UNIT-III
CORRELATION

Meaning, Types, Methods of Correlation, Uses, Scatter Diagram, Karl Pearson's coefficient of Correlation, Rank Correlation, Regression, Differences between Correlation and Regression, Regression Lines, Regression Equations, Uses and Limitations of Regression Analysis

CORRELATION

DEFINITION

- The variables are said to be correlated if the changes in one variable results in a corresponding change in the other variable. That is, when two variables move together we say they are correlated.
- Boddington states that " whenever some definite connection exists between the two or more groups, classes or series or data there is said to be correlation".
- Bowely defines correlation as, " when two quantities are so related that the fluctuations in one are in sympathy with the fluctuations of the other, that an increase or decrease of the one is found in connection with the increase or decrease of the other and greater the magnitude of change in the other, the quantities are said to be correlated"

- According to A. M Tuttle, “correlation is an analysis of the association between two or more variables.
- Simply, correlation may be defined as the degree of relationship between two variables.
- “Correlation analysis” the purpose of which is the determination of degree of relationship between the variables
- The method of correlation is developed by FRANCIS GALTON in 1885.

TYPES OF CORRELATION

The different types of correlation are

- Positive and Negative correlation
- Linear and Non-linear correlation
- Simple, Multiple and Partial correlation.

- Positive Correlation

  When the values of two variables move same direction, correlation is said to be positive

  ie; an increase in the value of one variable results into an increase in the other variable also or if decrease in the value of one variable results into a decrease in the other variable also correlation is said to be positive.

  Eg. Temperature and volume

- Negative correlation

  When the values of two variables move opposite direction, correlation is said to be negative.

ie; an increase in the value of one variable results into an decrease in the other variable also or if decrease in the value of one variable results into a increase in the other variable also correlation is said to be positive.

Eg. Pressure and volume

- Linear Correlation

  When the amount of change in one variable leads to a constant ratio of change in the other variable, correlation is said to be linear.

- Non linear Correlation

  Correlation is said to be non linear (curve linear) when the amount of change in one variable is not in constant ratio to the change in the other variable.

- Simple correlation

In the study of relationship between the variables, if there are only two variables, the correlation is said to be simple.

When one variable is related to a number of others, the correlation is not simple.

- Multiple correlation

In the study of multiple correlation we measure the degree of association between one variable on one side and all the other variable together on the other side.

- Partial correlation

In partial correlation we study the relationship of one variable with one of the other variables presuming that the other variable remains constant.

Degree of correlation

The degree or the intensity of the relationship between two variables can be ascertained by finding the value of coefficient of correlation. The degree of correlation can be classified into

- Perfect correlation

  When the change in the two variables is such that with an increase in the value of one, the value of the other increases in a fixed proportion, correlation is said to be perfect. The perfect correlation may be positive or negative. Coefficient of correlation is +1 for perfect positive correlation and it is –1 for perfect negative correlation.

- No correlation

  If the changes in the value of one variable are in association with the changes in the value ot other variablether will be no correlation. When there is no correlation the coefficient of correlation is zero.

- Limited degree of correlation

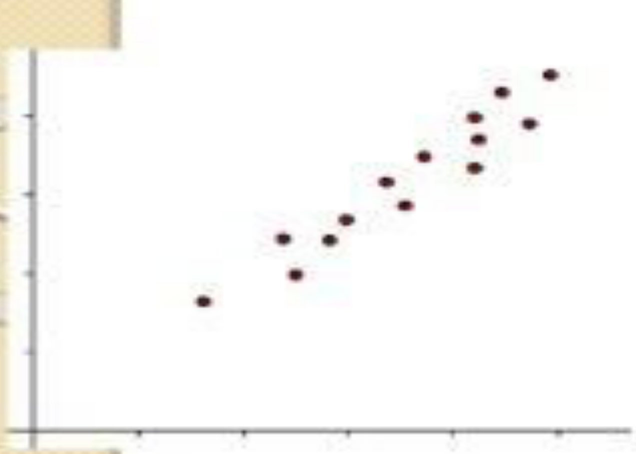
  In between no correlation and perfect correlation there may be limited degree of correlation. It may also be positive or negative. Limited degree of correlation may be termed as high, moderate or low. For limited degree of correlation the coefficient of correlation lies between 0 and 1 numerically.

METHODS FOR STUDYING CORRELATION

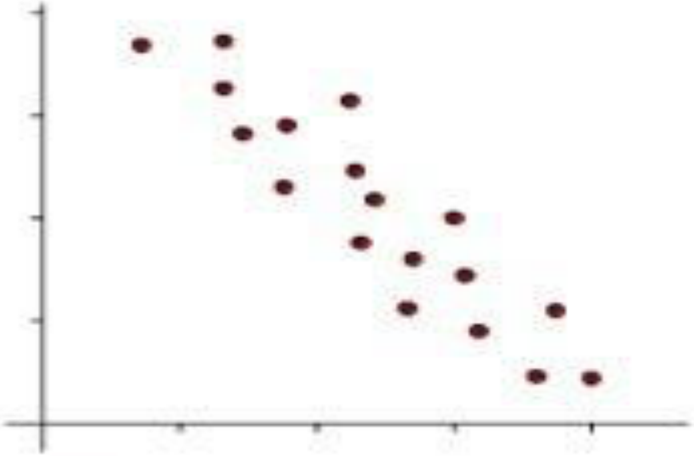
Correlation between two variables can be measured by both graphic and algebraic method. **Scatter diagram** and **correlation graph** are the two important graphic methods while **coefficient of correlation** is an algebraic method used for measuring correlation.

a) Scatter diagram

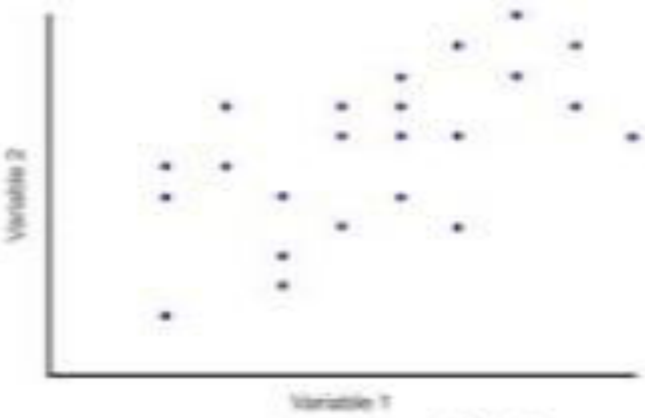
This is a graphical method of studying the correlation between two variables. One of the variable is shown on the X- axis and the other on the Y-axis. Each pair of values is plotted on the graph by means of a dot mark. After all the items are plotted we get as many dots on the graph paper as the number of points. If these points show some trend either upward or downward, the two variables are said to be correlated. If the point do not show any trend, the two variables are not correlated.



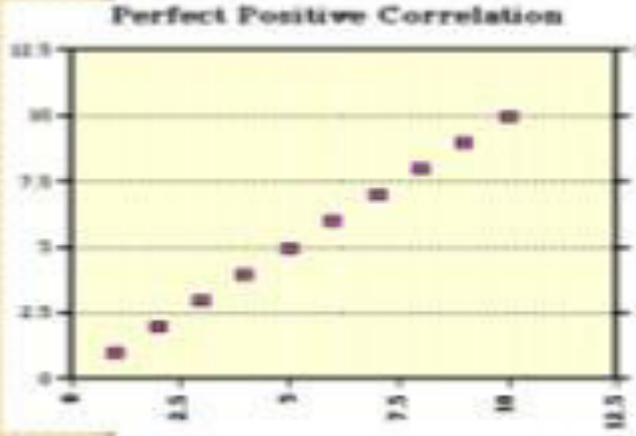
positive correlation



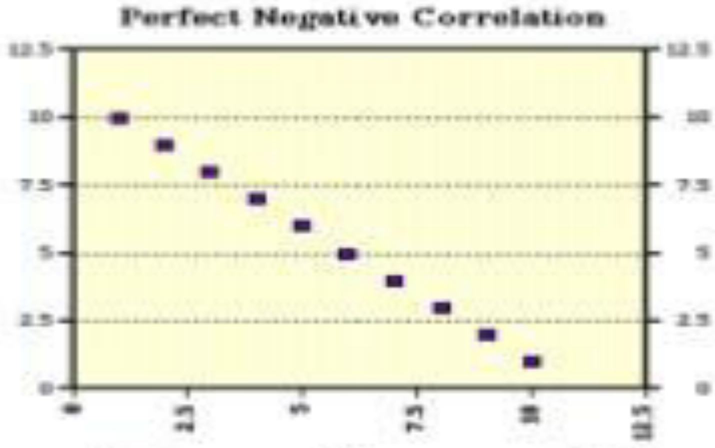
Negative correlation



zero correlation



perfect positive correlation



perfect negative correlation

b) Correlation Graph

Under this method, separate curves are drawn for the X variable and Y variable on the same graph paper. The values of the variable are taken as ordinates of the points plotted. From the direction and closeness of the two curves we can infer whether the variables are related. If both the curves are move in the same direction(upward or downward), correlation is said to be positive. If the curves are moving in the opposite direction correlation is said to be negative.

c) Coefficient of correlation

- Coefficient of correlation is an algebraic method of measuring correlation.
- Under this method, we measure correlation by finding a value known as the coefficient of correlation using an appropriate formula.
- Coefficient of correlation is a numerical value. It shows the degree or the extent of correlation between two variables.

- Coefficient of correlation is a pure number lying between -1 and +1.
- When the correlation is negative, it lies between -1 and 0.
- When the correlation is positive, it lies between 0 and 1.
- When the correlation of coefficient is zero, it indicates that there is no correlation between the variables.
- When the correlation coefficient is 1,there is perfect correlation.
- Between no correlation and perfect correlation there are varying degree of correlation.

Coefficient of correlation can be computed by applying the methods given below

- Karl Pearson's method
- Spearman's method
- Concurrent deviation method

PROPERTIES OF COEFFICIENT OF CORRELATION

1. Correlation coefficient has a well defined formula
2. Correlation coefficient is a pure number and is independent of its units of measurement.
3. It lies between -1 and +1.
4. Correlation coefficient does not change with reference to change of origin or change of scale.
5. Correlation of coefficient between x and y is same as that between y and x.

IMPORTANCE OF CORRELATION

- Correlation helps to study the association between two variables.
- Coefficient of correlation is vital for all kinds of research work.
- It helps in establishing Validity or Reliability of an evaluation tool.
- It helps to ascertain the traits and capacities of pupils while giving guidance or counselling.
- Correlation analysis helps to estimate the future values.

? What would be your interpretation if the correlation coefficient equal to

1) r = 0

Ans : There is no correlation between the variables

2) r = -1

Ans: negative perfect correlation

3) r =0.2

Ans: low positive correlation

4) r = 0.9

Ans: high positive correlation

5) r = -0.3

Ans: low negative correlation

6) r = - 0.8

Ans: High negative correlation

COMPUTATION OF COEFFICIENT OF CORRELATION

There are two different methods of computing coefficient of correlation .

They are ,

- ✓ RANK DIFFERENCE METHOD
- ✓ PRODUCT MOMENT METHOD

PRODUCT MOMENT METHOD

- Most widely used measure of correlation is the Pearson's Product moment Correlation Coefficient.
- This method is also known as Pearson's product moment method in honour of Karl Pearson , who is said to be the inventor of this method.
- The coefficient of correlation computed by this method is known as the product moment coefficient of correlation or Pearson's correlation coefficient.
- It is represented as 'r' .

The standard formula used in the computation of Pearson’s product moment correlation coefficient is as follows :

$$\mathrm{r} = \frac{\mathrm{N}\sum(XY) - \sum X \sum Y}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

Where,

- N - the no: of pairs of data
- Σ - the summation of the items indicated
- ΣX - the sum of all X scores
- ΣX^2 - each X score should be squared and then those squares summed{the sum of the X squared scores}
- $(\Sigma X)^2$ - X scores should be summed and the total squared(the squares of the sum of all the X scores)

- ΣY – the sum of all Y scores
- ΣY^2 - each Y score should be squared and then those squares summed
- $(\Sigma Y)^2$ - Y score should be summed and the total squared

CALCULATE THE CORRELATION OF THE FOLLOWING DATA

SUBJECT	SCORES IN TEST 1	SCORES IN TEST 2
A	5	12
B	3	15
C	2	11
D	8	10
E	6	18

SUBJECT	SCORES IN TEST 1 (X)	SCORES IN TEST 2 (Y)	XY	X^2	Y^2
A	5	12			
B	3	15			
C	2	11			
D	8	10			
E	6	18			
N=	$\Sigma X=$	$\Sigma Y=$	$\Sigma XY=$	$\Sigma X^2=$	$\Sigma Y^2=$

SUBJECT	SCORES IN TEST 1 (X)	SCORES IN TEST 2 (Y)	XY	X^2	Y^2
A	5	12	60	25	144
B	3	15	45	9	225
C	2	11	22	4	121
D	8	10	80	64	100
E	6	18	108	36	324
N=5	$\Sigma X=24$	$\Sigma Y=66$	$\Sigma XY=315$	$\Sigma X^2=138$	$\Sigma Y^2=914$

$$r = \frac{}{\sqrt{(5\times138-24^2)}\sqrt{(5\times914-66^2)}}$$

$$r = \frac{1575-1584}{\sqrt{(690-576)(4570-4356)}}$$

$$r = \frac{-75}{\sqrt{114\times214}}$$

$$r = \frac{-75}{\sqrt{24396}}$$

$$r = \frac{-75}{156.2}$$

$$r = -0.480$$

r= -0.480

ie, product moment correlation coefficient= -0.48

HOW TO EVALUATE A CORRELATION

- The values of ‘r’ always fall between -1 and +1 and the value does not change if all values of either variable are converted to a different scale.
- For eg. If the weights of the students were given in pounds instead of kilograms , the value of ‘r’ would not change.

INTERPRETATION OF CORRELATION COEFFICIENT

CORRELATION VALUE	INTERPRETATION
≤0.50	Very low
0.51 to 0.79	Low
0.80 to 0.89	Moderate
≥0.90	High (Good)

Spearman’s
Rank Correlation Coefficient

- The usual way of writing Spearman Rank Coefficient is:

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

Where:

d: differences between the ranks of the two variables

n: number of samples

Example:

Calculate Spearman's Rank Correlation

	Marks									
English	56	75	45	71	62	64	58	80	76	61
Maths	66	70	40	60	65	56	59	77	67	63

45	40	10	10	0	0
71	60	4	7	3	9
62	65	6	5	1	1
64	56	5	9	4	16
58	59	8	8	0	0
80	77	1	1	0	0
76	67	2	3	1	1
61	63	7	6	1	1

$$\sum d_i^2 = 25 + 1 + 9 + 1 + 16 + 1 + 1 = 54$$

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

$$\rho = 1 - \frac{6 \times 54}{10(10^2 - 1)}$$

$$\rho = 1 - \frac{324}{990}$$

$$\rho = 1 - 0.33$$

$$\rho = 0.67$$

as n = 10. Hence, we have a ρ (or r_s) of 0.67. This indicates a strong positive relationship between the ranks individuals obtained in the maths and English exam. That is, the higher you ranked in maths, the higher you ranked in English also, and vice versa.

Example 4. Compute coefficient of correlation by Karl Pearson Method for the following data

X :	1800	1900	2000	2100	2200	2300	2400	2500	2600
f :	5	5	6	9	7	8	6	8	9

Solution

Let the A.M.s A_x and A_y be 2200 and 6 for X and Y series respectively

X	Y	dx	(i=100) dx	dy	dx^2	dy^2	dxdy
1800	5	–400	–4	–1	16	1	4
1900	5	–300	–3	–1	9	1	3
2000	6	–200	–2	0	4	0	0
2100	9	–100	–1	3	1	9	–3
2200	7	0	0	1	0	1	0
2300	8	100	1	2	1	4	2
2400	6	200	2	0	4	0	0
2500	8	300	3	2	9	4	6
2600	9	400	4	3	16	9	12
N = 9			$\Sigma dx = 0$	$\Sigma dy = 9$	$\Sigma dx^2 = 60$	$\Sigma dy^2 = 29$	$\Sigma dx\, dy = 24$

$$r = \frac{(9)(24) - (0)\,(9)}{\sqrt{(9)\,(60) - (0)^2}\,\sqrt{(9)\,(29) - (9)^2}} = \frac{216}{\sqrt{97200}} = .69$$

(**Note : We can also proceed dividing X by 100)**

Age (X)	Weight (Y)	X^2	Y^2	XY
1	3	1	9	3
2	4	4	16	8
3	6	9	36	18
4	7	16	49	28
5	12	29	144	60
15	32	55	254	117

As $r = \dfrac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$

$\therefore \; r = \dfrac{5 \times 117 - 15 \times 32}{\sqrt{5 \times 55 - (15)^2}\sqrt{5 \times 254 - (32)^2}}$

$$= \frac{585 - 480}{\sqrt{275 - 225}\sqrt{1270 - 1024}} = \frac{105}{\sqrt{50 \times 246}} = \frac{105}{\sqrt{12300}} = \frac{105}{110.90} = 0.9467 \text{ Ans}$$

Type II : It is direct formula to find r. This formula can effectively be u where $\overline{X}$ and $\overline{Y}$ is not in fractions. The formula is

$r = \dfrac{\Sigma\, xy}{\sqrt{\Sigma\, x^2 . \Sigma\, y^2}}$; where dx is the deviation of X variable from its $\overline{X}$

y is the deviation of Y variable from its $\overline{Y}$. ; xy is the product of the two ab dx^2 is the square of x ; y^2 is the square of dy.

Example 2. Calculate coefficient of correlation between death and bi rate for the following data.

Birth Rate	24	26	32	33	35	30
Death Rate	15	20	22	24	27	24

Solution

Birth Rate X	Death Rate Y	$(X - \overline{X}) = x$	$(Y - \overline{Y}) = y$	$(X - \overline{X})^2 = x^2$	$(Y - \overline{Y})^2 = y^2$	$(X - \overline{X})(Y - \overline{Y}) = x$

Question:

Marks obtained by 5 students in algebra and trigonometry as given below: Calculate the Pearson correlation coefficient.

Algebra	15	16	12	10	8
Trigonometry	18	11	10	20	17

Solution:

Construct the following table:

x	y	x^2	y^2	xy
15	18	225	324	270
16	11	256	121	176
12	10	144	100	120
10	20	100	400	200
8	17	64	289	136
$\sum x = 61$	$\sum y = 76$	$\sum x2 = 789$	$\sum y2 = 1234$	$\sum xy = 902$

REGRESSION ANALYSIS

Birinder Singh, Assistant Professor, PCTE

Regression

- Regression Analysis measures the nature and extent of the relationship between two or more variables, thus enables us to make predictions.
- Regression is the measure of the average relationship between two or more variables.

Utility of Regression

- Degree & Nature of relationship
- Estimation of relationship
- Prediction
- Useful in Economic & Business Research

Difference Between Correlation & Regression

- Degree & Nature of Relationship
  - Correlation is a measure of degree of relationship between X & Y
  - Regression studies the nature of relationship between the variables so that one may be able to predict the value of one variable on the basis of another.
- Cause & Effect Relationship
  - Correlation does not always assume cause and effect relationship between two variables.
  - Regression clearly expresses the cause and effect relationship between two variables. The independent variable is the cause and dependent variable is effect.

Birinder Singh, Assistant Professor, PCTE

Difference Between Correlation & Regression

- Prediction
  - Correlation doesn't help in making predictions
  - Regression enable us to make predictions using regression line
- Symmetric
  - Correlation coefficients are symmetrical i.e. $r_{xy} = r_{yx}$.
  - Regression coefficients are not symmetrical i.e. $b_{xy} \neq b_{yx}$.
- Origin & Scale
  - Correlation is independent of the change of origin and scale
  - Regression coefficient is independent of change of origin but not of scale

Birinder Singh, Assistant Professor, PCTE

Types of Regression Analysis

- Simple & Multiple Regression
- Linear & Non Linear Regression
- Partial & Total Regression

Simple Linear Regression

- Simple Linear Regression
  - Regression Lines
  - Regression Equations
  - Regression Coefficients

Regression Lines

- The regression line shows the average relationship between two variables. It is also called Line of Best Fit.
- If two variables X & Y are given, then there are two regression lines:
  - Regression Line of X on Y
  - Regression Line of Y on X
- Nature of Regression Lines
  - If $r = \pm 1$, then the two regression lines are coincident.
  - If $r = 0$, then the two regression lines intersect each other at 90°.
  - The nearer the regression lines are to each other, the greater will be the degree of correlation.
  - If regression lines rise from left to right upward, then correlation is positive.

Birinder Singh, Assistant Professor, PCTE

Regression Equations

- Regression Equations are the algebraic formulation of regression lines.
- There are two regression equations:
  - Regression Equation of Y on X
    - $Y = a + bX$
    - $Y - \bar{Y} = byx\,(X - \bar{X})$
    - $Y - \bar{Y} = r.\,\frac{\sigma_y}{\sigma_x}\,(X - \bar{X})$
  - Regression Equation of X on Y
    - $X = a + bY$
    - $X - \bar{X} = bxy\,(Y - \bar{Y})$
    - $X - \bar{X} = r.\,\frac{\sigma_x}{\sigma_y}\,(Y - \bar{Y})$

Birinder Singh, Assistant Professor, PCTE

Regression Coefficients

- Regression coefficient measures the average change in the value of one variable for a unit change in the value of another variable.
- These represent the slope of regression line
- There are two regression coefficients:
  - Regression coefficient of Y on X: $b_{yx} = r.\ \frac{\sigma_y}{\sigma_x}$
  - Regression coefficient of X on Y: $b_{xy} = r.\ \frac{\sigma_x}{\sigma_y}$

Properties of Regression Coefficients

- Coefficient of correlation is the geometric mean of the regression coefficients. i.e. r = $\sqrt{b_{xy}\,.\,byx}$
- Both the regression coefficients must have the same algebraic sign.
- Coefficient of correlation must have the same sign as that of the regression coefficients.
- Both the regression coefficients cannot be greater than unity.
- Arithmetic mean of two regression coefficients is equal to or greater than the correlation coefficient. i.e. $\frac{bxy+byx}{2} \geq r$
- Regression coefficient is independent of change of origin but not of scale

Birinder Singh, Assistant Professor, PCTE

Obtaining Regression Equations

Birinder Singh, Assistant Professor, PCTE

Regression Equations in Individual Series using Normal Equations

- This method is also called as Least Square Method.
- Under this method, regression equations can be calculated by solving two normal equations:
  - For regression equation Y on X: Y = a + bX
    - $\Sigma Y = Na + b\Sigma X$
    - $\Sigma XY = a\Sigma X + b\Sigma X^2$
  - Another Method
    - $b_{yx} = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma X^2 - (\Sigma X)^2}$ & $a = \bar{Y} - b\bar{X}$
- Here a is the Y – intercept, indicates the minimum value of Y for X = 0
- & b is the slope of the line, indicates the absolute increase in Y for a unit increase in X.

Birinder Singh, Assistant Professor, PCTE

Calculate the regression coefficient and obtain the lines of regression for the following data

X	1	2	3	4	5	6	7
Y	9	8	10	12	11	13	14

X	Y	X^2	Y^2	X^Y
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
$\sum X = 28$	$\sum Y = 77$	$\sum X^2 = 140$	$\sum Y^2 = 875$	$\sum XY = 334$

Table 9.7

$\overline{X} = \frac{\Sigma X}{N} = \frac{28}{7} = 4$,

$\overline{Y} = \frac{\Sigma Y}{N} = \frac{77}{7} = 11$

$$
\begin{aligned}
b_{xy} &= \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\ \Sigma Y^2 - (\Sigma Y)^2} \\
&= \frac{7(334) - (28)(77)}{7(875) - (77)^2} \\
&= \frac{2338 - 2156}{6125 - 5929} \\
&= \frac{182}{196} \\
b_{xy} &= 0.929
\end{aligned}
$$

$$X-\overline{X} = b_{xy}(Y-\overline{Y})$$

$$X-4 = 0.929(Y-11)$$

$$X-4 = 0.929Y-10.219$$

$\therefore$ The regression equation X on Y is $X = 0.929Y - 6.219$

Regression coefficient *of* Y on X

$$b_{yx} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\,\Sigma X^2 - (\Sigma X)^2}$$

$$= \frac{7(334) - (28)(77)}{7(140) - (28)^2}$$

$$= \frac{2338 - 2156}{980 - 784}$$

$$= \frac{182}{196}$$

$$\therefore \quad b_{yx} = 0.929$$

Regression equation of Y on X

$$Y - \overline{Y} = b_{yx}(X - \overline{X})$$

$$Y - 11 = 0.929\,(X - 4)$$

$Y = 0.929X – 3.716 + 11$

$= 0.929X + 7.284$

The regression equation of Y on X is $Y = 0.929X + 7.284$

MEANING OF INTERCEPT AND SLOPE:

In the equation of a straight line (when the equation is written as "$y = a + bx$"), the slope is the number “a" that is multiplied on the *x*, and "*b*" is the *y*-intercept (that is, the point where the line crosses the vertical *y*-axis). This useful form of the line equation is sensibly named the "slope-intercept form".

* make prediction

Regression Notations

$X \Rightarrow$ independent Variable
$Y \Rightarrow$ Dependent Variable
$\overline{X} \Rightarrow$ mean of x
$\overline{Y} \Rightarrow$ mean of y
$x = X - \overline{X}$
$y = Y - \overline{Y}$

formulas

$B_1 = \Sigma xy / \Sigma x^2$
$B_0 = \overline{Y} - B_1 \cdot \overline{X}$
$Y' = B_0 + B_1 X$
$r_{xy} = \Sigma xy / \sqrt{\Sigma x^2 \cdot \Sigma y^2}$
$R^2 = r_{xy}^2$

Regression output

B_0 = Y intercept
B_1 = Slope
Y' = Predicted Y
r_{xy} = Correlation between x and y
R^2 = Coefficient of determination

Maths scores (X)	95	85	80	70	60
Statistics grade (Y)	85	95	70	65	70

olution :

X	Y	Deviation $(X-\bar{X})$ x	$(Y-\bar{Y})$ y	x^2	y^2	xy
95	85	17	8	289	64	13
85	95	7	18	49	324	12
80	70	2	−7	4	49	−1
70	65	−8	−12	64	144	9
60	70	−18	−7	324	49	12
: 390	ΣY : 385			$\Sigma x^2 =$ 730	$\Sigma y^2 =$ 630	Σ
78	$\bar{Y} = 77$					

$= 77 - 0.64 \times 78$ | $= \frac{470}{730}$

$= 26.77$ | $= 0.64$

egression equation is

$$\boxed{= 26.77 + 0.64X}$$

ficient of determination (R^2)

Coefficient of determination (R^2)

es the proportion of Variance in Y that is

table from X.

* If R^2 is close to zero, don't use regression
* If R^2 is significantly Greater than zero, use regression

$$= 470/\sqrt{730 \times 630}$$

$$= 470/\sqrt{459900}$$

$$= 470/678.159273$$

$$r_{xy} = 0.69$$

Coefficient of determination (R^2) is

$$R^2 = r_{xy}^2$$

$$R^2 = (0.69)^2 = 0.48$$

that means 48% of the variation in dependent variable can be explained by independent variable.

The regression equation is

$$Y' = 26.77 + 0.64 \cdot X$$

$$Y' = 26.77 + 0.64 \times 75$$

$$Y' = 26.77 + 48$$

$$= 74.77$$

X	1	2	3	4	5	6	7	8
Y	80	90	92	83	94	99	92	104

Solution: Normal Equation is

$$\Sigma Y = na + b\Sigma X \text{ ----- } (1)$$
$$\Sigma XY = a\Sigma X + b\Sigma X^2 \text{ ----- } (2)$$

X	Y	X^2	XY
1	80	1	80
2	90	4	180
3	92	9	276
4	83	16	332
5	94	25	470
6	99	36	594
7	92	49	644
8	104	64	832
$N = 8$ $\Sigma X = 36$	$\Sigma Y = 734$	$\Sigma X^2 = 204$	$\Sigma XY = 3408$

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

$b = \frac{\Sigma xy}{\Sigma x^2}$

Regression equation is

$Y = a + bx$

Substitute the values in equation

$$734 = 8a + 36b \text{ ----- } (1)$$
$$3408 = 36a + 204b \text{ ----- } (2)$$

in order to solve the equation multiply the equation (1) by 9 and equation (2) by 2.

$$-210 = -84b$$

$$\therefore \quad b = \frac{210}{84} = 2.5$$

Substitute 'b' value in equation ① to find 'a' value.

$$734 = 8a + 36(2.5)$$
$$734 = 8a + 90$$
$$8a = 734 - 90$$
$$8a = 644$$
$$a = \frac{644}{8} = 80.5$$

The regression equation is $Y = a + bX$

so $\boxed{Y = 80.5 + 2.5X}$

Profit (Rs. lakhs)	10	13	15	20	22	2

tion: Calculations for the straight line

Year (X)	Profit (Y)	Deviation $x = X - \text{Midvalue}$ $(X - 2017.5)$	x^2	xy
2015	10	-2.5	6.25	-25
2016	13	-1.5	2.25	-19.5
2017	15	-0.5	0.25	-7.5
2018	20	0.5	0.25	10
2019	22	1.5	2.25	33
2020	28	2.5	6.25	70
	$\Sigma Y = 108$		$x^2 : 17.5$	Σxy 61

Straight line equation

$$= \frac{\Sigma Y}{n} = \frac{108}{6} : 18$$

$$= \frac{\Sigma xy}{\Sigma x^2} = \frac{61}{17.5} = 3.49$$

production (in tonnes)	80	90	[illegible]	[illegible]

a) Fit a stright line trend by the method of least square

b, Estimate the production for the year 2021

c, plot the actual and trend value on a graph.

<u>Solution:</u> a) stright line trend $Y_c = a + bx$ ...... ①

Normal equations are $\Sigma y = Na + b\Sigma x$ .... ②

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

$X = x - \text{origin}$

$N \Rightarrow$ Number of years

When $\Sigma x = 0$ from ② we get $\Sigma y = Na \Rightarrow \boxed{a = \frac{\Sigma y}{N}}$

from ③ we get $\Sigma xy = b\Sigma x^2 \Rightarrow \boxed{b = \frac{\Sigma xy}{\Sigma x^2}}$

Year x	production Y	$X = x - 2017$	x^2	xy	Trend Value $Y_c = 90 + 2x$
2014	80	-3	9	-240	$90 + 2(-3) = 84$
2015	90	-2	4	-180	$90 + 2(-2) = 86$
2016	92	-1	1	-92	$90 + 2(-1) = 88$
2017	83	0	0	0	$90 + 2(0) = 90$
2018	94	1	1	94	$90 + 2(1) = 92$
[illegible]	[illegible]	2	4	198	$90 + 2(2) = 94$
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	$90 + 2(3) = 96$

The estimated production for 2021 is 98 tonnes

c) Graph: