QUANTITATIVE ANALYSIS

Dr. Archana B. Patankar | Computer Engg. Dept.| TSEC

Course Code:	Course Title	Credit
CSDLO6013	Quantitative Analysis	3

Prerequisite: Applied Mathematics	
Course Objectives:	
1	Introduction to the basic concepts in Statistics
2	Understand concept of data collection & sampling methods.
3	Introduction to Regression, Multiple Linear Regression
4	Draw interference using Statistical inference methods
5	Tests of hypotheses
Course Outcomes:	
1	Recognize the need of Statistics and Quantitative Analysis
2	Apply the data collection and the sampling methods.
3	Analyze using concepts of Regression, Multiple Linear Regression
4	Formulate Statistical inference drawing methods.
5	Apply Testing of hypotheses

Module	Content	Hrs
1	**Introduction to Statistics**	6
	Functions – Importance – Uses and Limitations of Statistics. Statistical data– Classification, Tabulation, Diagrammatic & Graphic representation of data	
2	**Data Collection & Sampling Methods**	6
	Primary & Secondary data, Sources of data, Methods of collecting data. Sampling – Census & Sample methods –Methods of sampling, Probability Sampling and Non-Probability Sampling.	
3	**Introduction to Regression**	8
	Mathematical and Statistical Equation – Meaning of Intercept and Slope – Error term – Measure for Model Fit –R2 – MAE – MAPE.	
4	**Introduction to Multiple Linear Regression**	8
	Multiple Linear Regression Model, Partial Regression Coefficients, Testing Significance overall significance of Overall fit of the model, Testing for Individual Regression Coefficients	
5	**Statistical inference**	6
	Random sample -Parametric point estimation unbiasedness and consistence - method of moments and method of maximum likelihood.	
6	**Tests of hypotheses**	5
	Null and Alternative hypotheses. Types of errors. Neyman-Pearson lemma-MP and UMP tests.	

Textbooks:	
1	Agarwal, B.L. (2006):-Basic Statistics. Wiley Eastern Ltd., New Delhi
2	Gupta, S. P. (2011):-Statistical Methods. Sultanchand&Sons, New Delhi
3	Sivathanupillai, M &Rajagopal, K. R. (1979):-Statistics for Economics Students.
4	Hogg ,R.V. and Craig, A.T.(2006), An introduction to mathematical statistics, Amerind publications.

References:	
1	Arora, P.N., SumeetArora, S. Arora (2007):- Comprehensive Statistical Methods. Sultan Chand, New Delhi
2	Montgomery,D.C. ,Peck E.A, & Vining G.G.(2003). Introduction to Linear Regression Analysis. John Wiley and Sons,Inc.NY
3	Mood AM, Graybill FA, and Boes, D.C.(1985), Introduction to the theory of statistics, McGrawhill Book Company, New Delhi.
4	Kapur, J.N. and Saxena,H.C.(1970), Mathematical statistics, Sultan Chand & company, New Delhi..

Assessment:	
Internal Assessment:	
Assessment consists of two class tests of 20 marks each. The first class test is to be conducted when approx. 40% syllabus is completed and second class test when additional 40% syllabus is completed. Duration of each test shall be one hour.	
End Semester Theory Examination:	
1	Question paper will comprise of total six questions.
2	All question carries equal marks
3	Questions will be mixed in nature (for example supposed Q.2 has part (a) from module 3 then part (b) will be from any module other than module 3)
4	Only Four question need to be solved.
5	In question paper weightage of each module will be proportional to number of respective lecture hours as mention in the syllabus.

1. INTRODUCTION TO STATISTICS

- **Functions**
- **Importance**
- **Uses and limitations of statistics**

- **Statistical data**
  - **Classification**
  - **Tabulation**
  - **Diagrammatic & graphical representation of data**

LEARNING OBJECTIVES

After going through this unit, you will be able to:

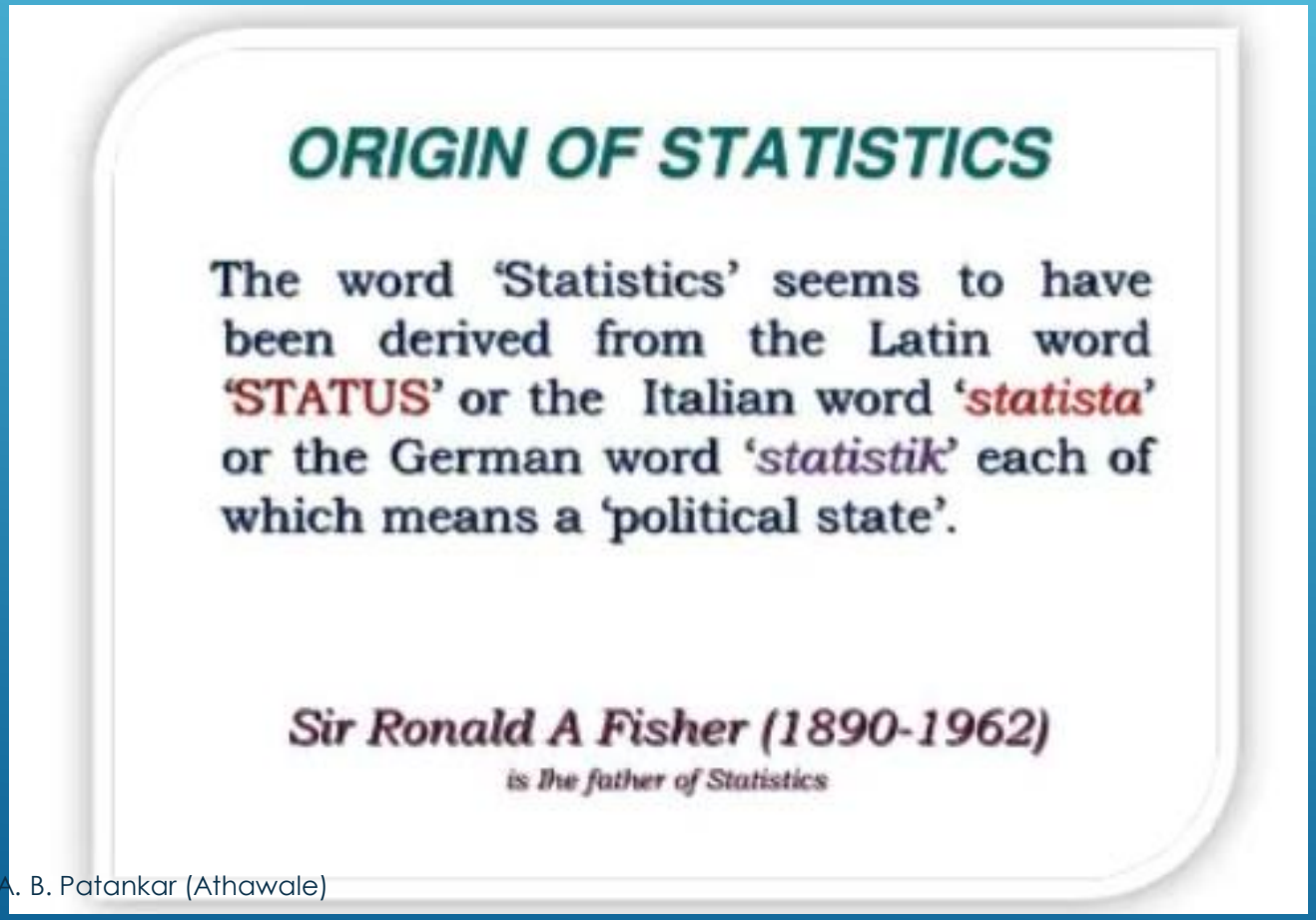
- define what is Statistics
- describe the uses of Statistics
- explain the concept of frequency distribution
- analyse classification and tabulation of data along with the importance of classification and tabulation etc
- explain graphical presentation of data.

Definitions, Scope and Limitations

INTRODUCTION

- In the modern world of computers and information technology, the importance of statistics is very well recognized by all the disciplines.
- Statistics has originated as a science of statehood and found applications slowly and steadily in
  - Agriculture,
  - Economics,
  - Commerce,
  - Biology,
  - Medicine,
  - Industry,
  - Planning, education and so on.
- As on date there is no other human walk of life, where statistics cannot be applied.

ORIGIN AND GROWTH OF STATISTICS



ORIGIN AND GROWTH OF STATISTICS

- The word ' Statistics' and ' Statistical' are all derived from the Latin word Status, means a political state.

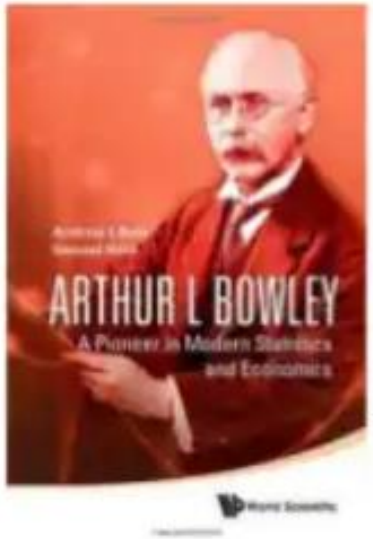
- Statistics is concerned with scientific methods for
  - collecting,
  - organising,
  - summarising,
  - presenting and analysing data
  - as well as deriving valid conclusions and making reasonable decisions on the basis of this analysis.

MEANING OF STATISTICS

- The word ‘ statistic’ is used to refer to
  - Numerical facts, such as the number of people living in particular area.
  - The study of ways of collecting, analysing and interpreting the facts.

Definition by A.L.Bowley

- "Statistics are numerical statement of facts in any department of enquiry placed in relation to each other."
- "Statistics may be rightly called the scheme of averages."

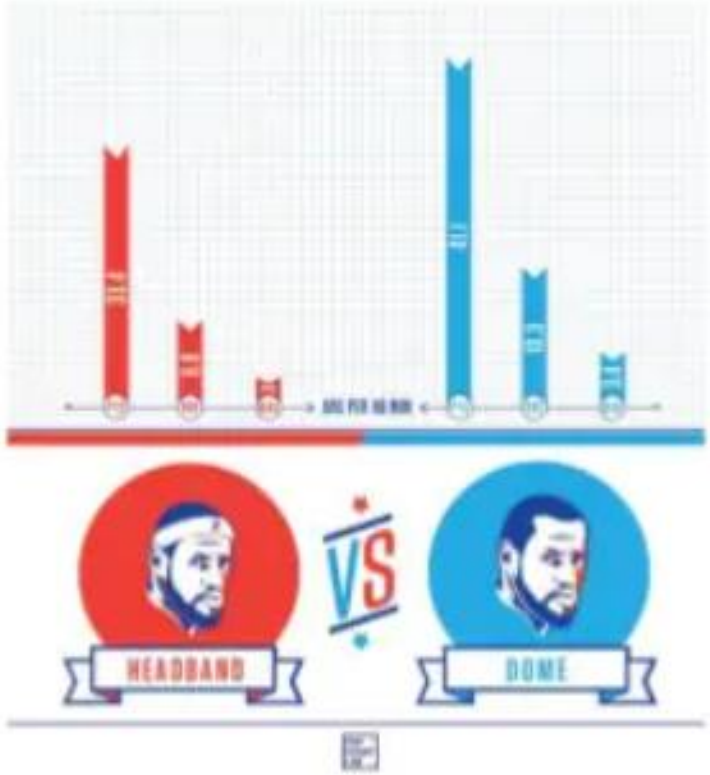


Definition by Croxton and Cowden

- "Statistics may be defined as the science of collection, presentation analysis and interpretation of numerical data from the logical analysis."
  - 1. Collection of Data
  - 2. Presentation of data
  - 3. Analysis of data
  - 4. Interpretation of data

Functions of Statistics

- 1. Condensation
- 2. Comparison
- 3. Forecasting
- 4. Estimation



FUNCTIONS

- Presentation of facts
- Simplification of complexities
- Facilitating comparisons
- Facilitating the formulation of policies
- Widening of human knowledge
- Useful in testing the laws of other sciences
- Facilitates the forecasting
- Establishment of correlation between two facts

Scope of Statistics

- 1. Statistics and Industry
- 2. Statistics and Commerce
- 3. Statistics and Agriculture
- 4. Statistics and Economics
- 5. Statistics and Education
- 6. Statistics and Planning
- 7. Statistics and Medicine
- 8. Statistics and Modern applications

SCOPE AND IMPORTANCE OF STATISTICS

- Useful to bankers
- Useful to insurance company
- Useful to railways and other transport agencies
- Useful to business
- Useful to economists
- Useful to planning

Limitations of Statistics

- Statistics is not suitable to the study of qualitative phenomenon
- Statistics does not study individuals
- Statistics laws are not exact
- Statistics table may be misused
- Statistics is only, one of the methods of studying a problem

LIMITATION

- Statistics does not study individuals
- Statistics deals with quantitative facts
- Statistics is true only to its averages
- Statistics may lead to fallacious conclusion
- Only experts can make use of statistics
- Homogeneity and uniformity is must

FREQUENCY DISTRIBUTION

A ***frequency distribution*** is a tabular presentation that generally organizes data into classes in terms of class intervals and shows the number of observations, called frequency, falling into each of these classes. An interval like 'a – b' (read as a to b) defining a class is called a class interval. The number of times an observation or value is repeated is called the ***simple frequency*** or frequency. On the other hand, the number of values included in a class interval is called the frequency of that class interval. Again when we add to the frequency of an observation (class interval), the frequencies of the previous observations (class intervals) then the sum is called the ***cumulative frequency*** of that observation (class interval).

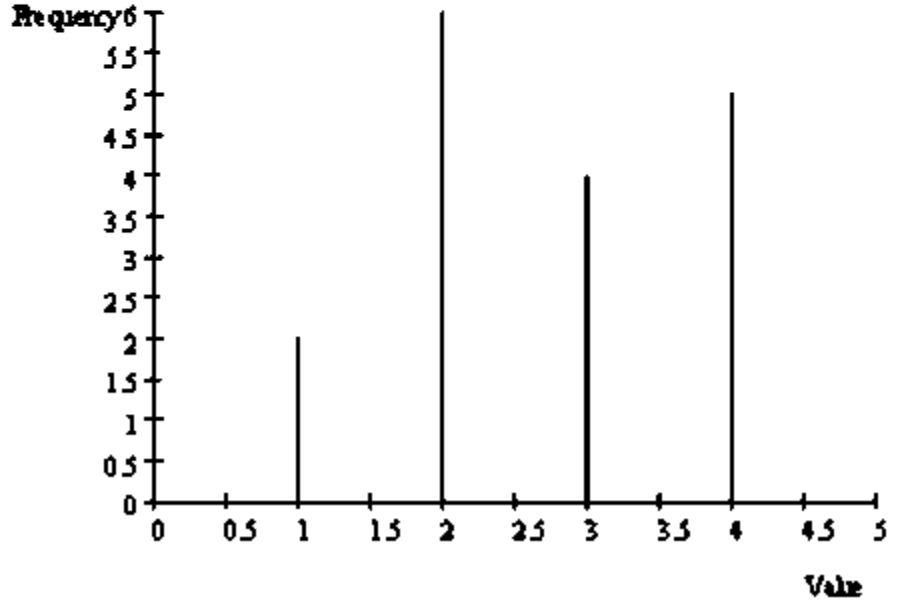
FREQUENCY DISTRIBUTION…

A frequency distribution with one variable is called a univariate frequency distribution. For example, if the students in a class are classified on the basis of their height then it will be a univariate distribution. Here the variable (variate) is the height of the students. There are two other distributions called bivariate frequency distribution and multivariate frequency distribution.

1. UNIVARIATE FREQUENCY DISTRIBUTION

X	f
4	5
3	4
2	6
1	2

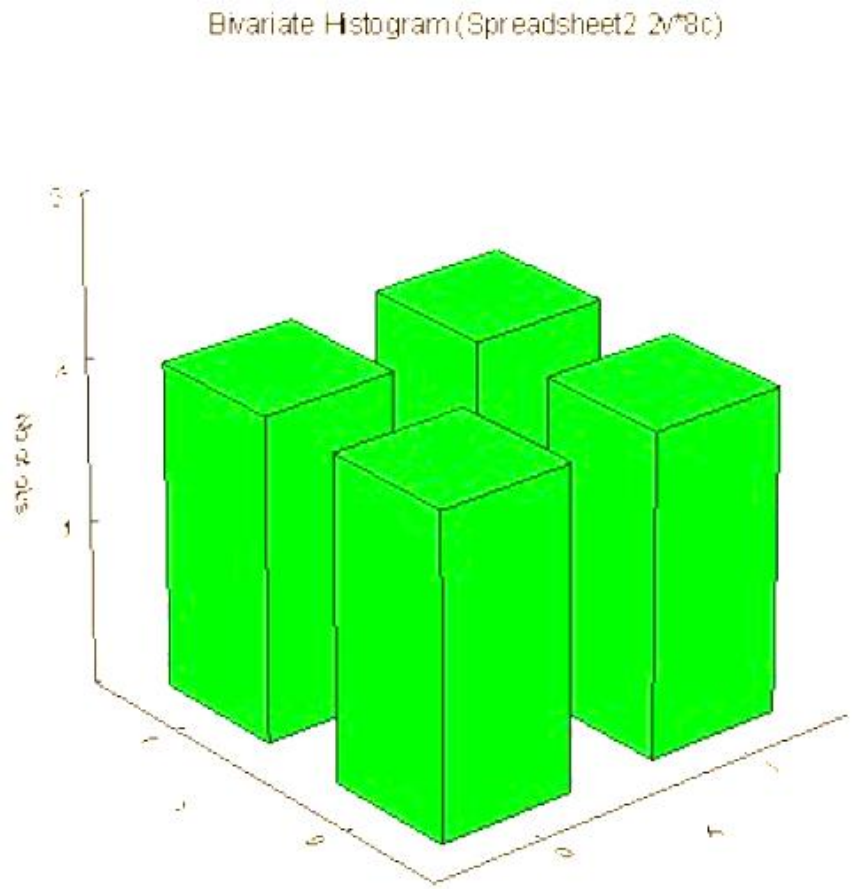
We need two dimensions to table or plot a univariate (1-variable) frequency distribution – one dimension for the value, one dimension for its frequency.



2. BIVARIATE/TWO-WAY FREQUENCY DISTRIBUTION

For example, suppose you throw two coins, X and Y, simultaneously and record the outcome as an ordered pair of values. Imagine that you threw the coin 8 times, and observed the following (1=Head, 0 = Tail)

(X,Y)	f
(1,1)	2
(1,0)	2
(0,1)	2
(0,0)	2



TYPES OF UNIVARIATE FREQUENCY DISTRIBUTION

1. Individual Observation
2. Discrete Frequency Distribution
3. Continuous Frequency Distribution

TYPES OF UNIVARIATE FREQUENCY DISTRIBUTION

Individual Observation

Individual observation is a series where items are listed singly after observation, as distinguished from listing them in groups. The following table is an example.

Roll No.	1	2	3	4	5	6	7	8	9	10
Marks	40	33	27	38	41	48	44	51	39	55

The data in the above form is called raw or disorganised data. The above presentation does not give any useful information. A better presentation of the above raw data would be to arrange in an ascending or descending order of magnitude which is called 'arraying' of the data.

FREQUENCY DISTRIBUTION…

Discrete Frequency Distribution: A discrete variable is one which takes isolated values or specific values. A discrete variable is the result of counting. A frequency distribution prepared for a discrete variable is called a discrete frequency distribution. It may be without class intervals or with class intervals.

FREQUENCY DISTRIBUTION…

Ungrouped Frequency Distribution

Consider the marks scored by 25 students:

7	5	3	5	6	4	6	2	8
1	7	2	8	5	5	6	3	
4	5	9	6	10	5	5	10	

Marks	Tally-sheet	Number of students (Frequency)
1	I	1
2	II	2
3	II	2
4	II	2
5	𝍸 II	7
6	IIII	4
7	II	2
8	II	2
9	I	1
10	II	2
	Total	25

FREQUENCY DISTRIBUTION…

An example of discrete frequency distribution without class intervals is given below:

Marks obtained by students (x)	30	33	45	55	63
No. of students (f)	5	7	4	5	6

The following is an example of discrete frequency distribution with class intervals:

Family size	1 - 2	3 - 4	5 - 6	6 - 7	7 - 8
No. of families	5	8	3	2	3

Although the above distribution looks like a continuous distribution yet is not so. The number of members in a family i.e. family size cannot be a fractional number like 1½, 3¾ etc. Hence the variable representing the family size is a discrete variable.

Ungrouped Frequency Distribution

FREQUENCY DISTRIBUTION…

Continuous Frequency Distribution: A continuous variable is one which can take any value integral or fractional within a specified range of values. A continuous variable is the result of measurement. A frequency distribution prepared for a continuous variable in terms of class intervals is called a continuous frequency distribution. **A discrete frequency distribution without class intervals** is called a simple or ungrouped frequency distribution and a frequency distribution (discrete or continuous) **with class intervals** is also called a grouped frequency distribution.

FREQUENCY DISTRIBUTION…

An example of continuous frequency distribution is the following:

Height of students (in cm.)	120-130	130-140	140-150
No. of students	10	17	10
Height of students (in cm.)	150-160	160-170	
No. of students	11	9	

Grouped Frequency Distribution

In the above example the variable is the height of students and it can take any value between 120cm. and 170cm. For example, the height of a student may be 120.17 cm., 133,255 cm, etc. Hence the variable representing the height of students is a continuous variable.

FREQUENCY DISTRIBUTION…

We have mentioned above that an interval defining a class is called a class interval. For example, in case of a frequency distribution like 0 - 9, 10 - 19, 20 - 29, ……or, 0 - 5, 5 - 10, 10 - 15, …….each is called a class interval. The end numbers of a class interval are called class limits. The smaller number is called the ***lower class limit*** and the larger number is called the ***upper class limit***.

The length of a class interval is the difference between the lower limit of the class interval and the lower limit of the immediately succeeding class interval. It may be noted that 'length of class interval' is also called 'width of class interval', or 'size of class interval'. If in a frequency distribution, whether discrete or continuous, the length of each class interval is the same then the frequency distribution is said to be a frequency distribution with equal class intervals. For example, frequency distributions involving class intervals 0 - 9, 10 - 19, 20 - 29, 30 - 39, 40 - 49, 50- 59 and 0 - 10, 10 - 20, 20 - 30, 30 - 40, 40 - 50, 50 - 60, frequency distributions with equal class intervals, the length (or width) of the class intervals in each distribution being 10. On the other hand, if in a frequency distribution the lengths of all the class intervals are not the same then such a distribution will be called a frequency distribution with unequal class intervals. An example of such a frequency distribution is : 0 - 10, 10 - 20, 20 - 35, 35 - 55, 55 - 70.

In a discrete frequency distribution with class intervals, say, 0 - 9, 10 - 19, 20 - 29,.....both the lower and the upper limits of a class (defined by an interval) are included in that class. For instance, in the aforementioned frequency distribution both the lower and the upper limits of any class (class interval) are included in that class. Note that, both the figures 10 and 19 which are respectively the lower and the upper limits of the class defined by the class interval 10 - 19 are included in that class. Here the class defined by the interval 10 - 19 includes the numbers greater than or equal to 10 but less than or equal to 19. The method of classification of data in terms of class intervals in which both the lower limit and the upper limit of any class (class interval) are included in that class (class interval) is known as the inclusive method of classification. On the other hand, in case of a continuous frequency distribution like 0 - 10, 10 - 20, 20 - 30.....while the lower limit of a class is included in that class, the upper limit is not included. For example, the class defined by the interval 0 - 10 includes the numbers which are greater than or equal to zero but less than 10. The method of classification of data in terms of class intervals in which the lower limit of any class included in that class but the upper limit is not included is known as the ***exclusive method of classification.***

There are two methods of class interval (a) Exclusive Method (b) Inclusive Method.

(a) Exclusive Method (Overlapping) Under this method, the upper limit of one class is the lower limit of the next class. For example,

Marks	10–20	20–30	30–40	Total
No. of Students	15	20	10	45

This method ensures continuity of data. A student whose mark is between 10 and 19.9 would be included in the 10–20 class. A student whose mark is 20 would be included in the class 20–30.

(b) Inclusive Method (Non-Overlapping) In this method, the upper limit of one class is included in that class itself.

Marks	10–19	20–29	30–39	40–49	Total
No. of Students	17	15	12	10	54

Here a student getting 29 marks is included in 20–29 class interval. Here the confusion is avoided because the upper limit of class is not the lower limit of the next class.

A frequency distribution with inclusive method of classification is called an ***inclusive frequency distribution*** and a frequency distribution with exclusive method of classification is called an ***exclusive frequency distribution.***

Thus, a discrete frequency distribution with class intervals is an inclusive frequency distribution and a continuous frequency distribution is an exclusive frequency distribution.

Class Boundaries or Real (True) Limits of a Class Interval: In most cases of measurement of continuous variables, all data are recorded nearest to some unit. For example, when the height of a person is recorded as 160cm., it means that the actual height of a person may be greater than or equal to 159.5cm, but less than 160.5cm. If the data recorded in integral values only are classified by inclusive method, viz., 155 - 159, 160 - 164, 165 - 169, etc., then the group 160 - 164 will include all persons with actual height being greater than or equal to 159.5cm. but less than 164.5cm. The limits 159.5 and 164.5 are called the lower and the upper boundaries or the lower and the upper real limits of the class defined by the inclusive class interval 160-164. Thus while 160 and 164 are respectively the lower and the upper limits of the inclusive class 160 - 164, then 159.5 and 164.5 are respectively called the lower and the upper boundaries of the same inclusive class 160 - 164. It is obvious that in this example the unit of the limits and the boundaries is 'centimetre'. It is to be noted that the lower boundary of any inclusive type class interval is 0.5 less than its lower class limit and upper class boundary is 0.5 higher than the upper class limit. In case of continuous frequency distribution which is an exclusive frequency distribution the upper limit of one class is the lower limit of the next class and in this case, class limits of a class coincide with the class boundaries of that class. Conversion of class limits to class boundaries is necessary

when continuous variables recorded in discrete form are classified by using inclusive type class intervals. When class limits of an inclusive frequency distribution are converted into class boundaries then the inclusive frequency distribution is converted into an exclusive frequency distribution.

Marks scored by 15 students are given below.

21	35	28	27	33	28	13	22	40	21	33
27	28	35	10							

(a) Arrange the marks in ascending order

(b) Arrange the marks in descending order

(c) Convert the marks into a continuous series of a class-interval of 10.

Solution

(a) Marks arranged in ascending order

10	27	33
13	27	33
21	28	35
21	28	35
22	28	40

(b) Marks arranged in descending order

40	28	22
35	28	21
35	28	21
33	27	13
33	27	10

(c) Formation of continuous series

Marks	Tally Marks	Frequency
10–20	II	2
20–30	𝍸, III	8
30–40	𝍸	5
	Total	15

Continuous class frequency

Marks	
10–20	2
20–30	8
30–40	5
Total	15

Mid-Point or Mid-Value of a Class Interval: The mid-point or the mid-value of a class interval is the point or value exactly at the middle of the class interval. This point or value is obtained as follows:

Mid-point or Mid-value = ½ (Lower class limit + upper class limit)

or

Mid-point or Mid value = ½ (Lower class boundary + upper class boundary)

Thus for the class intervals 10 - 19 and 5 - 10, the mid-points or mid-values

are = $\frac{10+19}{2}$ = 14.5 and = = 7.5 respectively.

The mid-value may also be calculated as:

Mid-value = Lower limit + ½ (upper limit - lower limit)

Mid-value = Lower boundary + ½ (upper boundary - lower boundary)

Open-end Distributions: An open-end distribution is one in which one or two classes lack one class limit. It is possible that the first class interval may not have the lower limit and the last class interval may not have the upper limit in a frequency distribution. Such a distribution is known as open-end distribution. We present below such as distribution:

Marks in Mathematics	Less than 10	10-20	20-30	30-40	40 and above
No. of students	10	17	15	5	7

In the above distribution the first class interval is presented as "less than 10" and hence the first class does not possess the lower limit. Again, the last class is presented as "40 and above" and hence there is no upper limit of this class. Each of these two class intervals is called an open end class interval.

CUMULATIVE FREQUENCY (CF) DISTRIBUTION

Cumulative Frequency Distribution (Cf) Cumulative frequencies are derived by the cumulation of the frequencies of successive values. It represents the total frequency of all previous variables including the variable or the class.

The cumulation is started from the lowest size to the highest size.

Less than cumulative frequency is obtained by adding successively the frequencies of all the previous variable including the variable against which it is written. More than cumulative frequency distribution is obtained by finding the cumulation total of frequencies starting from the highest to the lowest variable.

CUMULATIVE FREQUENCY (CF) DISTRIBUTION

Marks	Frequency	Cumulative frequency less than		Cumulative Frequency more than
20–30	3	3		100
30–40	8	(3 + 8)11	(100–3)	97
40–50	10	(11 + 10)21	(97–8)	89
50–60	5	(21 + 5)26	(89–10)	79
60–70	14	(26 + 14)40	(79–5)	74
70–80	20	(40 + 20)60	(74–14)	60
80–90	28	(60 + 28)88	(60–20)	40
90–100	12	(84 + 12)100	(40–28)	12

CUMULATIVE FREQUENCY (CF) DISTRIBUTION

The above less than cumulative frequency distribution and more than cumulative frequency distribution can also be expressed in the following forms:

Less than c.f.d		**More than c.f.d**	
End values (upper limit)	c.f (less than)	End values (lower limit)	c.f (more than)
Less than 30	3	More than 20	100
Less than 40	11	More than 30	97
Less than 50	21	More than 40	89
Less than 60	26	More than 50	79
Less than 70	40	More than 60	74
Less than 80	60	More than 70	60
Less than 90	84	More than 80	40
Less than 100	100	More than 90	12

From the above table, it is easy and quick to find out the number of students who have scored marks less than or more than a particular mark; for example, if we want to know the number of students who scored marks more than 50, it can be known by looking at the cumulative frequency table. In the above illustration, the number of students who scored marks more than 50 is 79 and less than 50 is 21.

Less than cumulative frequency corresponds to the upper limit of the class and more than cumulative frequency corresponds to the lower limit.

(a) Make a frequency distribution with intervals of 10 from the following data.

(b) Also prepare less than cumulative frequency distribution.

(c) And prepare more than cumulative frequency distribution.

20	36	63	84	11	72	52	66	85	74	21	43
57	95	15	45	88	72	74	54				

Solution

(a) Frequency Distribution

Marks	Tally marks	Frequency
10–20	I	2
20–30	I	2
30–40	I	1
40–50	I	2
50–60	III	3
60–70	I	2
70–80	IIII	4
80–90	III	3
90–100	I	1
	Total	20

(b) Less than c.f.d

Marks	c.f
Less than 20	2
Less than 30	(2 + 2)4
Less than 40	(4 + 1)5
Less than 50	(5 + 2)7
Less than 60	(7 + 3)10
Less than 70	(10 + 2)12
Less than 80	(12 + 4)16
Less than 90	(16 + 3)19
Less than 100	(19 + 1)20

(c) More than c.f.d

Marks	c.f
More than 10	20
More than 20	(20–2)18
More than 30	(18–2)16
More than 40	(16–1)15
More than 50	(15–2)13
More than 60	(13–3)10
More than 70	(10–2)8
More than 80	(8–4)4
More than 90	(4–3)1

Present the following data of the percentage marks of 20 students in the form of frequency table with ten classes of equal width, one class being 0–9.

06	17	28	63	70	60	80	75	92	11	09	32
43	58	73	55	23	82	33	88				

Solution

Frequency distribution of the marks of 20 students.

Marks	Tally Marks	Frequency
0–9	||	2
10–19	||	2
20–29	||	2
30–39	||	2
40–49	|	1
50–59	||	2
60–69	||	2
70–79	|||	3
80–89	|||	3
90–99	|	1
	Total	20

Two-way Frequency Distribution (Bivariate) A frequency table where two variables have been measured in the same set of items through cross classification is known as bivariate frequency distribution or two-way frequency distribution.

For example, marks obtained by students on two subjects, ages of husbands and wives, weights and heights of students etc.

The following data represent the marks in Statistics (x) and Commerce (y) of 25 students. Prepare a bivariate table from the following data.

x	25	18	20	8	30	35	17	26	42	28
y	46	32	30	20	50	52	8	42	64	50
x	30	28	25	25	60	7	32	36	25	25
y	50	46	42	38	70	8	48	55	38	40
x	17	22	31	36	48					
y	28	36	58	64	78					

Steps for construction

1. Determine the class interval of each of the variables.
2. Write one of the variables on the left hand side of the table and the other at the top.
3. The first student gets 25 in Statistics and 46 in Commerce. A tally mark has to be put in the cell where the column showing 25–35 marks in Statistics intersects the row showing 45–55 marks in Commerce.
4. Repeat the procedure for all the 25 students.
5. Total the tallies at the bottom and to the right side.
6. Totals at the right at the extreme column are for Commerce and those at the bottom row are for Statistics.

Bivariate Frequency Distribution

x→ / Y↓	5–15	15–25	25–35	35–45	45–55	55–65	Total
5–15	|	|					2
15–25	|						1
25–35		‖					3
35–45		|	𝍸				6
45–55			𝍸 |	|			7
55–65			|	‖			4
65–75						|	1
75–85					|		1
Total	2	5	12	4	1	1	25

15 pairs of values of two variables, x and y are given below. Form a two-way table.

x	16	22	35	43	29	24	14	
y	150	244	298	516	387	440	120	
x	39	41	19	28	38	39	29	44
y	481	453	247	415	387	451	512	415

Take class intervals of x as 10–20, 20–30 etc. and that of y as 100–200, 200–300 etc.

Solution

Formation of two-way table:

$x\rightarrow$ / $y\downarrow$	10–20	20–30	30–40	40–50	Total
100–200	II				2
200–300	I	I	I		3
300–400		I	I		2
400–500		II	II	II	6
500–600		I		I	2
Total	3	5	4	3	15

BIVARIATE/TWO-WAY CLASSIFICATION OF DATA

The following figures indicate income (x) and percentage expenditure on food (y) of 25 families. Construct a bivariate frequency table classifying x into intervals 200 – 300, 300 – 400, . . . and y into 10 – 15, 15 – 20, . . .

SR	*x(income)*	*y(% of exp)*
1	550	12
2	623	14
3	310	18
4	420	16
5	600	15
6	225	25
7	310	26
8	640	20
9	512	18
10	690	12
11	202	29
12	255	27
13	492	18

SR	*x*	*y*
14	587	21
15	643	19
16	689	11
17	523	12
18	317	18
19	384	17
20	400	19
21	680	13
22	300	25
23	425	16
24	555	15
25	325	23

	min	max
y	11	29
x	202	690

BIVARIATE/TWO-WAY CLASSIFICATION OF DATA

	Income (x)				
Expenditure (y) **(Percentage)**	200-300	300-400	400-500	500-600	600-700
10-15					
15-20					
20-25					
25-30					

MEASURES OF CENTRAL TENDENCY
OR

TYPES OF AVERAGES

The following are the three important types of averages:

1. Arithmetic mean
2. Median
3. Mode

Mean is a fundamental concept in mathematics and statistics. In statistics, it is defined as the measure of the central tendency of a probability distribution with median and mode. It is also recognized as the expected value.

Mean = (Sum of all the observations/Total number of observations)

Example: What is the mean of 2, 8, 10, 6 and 14?

$$Step\ 1:\ First\ add\ all\ the\ numbers.$$

$$2 + 8 + 10 + 6 + 14 = 40$$

$$Step\ 2:\ Now\ divide\ by\ 5\ (here\ 5\ is\ the\ total\ number\ of\ observations).$$

$$Mean = \frac{40}{5} = 8$$

ARITHMETIC MEAN

Question: Find the mean of the below distribution.

x	6	4	15	9	10
f	10	5	8	10	7

ARITHMETIC MEAN

Solution:

Calculation table for arithmetic mean:

x_i	f_i	x_if_i
6	10	60
4	5	20
15	8	120
9	10	90
10	7	70
–	$\sum f_i = 40$	$\sum x_i\, f_i = 360$

$$Mean = \overline{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{360}{40} = 9$$

$$Therefore\ Mean = 9$$

The median of a set of data values is the middle value of the data set when it has been arranged in ascending order. That is, from the smallest value to the highest value.

Example:

The marks of nine students in a geography test that had a maximum possible mark of 50 are given below:

47	35	37	32	38	39	36	34	35

Find the median of this set of data values.

Median is central value in orderd data so 5th value

n= 9

if n is ODD central value is value at (n +1)/2 th position

(n+1) /2 = (9 + 1)/2 = 5 th postion

After arranging in ascending order

Postion	1	2	3	4	5	6	7	8	9
Value	32	34	35	35	36	37	38	39	45
	4 values below					4 values above			

eg. The marks of 10 students are

36	34	25	34	44	44	47	31	33	48

n= 10

if n is even central value is the average of the values at (n/2) th and (n/2) +1 th position

Postion	1	2	3	4	5	6	7	8	9	10
Value	25	31	33	34	34	36	44	44	47	48

n/2= 5 value at 5 th postion = 34

n/2 + 1 = 6 value at 6 th postion = 36

Average of 34 & 36 = 35

Find the Median of the following Data

marks(x)	No of students (f)
1	5
2	8
3	10
4	15
5	20
6	16
7	14
8	13
9	10
10	9

Marks	No. of Students	cf	
1	5	5	
2	8	13	
3	10	23	
4	15	38	
5	20	58	
6	16	74	
7	14	88	
8	13	101	
9	10	111	
10	9	**120**	
	N = 120		
N/2 =	**60**		Median =6

FIND MEDIAN OF THE GIVEN DATA

Marks	No. of Students
1	36
2	34
3	25
4	34
5	44
6	44
7	47
8	31
9	33
10	48

From the following frequency distribution, find the median class: (2015)

Cost of living index	*No. of weeks*
1400-1550	8
1550-1700	15
1700-1850	21
1850-2000	8

Solution:

Cost of living index	*No. of weeks (f)*	*c.f.*
1400–1550	8	8
1550–1700	15	23
1700–1850	21	44
1850–2000	8	52
	52	

Here, $n = 52$; $\frac{n}{2} = \frac{52}{2} = 26$

∴ Median class 1700 – 1850.

MODE

The mode is that value of an observation which occurs most frequently in the data set, that is, the point (or class mark) with the highest frequency.

The concept of mode is of great use to large scale manufacturers of consumable items such as ready-made garments, shoe-makers, and so on. In all such cases it is important to know the size that fits most persons rather than 'mean' size.

Calculate mode from the following data

Calculate mode from the following data

ID	X
1	20
2	20
3	30
4	20
5	15
6	20
7	30
8	15
9	20
10	20

X	freq
20	6
30	2
15	2

mode =20

ID	X
1	20
2	20
3	30
4	30
5	30
6	20
7	30
8	15
9	20
10	15

X	freq
20	4
30	4
15	2

bimodal

ID	X
1	20
2	30
3	40
4	50
5	60
6	70
7	90
8	110
9	125
10	80

mode is illdefined

Calculate the mode from the following distribution.

Marks	No. of Students
10–20	14
20–30	17
30–40	25
40–50	21
50–60	16
60–70	19
70–80	14
80–90	13

$$Mode = l + \frac{h * (f1 - f0)}{2f1 - f0 - f2}$$

where,

l = lower limit of Modal class,

h = width of the Modal class ,

$f0$ = frequency of the **Pre − Modal** class,

$f1$ = frequency of the **Modal** class,

$f2$ = frequency of the **Post − Modal** class,

modal class: 30-40

l =	30
h =	10
f0=	17
f1=	25
f2=	21

$$\text{mode} = 30 + \frac{10*(25-17)}{2*25-17-21}$$

30+ 6.667

36.67

Find the mode of the following frequency distribution: (2013)

Class	Frequency
0–10	8
10–20	12
20–30	10
30–40	11
40–50	9

Solution:

Class	Frequency	
0-10	8	f_0
10-20	12	f_1
20-30	10	f_2
30-40	11	
40-50	9	

Maximum frequency is **12**

$\therefore$ Modal class is 10–20

$$\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h\right)$$

$$= 10 + \frac{12 - 8}{24 - 8 - 10} \times 10$$

$$= 10 + \frac{4 \times 10}{6}$$

$$= 10 + 6.\overline{6} = \mathbf{16.\overline{6}} \text{ or } \mathbf{16.67 \text{ (app.)}}$$

CLASSIFICATION OF DATA

Classification of raw data implies the arrangement of raw data into groups or classes on the basis of their similarities and resemblances. Below we quote two definitions of classification:

"Classification is the process of arranging data into sequences and groups according to their common characteristics, or separating them into different or related parts." - Secrist.

"A classification is a scheme for breaking a category into a set of parts, called classes, according to some precisely defined differing characteristics possessed by all the elements of the category". - Tuttle, A.M.

CLASSIFICATION OF DATA

Classification makes the data understandable and it reveals the true significance of the data. In order to analyse the collected data classification is a must. For example, if we classify the data pertaining to the workers working in a factory with respect to gender (male and female), age (between 0 - 10 years, 10 - 20 years, etc.), qualification (below H.S.L.C., H.S.L.C. etc.) etc. then only the data become suitable for analysis.

ADVANTAGES & IMPORTANCE OF CLASSIFICATION

(i) Classification helps in presenting the raw data in a coincise and simple form.

(ii) It facilities comparison by dividing the raw data on the basis of their similarities and resemblances.

(iii) It provides basis for tabulation and analysis of data.

(iv) Classification enables us to identify the possible characteristics in the data.

TYPES OF CLASSIFICATION

There are four types of classification of data which we are going to discuss now–

(i) **Geographical Classification:** In geographical classification data are classified on the basis of places or geographical locations. For example, if we write the population figures of the different districts of Assam against the corresponding districts then such type of classification of population data will be called geographical classification. Series of data arranged on the basis of places or geographical locations is called spatial series.

Example of geographical classification:

City	: Mumbai	Kolkata	Delhi	Chennai
Population density (per square km.) :	654	685	423	205

TYPES OF CLASSIFICATION

(ii) **Chronological classification:** When the data are classified on the basis of time then it is known as chronological classification. Series of data arranged with respect to time is known as time series. If we write down the savings of a country in different years against the respective years then such type of classification of data relating to the savings of the country will be called chronological classification.

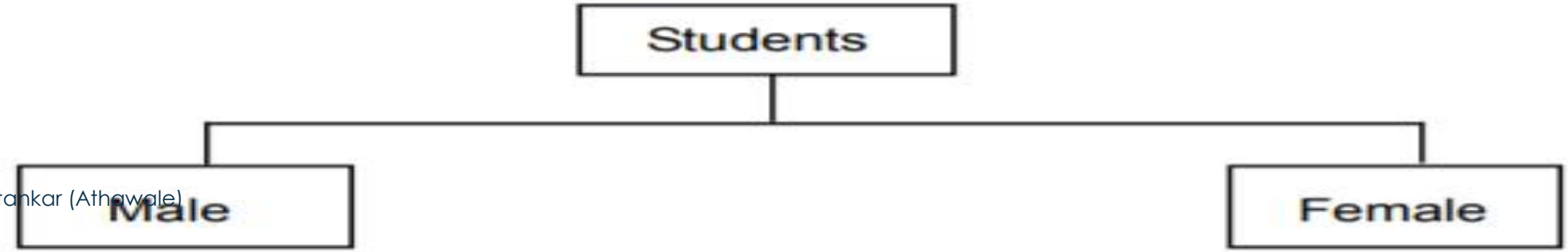
Example of chronological classification:

The following is an example of chronological classification

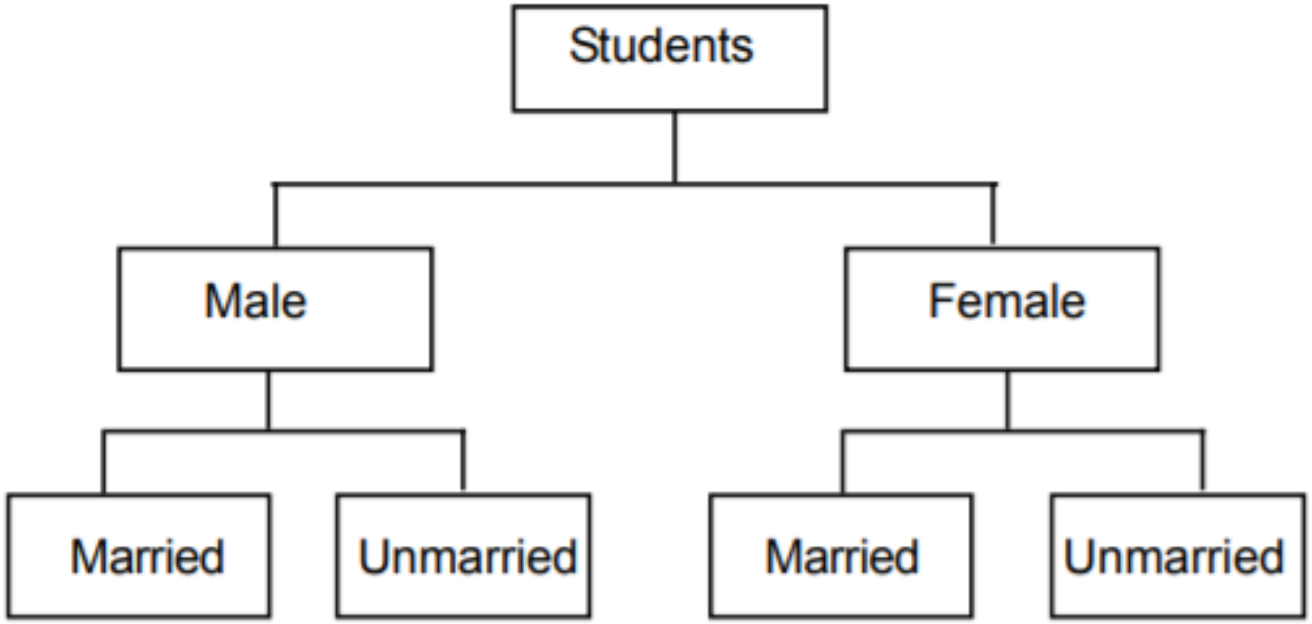
Year :	1941	1951	1961	1971	1981	1991	2001
Population : (crore)	31.9	36.9	43.9	54.7	75.6	85.9	98.6

(iii) Qualitative classification or classification according to attributes: An attribute is a non-measureable characteristic like education, sex, religion, caste, etc. If items or individuals are classified on the basis of one or more than one attributes then such a classification will be called **classification according to attributes**. There are two types of classifications according to attributes. If classification of items or individuals is made one the basis of a single attribute then such a classification is called **simple classification**. On the other hand, if items or individuals are classified on the basis of more than one attributes then such a classification is called **manifold classification**.

If we classify the students of a college on the basis of sex then it will be a case of simple classification. This is shown below:



Again, if we classify the workers in a factory on the basis of their sex and marital status as shown below then it will be a case of manifold classification:



(iv) **Quantitative classification or classification according to variables or class intervals:** If data are classified on the basis of the measureable characteristics called variables such as height, weight, marks, distance, etc. then such a classification will be called classification according to variables. If in this form of classification data are organized into classes showing the number of observations falling in each class then such a classification may also be called 'classification according to class intervals'. In this form of classification data are presented in the form of a frequency distribution.

TABULATION OF DATA

The collected data, after classification, are recorded in rows and columns to give them tabular form. Tabular presentation of data, more conveniently known as tabulation, may be defined as "the orderly or systematic presentation of numerical data in rows and columns designed to clarify the problem under consideration and to facilitate the comparison between the figures".

ADVANTAGES OR IMPORTANCE OF TABULATION

The following are the advantages of tabulation:

(i) Tabulation simplifies complicated data.

(ii) Tabulation presents quantitative data in a coincise and condenced form.

(iii) It facilitates comparison of data.

(iv) Presentation of data in tabular form provides a basis for analysis and interpretation of such data.

Objectives of Tabulation:

The objectives of tabulation are inherent in advantages of tabulation. Thus the objectives of tabulation are: (i) To simplify complicated data, (ii) To present quantitative data in a coincise and condenced form, etc.

DIFFERENT PARTS OF A TABLE

A table consists of the following parts. These may be considered as the essentials of a satisfactory table.

(i) **Table Number:** Every table should be identified by a number. It facilitates easy reference. The table number may be given at the beginning of the title of the table, or can be centered above the title of the table.

(ii) **Title**: A table must have a title which is to be written either below the table number or after the table number in the same line. The title should convey the full description of the contents in the table.

(iii) **Stub:** The extreme left hand column of the table which contains the headings of the rows is called stub.

(iv) **Caption:** Caption is the headings for the columns. It is the upper part of the table. There may be sub-heads or sub-captions in each caption.

(v) **Body:** It is the main part of the table containing the numerical figures.

(vi) **Totals:** The totals and sub-totals of all the rows and columns should be given in the table.

(vii) **Footnote:** Any explanatory note concerning the table itself, written directly beneath the table, is called 'footnote'. The purpose of footnote is to clarity some of the specific items given in the table.

(viii) **Source:** The source or sources of the data embodied in the table should be mentioned beneath the table if data are collected from secondary sources. It is given below the footnote.

TYPES OF TABLES

Tables are classified as : (i) Simple Tables and (ii) Complex Tables.

Simple Tables: Tables which are prepared on the basis of only one characteristic of the collected data are known as simple tables. Simple tables are also called one-way tables.

Example 1.

Table

Department wise classification of 100 university students

Department	No. of Students
Physics	30
Chemistry	25
Mathematics	20
Statistics	25
Total	100

Complex Tables: Tables which show more than one characteristic of the data are known as complex tables. We may have a two-way complex table showing two characteristics, a three-way complex table showing three characteristics and a multi-way complex table showing several characteristics. A two-way, three-way and multi-way complex table may also be called a two-fold, three-fold and multi-fold complex tables respectively.

Example 2. Two-way complex table:

Table

Department wise classification of 100 university students

Department	No. of Students		
	Male	Female	Total
Physics	20	10	30
Chemistry	18	7	25
Mathematics	15	5	20
Statistics	12	13	25
Total	65	35	100

TYPES OF TABLES

Example 3. Three-way complex table:

Table

Department wise classification of 100 university students

Department	Male			Female			Total		
	Hindu	Others	Total	Hindu	Others	Total	Hindu	Others	Total
Physics	16	4	20	8	2	10	24	6	30
Chemistry	14	4	18	6	1	7	20	5	25
Mathematics	13	2	15	5	0	5	18	2	20
Statistics	6	6	12	10	3	13	16	9	25
Total	49	16	65	29	6	35	78	22	100

TYPES OF TABLES

Example 4. In a sample study about tea habit in two towns the following information was received.

Town A: Females were 40%; total tea drinkers were 45% and male non-tea drinkers were 20%.

Town B: Males were 55%; male non-tea drinkers were 30%, and female tea drinkers were 15%.

Supplying the missing information, tabulate the above data.

Solution: Let us assume that there were 100 people in each town. We now tabulate the given data below.

Department	Town A			Town B		
	Male	Female	Total	Male	Female	Total
No. of people with tea habit			45		15	
No. of people without tea habit	20			30		
Total		40		55		

CALCULATION OF CLASS FREQUENCIES AND FORMATION OF FREQUENCY DISTRIBUTION TABLE

The number of class intervals and the length of class intervals in connection with a frequency distribution to be prepared for raw data is to be determined on the basis of the range of the data. Usually the numbers of class intervals should not be made less than 5 or more than 15. The range of the data is the difference between the largest and the smallest data. The lower limit of the first class interval is usually taken at zero or at a multiple of 5 in case of non-negative observations. For instance, if the smallest observation is zero or a number greater than zero but less than 5 then in case of class intervals of length 5 the first class interval may be taken as 0 - 5 in case of exclusive frequency distribution or as 0 - 4 in case of inclusive frequency distribution. Similarly if the smallest observation in an integer greater than or equal to 30 but less than 40 and if the length of the class interval is chosen to be 10 (so that the number of class intervals lies between 5 and 15) then the first class interval is to be taken as 30 - 39 (in case of inclusive frequency distribution) or as 30 - 40 (in case of exclusive frequency distribution).

Following are the steps that are necessary in forming a frequency distribution table from raw data:

(i) To form the table of frequency distribution involving three columns;

(ii) To form the class intervals and to place them in the first column of the frequency distribution table;

(iii) To make a tally bar or tally mark (a vertical bar) for each observation in the second column (headed by Tally Bars) against the corresponding class interval;

(iv) To draw the fifth tally bar after four tally bars have been drawn in either way IIIII or IIII.

(v) To count the number of tally bars against all class intervals after all the observations have been exhausted and to place these numbers against the respective class intervals in the third column (headed by the word 'frequency') of the frequency distribution table.

(vi) To add all the (simple) frequencies of the classes and to denote this sum by the letter N.

Example 5: In a particular subject 100 students of a particular class have got the following marks. Form a suitable frequency distribution.

75	80	66	56	50	82	66	45	68	56
80	69	87	74	80	76	81	64	57	68
58	60	73	74	66	87	84	98	79	64
80	63	58	87	96	77	72	74	79	78
86	66	76	76	80	82	63	75	83	34
76	68	74	41	68	40	59	90	35	94
65	88	85	84	79	87	76	65	76	80
75	82	96	95	73	60	52	80	88	63
76	83	60	79	72	77	57	92	67	52
72	86	72	86	73	78	78	79	75	60

Table

Mark wise distribution of students

Class interval	Tally bars	Frequency
30–39	||	2
40–49	|||	3
50–59	𝍸 𝍸 |	11
60–69	𝍸 𝍸 𝍸 𝍸	20
70–79	𝍸 𝍸 𝍸 𝍸	32
80–89	𝍸 𝍸 𝍸 𝍸 𝍸	25
90–99	𝍸 ||	7
	Total	N = 100

(i) How many students are getting marks between 40 and 49?

(ii) How many students are getting marks between 70 and 79?

(iii) How many students are getting marks between 50 and 79?

(iv) How many students are getting marks 40 or more?

(v) How many students are getting marks less than 60?

(vi) What is the percentage of students getting marks 80 or more?

Table

Mark wise distribution of students

Class interval	Tally bars	Frequency
30–39	||	2
40–49	|||	3
50–59	𝍸 𝍸 |	11
60–69	𝍸 𝍸 𝍸 𝍸	20
70–79	𝍸 𝍸 𝍸 𝍸	32
80–89	𝍸 𝍸 𝍸 𝍸 𝍸	25
90–99	𝍸 ||	7
	Total	N = 100

METHODS OF PRESENTING STATISTICAL INFORMATION

Diagrams and graphs are the methods which are devised to present huge mass of quantitative data so that they become at once comparable and give pleasure to the eye besides being interesting and easily understandable.

Usually data are presented with the help of the following diagrams:

(a) Bar Diagram (Chart)

(b) Rectangular and Square Diagrams (Chart)

(c) Circular and Pie Diagrams (or Pie Chart)

METHODS OF PRESENTING STATISTICAL INFORMATION

(a) Bar Diagram: In bar diagram only the length is considered, the breadth may be of any finite magnitude. The bar diagrams are divided into the following three categories:

(i) Simple Bar Diagram

(ii) Multiple Bar Diagram

(iii) Sub-Divided or Compound Bar Diagram

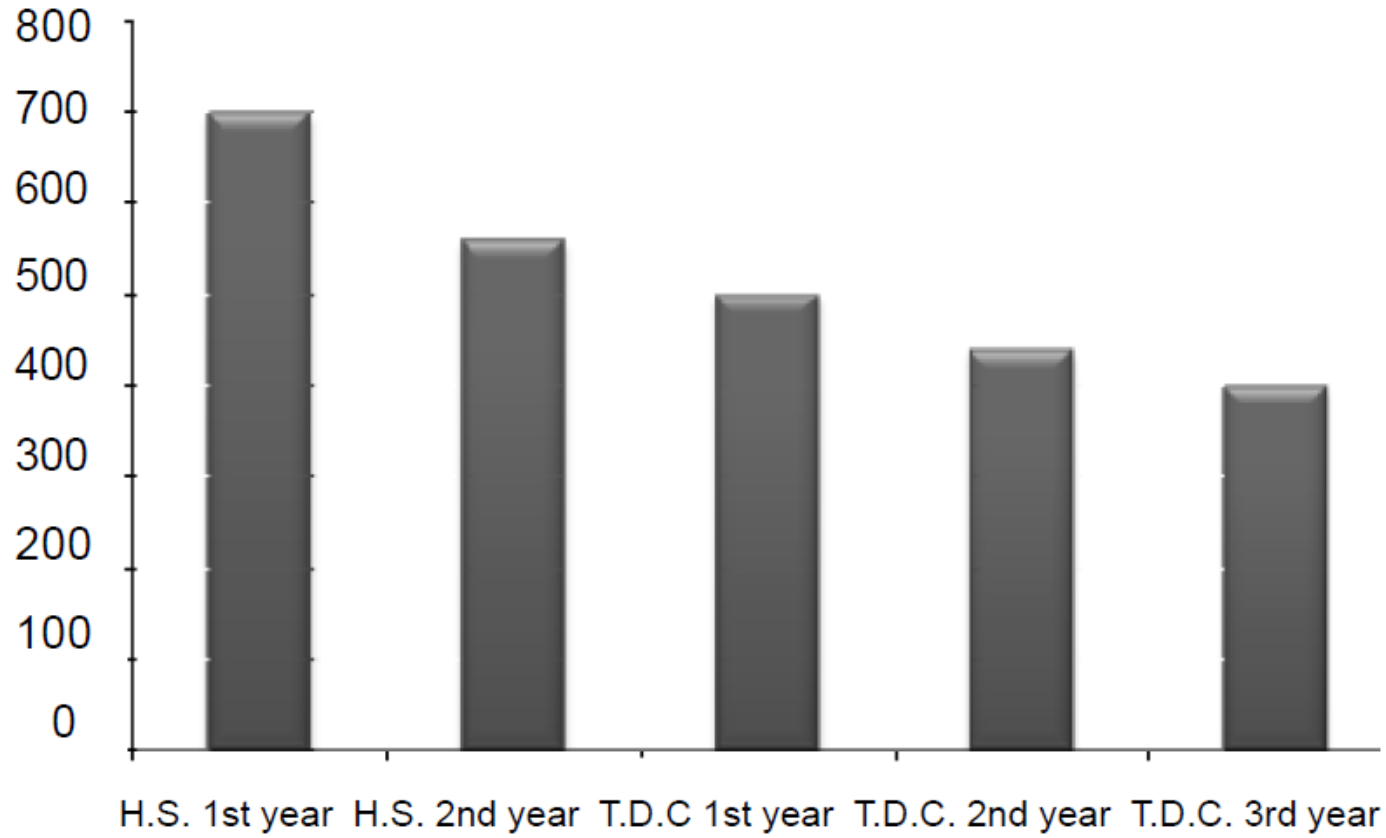
METHODS OF PRESENTING STATISTICAL INFORMATION

(i) Simple Bar Diagram: Only one type of data are presented with the help of simple bar diagram. For example, the volume of production of rice in Assam during the last five years can be presented with the help of simple bar diagram. In order to draw simple bar diagram, a bar is drawn for each datum. All the bars are on the same general base. The heights of the bars will be as per the magnitudes of the data. The breadth of each bar must be same and the gaps among the bars must be uniform. Generally the gap between two consecutive bars should not be less than half the breadth of a bar. The bars are drawn either on a common horizontal or on a common vertical base. Data can be easily compared with the help of the heights of the bars.

METHODS OF PRESENTING STATISTICAL INFORMATION

Example 1. Draw a vertical bar-chart for the number of students of a college: H.S. 1st year - 700, H.S. 2nd Year - 560, TDC 1st year - 500, T.D.C. 2nd year - 440 and T.D.C. 3rd year - 400

* ***Solution:*** In order to represent the students of the various classes of the college the following vertical simple bar diagram has been drawn:



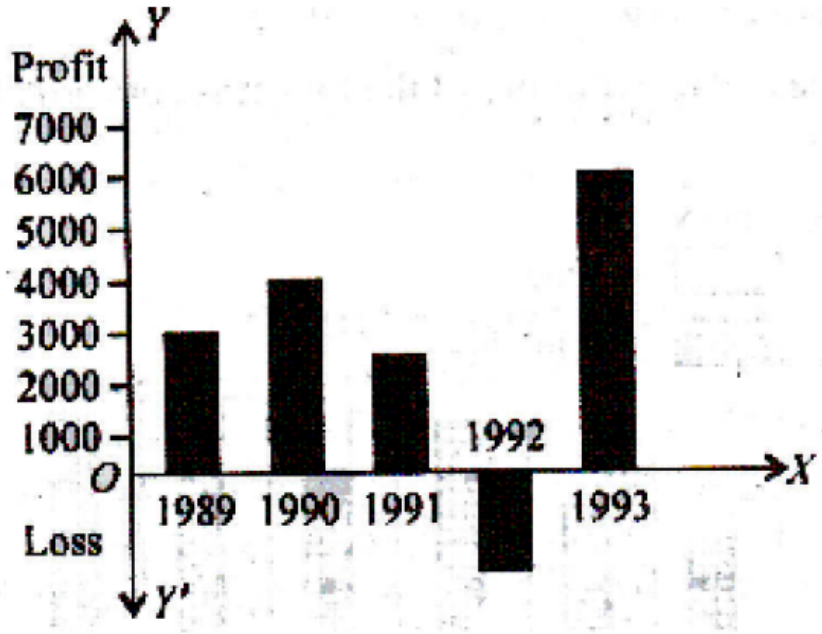
METHODS OF PRESENTING STATISTICAL INFORMATION

Example 2: The profits and losses of a business concern for the years 1989-1903 are given below:

Year :	1989	1990	1991	1992	1993
Profits (Rs.) :	3000	4000	2500	--	6000
Loss (Rs.) :	--	--	--	2000	--

Represent the above data by a bar-chart

* ***Solution:*** In order to represent the profit and loss of the business concern we draw the following simple bar diagram.



METHODS OF PRESENTING STATISTICAL INFORMATION

(ii) Multiple Bar Diagram: With the help of multiple bar diagram (multiple bar chart) more than one type of data can be depicted at a time side by side to represent 2, 3 or 4 series of values for comparison. For example, in order to show the profits and losses of a business concern for a few years, a pair of bar diagrams indicating profit and loss of each year are to be drawn side by side. For convenience the adjacent bars are differently shaded. In order that the diagram becomes distinct, gap is maintained between different groups of adjacent bars.

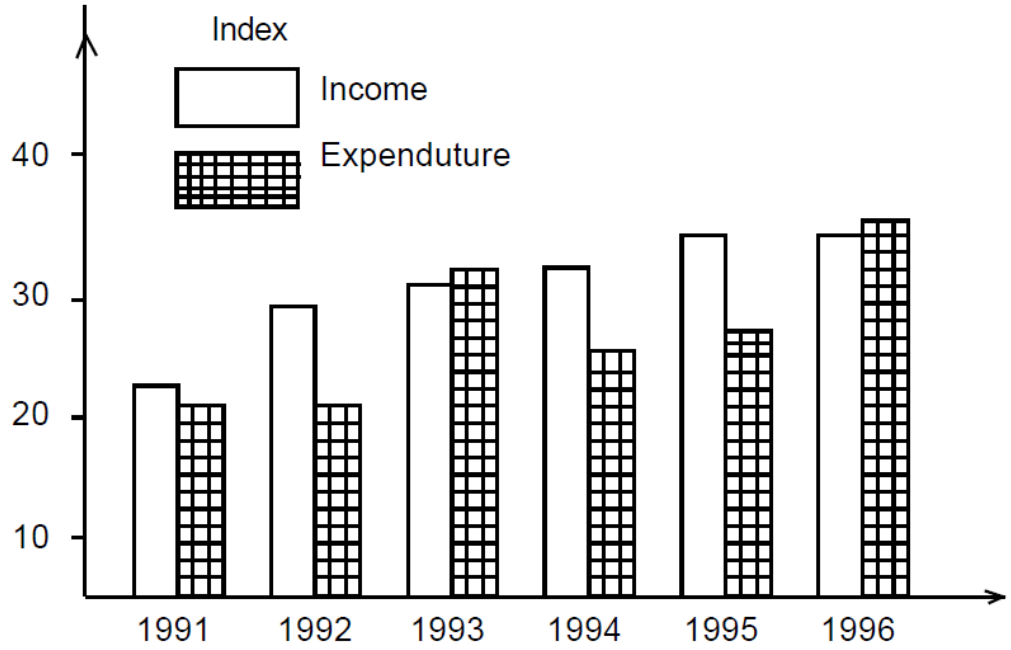
METHODS OF PRESENTING STATISTICAL INFORMATION

Example 3: The income and expenditure of a business concern (in thousand rupees) have been given below:

Year	1991	1992	1993	1994	1995	1996
Income	22.0	27.3	28.2	30.3	32.7	33.3
Expenditure	19.5	21.7	30.0	25.6	26.1	34.2

Represent the given data by a multiple bar-diagram.

* ***Solution:*** The income and expenditure of the business concern are represented by a multiple bar diagram as shown below:



Scale : Length of one small square on the Y – axis = Rs. 2000

METHODS OF PRESENTING STATISTICAL INFORMATION

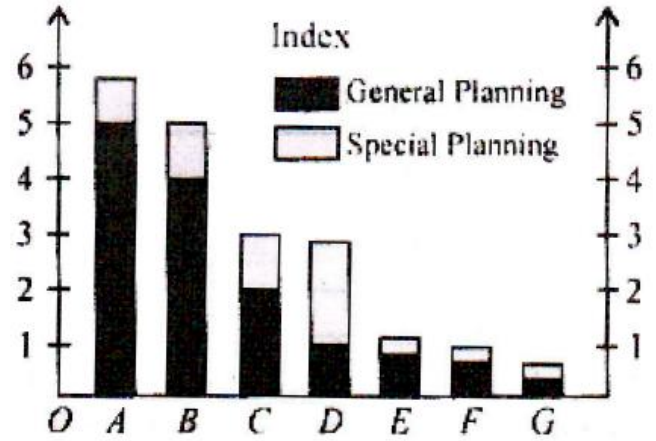
(iii) Sub-divided Bar Diagram: Sub-divided bar diagrams are used to present data which are to be shown in parts or which are totals of various subdivisions. Here a bar is drawn indicating the total corresponding to a particular item and then the bar is sub-divided into various parts or components in proportion to the various subdivisions of the quantity of the item. To distinguish various components from one another, different colours or shades may be given. The diagram may be drawn on absolute figures or percentages.

METHODS OF PRESENTING STATISTICAL INFORMATION

Example 4: The list of expenditure on various departmental heads during the Forth Five Year Plan of a certain state is given below. Represent these data by a sub-divided bar diagram.

	Items	General planning (Rs in lakh)	Special Planning (Rs. in lakh)	Total
A.	Irrigation and Power	5074	226	5300
B.	Social Welfare	3647	1245	4892
C.	Agriculture	2170	1030	3200
D.	Communication and			
	Transportation	1075	1740	2815
E.	Industry and			
	Minerals	1210	100	1310
F.	Cooperatives	738	377	1115
G.	Miscellaneous	86	212	368

* ***Solution:*** The following sub-divided bar diagram is drawn for the given data.



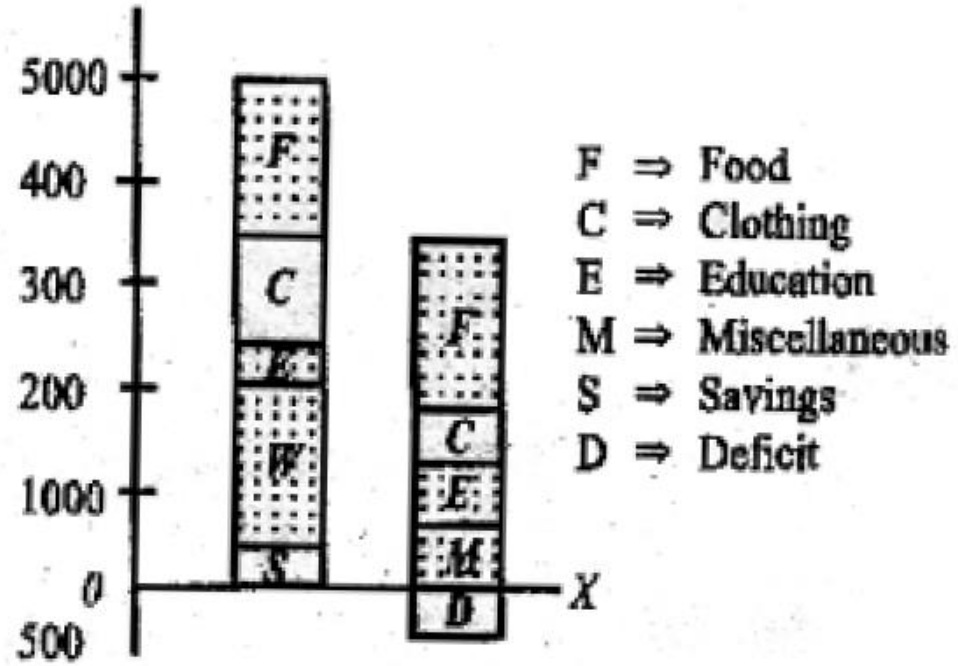
A : Irrigation and Power,
B : Social Welfare,
C : Agriculture,
D : Communication and Transportation,
E : Industry and Minerals,
F : Cooperatives,
G : Miscellaneous.

METHODS OF PRESENTING STATISTICAL INFORMATION

Example 5: Represent the following data by a suitable diagram.

Items of expenditure	Family A (Income Rs. 5000)	Family B (Income Rs. 3000)
Food	1500	1500
Clothing	1250	600
Education	250	500
Miscellaneous	1900	700
Savings or Deficit	+ 100	- 300

* ***Solution:*** The given data may be represented by a sub-divided bar diagram as shown below:



METHODS OF PRESENTING STATISTICAL INFORMATION

Pie Diagram (Pie Chart): A pie-diagram is a circle of radius neither too larger nor too small whose area is divided into as many different sectors as there are components of the whole data. This is done by drawing straight lines from the centre to the circumference of the circle. The area of the circular lamina represents the whole data and it is equivalent to 3600 at the centre. The area of each sector is proportional to the value of the corresponding components of the data. The area of a sector is proportional to the angle at the centre.

METHODS OF PRESENTING STATISTICAL INFORMATION

Example 6: The proposed expenditures on various items during a five year plan of a government are as follows:

Head	Amount of money (Rs. in crores)
Agriculture	12,000
Industry and Mines	9,000
Irrigation and Power	6,000
Education	8,000
Communication	5,000

Draw a pie chart for the data.

METHODS OF PRESENTING STATISTICAL INFORMATION

Solution: We have calculated below the various angles of pie-diagram:

Expenditure heads	Amount of expenditure (Rs. in crores)	Magnitude of angle (in degree)
Agriculture	12,000	108
Industry and Mines	9,000	81
Irrigation and Power	6,000	54
Education	8,000	72
Communication	5,000	45
Total	40,000	360

Calculation of angles:

Angle representing Rs. 40,000 = 360^0

$\therefore$ " " Rs. 1,000 = = 9^0

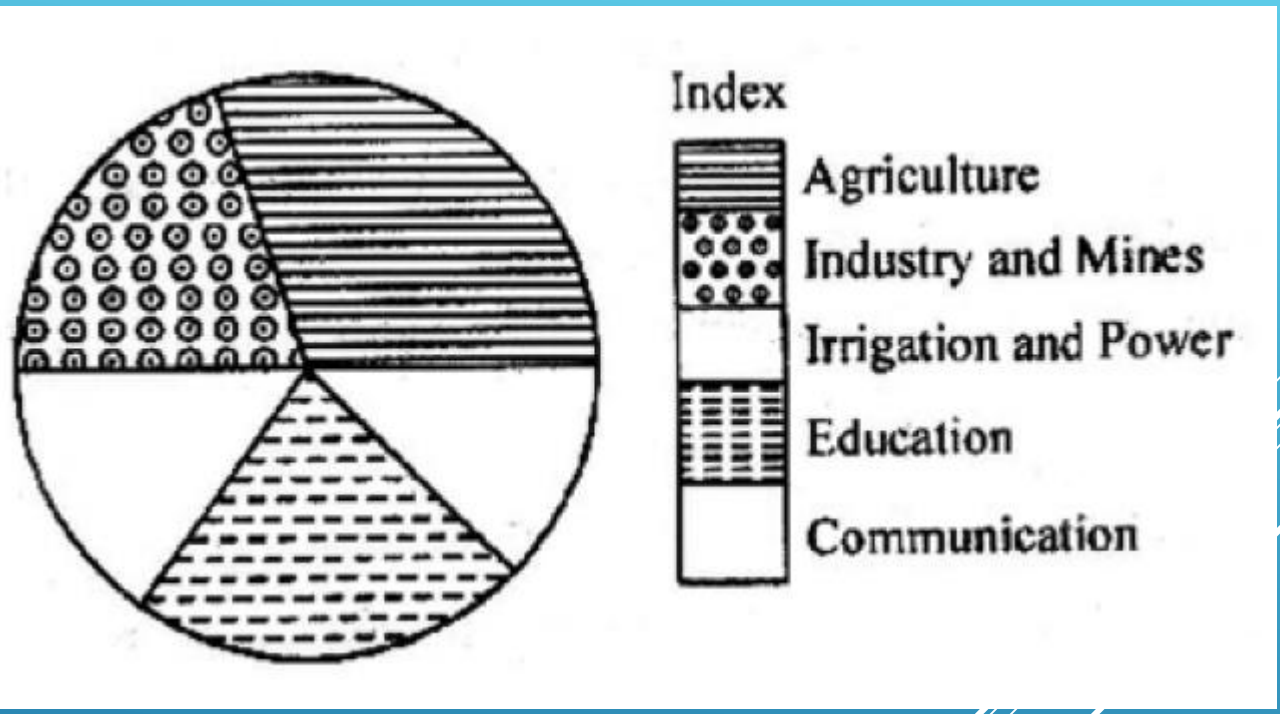
$\therefore$ " " Rs. 12,000=9 x12=108^0

" " Rs. 9,000=9 x 9=81^0

" " Rs. 6,000=9 x 6=54^0

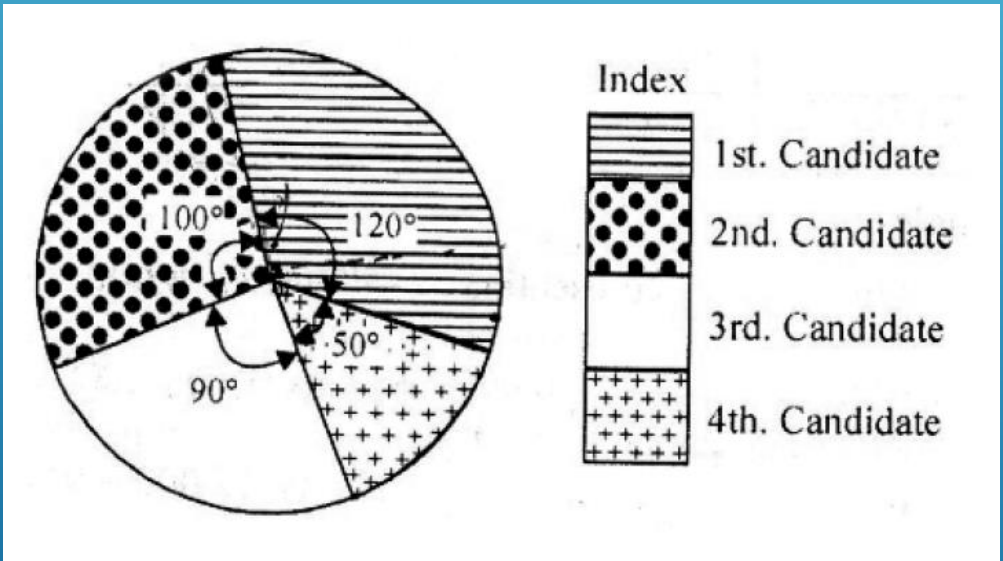
" " Rs. 8,000=9 x 8=72^0

" " Rs. 5,000=9 x 5=45^0



METHODS OF PRESENTING STATISTICAL INFORMATION

Example 7: In an election 72,000 votes were casted. Out of four candidates the first got 24,000 votes, the second got 20,000 votes, the third got 18,000 votes and the fourth got 10,000 votes. Draw a pie-chart for these data.



Advantages of Diagrams:

(i) Diagrammatic representation of statistical data makes it easy to understand the characteristics of such data.

(ii) A diagram or chart sometimes clarifies a complex problem and other reveals hidden facts which cannot be otherwise easily detected.

(iii) Many complex theories of economics can be easily explained with the help of diagrams.

Limitations of Diagrams:

(i) Diagrams do not show details. In a table one may show information covering a large number of items or on related topics by providing additional rows and columns. This is not possible in case of a diagram.

(ii) In a table statistical data can be shown exactly. But statistical data can be shown only approximately in a diagram.

Uses of Diagrams:

Since representation of statistical data by diagrams become appealing to the eyes and easily understandable as such diagrammatic representation of data has gained much importance in Economics, Business, and Commerce. Diagrams have become indispensible in various advertisements published by various business establishments. For busy traders diagrams are very important because they do not find time to study the characteristics of statistical data by other means. They can have a bird's eye view of the business situation by having a look on the diagrams.

Presentation of Data by Graphs : Frequency distributions can be described with the help of graphs. Graph is more distinct in comparison to diagram and it is comparatively easier to draw graph than diagram. Data of frequency distributions and time series data are generally represented by graphs. Generally diagrams are drawn to make comparative study. But diagrams do not depict the relationships among the variables to be studied. With the help of graphs the relationships among the variables can be studied. The importance of graph is more than that of diagram in statistical analysis.

In the practical field various types of graphs are used. These graphs can broadly be divided into two groups:

(a) Graphs of frequency distribution

(b) Graphs of time series

Graphs of frequency distribution: In order to make frequency distribution easily understandable three types of graphs are usually drawn. These are:

(i) Histogram or column diagram

(ii) Frequency Polygon

(iii) Cumulative frequency curve or Ogive.

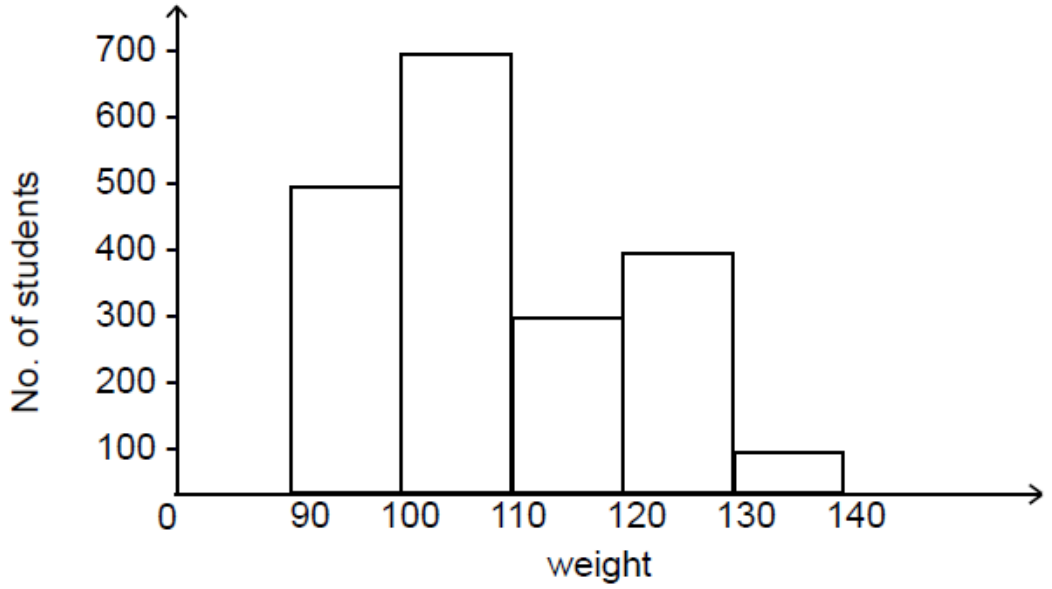
(i) Histogram or column diagram: The graph by which the frequencies of various class intervals of a frequency distribution with the help of adjacent vertical rectangles are shown is called a histogram or column diagram. First of all, the actual class-intervals are to be marked on the x-axis choosing a suitable scale. Then taking these as bases, rectangles are to be drawn continuously on this basis. When the class intervals are equal then the heights of the rectangles will be proportional to the frequencies of the corresponding class intervals, but if the class intervals are not equal then the heights of the rectangles will be proportional to the ratios between the corresponding class frequencies and the lengths of the class intervals.

Example 8: Draw a histogram to represent the following distribution:

(Frequency distribution of weights of 2000 students)

Weights (in lbs)	90-100	100-110	110-120	120-130	130-140
No. of students	500	700	300	400	100

* **Solution:** The class intervals of the above frequency distribution are of equal length. First of all, the class intervals are marked on the X-axis and then rectangles of heights proportional to the frequencies of the class intervals are drawn on them.



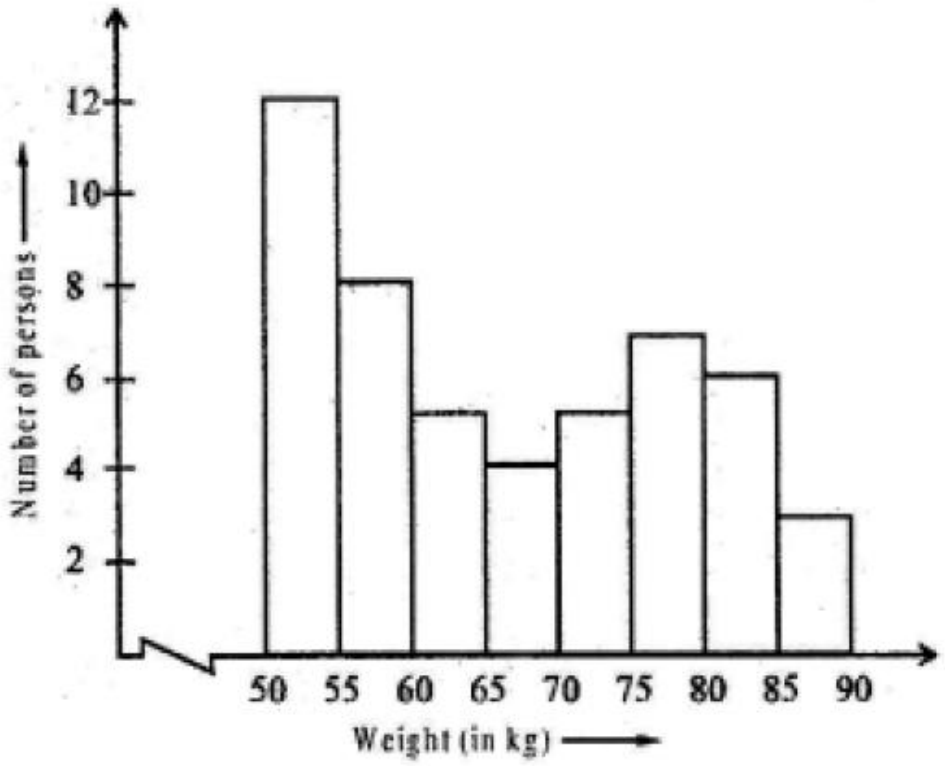
Example 9: The following is the distribution of weights (in kg) of 50 persons:

Weights (in kgs)	50-55	55-60	60-65	65-70	70-75	75-80	80-85	85-90
Number of persons	12	8	5	4	5	7	6	3

Draw a histogram for the above data.

* **Solution:** We represent the class limits along the X-axis on a suitable scale and the frequencies along the Y-axis on a suitable scale.

Since the scale on the X-axis starts at 50, a kink (break) is indicated near the origin to signify that the graph is drawn to scale beginning at 50, and not at the origin.



(ii) **Frequency Polygon:** The graph obtained by joining the mid-points of the upper horizontal sides of the adjacent rectangles of a histogram by line segments is called a frequency polygon. Frequency polygon can be drawn directly without obtaining indirectly from histogram. For this we are to take the mid-points of the class intervals on the X-axis and the frequencies of the class intervals are to be plotted against the corresponding class mid-points along the Y-axis. Then these plotted points are to be joined by line segments. Then the end points of the graph so obtained are to be joined with the mid-points of the two frequency less class intervals which preceed and succeed the first and the last class intervals respectively. They graph obtained thereby is the desired frequency polygon.

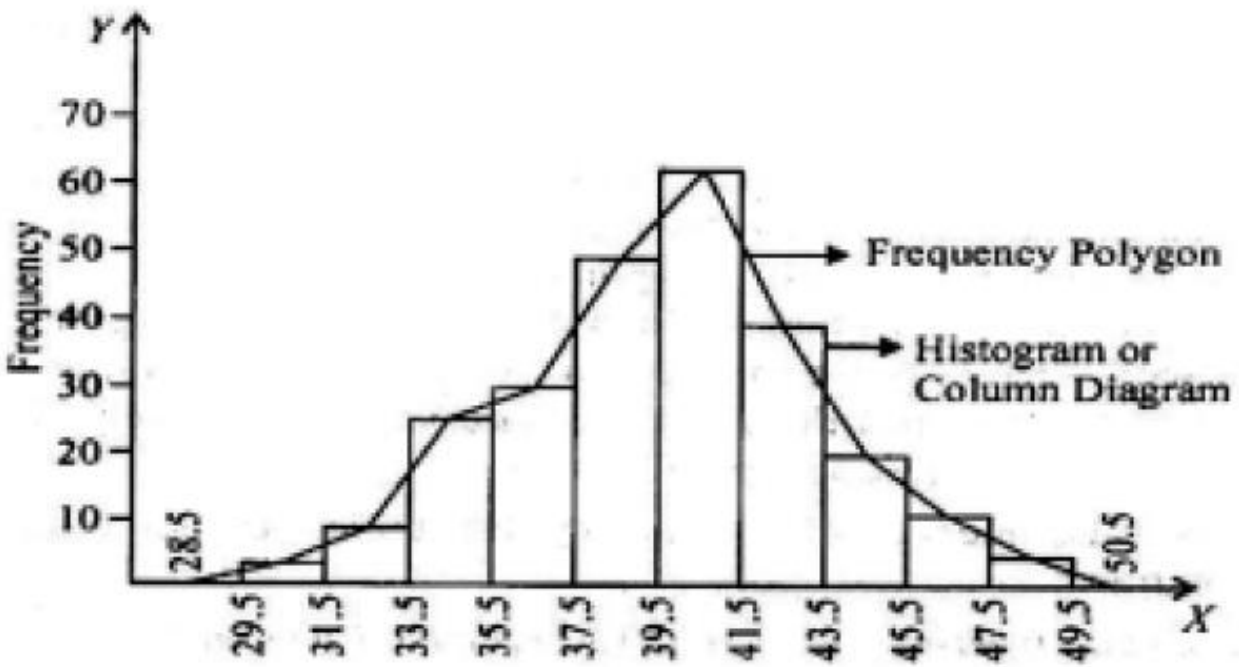
Note: In case the class intervals of a frequency distribution are not all equal then frequency polygon is usually not drawn.

Example 10: Draw (i) Histogram and (ii) Frequency Polygon for the following frequency distributions:

Weekly Wages (Rs.)	No of Workers
30-31	2
32-33	9
34-35	25
36-37	30
38-39	49
40-41	62
42-43	39
44-45	20
46-47	11
48-49	3

* **Solution:** The following table is to be prepared before drawing histogram and frequency polygon.

Class intervals	Real class limits	Mid-point	Frequency
30-31	29.5-31.5	30.5	2
32-33	31.5-33.5	32.5	9
34-35	33.5-35.5	34.5	25
36-37	35.5-37.7	36.5	30
38-39	37.5-39.5	38.5	49
40-41	39.5-41.5	40.5	62
42-43	41.5-43.5	42.5	39
44-45	43.5-45.5	44.5	20
46-47	45.5-47.5	46.5	11
48-49	47.5-49.5	48.5	3



Note: Here we have drawn the frequency polygon by joining the mid-pints of the upper horizontal sides of the rectangles by line segments. In order to draw the frequency polygon directly we are to take the mid-points of the class intervals on the X-axis and the frequencies of the corresponding class intervals on the Y-axis. Then we are to plot the points (30.5, 2), (32.5, 9), ...(48.5, 3) on graph paper and then we are to join these points by line segments. In both the methods, the end points of the graph so drawn are to be joined with the mid-points of the class intervals preceeding the first class and succeeding the last class. (The frequency of each of these two class intervals will be zero). The graph so obtained is the frequency polygon to be drawn.

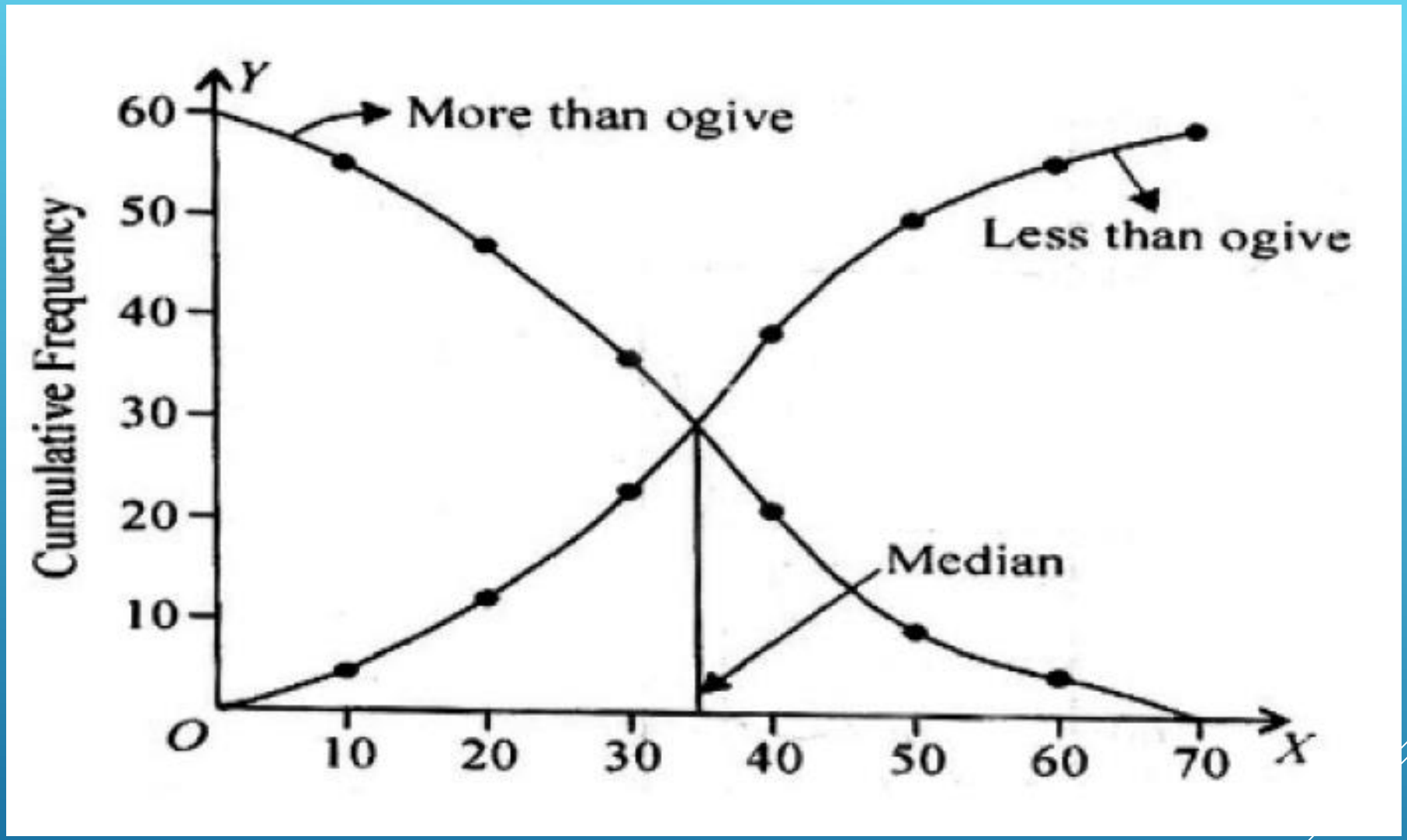
(iii) **Cumulative Frequency Curve or Ogive:** The graphical representation of a cumulative frequency distribution is known as the Cumulative frequency curve or ogive. First of all, from the given raw data or from the given frequency distribution, a cumulative frequency table is constructed. Then the upper class limits (or the upper class boundaries) of the class intervals are marked on the X-axis and the points are plotted by showing the corresponding cumulative frequencies on the Y-axis. Then these points are connected by drawing smooth curve with free hand. This curve is the cumulative frequency curve or the ogive. Depending upon the pattern of drawing the graph we get two types of ogives namely, less-than ogive and more-than ogive. The methods of drawing less-than ogive and more-than ogive are explained below. With the help of ogive one can determine median, quartiles, deciles etc.

Example 11: Draw (i) less than ogive and (ii) more than ogive for the following frequency distributions:

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of students	4	8	11	15	12	6	3

* ***Solution:*** The class intervals of the given frequency distribution are as per the exclusive method. Hence the lower and the upper limits of that class.

Intervals	Frequency	Cumulative frequency	
		Less-than	More-than
0-10	4	4	59
10-20	8	12	55
20-30	11	23	47
30-40	15	38	36
40-50	12	50	21
50-60	6	56	9
60-70	3	59	3


Y
60
50
40
30
20
10
Cumulative Frequency
More than ogive
Less than ogive
Median
O
10
20
30
40
50
60
70
X

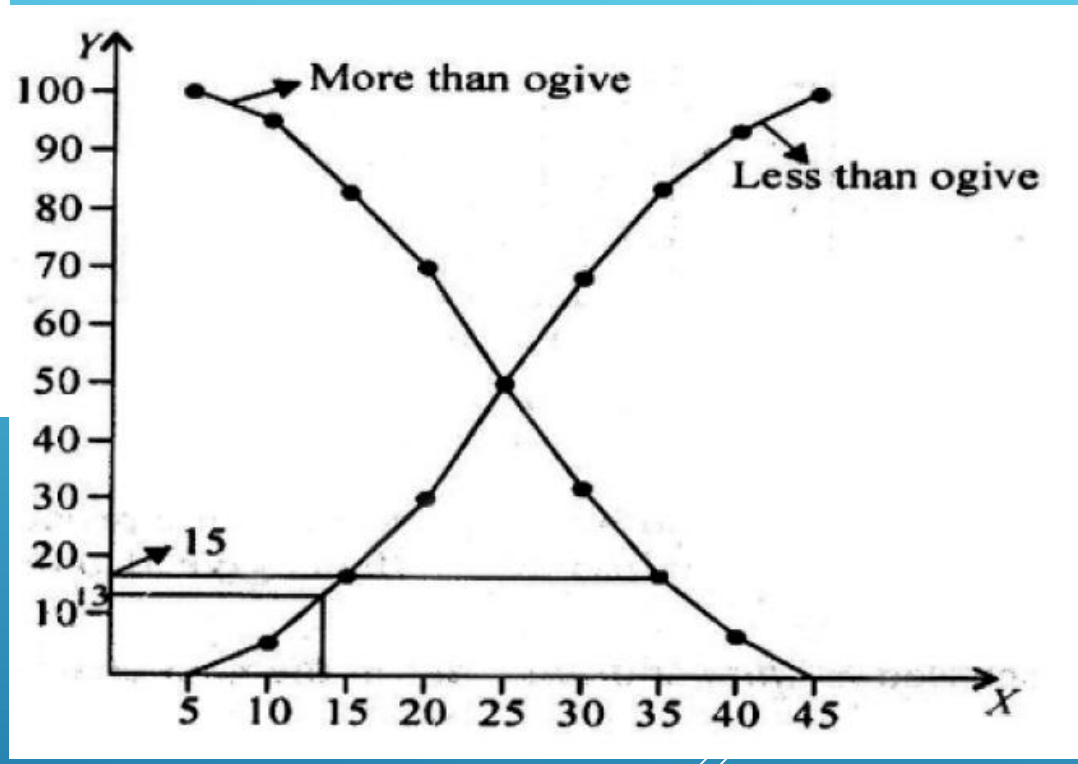
Example 12: The following data relate to sales of 100 companies. Draw the two ogives. Determine the number of companies whose sales are (i) les than Rs, 13 lakhs, (ii) more than Rs. 36 lakhs.

Sales (Rs. in lakhs)	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45
No. of Companies	5	12	13	20	18	15	10	7

Class Intervals	Frequency	Cumulative frequency	
		Less-than	More-than
5-10	5	5	100
10-15	12	17	95
15-20	13	30	83
20-25	20	50	70
25-30	18	68	50
30-35	15	83	32
35-40	10	93	17
40-45	7	100	7

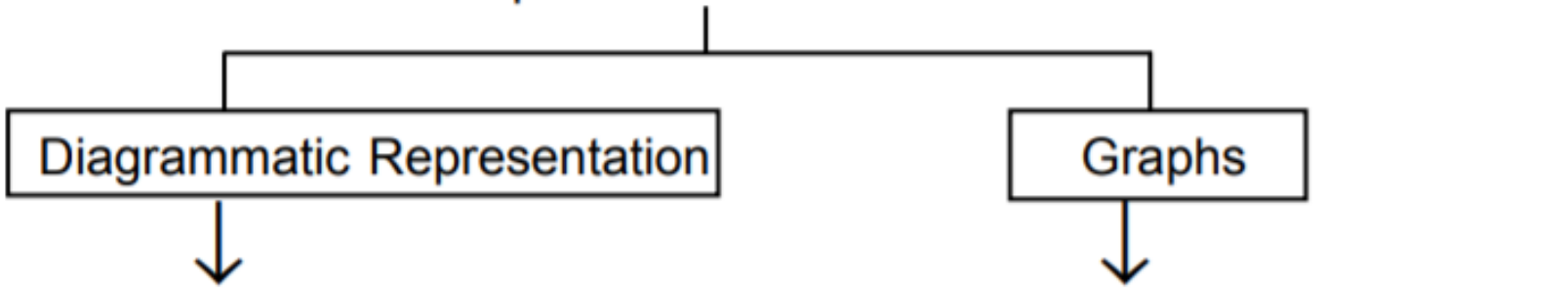
In case of less-than ogive the cumulative frequencies are plotted on a graph paper against the upper limits of the corresponding classes and in case of More than ogive the cumulative frequencies are plotted on a graph paper against the lower limits of the corresponding classes.

Taking the length of each small square on the X-axis as one unit and the length of each small square on the Y-axis as 5 units, the two ogives (Less-than and More than) have been drawn. It has been found that (i) the number of companies with sale proceeds more than 13 lakhs = 13, and (ii) the number of companies with sale proceeds more than 36 lakhs = 15.



SUMMARY

- Statistics is the numerical representation of data.
- Statistics can be used in economics, insurance companies, business, banks etc.
- Frequency distribution is a tabular presentation of data.
- Data can be classified on geographical, chronological, qualitative and quantitative factors.
- Statistical data representation



(a) Bar Diagram

(b) Pie Diagram

(a) Histogram

(b) Frequency Polygon

(c) Cumulative Frequency Curve