1

NATURE OF DATA AND CENTRAL TENDENCY

Unit Structure :

1.1 OBJECTIVE

- To understand the need & importance of the statistical techniques.
- To understand classification & tabulation of data.
- To study statistical techniques.

1.2 INTRODUCTION

We are in the 21st century. In the era of Globalization - Scientific inventions & Widespread use of technology lot of information is available in different forms. This information can be in the qualitative form e.g. Information about the people, area, subject etc. or it can be in the quantitative form i.e. Numerical data.

Quantitative information is more precise than the qualitative information. Hence it is widely use in research & development processes. Statistical techniques help us to store, classify and analyse data so that we can compare it & draw inferences or to make use of this data for our projects. Hence it is necessary & very interesting to study various statistical techniques.

1.3 NEED AND IMPORTANCE OF THE STATISTICAL TECHNIQUES

Last century witnessed large scale development in the field of science and technology. Modern machines like calculator and computers have become common and are part of our everyday life. But these are just machines. They can do analysis of data but they require suitable programme or software for all operations. It is necessary to understand various statistical techniques so that we can select suitable programme for the analysis of our data. Statistical techniques help us to compress large amount of data and help us in the analysis of the data so that we can interpret results and can take proper decision instantly.

Statistical operations form basis of the entire field of science & commerce. These are very essential for our development.

1.4 NATURE OF THE STATISTICAL DATA

Information or data can be in different forms.

a) Qualitative data - Descriptive data e.g. Biography of a person, decription of a project etc.

b) Quantitative data - Numerical data e.g. Amount of rainfall in different regions; Agricultural production, population etc. various statistical techniques are designed for the analysis of the quantitative data.

Data can also be classified as

a) Spatial data

b) Temporal data

a) Spatial data - Data related to space, area, region, village, Town, Taluka, District, State, Nation etc.

b) Temporal data - Data related to time e.g. growth of population from 1901 to 2011, Production of wheat from 1961 to 2011

Statistical data can be obtained from various sources. On the basis of collection of data it can be classified as

a) Primary data

b) Secondary data

a) Primary data - As the name indicates, primary data are collected for the first time and are thus original in character.

Primary data can be collected in different forms.

1) Direct personal investigation

2) Indirect oral investigation

3) By schedules and questionnaires

4) By local reports

b) Secondary data - Secondary data are those which have already been collected and analysed by someone else. Secondary data may be either published or unpublished. The sources of published data (secondary data) are as follows -

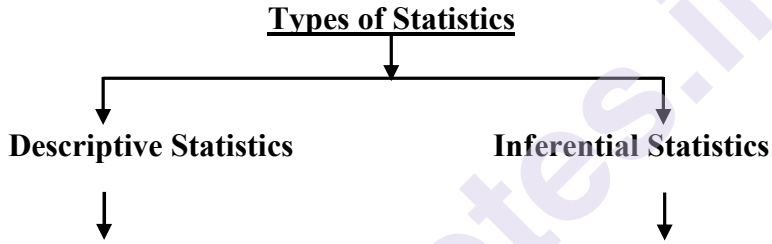
1) Central, State & local government publications.

2) United nation's reports, World Bank report etc.

3) Reports of companies, NGO & other organization.

4) Journals related to different subjects.

5) Research reports



Descriptive Statistics	Inferential Statistics
Designed to describe & explain the geographical situation by obtaining quantitative information	Helps to make probabilistic statements
	Regarding hypothesis,
e.g. an evaporation index price index	About relationships between variables
	About population from which samples are drawn
Mean	Correlation
Median	Regression
Mode	Hypothesis testing
Range	
Quartile Deviation	
Mean Deviation	
Standard Deviation	

1.5 FREQUENCY DISTRIBUTION

Frequency distribution helps us to classify large amount of data into 5-10 classes, so that it becomes more compact and can be used for further analysis. Let us understand following examples.

Q.1 Prepare frequency distribution table for the following data.

Amount of rainfall at 50 places of 'X' district. (Rainfall in cm)

25	210	420	170	370	290	310	185	280	240
125	310	30	470	220	110	40	410	75	127
490	90	320	140	22	175	130	130	190	60
410	138	410	95	360	380	80	170	45	260
330	470	140	280	130	160	420	230	270	140

Let us find out smallest and largest number in the given data.

Smallest number = 25

Largest number = 490

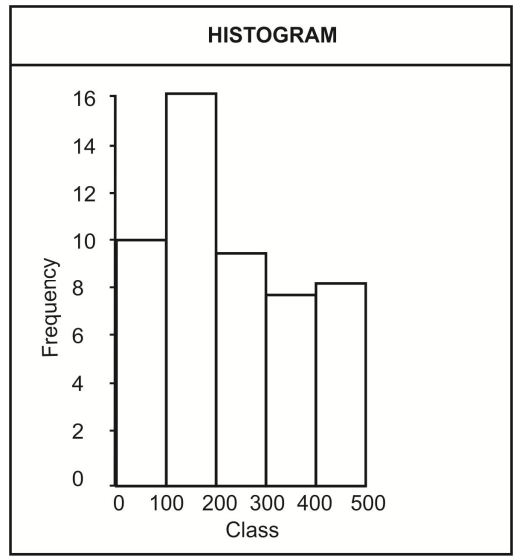
Now we can take fire classes to cover this data as a 0-100, 100-200, 200-300, 300-400 and 400-500.

Let us prepare frequency distribution table.

Class	Tally Score	Frequency
0 - 100	𝍸 𝍸 II	10
100 - 200	𝍸 𝍸 𝍸 I	16
200 - 300	𝍸 IIII	09
300 - 400	𝍸 II	07
400 - 500	𝍸 III	08
	Total	30

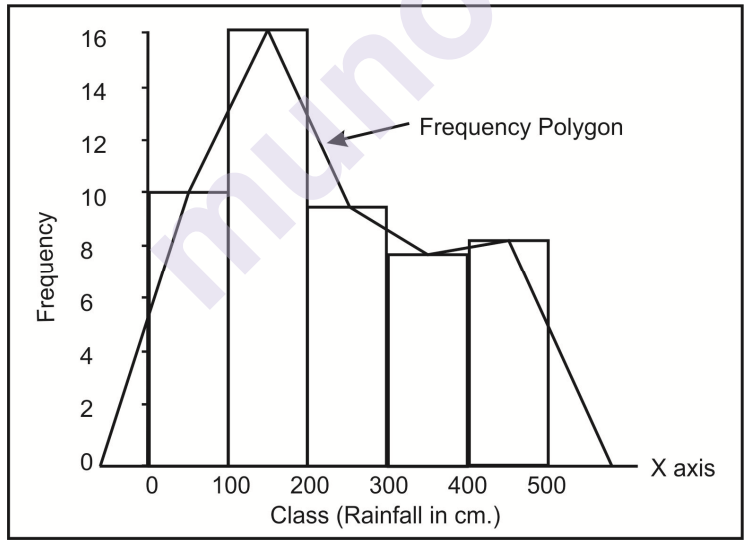
Tally score are the slant lines used for classification of data. A bunch of 5 is formed as it becomes easy to count numbers in the multiples of 5.

We can prepare Histogram for this frequency distribution table.



Histogram is a graphical representation of the frequency distribution table. Just by looking at the Histogram we get complete idea of the distribution of data in different classes; because bars in the histogram are drawn proportional to the number of variables in a particular class.

We can also draw frequency polygon for the same data. In order to prepare frequency polygon join mid points of the top portions of bars drawn in Histogram.



Join extreme ends of the frequency polygon to the 'X' axis as shown in the diagram.

Frequency Curve can be drawn for the same data. Procedure for drawing frequency curve is same as drawing of frequency polygon. But instead of joining points by straight line, these points are joined by curved line.

The statistical data obtained in primary survey, which is not grouped or classified is termed as the ungrouped data. e.g. Agricultural yield in 20 farms (in thousand tonnes).

500	290	290	180	470
100	800	750	980	850
420	960	700	775	500
150	300	400	375	400

When this data is grouped into different classes in frequency distribution table it is termed as group data. grouped data can be related to Discrete or Continuous series.

Discrete Series - In this type of data the items are capable of exact measurement. (No function) e.g. Number of persons, Number of Countries, Number of rivers etc.

Continuous Series - In this type of data the items are capable of division and can be measured in fractions of any size. E.g. Amount of rainfall, temperature, weight, height of the person etc.

Discrete series	
No. of Children per couple	**No. of couples**
2	40
3	10
4	05

Continuous series	
Height in cm.	**No. of persons**
140 - 150	20
150 - 160	15
160 - 170	22

Cumulative Frequency -

Consider following two examples.

1) You wish to distribute milk products for children whose age is less than 5 years.

2) You are preparing / updating list of persons whose age is more than 18 years, for the purpose of election.

We require less than or more than type of data frequently, for which cumulative frequency distribution table is prepared.

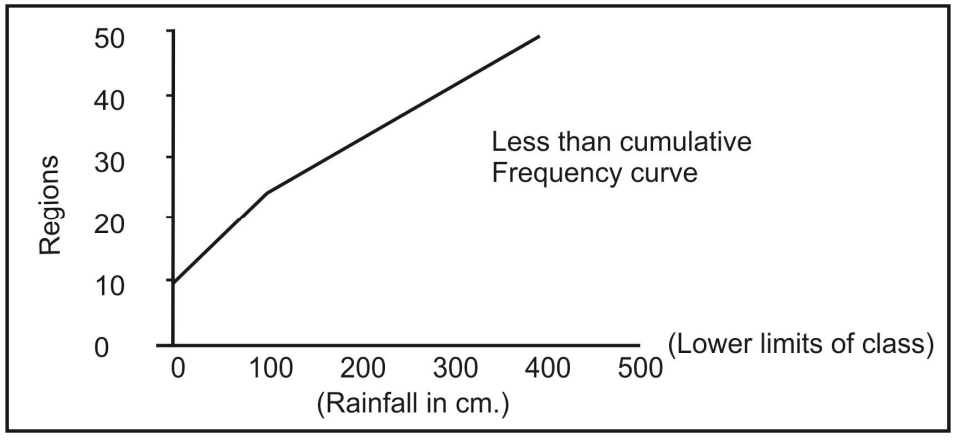
Q.1Prepare cumulative frequency distribution table for the following data.

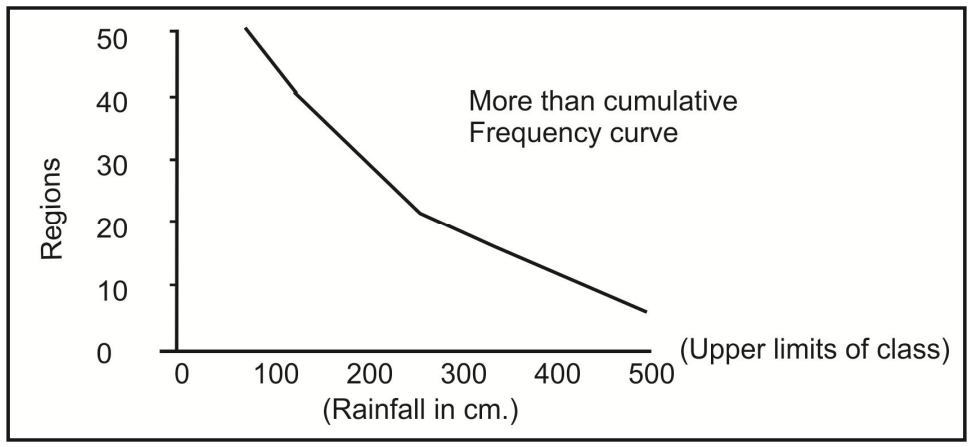
Rainfall (in cm)	No. of regions
0 - 100	10
100 - 200	16
200 - 300	09
300 - 400	07
400 - 500	08
Total	50

To prepare cumulative frequency table, values are progressively added or subtracted as shown in the following table.

Rainfall (in cm)	**Regions**	**Cumulative frequency less than**	**Cumulative frequency more than**
0 - 100	10	10	50
100 - 200	16	26	40
200 - 300	09	35	24
300 - 400	07	42	15
400 - 500	08	50	08
Total	50		

Cumulative frequency curve are also termed as ogive.





1.6 MEASURES OF CENTRAL TENDENCY

In order to compare one set of data (1000 values) with another set of data (1000 values) we require average or central number which represents the entire data.

"Average is an attempt to find one single figure to describe whole of figures."

- Clark

Average is normally value near to the middle value in the given data, so it is also called as the Central value. Some values in the data are less than the average value and same values are more than the average value.

e.g. Find out average of following numbers.

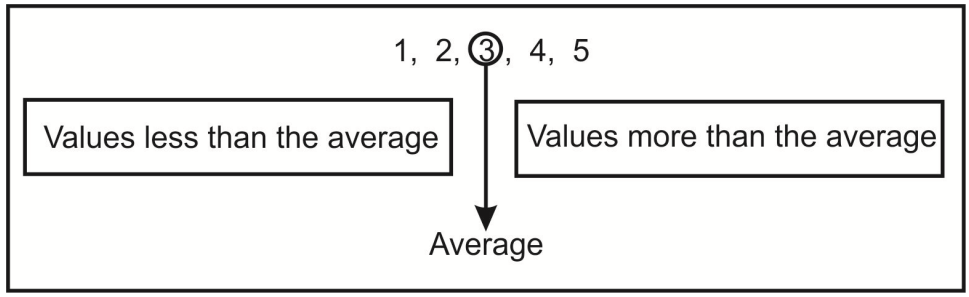
1, 2, 3, 4, 5

Let us add these numbers 1 + 2 + 3 + 4 + 5 = 15.

Total number of values = 5

$$\text{Average} = \frac{15}{5} = 3$$

Average of the given numbers is 3.



Average value, being central value is also termed as Measures of Central Tendency. Different types of measures of central tendency are

1) Mean

2) Median

3) Mode

1) Mean - It is also termed as 'average' or arithmetic mean. It is obtained by adding together all items and by dividing this total by the number of items.

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

Where $\bar{X} =$ Arithmetic average or mean

$\sum X =$ Total of all values in the data.

$n =$ Total number of values.

Find out mean of the following data 0, 10, 20

$\sum X = 0+10+20 = 30$ Total of all values in the data.

$n = 3$ Total number of values.

$$\bar{X} = \frac{\sum X}{n} = \frac{30}{3} = 10$$

$\bar{X}$ (Mean) = 10

Ungrouped data

Mean - 1, 2, 4, 6, 8, 10, 10

$$\sum X = 1+2+4+6+8+10+10 = 41$$

$$n = 7$$

$$\therefore \bar{X} = \frac{\sum X}{N}$$

$$= \frac{41}{7}$$

$$= 5.85$$

The average value of the given data according to the Mean is 5.85

Grouped data - discrete -

Mean -

Rainfall (in cm)	Regions
100	2
200	5
300	3

Rainfall (in cm) X	Regions f	fx
100	2	200
200	5	1000
300	3	900

$$\sum X = 2100$$
$$\sum f = 10$$
$$\bar{X} = \frac{\sum fX}{\sum f}$$
$$= \frac{2100}{10}$$
$$= 210\ cms$$

The average amount of rainfall per region according to mean is 210 cms.

Grouped data - continuous -

Mean -

Rainfall (in cm) X	Regions f	Mid point x	fx
0 - 100	4	50	200
100 - 200	10	150	1500
200 - 300	6	250	1500
	$\sum f = 20$		$\sum fX = 3200$

$$\sum fX = 3200$$
$$\sum f = 20$$

Mean $\bar{X} = \frac{\sum fX}{\sum f}$

$$= \frac{3200}{20}$$

$$\bar{X} = 160$$

The average agricultural yield per region according to mean is 160 thousand tonnes.

Merits of Arithmetic Mean -

1) It is central value, it is the centre of gravity balancing values on either side of it.

2) It is affected by the value of every item in the series.

3) It is easy to understand and calculate.

4) It is calculated by a rigid formula.

5) It is useful for further statistical analysis.

Limitations of Arithmetic Mean -

1) Extreme values of the data affect Mean -

e.g.

a) Average of 1, 2, 3 is 1 + 2 + 3 =6

$6 \div 3 = 2$

b) but average of 1, 2, 1002 is 1 + 2 + 1002 = 1005 ÷ 3 = 335

In the second example the extreme value affects Mean.

2) It can not be calculated for incomplete data. i.e. all values are required for calculation of mean.

1.7 MEASURES OF CENTRAL TENDENCY - MEDIAN

'Median' means middle value in a distribution (of data). Median splits the observation into two parts. (lower & higher values) median is also termed as a Positional average.

The term 'Position' means the place of value in a given data.

e.g.

Q.1 Find out median for the following data 1, 2, 4, 6, 8, 10, 10

Median -

1	Median value $= \frac{n+1}{2}$
2	
4	$= \frac{7+1}{2}$
6 ← Median	$= \frac{8}{2}$
8	$= 4^{th}$ *value*
10	
10	

∴ The average value according to median is 6

Q.2 Find out median for the following data 3, 4, 2, 1, 5, 7

Let us rearrange numbers in proper order - 1, 2, 3, 4, 5, 7

As the number is even (six) the mid point will be between 3rd & 4th value.

Hence Median $= \frac{3+4}{2} = \frac{7}{2} = 3.5$

Grouped data - discrete

Median -

Rainfall (in cms) - 100, 200, 300

Regions - 2, 5, 3

Rainfall (in cms)	**Regions f**	**Cumulative frequency less than**
100	2	2
200	5	7
300	3	10
	$\sum f = 10$	

Median $= \frac{n+1}{2} = \frac{11}{2} = 5.5^{th}$ *value*

As this number - (5.5^{th} value) is more than 2 (cumulative frequency) but less than 7, hence the median is located in the class whose cumulative frequency is 7. the rainfall amount of this class is 200 cm.

$\therefore$ The average amount of rainfall per region according to median is 200 cms.

Median -

Agricultural yield in thousand tons	Regions	Cumulative frequency less than
0 - 100	4	4
100 - 200	10	14
200 - 300	6	20
	$\sum f = 20$	

m = middle value $= \frac{n}{2} = \frac{20}{2} = 10$

$\therefore$ As number 10 is more than 4 but less than 14, median will be found in the class 100 - 200 - (median class)

$$Median = l_1 + \frac{l_2 - l_1}{f_1}(m - c)$$

$$l_1 = 100, l_2 = 200, f_1 = 10$$

$$= 4 \quad m = 10$$

$$= 100 + \frac{200 - 100}{10}(10 - 4)$$

$$= 100 + \frac{100}{10}(6)$$

$$= 100 + 10(6)$$

$$= 100 + 60$$

$$= 160$$

The average agricultural yield per region according to median is 160 thousand tonnes.

Median $= l_1 + \frac{l_2 - l_1}{f_1}(m - c)$

l_1 = lower limit of the class

l_2 = upper limit of the class

f_1 = Frequency of the median class

m = middle value

c = Cumulative frequency of the preceding class

Merits of Median -

1) Extreme values do not affect the median.
2) It is useful for open and data as only the position and not the values of items must be known.
3) It is easier to compute than the mean.
4) It can be used for qualitative data i.e. where ranks are given.
5) The value of median can be found out graphically.

Limitations of the median -

1) It is necessary to arrange data in proper order for calculation of median.
2) As it is a positional average, its value is not determined by each and every observation.
3) It is not much used for further statistical analysis.
4) The value of median is affected by sampling fluctuation than the value of the arithmetic mean.

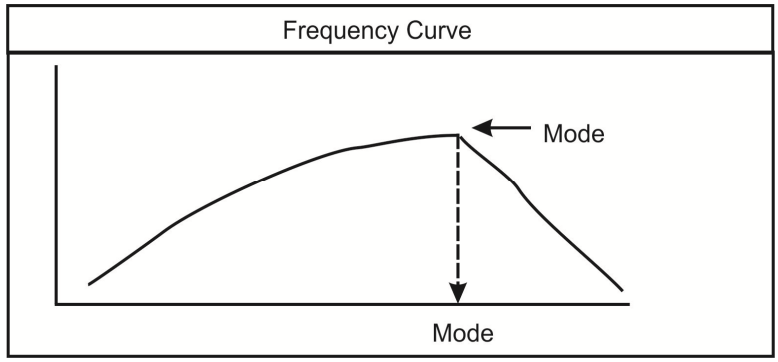
1.8 MEASURES OF CENTRAL TENDENCY - MODE

The mode or the modal value is that value in a series of observations which occurs with the greatest frequency.

Mode - 1, 2, 4, 6, 8, 10, 10

Mode =1 - as 10 is the most common number in the given data.

The Mode, this word is derived from the French word 'La Mode' means fashion. Where most of the people in the society use similar type of dress. Mode is at the highest peak of the frequency curve.



Mode - Discrete series

Rainfall (in cms)	Regions
100	2
200	5
300	3

← Maximum value - ∴ Modal class
Mode = 20 cms.

The average amount of rainfall received by each region according to mode is 200 cms.

Mode - Continuous Series

Agricultural production in thousand tons	Regions
0 - 100	4
100 - 200	10
200 - 300	6

← Maximum value - Modal class

$$\text{Mode} = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2}(l_2 - l_1)$$

$$= l_1 = 100, f_1 = 10, f_0 = 4, f_2 = 6$$

$$l_2 = 200$$

$$= 100 + \frac{10-4}{2(10)-4-6}(200-100)$$

$$= 100 + \frac{6}{20-10}(100)$$

$$= 100 + \frac{6}{10}(100)$$

$$= 100 + \frac{600}{10}$$

$$= 100 + 60$$

$$= 160$$

The average amount of agricultural production per region according to mode is 160 thousand tonnes.

The mode can also be obtained by using following formula based upon the relationship between mean, median & mode.

Mode = 3 Median - 2 Mean

Merits of mode -

1) It is easy to find mode in a given data.
2) It is not affected by the extreme values.
3) It can be used for the qualitative data.

   e.g. Most preferred colour of dress by girls.

1) Alka - Pink
2) Swati - Blue
3) Narendra - Red
4) Sonali - Pink
5) Soni - Pink
6) Kshama - Red

From the data it is clear that 3 out of 6 girls prefer Pink and so mode is pink colour. We can say that girls prefer pink dresses.

4) The value of mode can also be obtained from frequency curve (without doing calculation)

Limitations of Mode -

1) The value of mode cannot always be determined. - e.g. in the bimodal (Two modes) or multimodal frequencies.
2) It is not multi used in further statistical analysis.
3) It does not include all items of the data.
4) It is not much used.

❖❖❖❖

2

DISPERSION AND DEVIATION

Unit Structure :

2.1 OBJECTIVE

- To understand the concept of dispersion.
- To study various types of dispersion.
- To understand techniques of moving average & area mean.

2.2 INTRODUCTION

In measures of central tendency we get one representative number (Mean, Median or Mode) for the given data. hence we can compare two sets of data easily.

e.g. Set A = Mean = 80% marks

Set B = Mean = 40% marks

In this case distinction between two sets of data is very clear. So we can compare them & take decision.

Now consider following example.

Set X = 0, 10, 20

Set Y = 10, 10, 10

$$Set\ X = 0 + 10 + 20 = 30$$

$$\bar{X} = \frac{\sum X}{n} = \frac{30}{3} = 10$$

$$Set\ Y = 10 + 10 + 10 = 30$$

$$\bar{X} = \frac{\sum X}{n} = \frac{30}{3} = 10$$

Average value or mean for set X & set Y is same. Now it is difficult to compare these two sets.

Hence we require Measures of dispersion.

> It is degree to which numerical data tend to spread about an average value is called the variation of dispersion of the data.

Range is one of the type of dispersion technique.

Range = Max. Value - Min. Value

Set X = 0, 10, 20 = Maximum Value = 20

Min. Value = 0

Range = 20 - 0 = 20

Set Y = 10, 10, 10 = Maximum Value = 10

Min. Value = 10

Range = 10 - 10 = 0

Range is zero for Set Y. Which means the average value for Set 'Y' is more reliable than the average value for Set 'X'

2.3 RANGE

Range is the difference between the value of the smallest item and the value of the largest item included in the distribution.

Range = Max. Value - Min. Value

For ungrouped data

Find out range for the following data.

5, 1, 7, 8, 15, 20, 9, 10, 12, 14, 21, 3, 6, 5, 5, 2, 4, 10

Maximum Value = 21

Minimum Value = 1

Range = 21 - 1 = 20

Grouped data - discrete -

Range -

Rainfall (in cm)	Regions
100	2
200	5
300	3

Maximum amount of rainfall = 300 cms.

Minimum amount of rainfall = 100 cms

Range = 200 cms.

∴ Range, is 200 cms of rainfall. i.e. the amount of variation in the rainfall according to range is 200 cms.

Grouped data - Continuous -

Range -

Agricultural Production in thousand tons	Regions
0 - 100	4
100 - 200	10
200 - 300	6

Maximum agricultural production = 300 thousand tons

Minimum agricultural production = 0 thousand tons

Range = 300 thousand tons

Merits of Range -

1) Simplest & easiest method of dispersion.
2) It requires minimum time to calculate range.

Limitations of Range -

1) Range is not based on each and every item of the distribution.
2) Range is most unreliable measure of dispersion.

e.g.

2.4 QUARTILE DEVIATION

Quarter means 25%. In quartile deviation data is divided into four parts (25% each). First quarter is at 25%, Second quarter is at 50%, third quarter is at 75%, Fourth quarter is at 100%.

Difference in the third (Q3) and first (Q1) quartiles is termed as inter quartile range.

Inter quartile range = Q3 - Q1

Inter quartile range is reduced to the form or semi-inter quartile range or quartile deviation; by dividing it by 2.

Quartile Deviation -

$$Q.D. = \frac{Q3 - Q1}{2}$$

Quartile deviation gives the average amount by which the two quartiles differ from the median (50%)

Quartile deviation -

Rainfall (in cms)	**Regions**	**Cumulative frequency less than**	
100	2	2	
200	5	7	Q.1
300	3	10	Q3
	$\sum f = 10$		

$$Q1 = \frac{1}{4} \times 10 = 2.5\ region \ \therefore Q1 = 200$$

$$Q3 = \frac{3}{4} \times 10 = 7.5\ region \quad Q3 = 300$$

$$Quartile\ deviation = \frac{Q3 - Q1}{2}$$
$$= \frac{300 - 200}{2}$$
$$= \frac{100}{2}$$
$$= 50\ cms$$

Quartile deviation -

Agricultural yield thousand tons	Regions	Cumulative frequency less than
0 - 100	4	4
100 - 200	10	14
200 - 300	6	20
	$\sum f = 20$	

$Q1 = \frac{1}{4} \times 20 = 5 = 100 - 200$ Agricultural production thousand tons

$Q3 = \frac{3}{4} \times 20 = 15 = 200 - 300$ Agricultural production thousand tons

$$Median = 1, + \frac{1_2 - 1_1}{f_1}(m - c)$$
$$Q1 = 100 + \frac{200 - 100}{10}(5 - 4)$$
$$= 100 + \frac{100}{10}(1)$$
$$= 100 + 10$$
$$= 110$$
$$Q3 = 200 + \frac{300 - 200}{10}(15 - 14)$$
$$= 100 + \frac{100}{6}(1)$$
$$= 200 + 16.6$$
$$= 216.6$$

$$\text{Quartile deviation} = \frac{Q3 - Q1}{2}$$

$$= \frac{216.6 - 110}{2}$$

$$= 53.3$$

The variation in the agricultural yield among different regions according to quartile deviation is 53.3 thousand tonnes.

Merits of Quartile Deviation

1) It is superior to range as a measure of dispersion.

2) It can be used for open end distributions.

3) Quartile deviation is not affected by the extreme values.

Limitations of Quartile Deviation

1) Quartile deviation ignores 50% items. i.e. the first 25% and last 25%.

2) It is not much used for further statistical analysis.

3) It's value is affected by sampling fluctuations.

2.5 MEAN DEVIATION

The Mean deviation is also known as the average deviation. It is the average difference between the items in a distribution and the median or mean of that series.

In mean deviation of each item in the series is found out from the median. All deviations are added together (ignoring + or - signs). This total is divided by the number of observations.

$$\text{Mean Deviation} = \frac{\sum d}{n}$$

$\sum d$ = Sum of all deviations

n = number of observations / items

Calculate mean deviation for the following series.

X	10	11	12	13	14
F	3	12	18	12	3

Answer -

X	f	d	fd	c.f
10	3	2	6	3
11	12	1	12	15
12	18	0	0	33
13	12	1	12	45
14	3	2	6	48
	$n = 48$		$\sum fd = 36$	

Mean deviation $= \dfrac{\sum fd}{n}$

Median = size of $\dfrac{n+1^{th}}{2}$ item

$= \dfrac{48+1}{n} = 24.5^{th}$ item

Size of 24.5^{th} item is 12

Hence Median = 12

M.D. $= \dfrac{36}{48} = 0.75$

Mean Deviation - Continuous series

Q.2 Calculate the median and mean deviation of the following data.

Size	Frequency
0 - 10	7
10 - 20	12
20 - 30	18
30 - 40	25

Size	Frequency
40 - 50	16
50 - 60	14
60 - 70	8

Answer -

Size	f	c.f.	Mid point m	d = m - 35.2	fd
0 - 10	7	7	5	30.2	211.4
10 - 20	12	19	15	20.2	242.4
20 - 30	18	37	25	10.2	183.6
30 - 40	25	62	35	0.2	5.0
40 - 50	16	78	45	9.8	156.8
50 - 60	14	92	55	19.8	272.2
60 - 70	8	100	65	29.8	238.4
	$\sum f = 100$				$\sum fd = 1314.8$

Median = size of $\frac{n^{th}}{2}$ item $= \frac{100}{2} = 50^{th}$ item

Median lies in the class 30 - 40

$$\text{Median} = l_1 + \frac{n/2 - c.f.}{f} \times i$$

$l_1 = 30, n/2 = 50, c.f. = 35, f = 25, i = 10(40-30)$

$$\text{Median} = 30 + \frac{50-37}{25} \times 10$$

$$= 30 + 5.2 = 35.2$$

$$\text{Mean deviation} = \frac{\sum fd}{n} = \frac{1314.8}{100} = 13.148$$

Merits of Mean Deviation -

1) It is relatively simple to understand & compute.
2) It is based on each and every item of the data.
3) It is less affected by the extreme items of the data.

Limitations of M.D. -

1) Algebraic signs (+, -) are ignored in M.D.
2) It may not give us accurate results.
3) It is not capable of further statistical analysis.

2.6 STANDARD DEVIATION

Standard deviation is the square root of the arithmetic average of the squares of the deviations measured from the mean.

To find the S.D. the following steps are taken.

1) Find the deviations from the mean.

2) Square those deviations.

3) Find the mean of the sum of these deviations squared.

4) Find the square root of this mean.

Standard deviation-Grouped data - discrete.

Q. Find out S.D. for the following data.

Yield (in 000' kg) (X)	No. of regions (f)
40	10
45	15
50	25
55	30
60	28
65	13
70	9

Yield (X)	Regions (f)	fx
40	10	400
45	15	675
50	25	1250
55	30	1650
60	28	1680
65	13	845
70	9	630
	$\sum f = 130$	$\sum fx = 7130$

$$\text{Mean} = \frac{\sum fx}{\sum f} = \frac{7130}{130}$$

$$\text{Mean} = \bar{X} = 55$$

Ungrouped data -

Standard deviation -

1, 2, 4, 6, 8, 10, 10

X	d	d^2
1	- 4.9	24.0
2	- 3.9	15.2
4	- 1.9	3.6
6	0.1	0.01
8	2.1	4.4
10	4.1	16.8
10	4.1	16.8
$\sum x = 41$		$\sum d^2 = 80.8$

$$\text{Mean} = \frac{\sum x}{n}$$

$$= \frac{41}{7}$$

$$= 5.85$$

$$\bar{X} = 5.9$$

$$d = X - \bar{X}$$

$$\text{Standard deviation} = \sqrt{\frac{\sum d^2}{n}}$$

$$= \sqrt{\frac{80.8}{7}}$$

$$= \sqrt{11.54}$$

$$= 3.39$$

$\therefore$ Standard deviation of the given data is 3.39.

Standard deviation - ungrouped data.

Find out S.D. for the following data.

Height (in inches) 60, 60, 61, 62, 63, 63,63, 64, 64, 70

Height (in inches) X	Deviations from mean $(63°)$ d	d^2	$\text{Mean} = \frac{\sum x}{n}$
60	- 3	9	$= \frac{630}{10}$
60	- 3	9	
61	- 2	4	
62	- 1	1	

63	0	0	$= 63''$
63	0	0	
63	0	0	$= \left(X - \bar{X}\right)$
64	+1	1	
64	+1	1	
70	+7	49	
$\sum x = 630$		$\sum d^2 = 74$	

Standard deviation or $\sigma = \sqrt{\frac{\sum d^2}{n}}$

$$= \sqrt{\frac{74}{10}}$$
$$= \sqrt{7.4}$$
$$= 2.72''$$

$\therefore$ Deviation in height according to Standard deviation is $2.72''$.

Group data -

Standard deviation - Continuous series.

Find out S.D. for the following data.

Rainfall (in cm)	**Regions**
0 - 100	2
100 - 200	5
200 - 300	4
300 - 400	2

Rainfall (in cm)	Mid point X	f	*fx*	$\text{Mean} = \frac{\sum fx}{\sum f}$ $= \frac{2550}{13}$ $= 196.15$ $d = \left(X - \bar{X}\right)$
0 - 100	50	2	100	
100 - 200	150	5	750	
200 - 300	250	4	1000	
300 - 400	350	2	700	
		$\sum f = 13$	2550	

Rainfall (in cm)	Mid point X	f	Deviation from mean 196.15 (d)	*fd*	fd^2
0 - 100	50	2	- 146.15	292.3	42719.6
100 - 200	150	5	- 46.15	230.75	10649.1
200 - 300	250	4	53.85	215.4	11599.3
300 - 400	350	2	153.85	302.7	47339.6
		$\sum f = 13$		$\sum fd = 1046.2$	$\sum fd^2 = 112307.6$

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f}}$$

$$= \sqrt{\frac{112307.6}{13}}$$

$$= \sqrt{8639}$$

$$= 92.95\, cms$$

$\therefore$ Rainfall variability according to S.D. is 92.95 cms.

Standard deviation -

Rainfall (in cm) X	Regions f	fx
100	2	200
200	5	1000
300	3	900
	$\sum f = 10$	$\sum fx = 2100$

Mean $= \frac{\sum fx}{\sum f}$

$= \frac{2100}{10}$

$\bar{X} = 210$

$d = X - \bar{X}$

Rainfall (in cm) X	**Regions f**	**Deviation from mean =210 d**	fd	fd^2
100	2	- 110	- 220	24,200
200	5	- 10	- 50	500
300	3	90	270	24,300
	$\sum f = 10$			$\sum fd^2 = 49,000$

S.D. $= \sqrt{\frac{\sum fd^2}{\sum f}}$

$= \sqrt{\frac{49000}{10}}$

$= \sqrt{4900}$

S.D. = 70 cms. of rainfall

Variation in the amount of rainfall according to standard deviation is 70 cms.

Standard deviation -

Agricultural Production in thousand tonnes	Mid point X	Regions f	*fx*
0 - 100	50	4	200
100 - 200	150	10	1500
200 - 300	250	6	1500
		$\sum f = 20$	$\sum fx = 3200$

$$\text{Mean} = \frac{\sum fx}{\sum f} = \frac{3200}{20} = 160$$

$\bar{X} = 160$ thousand tons. $d = \left(X - \bar{X}\right)$

Agricultural Production in thousand tons	Mid point X	Regions f	Deviation from mean 160 d	*fd*	fd^2
0 - 100	50	4	- 110	440	48,400
100 - 200	150	10	- 10	100	1000
200 - 300	250	6	90	540	48600
		$\sum f = 20$			$\sum fd^2 = 98,000$

$$\text{Standard Deviation} = \sqrt{\frac{\sum fd^2}{\sum f}}$$

$$= \sqrt{\frac{98,000}{20}}$$

$$= \sqrt{4900}$$

$$= 70$$

The variation in the agricultural production among diff regions according to standard deviation is 70 thousand tons.

Merits of S.D.

1) It is the best method of deviation.
2) It is based on every item of the distribution.
3) It is used in further statistical analysis.

Limitations of S.D.

1) It is difficult & time consuming to calculate than other methods.
2) It gives more weight to extreme items & less to those which are near the mean.

2.7 MOVING AVERAE

In moving average method averages of three or five years are calculated, so that we are able to remove yearly fluctuations in the data & we can get general trend. The 3 yearly moving average shall be computed as follows.

$$\frac{a+b+c}{3}, \frac{b+c+d}{3}, \frac{c+d+e}{3} \ldots\ldots$$

The 5 yearly moving average shall be computed as follows :

$$\frac{a+b+c+d+e}{5}, \frac{b+c+d+e+f}{5}, \frac{c+d+e+f+g}{5} \ldots\ldots$$

Q.1 Calculate 3 yearly moving average of the production figures given below & draw trend line.

Year	Production
1973	15
1974	21
1975	30
1976	36
1977	42

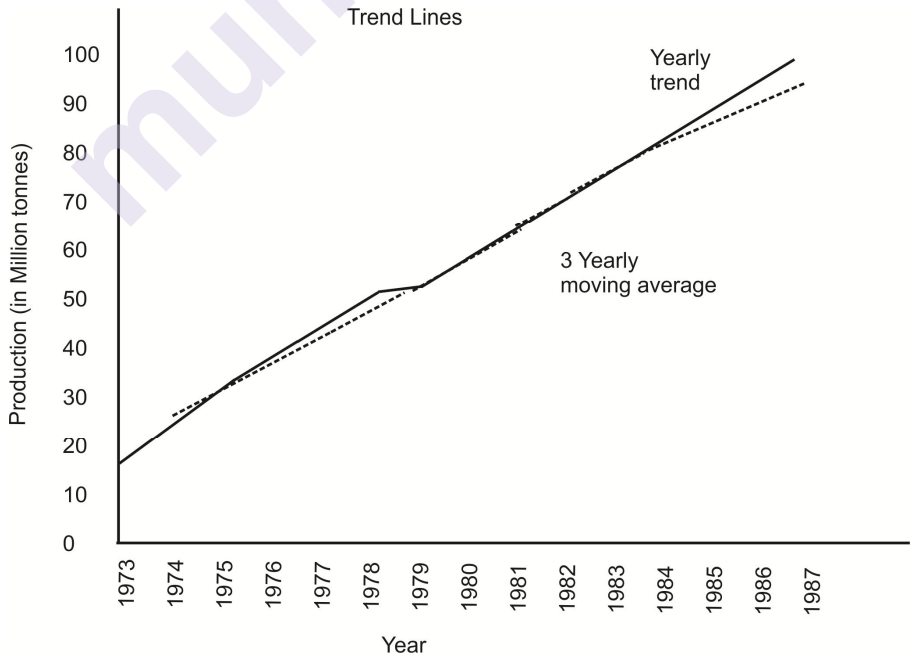
Year	Production
1978	46
1979	50
1980	56
1981	63
1982	70

Year	Production
1983	74
1984	82
1985	90
1986	95
1987	102

(Note : Production in million tonnes)

Answer :

Year	Production	3 Yearly total	3 yearly average
1973	15	--	--
1974	21	66	22
1975	30	87	29
1976	36	108	36
1977	42	124	41.33
1978	46	138	46
1979	50	152	50.67
1980	56	169	56.33
1981	63	189	63
1982	70	207	69
1983	74	226	75.33
1984	82	246	82
1985	90	267	89
1986	95	287	95.67
1987	102	--	



❖❖❖❖

3

CORRELATION, REGRESSION & HYPOTHESIS TESTING

Unit Structure :

3.1 OBJECTIVE

- To understand correlation between two series.
- To study the concept of Regression
- To understand X^2 test.
- To understand SPSS package.

3.2 INTRODUCTION

Correlation analysis deals with the association between two or more variables.

Correlation analysis attempts to determine the 'degree of relationship' between variables.

Regression analysis reveals average relationship between two variables and this makes possible estimation or prediction.

The X^2 test (pronounced as chi-square test) is one of the simplest & most widely used non-parametric test in statistics. The quantity X^2 describes the magnitude of the discrepancy between theory and observation.

SPSS means statistical package for social studies.

Correlation -

Association or correlation between two variables can be studied in different ways.

1) Scatter diagram

2) Rank correlation

Scatter diagram -

In this method dots are given on the paper with reference to 'X' and 'Y' axis. Each dot represents co-ordinates of that point. Alignment of dots (or scatter) represents correlation or association between two variables. Let us understand this concept with the help of following examples.

Q.1 Draw scatter diagrams for the data given below and decide nature of association between two variables.

Example 1

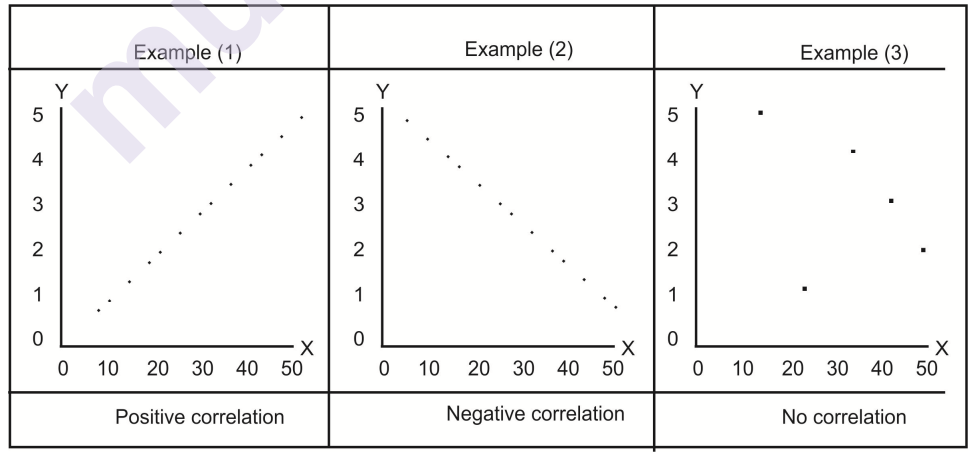
X	10	20	30	40	50
Y	1	2	3	4	5

Example 1

X	10	20	30	40	50
Y	5	4	3	2	1

Example 1

X	10	20	30	40	50
Y	5	1	4	3	2



3.3 RANK CORRELATION

This method of correlation was developed by the British psychologist Charles Edward Spearman in 1904.

In this method ranks are given to the values in 'X' and 'Y' sets of variables correlation is calculated using following formula.

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

Where $r =$ Rank correlation.

$d =$ Difference in the rank 'X' and rank 'Y'

$\sum d^2 =$ Total & square of all differences.

$n =$ number of pairs.

Q.1 Calculate the coefficient of correlation from the following data by the Spearman's Rank difference method.

Height (in cm)	140	145	150	155	160
Weight in kg	50	52	55	60	65

Answer -

Height (in cm)	Weight in kg	Rank X	Rank Y	d=Rank X - Rank Y	$n =$
140	50	1	1	0	0
145	52	2	2	0	0
150	55	3	3	0	0
155	60	4	4	0	0
160	65	5	5	0	0
					$\sum d^2 = 0$

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

$$= 1 - \frac{6 \times 0}{5\left(25 - 1\right)}$$

$$= 1 - \frac{6 \times 0}{5 \times 24}$$

$$= 1 - 0$$

$$= +1$$

Hence there is perfect positive correlation between height & weight which means if the height is more, weight is also more & if the height is less weight is also less.

Q.2 Calculate the coefficient of correlation from the following data by the Spearman's Rank method.

Price of Tea (Rs.)	Price of Coffee (Rs.)
75	120
88	134
95	150
70	115

Price of Tea (Rs.)	Price of Coffee (Rs.)
60	110
80	140
81	142
50	100

Answer -

Price of Tea (Rs.)	Price of Coffee (Rs.)	Rank X	Rank Y	d= X - Y	$n =$
75	120	4	4	0	0
88	134	7	5	2	4
95	150	8	8	0	0
70	115	3	3	0	0
60	110	2	2	0	0
80	140	5	6	-1	1
81	142	6	7	1	1
50	100	1	1	0	0
					$\sum d^2 = 6$

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

$$= 1 - \frac{6 \times 6}{8\left(64 - 1\right)}$$

$$= 1 - \frac{36}{504}$$

$$= 1 - 0.071$$

$$= +0.929$$

There is story positive correlation between price of tea & price of coffee.

3.4 REGRESSION

Regression is the measure of the average relationship between two or more variables. Hence it provides a mechanism for prediction or forecasting.

Regression analysis provide estimates of values of the dependent variable from values of the independent variable.

Regression lines -

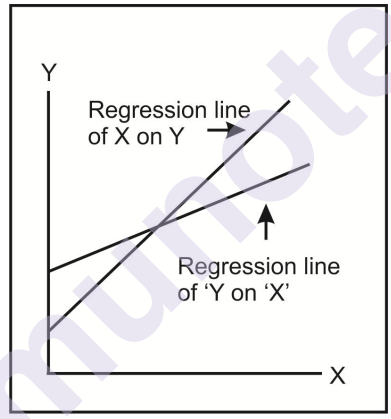
Normally we deal with 'X' and 'Y' variables in correlation & regression. We can draw two regression lines as

a) Regression of 'X' on 'Y' &

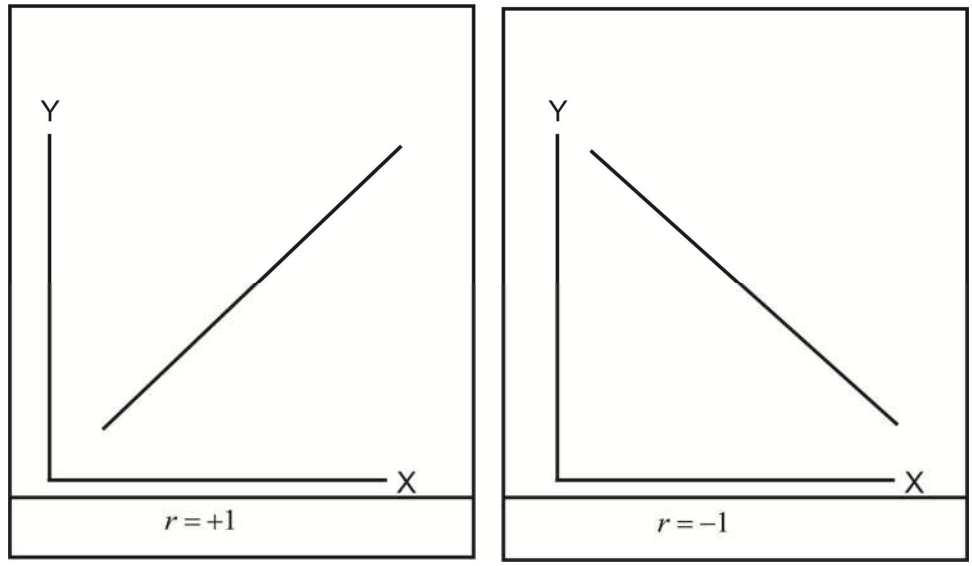
b) Regression of 'Y' on 'X'

Regression line of 'X' on 'Y' gives the most probable values of 'X' for given values of 'Y'.

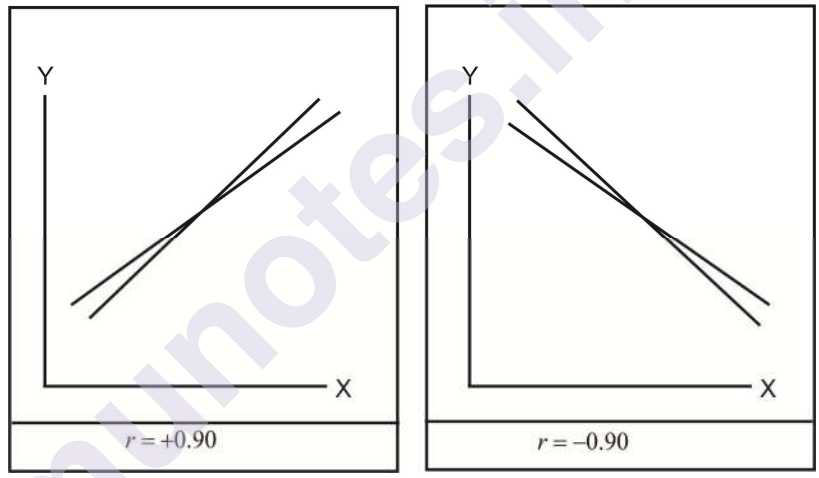
Similarly the regression line of 'Y' on 'X' gives the most probable value of 'Y' for given value of 'X'.



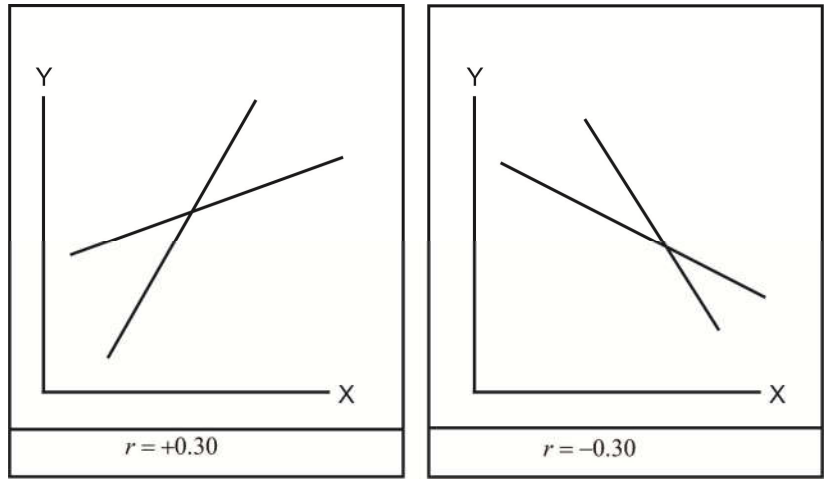
However when there is either perfect positive or perfect negative correlation between the two variables (r = +1 or r=-1) the regression lines will coincide. i.e. we will have only one line.



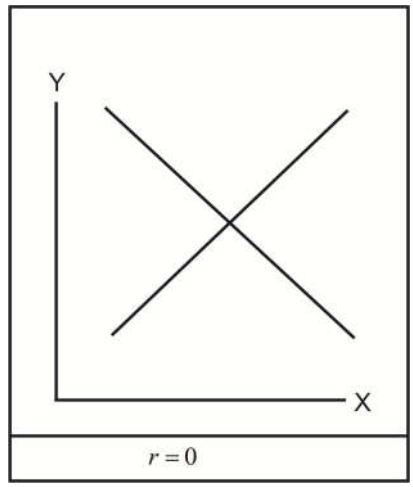
When two regression lines are near to each other, the degree of correlation will be higher.



When two regression lines are away from each other the degree of correlation will be lower.



If there is no correlation between the two variables (r=0) i.e. if the variables are independent, the regression lines are at right angles to each other.



Q.1 From the following data obtain two regression equations.

X	6	2	10	4	8
Y	9	11	5	8	7

Answer -

X	6	XY		
6	9	54	36	81
2	11	22	4	121
10	5	50	100	25
4	8	32	16	64
8	7	56	64	49
$\sum X = 30$	$\sum Y = 40$	$\sum XY = 214$	$\sum X^2 = 220$	$\sum Y^2 = 340$

Regression equation of Y on X $Y_0 = a + bX$

To find out the values of a and b the following two normal equations are used.

$$\sum Y = Na + b\sum x$$
$$\sum XY = a\sum X + b\sum X^2$$

Substituting the values

$40 = 5a + 30b$ _____ (1)

$214 = 30a + 220b$ _____ (2)

Multiplying equation (1) by $6240 = 3a + 180b$ _____ (3)

$214 = 30a + 220b$ _____ (4)

Deducting equation (4) from (3)

$$-40b = 26$$
$$b = -0.65$$

Substituting the value of b in equation (1)

$$40 = 5a + 30(-0.65)$$
$$5a = 40 - 19.5$$
$$= 59.5$$
$$a = 11.9$$

Putting he values of a and b in the equation, the regression of Y on X is $Y = 11.9 - 0.65X$.

Regression line of X on Y $X_0 = a + bY$ and the two normal equations are

$$\sum X = Na + b\sum Y$$
$$\sum XY = a\sum Y + b\sum Y^2$$

$30 = 5a + 40b$ _____ (1)

$214 = 40a + 340b$ _____ (2)

Multiplying equation (1) by 8

$240 = 40a + 320b$ _____ (3)

$214 = 40X + 340b$ _____ (4)

From equation (3) from (4)

$$-20b = 26$$
$$b = -1.3$$

Substituting the value of b in equation (1)

$$30 = 5a + 40(-1.3)$$
$$5a = 30 + 52$$
$$= 82$$
$$a = 16.4$$

Putting he values of a and b in the equation, the regression line of X on T is $X = 16.4 - 1.3Y$.

Drawing Regression Lines -

Steps for drawing regression lines are as follows.

1) Choose any two values (Preferably well apart) for the unknown variable on the right hand side of the equation.

2) Compute the other variable

3) Plot the two pairs of values.

4) Draw straight line through the plotted points.

a) Regression line of Y on X

(Y = 11.9 - 0.65 X)

1) Let $X = 2, Y = 11.9 - 0.65(2)$

$= 11.9 - 1.3$

$= 10.6$

2) Let $X = 10, Y = 11.9 - 0.65 \times (2)$

$= 5.4$

These points and the regression line through them can be represented on graph paper.

b) Regression line of X on Y.

$X = 16.4 - 1.3Y$

1) Let $Y = 10,$

$X = 16.4 - 1.3(10)$

$= 16.4 - 13$

$= 3.4$

2) Let $Y = 6,$

$X = 16.4 - 1.3(6)$

$= 16.4 - 7.8$

$= 8.6$

Let us plot these values on the graph.

a) Regression line of Y on X

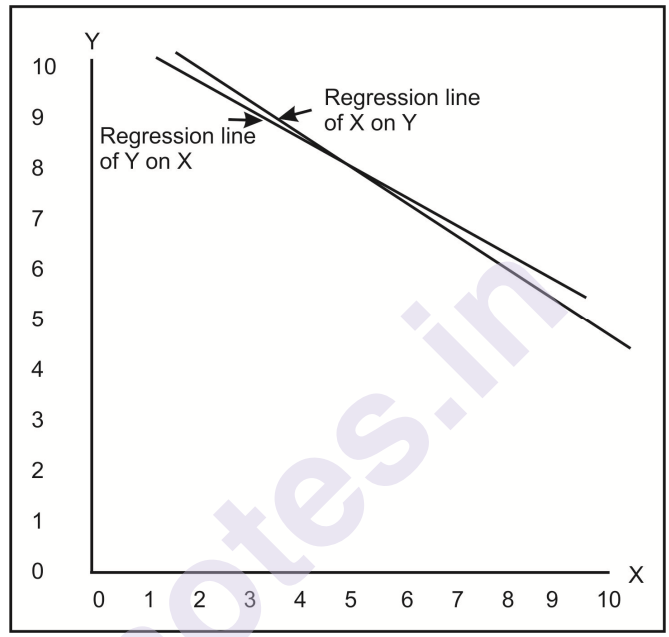
$X = 2, Y = 10.6$

$X = 10, Y = 5.4$

b) Regression line of X on Y

$Y = 10, X = 3.4$

$Y = 6, X = 8.6$



3.5 X^2 TEST

The X^2 test (pronounced as Chi-square test) is one of the simplest and most widely used non-parametric test.

The X^2 indicates the extent of difference between theory (expected) and observation (actual).

$$X^2 = \sum \frac{(O-E)^2}{E}$$

$x^2 =$ Chi-square test

$O =$ Observed frequencies

$E =$ Expected frequencies

Steps for determination of x^2

1) It is necessary to calculate expected frequencies.

$$E = \frac{CT \times RT}{n}$$

$E =$ Expected frequencies

$CT =$ Column total

$RT =$ Row total

$n =$ Total number of observations

2) Find out difference between observed and expected frequencies and calculate the squares of these differences i.e. obtain the values of $(O-E)^2$.

3) Divide the values of $(O-E)^2$ by the respective expected frequency and obtain the total $\sum\left[(O-E)^2/E\right]$. We get the value of x^2 which can be from zero to infinity.

$x^2 = 0$ means that the observed and expected frequencies completely coincide.

Greater the difference between the observed and expected frequencies greater will be the value of x^2.

The calculated value of x^2 is compared with the table value of x^2 for given degrees of freedom at a certain specified level of significance.

If at the stated level (normally 5% level is selected) the calculated value of x^2 is more than the table value of x^2, the difference between theory and observation is considered to be significant. Which means it could not have arisen due to fluctuations of simple sampling. If on the other hand the calculated value of x^2 is less than the table value, the difference between theory and observation is not considered as significant. Which means it is regarded as due to fluctuations of simple sampling and hence ignored.

Degrees of freedom -

Degrees of freedom means the number of classes to which the values can be assigned at will without violating the restrictions or limitations placed. e.g. If we wish to prepare a table which contains 5 numbers and total of all numbers is 100 then we can select four out of five numbers as per our wish but fifth number we will have to put after adding four numbers & subtracting total from 100.

1	2	3	4	5	
					= 100

Total of these five numbers is 100. We can put aby four numbers as per our hoice in these boxes.

1	2	3	4	5	
05	30	15	40		= 100

Now total of these four numbers is 05 + 30 15 + 40 = 90. We can not put fifth number as per our choice.

We can obtain fifth number by subtracting total of four numbers (90) from 100.

100 - 90 = 10

Hence fifth number is 10.

In this example the degrees of freedom is 4.

In this example our constraint is only one (1).

We can use following formula for calculation of degrees of freedom.

$V = n - K$

$V =$ degrees of freedom

$n =$ number of boxes / rows / column

$K =$ constraint

In our example $n = 5$ & $K = 1$. Hence the degrees of freedom is four.

$V = n - k$

$V = 5 - 1$

$V = 4$

For a contingency table used in x^2 test following formula is used for the calculation of degrees of freedom.

Degrees of freedom $= (c-1)(r-1)$

$C = Column$

$r = Rows$

If our table contains two columns & two rows.

5	20	Total
10	08	Total
Total	Total	Grand Total

This is 2×2 table.

The degrees of freedom for all cells is :

$$V = (c-1)(r-1)$$
$$= (2-1)(2-1)$$
$$= 1 \times 1$$
$$V = 1$$

For 3×3 table.

			Total
			Total
			Total
Total	Total	Total	

$$V = (c-1)(r-1)$$
$$= (3-1)(3-1)$$
$$= 2 \times 2$$
$$V = 4$$

It means only four expected frequencies need to be computed. The others are obtained by subtraction from normal totals.

Q.1 In an antimalarial campaign in a certain area, quinine was administered to 8/2 persons out of a total population of 3243. The number of fever cases is shown below.

Treatment	Fever	No Fever	Total
Quinine	20	792	812
No Quinine	220	2216	2436
Total	240l	3008	3248

Discuss the usefulness of quinine in checking malaria.

Answer -

It is necessary to prepare hypothesis. Hypothesis in this example is as follows.

Hypothesis = Quinine is not effective in checking malaria.

X^2 test - Expected frequency of first column & first row

$$= \frac{(column\ total) \times (Row\ total)}{(Total\ number\ of\ observations)}$$

$$= \frac{240 \times 812}{3248}$$

$$= 60$$

	Column 1	Column 2	Total
Raw 1	60		812
Raw 2			
Total	240		3248

	Column 1	Column 2	Total
Raw 1	60	752	812
Raw 2	180	2256	2436
Total	240	3007	3248

O	E	$(O-E)^2$	$(O-E)^2 / E$
20	60	1600	26.66
220	180	1600	3.88
792	752	1600	2.12
2216	2256	1600	0.70
			$\sum (O-E)^2 / E = 38.39$

$$x^2 = \left[\sum (O-E)^2 / E\right]$$

$$= 38.9$$

Degrees of freedom $V = (C-1)(r-1)$

$$= (2-1)(2-1)$$
$$= 1 \times 1$$
$$= 1$$

For degree of freedom $(V = 1)$, x^2 at 0.05 or 5% level of significance table value for x^2 is 3.84.

The calculated value of x^2 is greater than the table value, 33.39>3.84.

The hypothesis is rejected.

Hence quinine is useful in checking malaria.

❖❖❖❖

4

SAMPLING

Unit Structure

4.1 OBJECTIVES

By the end of this unit you will be able to:

- Understand Point sampling –Systematic and random
- Know the Line sampling – Systematic and random
- Learn Area sampling – Systematic and random

4.2 INTRODUCTION

In this chapter ,we are going to learn about sampling in geography After knowing what is the sampling we will learn about the different types of sampling li ke point, line and area sampling. Also we will learn the systmetic and random sampling.

4.3 SUBJECT-DISCUSSION

When you collect any sort of data, especially quantitative data, whether observational, through surveys or from secondary data, you need to decide which data to collect and from whom.
This is called the **sample**.

There are a variety of ways to select your sample, and to make sure that it gives you results that will be reliable and credible.

4.4 SAMPLE AND SAMPLE DESIGN IN GEOGRAPHY

What is sampling?

- A shortcut method for investigating a whole population
- Data is gathered on a small part of the whole parent population or sampling frame, and used to inform what the whole picture is like

Why sample?

In reality there is simply not enough; time, energy, money, labour/man power, equipment, access to suitable sites to measure every single item or site within the parent population or whole sampling frame.

Therefore an appropriate sampling strategy is adopted to obtain a representative, and statistically valid sample of the whole.

Sampling considerations

- Larger sample sizes are more accurate representations of the whole
- The sample size chosen is a balance between obtaining a statistically valid representation, and the time, energy, money, labour, equipment and access available
- A sampling strategy made with the minimum of bias is the most statistically valid
- Most approaches assume that the parent population has a normal distribution where most items or individuals clustered close to the mean, with few extremes
- A 95% probability or confidence level is usually assumed, for example 95% of items or individuals will be within plus or minus two standard deviations from the mean
- This also means that up to five per cent may lie outside of this - sampling, no matter how good can only ever be claimed to be a very close estimate

4.5 SAMPLING TECHNIQUES

Three main two of sampling strategy:

- Random
- Systematic

Within these types, you may then decide on a; point, line, area method.

Random sampling

- Least biased of all sampling techniques, there is no subjectivity - each member of the total population has an equal chance of being selected
- Can be obtained using random number tables
- Microsoft Excel has a function to produce random number

The function is simply:

- =RAND()

Type that into a cell and it will produce a random number in that cell. Copy the formula throughout a selection of cells and it will produce random numbers.

You can modify the formula to obtain whatever range you wish, for example if you wanted random numbers from one to 250, you could enter the following formula:

- =INT(250*RAND())+1

Where INT eliminates the digits after the decimal, 250* creates the range to be covered, and +1 sets the lowest number in the range.

Paired numbers could also be obtained using;

- =INT(9000*RAND())+1000

These can then be used as grid coordinates, metre and centimetre sampling stations along a transect, or in any feasible way.

Methodology

A. Random point sampling

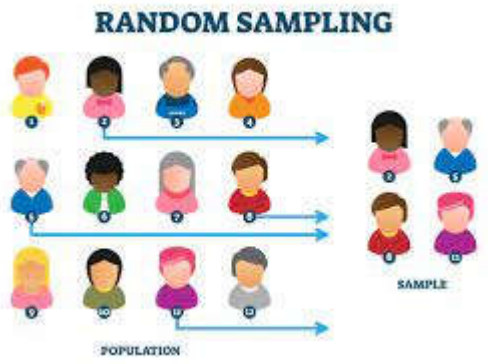
- A grid is drawn over a map of the study area
- Random number tables are used to obtain coordinates/grid references for the points
- Sampling takes place as feasibly close to these points as possible

B. Random line sampling

- Pairs of coordinates or grid references are obtained using random number tables, and marked on a map of the study area
- These are joined to form lines to be sampled

C. Random area sampling

- Random number tables generate coordinates or grid references which are used to mark the bottom left (south west) corner of quadrats or grid squares to be sampled



Advantages and disadvantages of random sampling

Advantages:

- Can be used with large sample populations
- Avoids bias

Disadvantages:

- Can lead to poor representation of the overall parent population or area if large areas are not hit by the random numbers generated. This is made worse if the study area is very large
- There may be practical constraints in terms of time available and access to certain parts of the study area

Systematic sampling

Samples are chosen in a systematic, or regular way.

- They are evenly/regularly distributed in a spatial context, for example every two metres along a transect line
- They can be at equal/regular intervals in a temporal context, for example every half hour or at set times of the day
- They can be regularly numbered, for example every 10th house or person

Methodology

A. Systematic point sampling

A grid can be used and the points can be at the intersections of the grid lines, or in the middle of each grid square. Sampling is done at the nearest

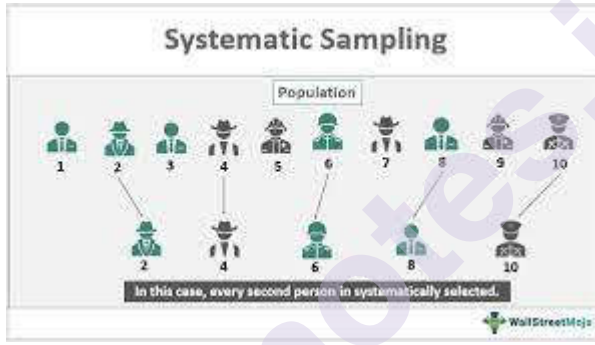
feasible place. Along a transect line, sampling points for vegetation/pebble data collection could be identified systematically, for example every two metres or every 10th pebble

B. Systematic line sampling

The eastings or northings of the grid on a map can be used to identify transect lines. Alternatively, along a beach it could be decided that a transect up the beach will be conducted every 20 metres along the length of the beach

C. Systematic area sampling

A 'pattern' of grid squares to be sampled can be identified using a map of the study area, for example every second/third grid square down or across the area - the south west corner will then mark the corner of a quadrat. Patterns can be any shape or direction as long as they are regular.



Advantages and disadvantages of systematic sampling

Advantages:

- It is more straight-forward than random sampling
- A grid doesn't necessarily have to be used, sampling just has to be at uniform intervals
- A good coverage of the study area can be more easily achieved than using random sampling

Disadvantages:

- It is more biased, as not all members or points have an equal chance of being selected
- It may therefore lead to over or under representation of a particular pattern

❖❖❖❖

5

FIELDWORK IN THE GEOGRAPHY OF ANY ONE PLACE/VILLAGE

In this unit, we will learn about geographical report writing. We understand how to write a geographical report, what methods are used, and what technique is useful for geographical report writing.

GEOGRAPHICAL FIELD REPORT:

Writing a report of the work carried on in the field is documentation of the fieldwork. This helps in the systematic reviewing of the work by students who accomplished the task and is a reference for future field trips. Field reports must be short, clear, and informative with supportive data, maps, sketches, photographs etc.

There are a number of steps involved in report writing. They are:

Title:

Identify the topic of investigation which is the purpose of field work. This is the title of the work and it has to be written in bold letters at the top of the report.

Introduction:

Every report should start with a brief introduction to the subject under study. It should explain what part of geography it relates to. For example if the study is about a stream, it falls under the branch of physical geography, more specifically geomorphology - an exogenetic agent of denudation. The time frame that was planned for the fieldwork can be elaborated. If the field work is extending for more than one day, then a clear timetable should be given.

Need for the Study:

The reason why the field work is undertaken can be mentioned. This explains the need for the field work.

The Study Area:

Details of the study area are explained here – starting with the absolute or geographical location of the study area, the choice of the study area and the physiography of the area. Other known physical and cultural details of the study area can be mentioned here. A copy of the map, satellite image etc. can be incorporated here.

Methodology Used:

The methods used to carry out the field work have to be mentioned here. The method of information collection varies according to the type of

study. It could be through observation, investigation, measurements; data collection from primary and secondary sources; field sketches, audio-video recording and photographs and GNSS surveys.

Data Analysis:

The data collected through fieldwork should be presented in a simple way for easy analysis. The method of representation of data should be according to the method of data collected. Example:

1. If the observation method is used in data collection then the data can be represented as photographs or field sketches.
2. If data is collected through surveys, it can be represented as a plan or map.
3. Data collected from secondary sources can be presented as tables, graphs, diagrams, or charts.
4. Data collected through GNSS surveys can be mapped.

The data represented in various forms have to be neatly labeled and indexed for easy identification and understanding. The photographs, diagrams, tables, maps etc. prepared during post field work have to be arranged in a sequential order so that they can provide an answer to the purpose of study and add more meaning and value to the report of work done in the field.

Conclusion:

The conclusion gives the gist of the field work – the aim, the results or findings and how it relates to existing knowledge and the addition of new knowledge through this field work. The conclusion has to present how the fieldwork has enhanced the theoretical knowledge gained in the class.

The table below gives a few steps in the preparation of field report for a few case studies under physical geography.

Steps involved in preparation of field report for field studies in physical geography

Sub topics	River	Hillock	Forest	Coast
Data Collection	Specify the method of data collection as primary / secondary source.	Specify the method of data collection as primary / secondary.	Specify the method of data collection as primary / secondary.	Specify the method of data collection as primary / secondary source.
Data Representation	Represent the data in any cartographic form such as sketch / chart / graph / map.	Represent the data in any cartographic form such as chart / graph / map / sketch.	Represent the data in any cartographic form such as chart / graph / map / sketch.	Represent the data in any cartographic form such as sketch / chart / graph / map.
Findings	From the representation list your findings.	From the representation list your findings.	From the representation list your findings.	From the representation list your findings.
Report - Writing Narrate the full work in simple language and submit.	Narrate the full work in simple language and submit.	Narrate the full work in simple language and submit.	Narrate the full work in simple language and submit.	
References	The report should have the details of references related to the study and source of data used for the study.	The report should have the details of references related to the study and source of data used for the study.	The report should have the details of references related to the study and source of data used for the study.	The report should have the details of references related to the study and source of data used for the study.

Steps involved in preparation of field report for field studies in physical geography

Sub topics	River	Hillock	Forest	Coast
Aim	To understand river as a natural resource.	To understand hillock as natural resource.	To understand forest as natural resource.	To understand coast as a natural resource.
Learning Objectives	➤ Identify the stage of river. ➤ Trace the source of the river. ➤ Assess the command area of the river. ➤ Analyse river as an ecosystem.	➤ Identify the geological history of the hillock. ➤ Determine the height of the hillock by simple measurement ➤ Draw the cross sections of it. ➤ Co-relate the vegetation with slope, supply of water and climate of the place.	➤ Identify the type of forest. ➤ List the role of forest in the life of the people. ➤ Indentify fauna and flora and their trophic level.	➤ Identify the type of coast and coastal features. ➤ List the role of coast in the life of the people. ➤ Indentify fauna and flora and their trophic level.
Study Area	Write about the river chosen, location of the village or town which is selected as study area.	Write about the hillock chosen, the village or town where the hill is located in the study area.	Write about the forest chosen, location of the village or town which is located in the study area.	Write about the coastal tract chosen and location of the village or town which is located in the tract.
Methodology	➤ With the theoretical knowledge gained to indentify the stages of a river. ➤ Trace the source of the river from published sources. ➤ Gather information about the area served by the river in terms of supplying water for irrigation, drinking purpose, industrial purpose and recreation. ➤ Observe and record the fauna and flora along the river side. ➤ Take photo/make field sketches for all your observations.	➤ Observe the agents of erosion responsible for the formation of the hillock. ➤ Using clinometer measure the height. ➤ Draw a sketch of the hillock. ➤ Collection information on cultural importance of the hillock religious / cave / paintings / resort. ➤ Study the varieties of biodiversity and correlate with the climate.	➤ Gather information about the type of trees present in the forest. ➤ Interact with local people and collect information about the resources available in terms of timber / fuel / herb / fruits and nuts / any other. ➤ Construct a trophic level diagram for the forest with the information your collected.	➤ Gather information about the area served by the coast in terms of supplying sea food, salt, power production, industrial purpose and recreation. ➤ Gather information About the type of fauna and flora along the coast and coastal water. ➤ Indentify the interaction of people with the resources available in terms of fuel/ food/fish weed / any other. ➤ Construct topic level diagram for the coastal ecosystem ➤ With your observation and gathered information, collect the historical facts about the coastal belt.
Limitation	Specify your limitations in terms of fund / time / study area selected.	Specify your limitations in terms of fund / time / study area selected.	Specify your limitations in terms of fund / time / study area selected.	Specify your limitation in terms of fund / time / study area selected.

(Continued)

Exercises

1. Measure your school's play ground and draw a plan of the same.
2. Arrange a field trip to a River line area to study the land, the direction of flow of water, trees and other plants in the area. Make a field sketch and prepare a short report.
3. Measure the daily temperature at 11.00 am and 4.00 pm and find the monthly average of maximum and minimum temperature.
4. Plan a field visit to a nearby hilly area to study the slope, gradient, trees and other plants in that area. Prepare a field sketch of the same and write a short report.

❖❖❖❖

QUESTION PAPER PATTERN (SEM - VI)
MARKS:-100 TIME:4 HRS

:

1. All questions are compulsory.
2. Figures to the right indicate marks to a sub-question.
3. Use of map stencils and simple calculator is allowed.

Q.1	Unit-I	16Marks
Q.2	Unit–II	16Marks
Q.3	Unit–III	16Marks
Q.4	Unit–IV	16Marks
Q.5	Unit–V	16Marks
Q.6	JournalandViva	20Marks