Co-Relation:

Co → Combine

Relation → Relationship.

Weight — calories intake.

Weight — Activity.

Weight — Metabolism.

⋮

etc.

weight ↑ = calories intake ↑

Direct | Strong Relationship.

CORRELATION:

⇒ **Definition:**

Correlation is a technique, which measures the strength of relationship b/w two variables. It is denoted by 'r'.

Relationship

Positive — Negative

Positive:

↑ ↑

↓ ↓

(Direct Relation)

Negative:

↑ ↓

↓ ↑

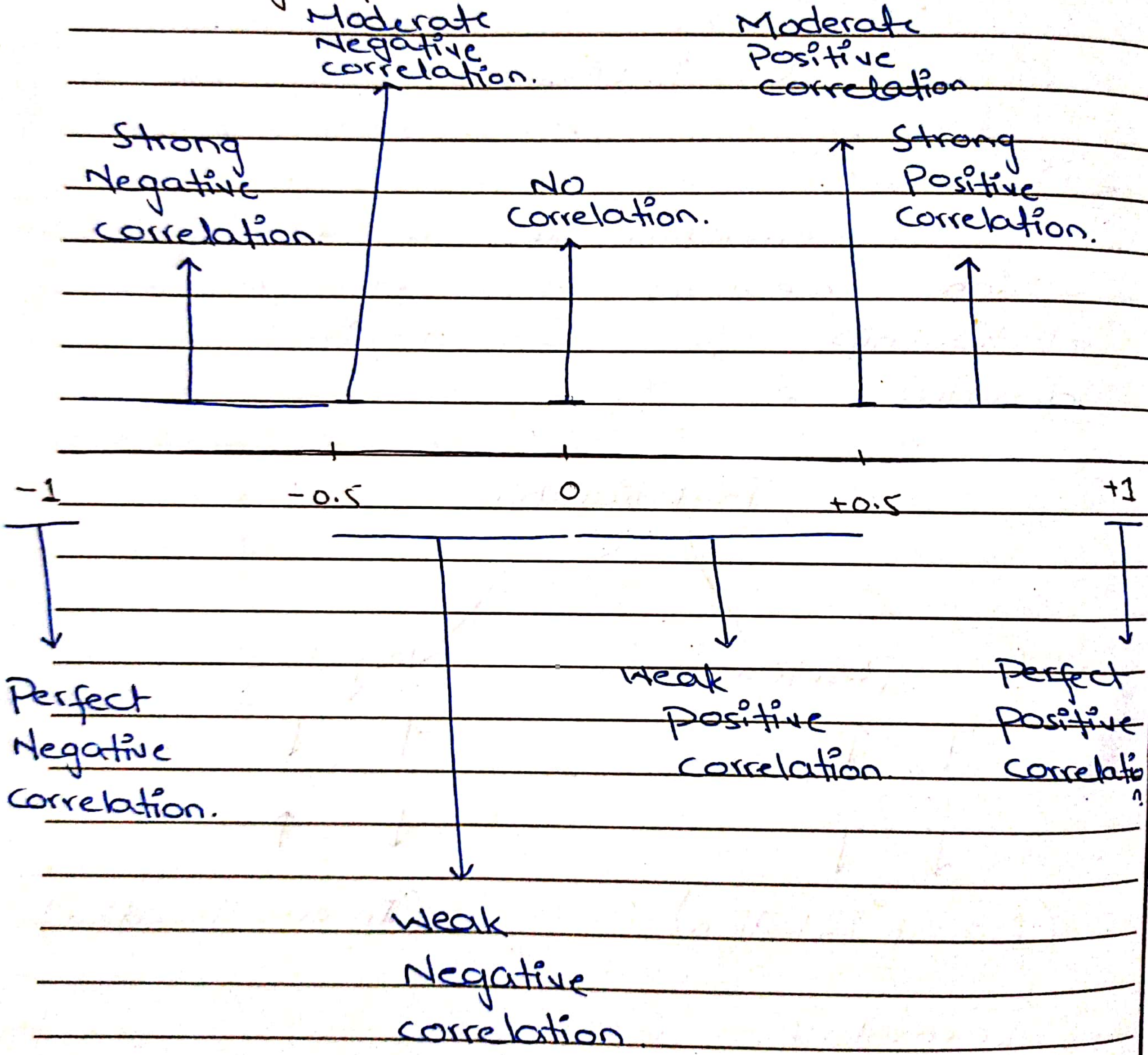
(Inverse Relation)

⇒ **Formula:**

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

r → Correlation coefficient.

→ Range Of Correlation :



⇒ Positive Correlation:

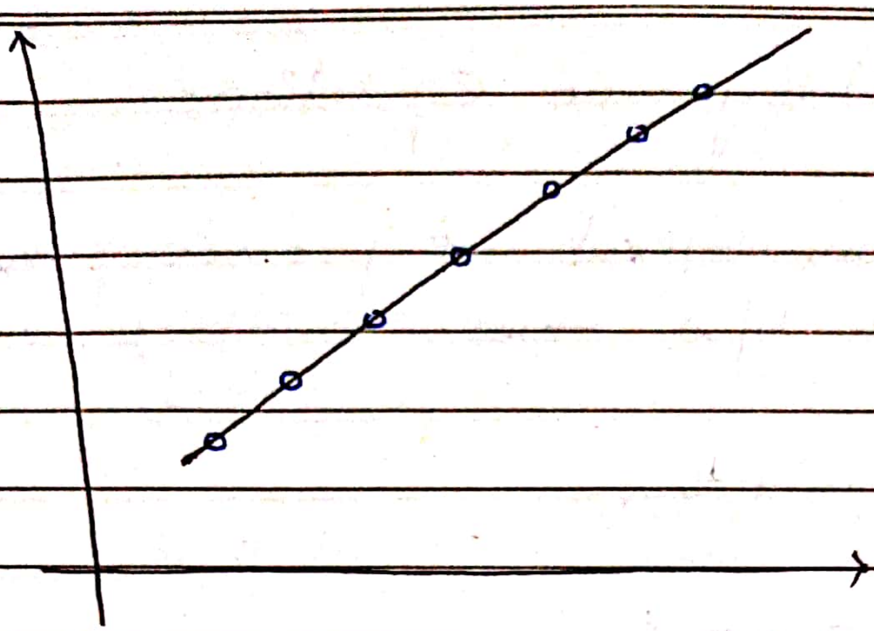
If two variables are related in such a way that they increase or decrease together. It means that their change is in same direction. Then such a correlation is called positive correlation or Direct correlation. e.g.,

Relation b/w age & weights of children

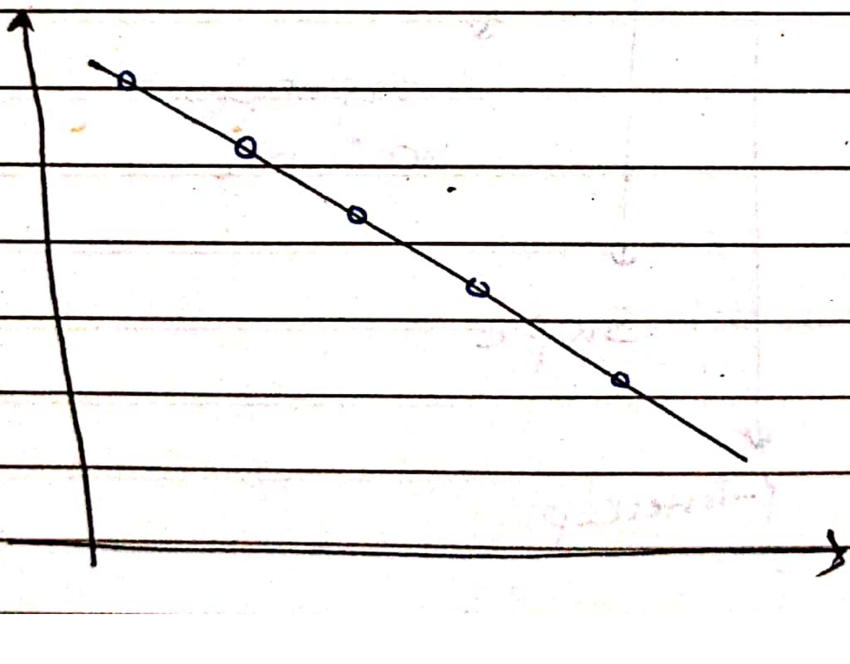
⇒ Negative Correlation:

If two variables are related in such a way that one variable increases & the other decreases. It means that the change of the variables are in opposite direction. Then such a relation b/w variables is called Negative correlation. e.g.,

Relation b/w supply & price will be negative.



Perfect Positive correlation



Perfect Negative correlation

⟹ Example:

calculate co-efficient of correlation b/w X & Y from the following data:

X	1	2	3	4	5
Y	2	5	3	8	7

$$\bar{X} = \frac{\Sigma X}{n} = \frac{15}{5}$$

$$\boxed{\bar{X} = 3}$$

$$\bar{Y} = \frac{\Sigma Y}{n} = \frac{25}{5}$$

$$\boxed{\bar{Y} = 5}$$

X	Y	$X-\bar{X}$	$Y-\bar{Y}$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
1	2	-2	-3	4	9	6
2	5	-1	0	1	0	0
3	3	0	-2	0	4	0
4	8	1	3	1	9	3
5	7	2	2	4	4	4
15	25	0	0	10	26	13

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

$$= \frac{13}{\sqrt{10 \times 26}}$$

$$r = 0.8$$

There is strong positive correlation b/w x & y.

→ Example:

Find coefficient of correlation b/w the variables X & Y represented in the following table

X	25	29	30	30	31	32	33	35	37	38
Y	20	22	24	29	23	31	29	31	30	31

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

$$= \frac{320}{10}$$

$$\boxed{\bar{X} = 32}$$

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

$$= \frac{270}{10}$$

$$\boxed{\bar{Y} = 27}$$

X	Y	$X-\bar{X}$	$Y-\bar{Y}$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
25	20	-7	-7	49	49	49
29	22	-3	-5	9	25	15
30	24	-2	-3	4	9	6
30	29	-2	2	4	4	-4
31	23	-1	-4	1	16	4
32	31	0	4	0	16	0
33	29	1	2	1	4	2
35	31	3	4	9	16	12
37	30	5	3	25	9	15
38	31	6	4	36	16	24
				138	164	123

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

$$= \frac{123}{\sqrt{138 \times 164}}$$

$$\boxed{\gamma = 0.82}$$

Strong Positive Correlation.

Scanned with CamScanner

→ Example :

Find coefficient of correlation b/w x & y

X	400	200	700	100	500	300	600
Y	50	60	20	70	40	30	10

$\bar{X} = 2800/7 = \boxed{400}$

$\bar{Y} = 280/7 = \boxed{40}$

X	Y	$X-\bar{X}$	$Y-\bar{Y}$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
400	50	0	10	0	100	0
200	60	-200	20	40000	400	-4000
700	20	300	-20	90000	400	-6000
100	70	-300	30	90000	900	-9000
500	40	100	0	10000	0	0
300	30	-100	-10	10000	100	1000
600	10	200	-30	40000	900	-6000
				280000	2800	-24000

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

$$= \frac{-24000}{\sqrt{280000 \times 2800}}$$

$$= -0.857$$

There is strong Negative correlation b/w x & y.

→ Example:

X	30	35	40	45	50	60	70	80	90
Y	2	4	5	5	8	15	24	30	32

$r = ?$

$$\bar{X} = \frac{500}{9} = \boxed{55.55}$$

$$\bar{Y} = \frac{125}{9} = \boxed{13.89}$$

X	Y	$(X-\bar{X})$	$(Y-\bar{Y})$	$(Y-\bar{Y})^2$	$(X-\bar{X})^2$	$(X-\bar{X})(Y-\bar{Y})$
30	2	−25.55	−11.89	141.37	652.80	303.7895
35	4	−20.55	−9.89	97.81	422.30	203.2395
40	5	−15.55	−8.89	79.03	241.80	138.24
45	5	−10.55	−8.89	79.03	111.30	93.79
50	8	−5.55	−5.89	34.69	30.80	32.69
60	15	4.45	1.11	1.23	19.80	4.94
70	24	14.45	10.11	102.21	208.80	146.09
80	30	24.45	16.11	259.53	597.80	393.89
90	32	34.45	18.11	327.97	1186.80	623.89
				1122.87	3472.2	1940.56

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

$$= \frac{1940.56}{\sqrt{3472.2 \times 1122.87}}$$

$$\boxed{r = 0.98}$$

There is strong positive correlation b/w x & y.

Scanned with CamScanner

→ Example:

$r = ?$

x	137	209	113	189	178	200	219
y	23	47	22	40	39	51	49

$\bar{x} = \dfrac{1245}{7} = 177.8$

$\bar{y} = \dfrac{271}{7} = 38.7$

x	y	$x-\bar{x}$	$(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
137	23	-40.8	-15.7	1664.64	246.49	640.56
209	47	31.2	8.3	973.44	68.89	258.96
113	22	-64.8	-16.7	4199.04	278.89	1082.16
189	40	11.2	1.3	125.44	1.69	14.56
178	39	0.2	0.3	0.04	0.09	0.06
200	51	22.2	12.3	492.84	151.29	273.06
219	49	41.2	10.3	1697.44	106.09	424.36
				9152.88	853.43	2693.7

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

$$= \frac{2693.7}{\sqrt{(9152.88)(853.43)}}$$

$$r = \frac{2693.7}{2794.88}$$

$$\boxed{r = 0.964}$$

Strong positive correlation.

Regression Analysis:

In Reg. analysis, we estimate one variable based on another variable.

- The variable being estimated is the dependent variable.
- The variable used to make the estimate or predict the value is the independent variable.
- The relationship b/w dep. & indep. variables is linear.

Regression Equation:

An equation that expresses the linear relationship b/w two variables is called Regression Equation.

$$Y = a + bX$$

independent variable.

Dependent variable.

slope (Regression Coefficient).

Y-intercept.

LEAST SQUARE'S PRINCIPLE:

The least square criterion is used to determin the regression equation. The least squares regression line is of the form:

$$\hat{Y} = a + bX.$$

$\hat{Y}$ → Estimated value of 'Y' for a selected value of 'X'.

⟹ Interpretation of 'a':

It is the value of 'Y' when $X = 0$. e.g.,

Expenditure = a + b (income)

Here 'a' is original value of dependent variable (Expenditure). i.e.,

Expenditure = a + b (zero=Income)

Expenditure = a

$a = \bar{Y} - b\bar{X}$

⟹ Interpretation of 'b':

- It show the amount of change in '$\hat{y}$' for a change of one unit in 'x'.
- A positive value of 'b' indicates a direct/positive relationship b/w two vari-ables & a Negative value of 'b' indicates an inverse/Negative relationship b/w two variables.

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

⇒ **Note :**

The sign of 'b' & sign of 'r' the correlation coefficient, are always the same.

→ Example :

compute regression equation of Y on X for the following data.

X	5	6	8	10	12	13	15	16	17
Y	16	19	23	28	36	41	44	45	50

$$Y = a + bX$$

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

$$\bar{x} = \frac{102}{9} = 11.33$$

$$\bar{y} = \frac{302}{9} = 33.56$$

Scanned with CamScanner

x	y	$(x-\bar{x})$	$(x-\bar{x})^2$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$
5	16	-6.33	40.07	-17.56	111.1548
6	19	-5.33	28.41	-14.56	77.60
8	23	-3.33	11.09	-10.56	35.16
10	28	-1.33	1.77	-5.56	7.39
12	36	0.67	0.45	2.44	1.63
13	41	1.67	2.79	7.44	12.42
15	44	3.67	13.47	10.44	38.31
16	45	4.67	21.81	11.44	53.42
17	50	5.67	32.15	16.44	93.21
			152.00		430.3

$$b = \frac{430.3}{152.0}$$

$$\boxed{b = 2.831}$$

$$a = \bar{y} - b\bar{x}$$

$$a = 33.56 - 2.831(11.33)$$

$$\boxed{a = 1.48}$$

So, Estimated Regression eq. is:

$$\hat{Y} = a + b(X)$$

$$\boxed{\hat{Y} = 1.48 + 2.831(X)}$$

Here;

$b = 2.831$ indicates that the value of Y increase by 2.831 units for one unit increase in X.

→ Example:

In an experiment to measure the stiffness of a spring, the length of the spring under different loads was measured as follows:

X = Loads	3	5	6	9	10	12	15	20	22	28
Y = length	10	12	15	18	20	22	27	30	32	3[illegible]

Find regression equation. (Y on X)

Y is dependent variable & X is indep. variable

$$Y = a + bX$$

$$b = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}$$

$$\bar{X} = 130/10 = 13$$

$$\bar{Y} = 220/10 = 22$$

Scanned with CamScanner

x	y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
3	10	-10	-12	100	120
5	12	-8	-10	64	80
6	15	-7	-7	49	49
9	18	-4	-4	16	16
10	20	-3	-2	9	6
12	22	-1	0	1	0
15	27	2	5	4	10
20	30	7	8	49	56
22	32	9	10	81	90
28	34	15	12	225	180
				598	607

$$b = \frac{607}{598}$$

$$\boxed{b = 1.015}$$

$$a = \bar{y} - b\bar{x}$$

$$= 22 - 1.015(13)$$

$$\boxed{a = 8.8}$$

Scanned with CamScanner

Estimated Regression (Reg.) Eq is:

$$\hat{Y} = a + bX$$

$$\boxed{\hat{Y} = 8.8 + 1.015(x)}$$

⟹ Example :

Find Reg. Eq. of Y on X for the following data:

X	Y	$(X-\bar{X})$	$(Y-\bar{Y})$	$(X-\bar{X})^2$	$(X-\bar{X})(Y-\bar{Y})$
75	82	-4.8	0.1		-0.48
80	78	0.2	-3.9		-0.78
93	86	13.2	4.1		54.12
65	72	-14.8	-9.9		146.52
87	91	7.2	9.1		65.52
71	80	-8.8	-1.9		16.72
98	95	18.2	13.1		238.42
68	72	-11.8	-9.9		116.82
84	89	4.2	7.1		29.82
77	74	-2.8	-7.9		22.12
				1041.6	688.8

$\bar{X} = 79.8$

$\bar{Y} = 81.9$

Scanned with CamScanner

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

$$= \frac{688.8}{1041.6}$$

$$\boxed{b = 0.661}$$

$$a = \bar{y} - b\bar{x}$$

$$= 81.9 - 0.661(79.8)$$

$$\boxed{a = 29.1522}$$

$$\hat{y} = a + bx$$

$$\boxed{\hat{y} = 29.1522 + 0.661(x)}$$

Scanned with CamScanner

→ Example:

Sales Representative	No. of Copiers Sold (Y)	No. of Sales calls (X)
Tom	30	20
Jeff	60	40
Brian	40	20
Greg Fish	60	30
Susan	30	10
Carlos	40	10
Rich	40	20
Mike	50	20
Mark	30	20
Soni	70	30

1) → $r = ?$

4) interpret 'b'.

2) → Reg. Eq = ?

3) → What is the expected No. of copiers Sold by a representative who made 20 calls?

$\bar{x} = 220 | 10 = 22$

$\bar{y} = 450 | 10 = 45$

x	y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
20	30	-2	-15			30
40	60	18	15			270
20	40	-2	-5			10
30	60	8	15			120
10	30	-12	-15			180
10	40	-12	-5			60
20	40	-2	-5			10
20	50	-2	5			-10
20	30	-2	-15			30
30	70	8	25			200
				760	1850	900

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

$$= \frac{900}{\sqrt{760 \times 1850}}$$

$$\boxed{r = 0.759}$$

Interpretation:

There is direct relationship b/w the no. of sales calls & the no. of copiers sold. There is strong positive correlation b/w x & y.

2)

$$Y = a + bX$$

$$b = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\Sigma(X - \bar{X})^2}$$

$$= \frac{900}{760}$$

$$\boxed{b = 1.1842}$$

$$a = \bar{Y} - b\bar{X}$$

$$= 45 - 1.1842(22)$$

$$\boxed{a = 18.9476}$$

$$\hat{Y} = a + bX$$

$$\boxed{\hat{Y} = 18.9476 + 1.1842(X)} \rightarrow (1)$$

3) put $x = 20$ in (1).

$$\hat{Y} = 18.9476 + 1.1842(20)$$

$$\hat{Y} = 42.6316.$$

If a salesperson makes 20 calles, he or she can expect to sell 42.6316 copiers.

4) $b = 1.1842.$

Here 'b' indicates that for each additional sales call made, the sales-representative can expect that there is 1.1842 increase in Y (No. of copiers sold).