22/04
MONDAY

* **Definition of correlation** : To study the degree of relati[on] b/w the variables.

* **Types of correlation**

  - Positive, negative and null correlation
  - Simple, partial and multiple correlation
  - Linear and non-linear correlation

* **Methods of studying correlation coefficient**

  - Scatter diagram method
  - karl pearsons coefficient of correlation
  - Spearmans Rank correlation coefficient
  - kelly's Correlation
  - kendals method
  - Concurrent deviation method

* **coorelation** : Coorelation is a statistical technique that is used to measure and describe the strength and direction of the relationship b/w two variable. Coorelation requires two scores from the same individu[al]

coorelation cooefficient is 'r'

The coorelation ranges from -1 to +1

Types of correlation

- **Positive, negative and null coorelation**

- positive correlation: The coorelation in the same direction. If one variable increases other is also increased and if one variable decreases the other should also decrease

For eg: the length of an iron bar will increase as temperature increases.

- negative coorelation

The correlation in opposite direction is negative coorelation, if one variable increases the other decreases and vice versa, for eg. the vol. of gas will decrease as the pressure increases.

- Nullcoorelation: If there is no relationship b/w the two variables such that the value of one variable change and the other variable remains constant is called null coorelation

- **Simple, Partial and multiple coorelation**

- Simple coorelation: When only two variables are studied it is simple coorelation.

- Partial Coorelation: Here we recognize more than two variables, but consider only two variables to be influencing each other, the effect of other influencing variables being kept constant.

eg:

- <u>Multiple Coorelation</u>: three or more variables are studied the it is multiple coorelation. eg: when one studies the relationship b/w the yield of rice per acre and both the amount of rainfall, amount of fertilizer etc.

• <u>Linear and non-linear coorelation</u>

- <u>Linear coorelation</u>: Coorelation is said to be linear if the ratio of change is constant. eg: The amount of output in a factory is doubled by doubling the no. of workers is a linear coorelation.

- <u>Non-linear coorelation</u>: Coorelation is said to be non linear if the ratio of change is not constant.

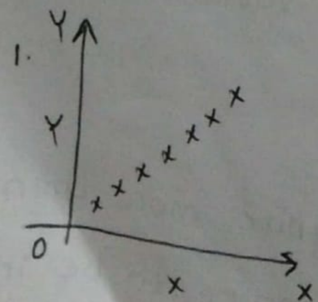
2016 <u>Methods</u>

<u>1. Scatter Diagram method.</u>

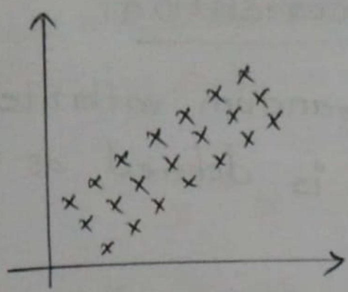
- It is a graphical method. (approximate method)

It is the first developed method.

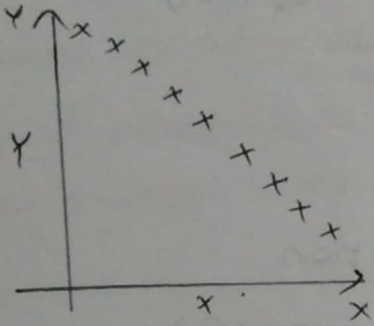
1.



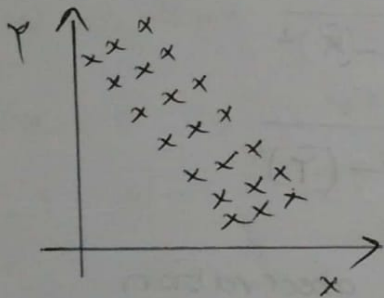
If the graph is forming a straight line moving from lower left corner to upper right corner then → perfect +ve coorelation

2. 

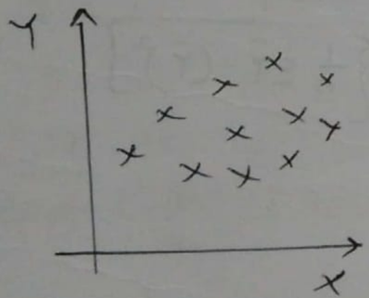
If the graph forms a band moving from lower left to upper right corner then -
positive coorelation, $r = 0$ to 1

3. 

If the graph forms a straight line moving from upper left to lower right then,
perfect -ve coorelation
$r = -1$

4. 

If graph is a band moving from upper ~~rig~~ left to lower right then,
~~per~~ negative coorelation
$r = -1$ to 0

5. 

Scattered graph
$r = 0$
Null coorelation.

2. Karl Pearson's coefficient of correlation

Let x and Y be two random variables, the coorelation coefficient b/w them is defined as follo..

$$r(x,Y) = \frac{Co.V(x,Y)}{\sqrt{Var.(x)\cdot Var(Y)}} \quad OR \quad \frac{Cov(x,y)}{\sigma_x \cdot \sigma_y}$$

Cov → co-variance

Var → variance

σ → standard deviation

where $cov(x,Y) = \frac{1}{n}\Sigma xY - (\bar{x})(\bar{Y})$

$$\sigma_x = \sqrt{\frac{1}{n}\Sigma x^2 - (\bar{x})^2}$$

$$\sigma_Y = \sqrt{\frac{1}{n}\Sigma Y^2 - (\bar{Y})^2}$$

$\bar{X} = \frac{\Sigma X}{n}$; $\bar{Y} = \frac{\Sigma Y}{n}$; n = no. of observation

then $$r(x,Y) = \frac{\frac{1}{n}\Sigma xY - (\bar{x})(\bar{Y})}{\sqrt{\left\{\frac{1}{n}\Sigma x^2 - (\bar{x})^2\right\}\left\{\frac{1}{n}\Sigma \bar{Y} - (\bar{Y})^2\right\}}}$$

Q.1 Compute the coefficient of correlation b/w x and Y using the following data

X :	1	3	5	7	8	10
Y :	8	12	15	17	18	20

Assume the variables are dependent.
no. of observations should be equal.
the method should be identified.

Use Karl Pearson's method.

Given that $n = 6$

X	Y	X^2	Y^2	XY
1	8	1	64	8
3	12	9	144	36
5	15	25	225	75
7	17	49	289	119
8	18	64	324	144
10	20	100	400	200
$\Sigma X = 34$	$\Sigma Y = 90$	$\Sigma X^2 = 248$	$\Sigma Y^2 = 1446$	$\Sigma XY = 582$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{34}{6} = 5.67$$

$$\bar{Y} = \frac{\Sigma Y}{n} = \frac{90}{6} = 15$$

$$\therefore r = \frac{\frac{1}{n}\Sigma XY - (\bar{x})(\bar{y})}{\sqrt{\left\{\frac{1}{n}\Sigma x^2 - (\bar{x})^2\right\}\cdot\left\{\frac{1}{n}\Sigma Y^2 - (\bar{y})^2\right\}}}.$$

$$= \frac{\frac{1}{6}\times 582 - 5.67\times 15}{\sqrt{\left\{\frac{1}{6}\times(248) - (5.67)^2\right\}\left\{\frac{1}{6}\times 1446 - (15)^2\right\}}}$$

$$\frac{97 - 85.05}{\sqrt{(9.18)\times 16}} = \frac{11.95}{12.119} = \underline{\underline{0.986}}$$

Q.2 Calculate the coorelation coefficient for the following ages of husbands (x) and wifes (Y).

X:	23	27	28	28	29	30	31	33	35	36
Y:	18	20	22	27	21	29	27	29	28	29

Given that $n = 10$

As The variables are dependent.

X	Y	X^2	Y^2	XY
23	18	529	324	414
27	20	729	400	540
28	22	784	484	616
28	27	784	729	756
29	21	841	441	609
30	29	900	841	870
31	27	961	729	837
33	29	1089	841	957
35	28	1225	784	980
36	29	1296	841	1044
$\Sigma X = 300$	$\Sigma Y = 250$	$\Sigma X^2 = 9138$	$\Sigma Y^2 = 6414$	$\Sigma XY = 7623$

$\bar{X} = \frac{\Sigma X}{n} = \frac{300}{10} = 30$

$\bar{Y} = \frac{\Sigma Y}{n} = \frac{250}{10} = 25$

$$r = \frac{\frac{1}{n}\Sigma XY - (\bar{X})(\bar{Y})}{\sqrt{\left\{\frac{1}{n}\Sigma X^2 - (\bar{X})^2\right\}\left\{\frac{1}{n}\Sigma Y^2 - (\bar{Y})^2\right\}}}$$

$$= \frac{\frac{1}{10}\times 7623 - 30\times 25}{\sqrt{\left\{\frac{1}{10}\times 9138 - 30^2\right\}\left\{\frac{1}{10}\times 6414 - 25^2\right\}}}$$

$$= \frac{12.3}{\sqrt{13.8 \times 16.4}} = \underline{\underline{0.8176}}$$

Scanned by CamScanner

Short-Cut Method

If the variables are 4-5 digits, use this method.

In this method convert X & Y to U, V by choosing any value as mean (A) & subtract others. ie changing the origin & scale.

$$r(U,V) = \frac{\frac{1}{n}\Sigma UV - (\bar{U})(\bar{V})}{\sqrt{\left\{\frac{1}{n}\Sigma U^2 - (\bar{U})^2\right\}\left\{\frac{1}{n}\Sigma V^2 - (\bar{V})^2\right\}}}$$

where, ~~X-A~~ $U = X - A$

$V = Y - A$

A - assumed mean.

or $U = \frac{X-A}{C}$, $V = \frac{Y-A}{C}$ } used when even after changing origin, if it is big, divide by a common scale (C)

Q. Compute the coorelation coefficient b/w X & Y using following data:

X :	65	67	66	71	67	70	68	69
Y :	67	68	68	70	64	67	72	70

Use short-cut method

Given n=8

X	Y	U = X − A	V = Y − A	U^2	Y^2	UV
65	67	−1	~~−1~~ ~~8~~ 0	1	0	0
67	68	1	~~−2~~ 1	1	1	1
A → 66	68	0	~~−2~~ 1	0	1	0
71	~~70~~ 70 → A	5	~~0~~ 3	25	9	15
67	64	1	−3	1	9	−3
70	67 → A	4	0	16	0	0
68	72	2	5	4	25	10
69	70	3	3	9	9	9
		$\Sigma U = 15$	$\Sigma V = 10$	$\Sigma U^2 = 57$	$\Sigma V^2 = 54$	$\Sigma UV = 32$

$\bar{U} = \frac{\Sigma U}{n} = \frac{15}{8} = 1.875$

$\bar{V} = \frac{\Sigma V}{n} = \frac{10}{8} = 1.25$

$$r = \frac{\frac{1}{n}\Sigma UV - (\bar{U})(\bar{V})}{\sqrt{\left\{\frac{1}{n}\Sigma U^2 - (\bar{U})^2\right\}\left\{\frac{1}{n}\Sigma V^2 - (\bar{V})^2\right\}}}$$

$$= \frac{\frac{1}{8}\times 32 - (1.875)(1.25)}{\sqrt{\left\{\frac{1}{8}\times 57 - 1.875^2\right\}\left\{\frac{1}{8}\times 54 - 1.25^2\right\}}}$$

$$= \frac{\cancel{1.943}\ 1.656}{\sqrt{3.609 \times \cancel{10.54}\ 5.187}} = \underline{\underline{0.382}}$$

Q.2 Obtain spearmans rank coorelation coefficient from the following data

x : 65 67 66 71 67 70 68 69

y : 67 68 68 70 64 67 72 70

Given that, $n=8$

x	Y	Rank of x R_1	Rank of y R_2	$d_i = R_1 - R_2$	d_i^2
65	67	8	6.5	1.5	2.25
67	68	5.5	4.5	1	1
66	68	7	4.5 $\left(\frac{4+5}{2}\right)$	2.5	6.25
71	70	1	2.5	-1.5	+2.25
67	64	5.5 $\left(\frac{5+6}{2}\right)$	8	-2.5	6.25
70	67	2	6.5 $\left(\frac{6+7}{2}\right)$	-4.5	20.25
68	72	4	1	3	9
69	70	3	2.5 $\left(\frac{2+3}{2}\right)$	0.5	0.25
					$\Sigma d_i^2 = 47.5$

When two ranks are repeated, the then take avg of coming two ranks

we know that,

$$P = 1 - \frac{6 \sum_{i=1}^{8} d_i^2 + c}{n(n^2-1)}$$; for repeated ranks

where $c = \frac{m(m^2-1)}{12} = \underset{\text{for (67 in x)}}{\frac{2(2^2-1)}{12}} + \underset{\text{for } Y=70}{\frac{2(2^2-1)}{12}} + \underset{\text{for } Y=68}{\frac{2(2^2-1)}{12}} + \underset{\text{for } Y=67}{\frac{2(2^2-1)}{12}}$

C : 2

$$P: 1 - \frac{6(47.5+2)}{8(8^2-1)}$$

$= \cancel{0.458}$ $\underline{0.411}$

Q.3 10 competetors in a beauty contest were ranked by three judges as follows:

~~Judges~~ Judges	Competitors 1	2	3	4	5	6	7	8	9	10
A	6	5	3	10	2	4	9	7	8	1
B	5	8	4	7	10	2	1	6	9	3
C	4	9	8	1	2	3	10	5	7	6

Discuss which pair of judges have the nearest approach to common taste of beauty.

Find relation b/w judge AB, BC, AC.

For AB

A R_1	B R_2	Rank of A R_3	$d_i = R_1 - R_2$	d_i^2
6	5		1	1
5	8		-3	9
3	4		-1	1
10	7		3	9
2	10		-8	64
4	2		2	4
9	1		8	64
7	6		1	1
8	9		-1	1
1	3		-2	4

$\Sigma d_i^2 = 158$

$$\therefore \quad P = 1 - \frac{6 \sum_{i=1}^{10} d_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 158}{10(10^2-1)} = \underline{\underline{0.0424}}$$

For BC

B R_1	C R_2	$d_i = R_1 - R_2$	d_i^2
5	4	1	1
8	9	-1	1
4	8	-4	16
7	1	6	36
10	2	8	64
2	3	-1	1
1	10	-9	81
6	5	1	1
9	7	2	4
3	6	-3	9

$\Sigma d_i^2 = 214$

$$P = 1 - \frac{6 \sum_{i=1}^{10} d_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 214}{10(10^2-1)} = -0.297$$

$\therefore \rho = 1 - \dfrac{6 \sum_{i=1}^{10} d_i^2}{n(n^2-1)}$

$= 1 - \dfrac{6 \times 26}{10(10^2-1)}$

$= \underline{\underline{0.842}}$

1/09/2016

Regression Analysis

* Regression

* Regression equations or regression lines

Let x and y be two random variables, the regression equation of x and Y is defined as

$$x - \bar{x} = r \frac{\sigma_x}{\sigma_y} (Y - \bar{y}) \rightarrow (1)$$

and the regression equation

$$Y - \bar{y} = r \cdot \frac{\sigma_y}{\sigma_x} (X - \bar{x}) \rightarrow (2)$$

then the regression coefficient of x on Y is defined as

$$b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y} \quad -(3)$$

and the regression coefficient of Y on x is defined as

$$b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x} \quad -(4)$$

Multiply the eqn (3) & (4) ~~(2) & (3)~~ we get.

$$b_{xy} \cdot b_{yx} = \left\{ r \cdot \frac{\cancel{\sigma_x}}{\cancel{\sigma_y}} \right\} \times \left\{ r \times \frac{\cancel{\sigma_y}}{\cancel{\sigma_x}} \right\}$$

$$b_{xy} \cdot b_{yx} = r^2$$

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

ie correlation coefficient is the geometric mean of two regression coefficients.

Q.1 Calculate the coefficient of correlation from the following data. Also obtain the regression lines and estimate the ~~value~~ value of y when x = 6.2

x:	1	2	3	4	5	6	7	8	9
y:	9	8	10	12	11	13	14	16	15

Given

n = 9

X	Y	X^2	Y^2	XY
1	9	1	81	9
2	8	4	64	16
3	10	9	100	30
4	12	16	144	48
5	11	25	121	55
6	13	36	169	78
7	14	49	196	98
8	16	64	256	128
9	15	81	225	135
$\Sigma X = 45$	$\Sigma Y = 108$	$\Sigma X^2 = 285$	$\Sigma Y^2 = 1356$	$\Sigma XY = 597$

Scanned by CamScanner

$$\text{cov}(x,y) = \frac{1}{n}\Sigma xy - (\bar{x})(\bar{y})$$

$$= \frac{1}{9}\times 597 - (5\times 12)$$

$$= \underline{6.33}$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{45}{9} = 5$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{108}{9} = 1$$

$$\sigma_x = \sqrt{\frac{1}{n}(\Sigma x^2) - (\bar{x})^2} = \sqrt{\frac{1}{9}\times 285 - 5^2} = 2.58$$

$$\sigma_y = \sqrt{\frac{1}{n}\Sigma y^2 - (\bar{y})^2} = \sqrt{\frac{1}{9}\times 1356 - 12^2} = 2.58$$

$$\therefore r = \frac{\text{cov}(x,y)}{\sigma_x \cdot \sigma_y} = \frac{6.33}{2.58\times 2.58} = \underline{0.95}$$

The regression eqn of x on y is

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

$$x - 5 = 0.95\times\frac{2.58}{2.58}\times(y-12)$$

$$x - 5 = 0.95(y-12)$$

$$x = 0.95y - 12\times 0.95 + 5$$

$$x = 0.95y - 6.4 \quad \text{— (1)}$$

Scanned by CamScanner

regression eqn of Y on X is

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

$$Y - 12 = 0.95 \times \frac{2.58}{2.58} \times (x - 5)$$

$$Y - 12 = 0.95x - 5 \times 0.95$$

$$Y = 0.95x - 4.75 + 12$$

$$Y = 0.95x + 7.25 \quad - (2)$$

If $x = 6.2$, then Y:

$$Y = 0.95 \times 6.2 + 7.25$$

$$Y = 13.14$$

Q.2 Obtain the eqn of regression lines from the following data. Hence find the coorelation b/w X and Y. Also estimate the value of (i) Y, when x = 38

(ii) x, when y = 18.

x :	22	26	29	30	31	31	34	35
Y :	20	20	21	29	27	24	27	31

Q.3 Obtain the eqn of the lines of regression from the following data:

(i)

x :	1	2	3	4	5	6	7
y :	9	8	10	12	11	13	14

Scanned by CamScanner

(ii) x: 45 55 56 58 60 65 68 70 75 80 85

y: 56 50 48 60 62 64 65 70 74 82 90

Q.2 Given, $n = 8$

X	Y	X^2	Y^2	XY
22	20	484	400	440
26	20	676	400	520
29	21	841	441	609
30	29	900	841	870
31	27	961	729	837
31	24	961	576	744
34	27	1156	729	918
35	31	1225	961	1085
$\Sigma X = 238$	$\Sigma Y = 199$	$\Sigma X^2 = 7204$	$\Sigma Y^2 = 5077$	$\Sigma XY = 6023$

$$\text{cov}(X,Y) = \frac{1}{n}\Sigma XY - (\bar{x})(\bar{y})$$

$$= \frac{1}{8} \times 6023 - (29.75 \times 24.87)$$

$$= \underline{\underline{12.99}}$$

$\bar{x} = \frac{\Sigma X}{n} = \frac{238}{8}$

$= 29.75$

$\hat{Y} = \frac{\Sigma Y}{n} = \frac{199}{8}$

$= 24.87$

$$\sigma_x = \sqrt{\frac{1}{n}(\Sigma x^2) - (\bar{x})^2} = \sqrt{\frac{1}{8} \times 7204 - 29.75^2}$$

$$= \underline{\underline{3.92}}$$

$\sigma_y = \sqrt{\frac{1}{n}(\Sigma Y^2) - (\bar{y})^2} = \sqrt{\frac{1}{8} \times 5077 - 24.87^2}$

$= 4.013$

$\therefore r = \dfrac{cov(x,y)}{\sigma_x \cdot \sigma_y} = \dfrac{12.99}{3.92 \times 4.013}$

$= \underline{\underline{0.82}}$

6i The regression eqn of x on Y is

$X - \bar{x} = r \cdot \dfrac{\sigma_x}{\sigma_y} (Y - \bar{y})$

$X - 29.75 = \dfrac{3.92}{4.013} \times 0.82 \times (Y - 24.87)$

$X - 29.75 = 0.8(Y - 24.87)$

$X - 29.75 = 0.8Y - 19.92$

$X = 0.8Y + 9.83 \quad — ①$

regression eqn of Y on X is

$Y - \bar{y} = r \cdot \dfrac{\sigma_y}{\sigma_x} (X - \bar{x})$

$Y - 24.87 = 0.82 \times \dfrac{4.013}{3.92} \times (X - 29.75)$

$Y - 24.87 = 0.839 (X - 29.75)$

$Y - 24.87 = 0.839X - 24.973$

$Y = 0.839X - 0.103 \quad — ②$

(i) when $x = 38$

$$Y = 0.839X - 0.103$$

$$= 0.839 \times 38 - 0.103$$

$$Y = \underline{\underline{31.78}}$$

(ii) when $Y = 18$

$$X = 0.8Y + 9.83$$

$$= 0.8 \times 18 + 9.83$$

$$\underline{\underline{X = 24.23}}$$

Q.3

(i)

x	y	x^2	Y^2	XY
1	9	1	81	9
2	8	4	64	16
3	10	9	100	30
4	12	16	144	48
5	11	25	121	55
6	13	36	169	78
7	14	49	196	98
$\Sigma X = 28$	$\Sigma Y = 77$	$\Sigma X^2 = 140$	$\Sigma Y^2 = 875$	$\Sigma XY = 334$

$cov(X,Y) = \frac{1}{n} \Sigma XY - (\bar{X})(\bar{Y})$

$\bar{X} = \frac{\Sigma X}{n} = \frac{28}{7} = 4$

$= \frac{1}{7} \times 334 - (4 \times 11)$

$\bar{Y} = \frac{\Sigma Y}{n} = \frac{77}{7} = 11$

$= 3.714$

$\sigma_x = \sqrt{\frac{1}{n} \Sigma X^2 - (\bar{X})^2} = \sqrt{\frac{1}{7} \times 140 - 4^2}$

$= 2$

$\sigma_y = \sqrt{\frac{1}{n} \Sigma Y^2 - (\bar{Y})^2} = \sqrt{\frac{1}{7} \times 875 - 11^2}$

$= 2$

$\therefore r = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y} = \frac{3.714}{2 \times 2} = \underline{\underline{0.9285}}$

The regression eqn of X on Y is

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

$X - 4 = 0.9285 \times \frac{2}{2} \times (Y - 11)$

$X - 4 = 0.9285 (Y - 11)$

$X - 4 = 0.9285Y - 10.213$

$X = 0.9285Y - 6.213$ — ①

The regression eqn of Y on X is

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

$$(Y - 11) = 0.9285 \times \frac{2}{2} \times (X - 4)$$

$$Y - 11 = 0.9285(X - 4)$$

$$Y - 11 = 0.9285X - 3.714$$

$$Y = 0.9285X + 7.286 \quad \text{--- (2)}$$

(ii)

x	y	x^2	y^2	xy
45	56	2025	3136	2520
55	50	3025	2500	2750
56	48	3136	2304	2688
58	60	3364	3600	3480
60	62	3600	3844	3720
65	64	4225	4096	4160
68	65	4624	4225	4420
70	70	4900	4900	4900
75	74	5625	5476	5550
80	82	6400	6724	6560
85	90	7225	8100	7650
$\Sigma X = 717$	$\Sigma Y = 721$	$\Sigma X^2 = 48149$	$\Sigma Y^2 = 48905$	$\Sigma XY = 48398$

Scanned by CamScanner

$$\text{cov}(x,y) = \frac{1}{n}\Sigma XY - (\bar{x})(\bar{y})$$

$$\bar{x} = \frac{\Sigma X}{n} = \frac{717}{11} = 65.18$$

$$= \frac{1}{11} \times 48398 - (65.18 \times 65.54)$$

$$= \underline{\underline{127.92}}$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{721}{11} = 65.55$$

$$\sigma_x = \sqrt{\frac{1}{n}\Sigma x^2 - (\bar{x})^2} = \sqrt{\frac{1}{11} \times 48149 - 65.18^2}$$

$$= 11.3467$$

$$\sigma_y = \sqrt{\frac{1}{n}\Sigma y^2 - (\bar{y})^2} = \sqrt{\frac{1}{11} \times 48905 - 65.55^2}$$

$$= 12.2109$$

$$\therefore r = \frac{\text{cov}(x,y)}{\sigma_x \cdot \sigma_y} = \frac{127.92}{11.3467 \times 12.2109} = \underline{\underline{0.923}}$$

The regression eqn of X on Y is

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

$$(X - 65.18) = 0.923 \times \frac{11.3467}{12.2109} \times (Y - 65.55)$$

$$X - 65.18 = 0.8576\,(Y - 65.55)$$

$$X - 65.18 = 0.8576Y - 56.22$$

$$X = \underline{\underline{0.858Y + 8.959}} \quad \text{— (1)}$$

Scanned by CamScanner

The eqn of Y on x

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

$$Y-65.55 = 0.923 \times \frac{12.2109}{11.3467} \times (x-65.18)$$

$$Y-65.55 = 0.9932(x-65.18)$$

$$Y-65.55 = 0.9932x - 64.74$$

$$\underline{\underline{Y = 0.9932x + 0.807}} \quad - (2)$$

Q.4 In a partially destroyed laboratory record of an analysis of ~~a~~ correlation data, the following results are only legible. variance of $x = 9$ ~~variance~~, regression eqns are

$8x - 10y + 66 = 0$ and $40x - 18y = 214$

What ~~where~~ were the (i) mean values of x and y

(ii) the correlation coefficient b/w x and y.

(iii) standard deviation of y.

Given that

variance of $x = 9$; ie $\sigma_x^2 = 9$, $\sigma_x = 3$

(i) Assume, the eqns (1) and (2) are passing through the points $(\bar{x}, \bar{y})$.

$$8x - 10y + 66 = 0 \quad - (1)$$

$$40x - 18y = 214 \quad - (2)$$

$8\bar{x} - 10\bar{y} = -66 \quad -(3)$

$40\bar{x} - 18\bar{y} = 214 \quad -(4)$

Solving (3) & (4)

$\bar{x} = 13 \quad , \bar{y} = 17$

$\therefore$ The mean values of x and y is $\underline{\underline{\bar{x} = 13 \text{ and } \bar{y} = 17}}$

(ii)

Take eqn (1) $\Rightarrow 8x - 10y = -66$

$-10y = -66 - 8x$

$10y = 66 + 8x$

$y = \frac{66}{10} + \frac{8x}{10} \quad (y = a + bx)$

$\therefore$ The regression coefficient of y on x is

$b_{yx} = \frac{8}{10}$

Take eqn (2) $\Rightarrow 40x - 18y = 214$

$40x = 214 + 18y$

$x = \frac{214}{40} + \frac{18}{\cancel{240}}y \quad (x = a + by)$

$\therefore$ The regression coefficient of x on y is

$b_{xy} = \frac{18}{40}$

Then w.k.t, $r = \sqrt{b_{xy} \cdot b_{xy}}$

$= \sqrt{\frac{8}{10} \times \frac{18}{\cancel{40}\,5}} \quad = \underline{0.6}$

(iii) we know that

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

$$\frac{8}{10} = 0.6 \times \frac{\sigma_y}{3}$$

$$\sigma_y = \frac{8 \times 3}{0.6 \times 10} = \underline{\underline{4}}$$

* while calculating, b_{yx} & b_{xy}, chelk that ~~one~~ if one value is < unity, then other should be > unity or else both should be < unity. based on that decide whether x ~~is~~ on LHS or y on LHS for both eqns.

Q.5 The eqns of lines of regression are given by $x+2y-5=0$ and $2x+3y-8=0$, and variance of x is 12 compute the values of $\bar{x}$, $\bar{y}$, σ_y^2 and r_{xy}

Given, Variance $\sigma_x^2 = 12$, $\sigma_x = 3.464$

(i) Assume the eqn ① & ② are passing through the points $\bar{x}$ and $\bar{y}$

$$x + 2y = 5 \quad -①$$

$$2x + 3y = 8 \quad -②$$

$$\bar{x} + 2\bar{y} = 5 \quad -③$$

$$2\bar{x} + 3\bar{y} = 8 \quad -④$$

Scanned by CamScanner

Solving ③ & ④

$\bar{x} = 1$, $\bar{y} = 2$

$\therefore$ The mean of x and y are ; $\bar{x} = 1$, $\bar{y} = 2$

(iii) Take eqn ① => $x + 2y - 5 = 0$.

$x = 5 - 2y$ $\quad\quad$ $x = a + by$

~~b~~

$\therefore$ The regression coefficient of y on x is

$b_{xy} = -2$

② => $2x + 3y - 8 = 0$.

~~2x~~ $3y = 8 - 2x$.

$y = \frac{8}{3} - \frac{2}{3}x$. $\quad (y = a + bx)$

$\therefore$ The regression coefficient of x on y is

$b_{yx} = \frac{-2}{3}$

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

$= \sqrt{-2 \times -\frac{2}{3}} = 1.15$

r value should be always ~~<~~ -1 to +1

$\therefore$ change assumption

Take eqn 1

$$x+2y-5=0$$

$$2y=5-x$$

$$y=\frac{5}{2}-\frac{x}{2} \quad \text{— (a)}$$

$y=a+bx$

$$\therefore b_{yx}=-0.5$$

take eqn 2

$$2x+3y-8=0$$

$$2x=8-3y$$

$$x=4-\frac{3}{2}y \qquad (x=a+by)$$

$$\therefore b_{xy}=-1.5$$

~~If~~ * b_{xy} & b_{yx} should have same sign

$$\therefore r=\sqrt{b_{xy}\times b_{yx}}=\sqrt{-0.5\times-1.5}$$

$$=\underline{\underline{0.866}}$$

we know that

$$b_{yx}=r.\frac{\sigma y}{\sigma x}$$

$$-0.5=0.866\times\frac{\sigma y}{3.464}$$

$$\sigma y=-2$$

$$b_{xy}=r.\frac{\sigma x}{\sigma y}$$

$$-1.5=0.866\times\frac{3.464}{\sigma y}$$

Scanned by CamScanner

$\sigma_y^2 = 4$

Q.6 Variance of $x = 1$, the regression eqns are $3x + 2y = 26$ and $6x + y = 31$. Find (i) mean values of x and y (ii) the standard deviation of y (iii) corelation coefficient b/w x and y

Given variance of $x = 1$. $\sigma_x^2 = 1$, $\sigma_x = 1$

~~The~~ (i) Assume the eqn ① & ② are passing through the points $\bar{x}$ and $\bar{y}$

$$3x + 2y = 26 \quad -①$$
$$6x + y = 31 \quad -②$$

$$3\bar{x} + 2\bar{y} = 26 \quad -③$$
$$6\bar{x} + \bar{y} = 31 \quad -④$$

solving ③ & ④, $\bar{x} = 4$, $\bar{y} = 7$

$\therefore$ The mean values of x & y are $\bar{x} = 4$ and $\bar{y} = 7$

(iii) take eqn ①

$$3x + 2y = 26$$
$$3x = 26 - 2y$$
$$x = \frac{26}{3} - \frac{2}{3}y \qquad (x = a + by)$$

$$\therefore b_{xy} = -\frac{2}{3}$$

take eqn (2)

$6x+y=31$

$y=31-6x$ $\quad (y=a-bx)$

$\therefore b_{yx}=-6$

$\therefore r=\sqrt{b_{xy}\times b_{yx}}=2 \quad r>2 \quad \therefore$ change assumpti

take eqn (1)

$3x+2y=26$

$2y=26-3x$

$y=26-\frac{3x}{2} \quad (y=a+bx)$

$\therefore b_{yx}=-1.5$

take eqn (2)

$6x+y=31$

$6x=31-y$

$x=\frac{31}{6}-\frac{y}{6}$

$\therefore b_{xy}=-\frac{1}{6}$

$\therefore r=\sqrt{b_{xy}\times b_{yx}}=\sqrt{-1.5\times-\frac{1}{6}}$

$\therefore r=\underline{\underline{0.5}}$

(iii) w.k.t

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

$$-1.5 = 0.5 \times \frac{\sigma_y}{1}$$

$$\sigma_y = -3$$

1/09/2016 **Method of Least Squares**

This is the most widely used method of obtaining trend.
If we fit a straight line using the method of least squares, then the sum of deviations of points on either side of the line is equal to zero. (ie. sum of the deviations of the computed values from the actual value is zero.). Also the sum of the squares of these deviations will be least as compared to those obtained using other lines. Hence the name "method of least squares". The straight line obtained using this method is called the line of best fit.

Fitting a straight line using the method of least square

$y = a + bx$

$\Sigma y = na + b\Sigma x$

$\Sigma xy = a \cdot \Sigma x + b\Sigma x^2$

where 'n' represents the no. of years of which the data are given.

Q.1 By the method of least square find the best fitting straight line for the data given below.

x:	5	10	15	20	25
y:	15	19	23	26	30

We know that the straight line

$$y = a + bx$$

and the normal eqns are

$$\Sigma y = na + b\Sigma x$$

$$\Sigma xy = a \cdot \Sigma x + b\Sigma x^2$$

x	y	x^2	xy
5	15	25	75
10	19	100	190
15	23	225	345
20	26	400	520
25	30	625	750
$\Sigma x = 75$	$\Sigma y = 113$	$\Sigma x^2 = 1375$	$\Sigma xy = 1880$

$n = 5$

$5a + b \times 75 = 113$ — ①

~~1800.~~

$75a + 1375b = 1880$ — ②

solving ① & ② $\cancel{a = 11.5}$ $\cancel{b = 0.74}$

$a = 11.5$

$b = 0.74$

$\therefore$ The best fitting straight line is $y = a + bx$

ie $y = 11.5 + 0.74x$

Q2 Fit a straight to the data given below. Also estimate the value of y at $x = 2.5$.

x :	0	1	2	3	4
y :	1	1.8	3.3	4.5	6.3

x	y	x^2	xy
0	1	0	0
1	1.8	1	1.8
2	3.3	4	6.6
3	4.5	9	13.5
4	6.3	16	25.2
$\Sigma x = 10$	$\Sigma y = 16.9$	$\Sigma x^2 = 30$	$\Sigma xy = 47.1$

Scanned by CamScanner

$n = 5$

we know that the straight line

$y = a + bx$

and the normal eqns are

$\Sigma y = na + b\,\Sigma x$ — ①

$\Sigma xy = a\,\Sigma x + b\,\Sigma x^2$ — ②

① => $5a + 10b = 16.9$

② => $10a + 30b = 47.1$

$a = 0.72$

$b = 1.33$

∴ The best fitting straight line is $\underline{\underline{y = 0.72 + 1.33x}}$

at, $x = 2.5$

$y = 0.72 + 1.33 \times 2.5$

$= \underline{\underline{4.045}}$

Q.3 Fit a straight line for the following data and also estimate the value of y at $x = 70$

x:	71	68	73	69	67	65	66	67
y:	69	72	70	70	68	67	68	64

~~x~~ ~~y~~ ~~x^2~~ ~~xy~~

<u>shortcut method</u>

$X = y - c$

$Y = x - c$

c - any assumed value

h

$X = x - 67$	$Y = y - 68$	~~X^2~~	XY	X^2
$71-67=4$	$69-68=1$	~~16~~	4	16
$68-67=1$	$72-68=4$	~~1~~	4	1
$73-67=6$	$70-68=2$	~~36~~	12	36
$69-67=2$	$70-68=2$	~~4~~	4	4
$67-67=0$	$68-68=0$	~~0~~	0	0
$65-67=-2$	$67-68=-1$		2	4
$66-67=-1$	$68-68=0$		0	1
$67-67=0$	$64-68=-4$		0	0
$\Sigma X = 10$	$\Sigma Y = 4$		$\Sigma XY = 26$	$\Sigma X^2 = 62$

Given

$n = 8$

W.k.t the straight line

$y = a + bx$

and the normal eqns are

$\Sigma y = na + b\Sigma x$

$\Sigma xy = a \cdot \Sigma x + b \cdot \Sigma x^2$

$8a + 10b = 4$ — (1)

$10a + 62b = 26$ — (2)

$\therefore$ ~~$a = 0.3276$~~ $a = -0.03$

~~$b = 0.388$~~ $b = 0.424$

The best fitting straight eqn is $y = a + bx$

$\cancel{X = x - 67}$ $\quad X = x - 67 \quad Y = y - 68$

$(y-68) = a + b\times(x-67)$

$y - 68 = a + bx - 67b$

$y = a + bx - 67b + 68$

$\quad - 0.03 + 0.424(x-67) + 68$

$= a + b(x-67) + 68 \Rightarrow \cancel{0.316 + 0.368(x-67) + 68}$

$= \cancel{68.316 + 0.368x - 24.656} \Rightarrow \cancel{43.66 + 0.368x}$

$\therefore$ at $x = 70$, $\cancel{y = 43.66 + 0.368 \times 70}$

$y = 67.97 + 0.424x - 28.408$

$= \cancel{69.424} \quad 69.24$

Fitting A Second Degree Parabola

Q. Fit a parabola by the method of least square to the following data; Also estimate y, at $x = 6$

x :	1	2	3	4	5
y :	5	12	26	~~7~~60	97

We know that the parabola be

$$y = a + bx + cx^2$$

and the normal eqns are

$$\Sigma y = na + b \cdot \Sigma x + c \cdot \Sigma x^2$$

$$\Sigma xy = a \cdot \Sigma x + b \cdot \Sigma x^2 + c \cdot \Sigma x^3$$

$$\Sigma x^2 y = a \cdot \Sigma x^2 + b \Sigma x^3 + c \cdot \Sigma x^4$$

x	y	x^2	x^3	x^4	xy	$x^2 y$
1	5	1	1	1	5	5
2	12	4	8	16	24	48
3	26	9	27	81	78	234
4	60	16	64	256	240	960
5	97	25	125	625	485	2425
$\Sigma x = 15$	$\Sigma y = 200$	$\Sigma x^2 = 55$	225	979	832	3672

$$5a + 15b + 55c = 200 \quad \text{—} (1)$$

$$15a + 55b + 225c = 832 \quad \text{—} (2)$$

$$55a + 225b + 979c = 3672 \quad \text{—} (3)$$

Solving (1), (2) & (3)

a = ~~[illegible]~~ 10.4

b = − ~~[illegible]~~ −11.08

c = ~~8~~ 5.71

The best fitting 2nd degree parabola eqn is

$~~y = 26.4 + (-24.8)x + 8x^2~~$

$~~y = 26.4 - 24.8x + 8x^2~~$

~~at x = 6~~

$~~y = 26.4~~$

$y = 10.4 + (-11.08)x + 5.71x^2$

at $x = 6$

$y = 10.4 - (11.08 \times 6) + 5.71 \times 6^2$

$y = 149.48$

Q.2 Fit a second degree parabola with the data

x : 1929 1930 1931 1932 1933 1934 1935

y : 352 356 357 358 360 361 361

Apply shortcut method

$X = x - 1932$

$Y = y - 358$

X	Y	X^2	X^3	X^4	XY	X^2Y
-3	-6	9	-27	81	18	-54
-2	-2	4	-8	16	4	-8
-1	-1	1	-1	1	1	-1
0	0	0	0	0	0	0
1	2	1	1	1	2	2
2	3	4	8	16	6	12
3	3	9	27	81	9	27
$\Sigma X = 0$	$\Sigma Y = -1$	$\Sigma X^2 = 28$	$\Sigma X^3 = 0$	$\Sigma X^4 = 196$	$\Sigma XY = 40$	$\Sigma X^2Y = -22$

We know that the parabola eqn is

$$y = a + bx + cx^2$$

and normal eqn are:

$$\Sigma Y = na + b\Sigma X + c\,\Sigma X^2$$

$$\Sigma XY = a\,\Sigma X + b\Sigma X^2 + c\,\Sigma X^3$$

$$\Sigma X^2 Y = a\,\Sigma X^2 + b\,\Sigma X^3 + c\,\Sigma X^4$$

$7a + 0b + 28c = -1$ — (1)

$0a + 28b + 0c = 40$ — (2)

$28a + 0b + 196c = -22$ — (3)

(1) & (2)

$7a + 0b + 28c = -1$

$0a + 28b + 0c = 40$

$7a + 28c = -1$ — (a)

$28b = 40$ — (b)

$28a + 196c = -22$ — (c)

$$7a + 0b + 28c = -1 \quad — ①$$

$$0a + 28b + 0c = 40 \quad — ②$$

$$28a + 0b + 196c = -22 \quad — ③$$

$$28b = 40$$

$$\underline{b = 1.428}$$

$$7a + 28c = -1 \quad — ①$$

$$28a + 196c = -22 \quad — ③$$

$$① \times 4 = 28a + \overset{112}{\cancel{28}}c = -4$$

$$- 28a + 196c = -22$$

$$\overset{-84}{\cancel{588}}c = \cancel{+28}$$

$$\underline{c = \cancel{-0.044}} \quad -0.214$$

$$7a + 28 \times \overset{-0.214}{\cancel{-0.044}} = -1$$

$$\underline{a = 0.713}$$

$$a = 0.713, \ b = 1.428, \ c = -0.214$$

Parabola eqn is

$$(y-358) = 0.713 + 1.428(x-1932) + $$
$$-0.214(x-1932)^2$$

$$(y-358) = 0.713 + 1.428x - 2758.890 - 0.214(x^2 - 38646$$
$$+ 3732624$$