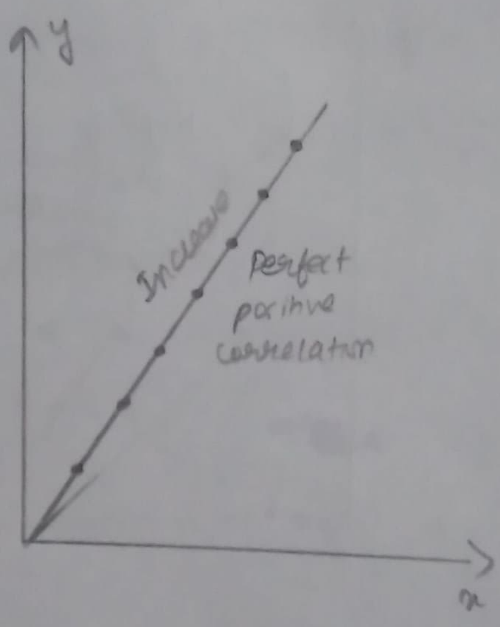


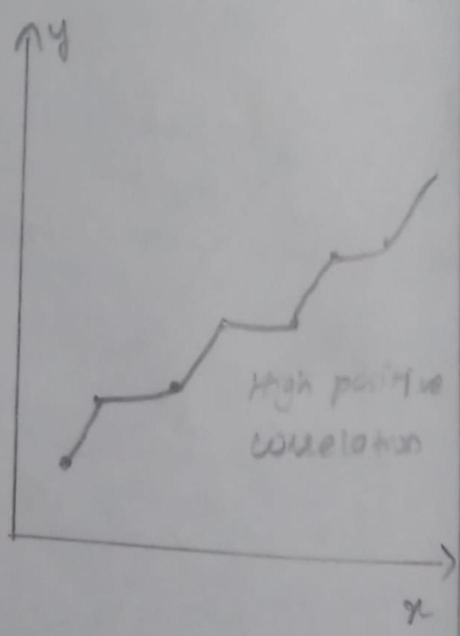

y
Low
negative
correlation
x

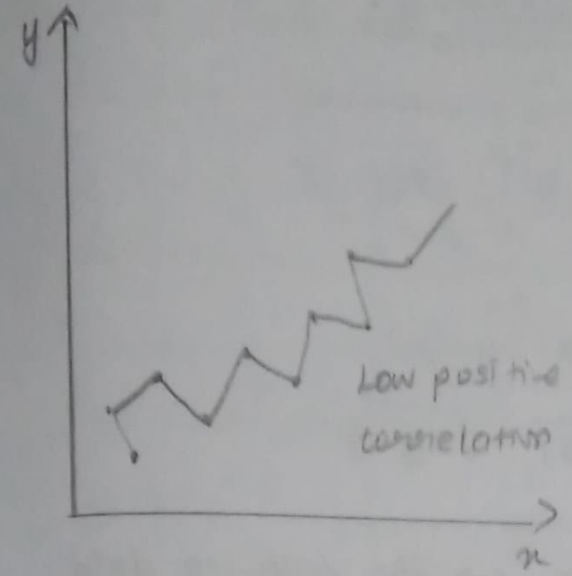

y
NO correlation
x


y
NO correlation
x


y
NO
correlation
x


y
Increase
Perfect
positive
correlation
x


y
High positive
correlation
x



Karl Pearson's coefficient of correlation:

* It is denoted as 'r'

* It is the ratio between the co-variance of the 2 variables x and y and the product of the standard deviation x and the product of the standard deviation y.

* In symbols.

$$r = \frac{cor(x,y)}{\sigma x, \sigma y}$$

* variance is nothing ~~by~~ but (standard deviation)^2

$$\text{variance} = \sigma^2$$

$$\text{variance} = \frac{\Sigma (x-\bar{x})^2}{n}$$

$$\text{variance} = \frac{\Sigma (x-\bar{x})(x-\bar{x})}{n}$$

co variance = $\dfrac{\dfrac{\sum (x-\bar{x})(x-\bar{x})}{n}}{\sqrt{\dfrac{\sum (x-\bar{x})^2}{n}} \cdot \sqrt{\dfrac{\sum (y-\bar{y})^2}{n}}}$

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

Calculate correlation for the following data

x	5	9	7	8	3	4	6
y	15	27	21	24	9	12	18

Solution:

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

$$\bar{x} = \frac{\sum x}{n} = \frac{42}{7} = \boxed{6}$$

$$\bar{y} = \frac{\sum y}{n} = \frac{126}{7} = \boxed{18}$$

[see the table]

$$r = \frac{84}{\sqrt{28} \times \sqrt{252}}$$

$$r = \frac{84}{5.291 \times 15.87} = \frac{84}{84} = \boxed{1}$$

THERE IS A perfect negative correlation BETWEEN VARIABLES x, y

$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x}).(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
-1	-3	3	1	9
3	9	27	9	81
1	3	3	1	9
2	6	12	4	36
-3	-9	27	9	81
-2	-6	12	4	36
0	0	0	0	0
		$\Sigma(x-\bar{x}).(y-\bar{y}) = 84$	$\Sigma(x-\bar{x})^2 = 28$	$\Sigma(y-\bar{y})^2 = 252$

Example 2:

x	6	4	3	8	7	9	5
y	15	20	22.5	10	12.5	7.5	17.5

Solution

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

$$\bar{x} = \frac{\Sigma x}{n} = \frac{42}{7} = \boxed{6}$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{105}{7} = \boxed{15}$$

$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x}).(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
0	0	0	0	0
-2	5	-10	4	25
-3	7.5	-22.5	9	56.25
2	-5	-10	4	25
1	-2.5	-2.5	1	6.25
3	-7.5	-22.5	9	56.25
-1	2.5	-2.5	1	6.25
		$\Sigma(x-\bar{x}).(y-\bar{y}) = -70$	$\Sigma(x-\bar{x}) = 28$	$\Sigma(y-\bar{y}) = 175$

$$r = \frac{-70}{\sqrt{28} \times \sqrt{175}}$$

$$= \frac{-70}{5.291 \times 13.22} = \frac{-70}{70} = \boxed{-1}$$

There is a perfect negative correlation.

Example : 3

x	6	4	3	8	7	9	5
y	15	15	15	15	15	15	15

Solution

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

$\bar{x} = \frac{\Sigma x}{n} = \frac{42}{7} = \boxed{6}$

$\bar{y} = \frac{\Sigma y}{n} = \frac{105}{7} = \boxed{15}$

$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
0	0	0	0	0
-2	0	0	4	0
-3	0	0	9	0
2	0	0	4	0
1	0	0	1	0
3	0	0	9	0
-1	0	0	1	0
		$\Sigma(x-\bar{x})(y-\bar{y}) = 0$	$\Sigma(x-\bar{x})^2 = 28$	$\Sigma(y-\bar{y})^2 = 0$

$r = \frac{0}{\sqrt{28} \times \sqrt{0}}$

$= \frac{0}{5.291 \times 0} = \frac{0}{0} = \boxed{0}$

NO correleation

Spearmen's Rank correlation :

* It is denoted as (Rho)

symbol = ρ

$$(\text{Rho})\ \rho = 1 - \left\{ \frac{6 \Sigma D^2}{N(N^2-1)} \right\}$$

Calulate spearman's Rank correlation for the following data

x	5	9	7	8	3	4	6
y	15	27	21	24	9	12	18

Rx	Ry	D (Rx − Ry)	D^2
5	5	0	0
1	1	0	0
3	3	0	0
2	2	0	0
7	7	0	0
6	6	0	0
4	4	0	0

$$\rho = 1 - \left\{ \frac{6 \times 0}{7(7^2-1)} \right\}$$

$$= 1 - \left\{ \frac{0}{7(49-1)} \right\}$$

$$= 1 - \left\{ \frac{0}{7(48)} \right\}$$

$= 1 - \left\{ \frac{0}{336} \right\} = 1 - 0 = \boxed{1}$

perfect positive correlation

Example

x	6	4	3	8	7	9	5
y	15	20	22.5	10	12.5	7.5	17.5

Rx	Ry	D $(Rx - Ry)$	D^2
4	4	0	0
6	2	4	16
7	1	6	36
2	6	-4	16
3	5	-2	4
1	7	-6	36
5	3	2	4
			$\Sigma d^2 = 112$

$$P = 1 - \left\{ \frac{6\,(112)}{7(7^2-1)} \right\}$$

$$= 1 - \left\{ \frac{672}{7(49-1)} \right\} = 1 - \left\{ \frac{672}{7(48)} \right\}$$

$$= 1 - \left\{ \frac{672}{336} \right\}$$

$$= 1 - 2 = \boxed{-1}$$

Perfect Negative correlation.

Example

x	5	4	3	8	7	9	5
y	15	20	12.5	10	12.5	7.5	12.5

Solution.

$$p = 1 - \left\{ \frac{6\left\{\Sigma d^2 + \frac{m(m^2-1)}{12} + \frac{m(m^2-1)}{12}\right\}}{N(N^2-1)} \right\}$$

Rx	Ry	D (Rx − Ry)	D^2
4.5	2	2.5	6.25
6	1	5	25
7	4	3	9
2	6	−4	16
3	4	−1	1
1	7	−6	36
4.5	4	0.5	0.25
			$\Sigma d^2 = 93.5$

$$p = 1 - \left\{ \frac{6\left\{93.5 + \frac{2(2^2-1)}{12} + \frac{3(32-1)}{12}\right\}}{7(7^2-1)} \right\}$$

$$= 1 - \left\{ \frac{6(93.5 + \frac{2(4-1)}{12} + \frac{3(9-1)}{12})}{7(49-1)} \right\}$$

$$= 1 - \left\{ \frac{6 \left\{ 93.5 + \frac{2(3)}{12} + \frac{3(8)}{12} \right\}}{7(48)} \right\}$$

$$= 1 - \left\{ \frac{6 \left(93.5 + \frac{6}{12} + \frac{24}{12} \right\}}{336} \right\}$$

$$= 1 - \left\{ \frac{6 \ (93.5 + 0.5 + 2)}{336} \right\}$$

$$= 1 - \left\{ 6 \left\{ \frac{96}{336} \right\} \right\}$$

$$= 1 - \{ 6 (0.285) \}$$

$$= 1 - 1.71 = \boxed{-0.71}$$

High negative correlation.

Example

x	73.2	85.8	78.9	75.8	77.2	81.2	83.8
y	97.8	99.2	98.8	98.3	98.3	96.7	97.1

solution:

$$p = 1 - \left\{ \frac{6 \left(\Sigma d^2 + \frac{m(m^2-1)}{12} + \frac{m(m^2-1)}{12} \right)}{N(N^2-1)} \right\}$$

Rx	Ry	D $(Rx-Ry)$	D^2
7	5	2	4
1	1	0	0
4	2	2	4
6	3.5	2.5	6.25
5	3.5	1.5	2.25
3	7	-4	16
2	6	-4	16
			$\Sigma d^2 = 48.5$

$$p = 1 - \left\{ \frac{6\left(48.5 + \frac{2(2^2-1)}{12}\right)}{7(7^2-1)} \right\}$$

$$= 1 - \left\{ \frac{6\left\{48.5 + \frac{6}{12}\right\}}{336} \right\}$$

$$= 1 - \left\{ \frac{6(48.5 + 0.5)}{336} \right\}$$

$$= 1 - \left\{ 6\left(\frac{49}{336}\right) \right\}$$

$$= 1 - \{6 \times (0.145)\}$$

$$= 1 - 0.87$$

$$\boxed{= 0.13}$$

Regression:

+ Average relationship between two variables is known as regression

+ x, y (2 variables)

There are two regression equations namely:

1. Regression equation of x on y
2. Regression equation of y on x

+ To estimate x value when y is given we use x on y

+ To estimate y value when x is given we use y on x

Formula:

Regression equation of y on x is

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

$$= r\frac{\sigma_y}{\sigma_x}(x-\bar{x})$$

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

$$b_{yx} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

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

$$\frac{\text{Covariance}(x,y)}{\sigma_x \cdot \sigma_y}, \frac{\sigma_y}{\sigma_x} = \frac{\text{Covari}(x,y)}{\sigma_x^2}$$

Regression equation of x on y

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

$$= r\,\frac{\sigma x}{\sigma y}(y-\bar{y})$$

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

1. where b_{yx} is the regression coefficient of y on x
2. where b_{xy} is the regression coefficient of x on y

r = Correlation of coefficient

σx = Standard deviation of x

σy = Standard deviation of y

$\bar{x}$ = Mean of x

$\bar{y}$ = Mean of y

Note:

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

* Correlation coefficient is (r) the geometric Mean of 2 regression coefficients.

Example:

construct regression equation of x and y and y on x from the following data.

x	10	12	15	20	23
y	14	17	23	21	25

A) estimate y when $x = 13$

B) estimate x when $y = 22$

c) find r using regression coefficients.

Regression equation x on y

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

$$\bar{x} = \frac{\Sigma x}{n} = \frac{80}{5} = 16$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{100}{5} = 20$$

$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(y-\bar{y})^2$	$(x-\bar{x})^2$
-6	-6	36	36	36
-4	-3	12	9	16
-1	3	-3	9	1
4	1	4	1	16
7	5	35	25	49
		$\Sigma(x-\bar{x}).(y-\bar{y}) =$ 84	$\Sigma(y-\bar{y})^2 =$ 80	$\Sigma(x-\bar{x})^2 =$ 118

$(x-16) = 1.05\,(y-20)$

$(x-16) = 1.05y - 21$

$x = 1.05y - 21 + 16$

$\boxed{x = 1.05y - 5}$

Regression Equation x on y

Regression equation of y on x

$= (y-\bar{y}) = \dfrac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (x-\bar{x})^2}(x-\bar{x})$

$= (y-20) = \dfrac{84}{118} \times (x-16)$

$= y - 20 = 0.71 \times (x-16)$

$(y-20) = 0.7x - 11.36$

$y = 0.71x - 11.36 + 20$

$\boxed{y = 0.71x + 8.64}$

Regression equation y on x

$\boxed{b_{yx} = 0.71}$, $\boxed{b_{xy} = 1.05}$

(1) estimate y

$y = 0.71x + 8.64$

$= 0.71(13) + 8.64$

$= 9.23 + 8.64$

$y = 17.87$

② estimate x

$$x = 1.05y - 5$$
$$= 1.05(22) - 5$$
$$= 23.1 - 5$$
$$\boxed{x = 18.1}$$

3. find r

$$r = \pm\sqrt{b_{xy} \cdot b_{yx}}$$
$$= \pm\sqrt{1.05 \cdot 0.71}$$
$$= \pm\sqrt{0.7455}$$
$$\boxed{r = 0.863}$$

Notes

$$= \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2} \cdot \sqrt{\Sigma(y-\bar{y})^2}}$$
$$= \frac{84}{\sqrt{80} \times \sqrt{118}}$$
$$= \frac{84}{8.94 \times 10.86}$$
$$= \frac{84}{97.0884}$$
$$= 0.8651$$

Example:

$\bar{x} = 53.2$, $b_{yx} = -1.5$, $\bar{y} = 27.9$

what is the most likely value of y when $x = 60$.

solution:

$$(y - \bar{y}) = b_{yx}(x - \bar{x})$$
$$(y - 27.9) = 1.5(x - 53.2)$$
$$(y - 27.9) = -1.5x + 79.8 + 27.9$$

$$\boxed{y = -1.5x + 107.7}$$

Regression equation y on x.

what is the most likely value of y when $x = 60$

$$y = -1.5x + 107.7$$
$$= -1.5(60) + 107.7$$
$$= -90 + 107.7$$
$$\boxed{y = 17.7}$$

Example

construct the regression equation of x on y and y on x from the following data

x	45	48	50	55	65	70	75	72	80	85
y	25	30	35	30	40	50	45	55	60	65

Solution:

Regression equation x on y

$$(x - \bar{x}) . \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2} (y - \bar{y})$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{645}{10} = 64.5$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{435}{10} = 43.5$$

$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x}).(y-\bar{y})$	$(y-\bar{y})^2$	$(x-\bar{x})^2$
-19.5	-18.5	360.75	342.25	380.85
-16.5	-13.5	222.75	182.25	272.25
-14.5	-8.5	123.25	72.25	210.25
-9.5	-13.5	128.25	182.25	90.25
-0.5	-3.5	1.75	12.25	0.25
5.5	6.5	35.75	42.25	30.25
10.5	1.5	15.75	2.25	110.25
7.5	11.5	86.25	132.25	56.25
15.5	16.5	255.75	272.25	240.25
20.5	21.5	440.75	462.25	420.25
		$\Sigma(x-\bar{x})(y-\bar{y}) =$ 1671	1702.5	1810.5

$$(x-64.5) = \frac{1671}{1702.5} \times (y-43.5)$$

$$(x-64.5) = 0.98 (y-43.5)$$

$$(x-64.5) = 0.98y - 42.63$$

$$x = 0.98y - 42.63 + 64.5$$

$$\boxed{x = 0.98y + 21.87}$$

Regression equation x on y.

Regression equation y on x

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

$$(y-43.5) = \frac{1671}{1810.5}(x-64.5)$$

$$(y-43.5) = 0.92(x-64.5)$$

$$(y-43.5) = 0.92x - 59.34$$

$$y = 0.92x - 59.34 + 43.5$$

$$y = 0.92x - 15.84$$

Regression equation y on x.