UNIT 1: INTRODUCTION TO STATISTICS

UNIT STRUCTURE

1.1 Learning Objectives

1.2 Introduction

1.3 Meaning and Definition of Statistics

1.4 Frequency Distribution

- 1.4.1 Classification of Data
- 1.4.2 Types of Classification

1.5 Tabulation of Data

- 1.5.1 Advantages or Importance of Tabulation
- 1.5.2 Different Parts of Table
- 1.5.3 Types of Tables
- 1.5.4 Calculation of Class Frequencies and Formation of Frequency Distribution Table

1.6 Methods of Presenting Statistical Information

1.7 Let Us Sum Up

1.8 Further Readings

1.9 Answer To Check Your Progress

1.10 Model Questions

1.1 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.

1.2 INTRODUCTION

In this unit we will discuss the importance of statistics in business. It has been ascertained that the word 'statistics' has been derived from the Latin word 'status' or the Italian word 'statista' or the German word 'statistic'. Each of these words means 'a political state'. In old days, the word 'statistics' was used to indicate the quantities like number of soldiers in a state, volume of arms and ammunition, volume of taxes, etc. connected with administrative and political activities of a state. In those days statistics was also called 'The Science of Kings'. But in modern age the scope of statistics is wide enough. Now- a- days the use of statistics is extended to various fields such as agriculture, economics, sociology, psychology, business, management, etc. Below we discuss certain definitions of statistics.

1.3 MEANING AND DEFINITION OF STATISTICS

The word 'statistics' has been defined differently by different authors in different times. This word conveys two different meanings: one when defined in plural sense and the other when defined in singular sense. In plural sense it refers to numerical statements of facts or data. In singular sense it refers to the subject statistics like other subjects such as mathematics, economics, physics, etc. For example, the data relating to the volume of production in different years, the volume of exports and imports in different years, the price of a commodity at different places, etc., refer to statistics used in plural sense. The word 'statistics' when used in singular sense implies statistical methods or the science of statistics.

Defining Data: Since the complexity of business environment makes the process of decision-making difficult thus for decision- makers it is important to develop the ability to extract meaningful information from raw data in order to make better decisions. Decisions have to be based upon data which show relationship between pair of variables, indicate trends and show rates of change in various relevant variables. Thus data can be defined as observations expressed in numerical figures obtained by

measuring or counting is called data.

Below we mention some of the definitions of statistics as a subject.

1. "Statistics is the science which deals with the methods of collecting, classifying, presenting, comparing and interpreting numerical data collected to throw some light on any sphere of enquiry." ..... Selligman
2. "Statistics may be regarded as a body of methods for making wise decisions in the face of uncertainty." ..... Wallis and Roberts.
3. "Statistics may be defined as the science of collection, presentation, analysis and interpretation of numerical data." ..... Croxton and Cowden.

Useful of statistics in Business: Statistics is an aid to commerce and industry. It helps the businessmen and industrialists to decide things more precisely and objectively.

Success of a business depends on accurate forecasts of sales. That is, a producer or dealer should first estimate the demand for his products, analysis the possible effects of factors like changes in tastes, fashions etc. This is called business forecasting. Statistical methods are very helpful in business forecast. For instance, a forecast of sales can be made by studying time series of present prospective conditions of the conditions of the concern and related business.

Statistics is also useful in conducting market research by the businessmen. A skillful analysis of data on population, purchasing power, customs of people, competition, transportation costs etc., should be made before launching a new product. Again in building up and maintaining an extensive market, it is important to keep accurate records of its present geographical distribution.

Again cost accounting is entirely statistical in outlook and it is with the help of this technique that producers are in a position to decide about the prices of various commodities.

Statistical quality control is a standard procedure for improving the quality of manufactured goods. The number of defective articles produced by a machine is recorded continuously on a chart, and the statistician,

looking at the chart, can tell when the production is becoming worse, and thus he maintains the quality of the manufactured articles at a uniformly high level.

Statistics is also very helpful in business management. With the growing size of business enterprises centralized management of all the activities of an industrial concern from production to sales has become impossible. Separate business departments are to be created on the basis of function they serve such as planning, material control, purchasing, sales promotion, accounting, personnel, etc. For the proper execution of these functions, the management has to depend upon statistical data and accounting records

1.4 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).

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.

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.

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.

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.

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	

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.

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.***

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.

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.



CHECK YOUR PROGRESS

Q 1: What is frequency distribution?

.....

.....

Q 2: Identity the discrete and continuous frequency distribution.

(i)	x :	5 – 10	11 – 16	17 – 22	23 – 28
	f :	5	2	8	9
(ii)	x :	2 – 8	8 – 14	14 – 20	20 – 26
	f :	4	7	9	10

1.4.1 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 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 or 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.

1.4.2 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

(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.

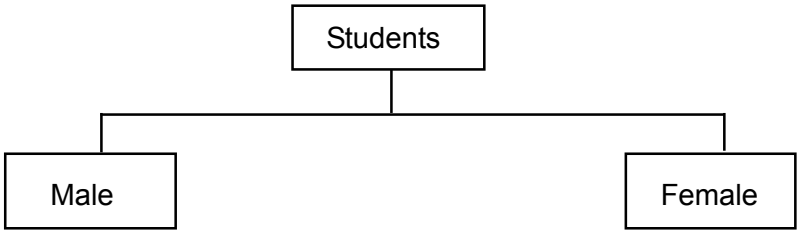
Example of chronological classification:

The following is an example of chronological classification

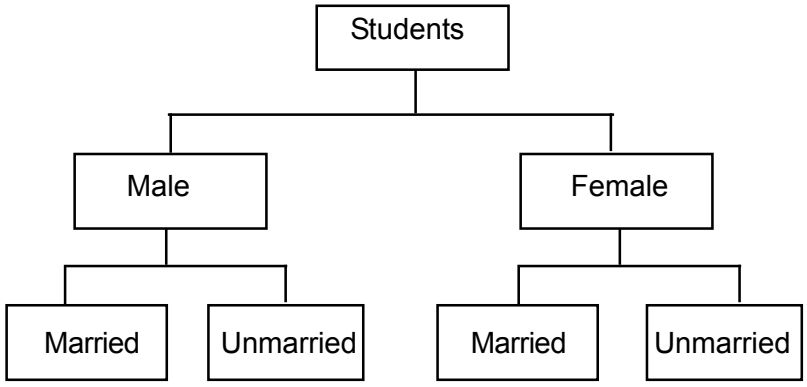
Year :	1941	1951	1961	1971	1981	1991	2001
Population : (crore)	31.9	36.9	43.9	54.7	75.6	85.9	98.6

(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.

1.5 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".

1.5.1 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.

1.5.2 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.

1.5.3 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

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

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	40	5	45	25	15	40
No. of people without tea habit	20	35	55	30	30	60
Total	60	40	100	55	45	100

Explanation: Let us first consider town A. Since we have assumed total number of people to be 100, hence 40% female will imply 40 females, i.e., there are 40 females in town A. Therefore, the number of males in town A = 100 - 40 = 60. Again, the total number of people (including male and female) in town A with tea habit is 45% i.e., 45. Hence the number of people without tea habit is 20%, i.e., 20. Hence the number of females without tea habit = 55 - 20 = 35. Since the total number of males is 60 and the number of males without tea habit is 20 hence the number of males with tea habit = 60-20 = 40. Similarly we find the data relating to town B as shown in the table.

1.5.4 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

Solution: [The above data are called ungrouped or unclassified or raw data. From these data it is not easy to answer such questions as (i) how many students have passed, (ii) how many students have got marks between 70 - 79, (iii) how many students have got letter marks, etc.]

We observe from the given data that the maximum mark is 98 and the minimum mark is 34. For convenience we insert these numbers within two small circles. Now we form the class intervals as 30 - 39, 40 - 49, 50 - 59, ... keeping in mind that the number of class intervals should not be less than 5 and more than 15. (We may also form the class intervals as 30 - 40, 40 - 50, 50 - 60, ... When all the figures are integers we may form the class intervals either by inclusive method or by exclusive method). In case there are some fractional numbers then of course, we must form the class intervals by exclusive method. Now, following the steps in forming a frequency distribution as mentioned in 5.5.4, we get the following frequency distribution table.

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

From the above table of frequency distribution we can very easily answer the following type of questions:

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

[Ans. 3]

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

[Ans. 32]

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

[Ans. 11+20+32=53]

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

[Ans. 98]

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

[Ans.16]

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

[Ans. 32%]



CHECK YOUR PROGRESS

Q 3: State whether the following statements are true or false:

(i) An attribute can be measured.

(ii) A discrete variable can take any fractional value.

(iii) A continuous variable may assume any value between two specified limits.

(iv) The smallest possible measurement in a class is called the lower limit of that class.

(v) Class limits and class boundaries of an exclusive class are the same.

(vi) Class limits and class boundaries of an inclusive class are the same.

(vii) An open-end distribution creates problems in the calculation of mid-points.

(viii) The mid-point of a class interval is equal to the lower limit of that class plus half the difference between the lower limit and the upper limit of that class.

(ix) A discrete variable is the result of counting.

(x) A continuous variable is the result of counting.

1.6 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)

We shall limit our discussions to various types of bar diagrams as

well as pie diagram only as these are the most commonly used diagrams.

(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

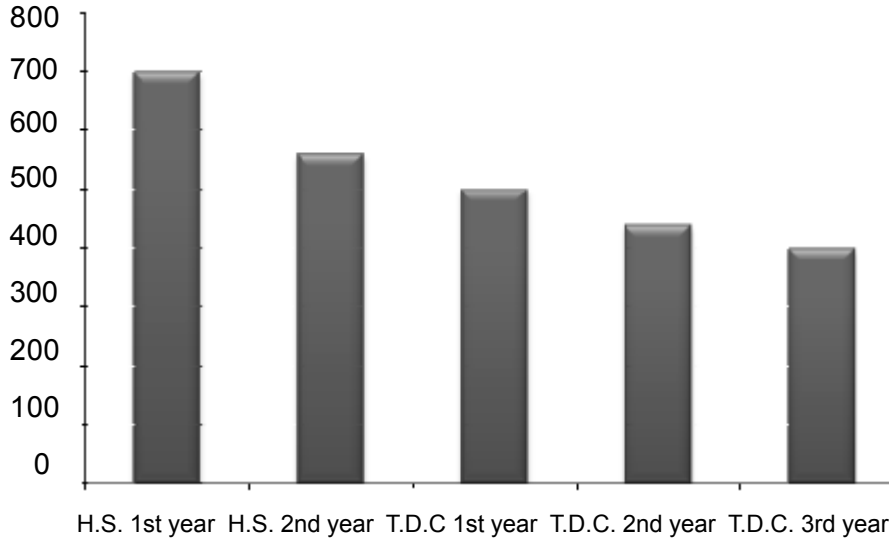
(ii) Multiple Bar Diagram

(iii) Sub-Divided or Compound Bar Diagram

(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.

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:

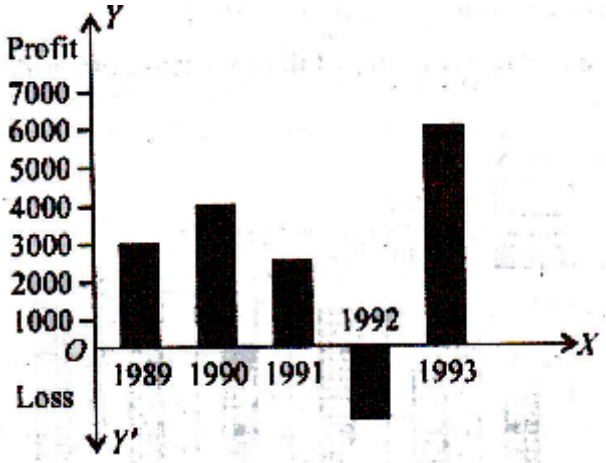


Example 2: The profits and losses of a business concern for the years 1989-1903 are given below:

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.



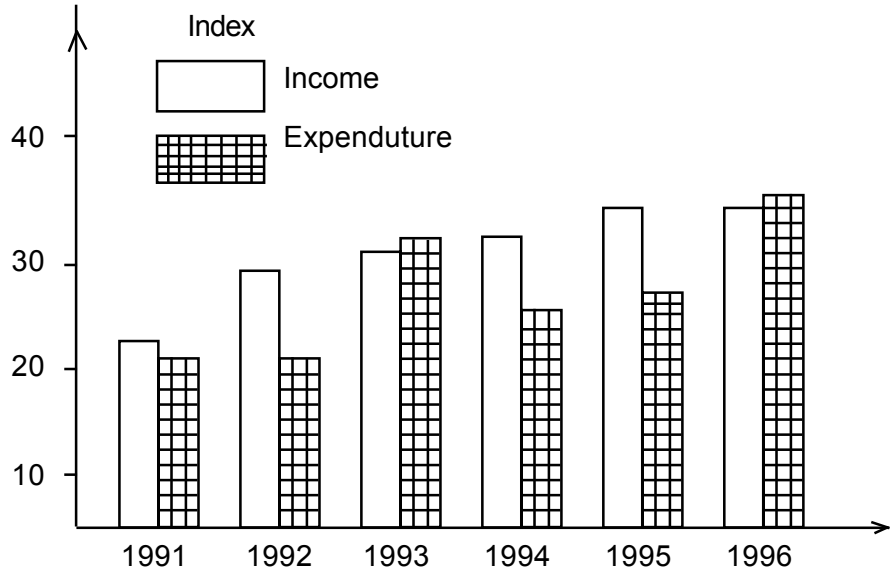
(ii) Multiple Bar Diagram: With the help of multiple bar diagram (multiple bar chart) more than one type of data can be depicted at a time side by side to represent 2, 3 or 4 series of values for comparison. For example, in order to show the profits and losses of a business concern for a few years, a 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.

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

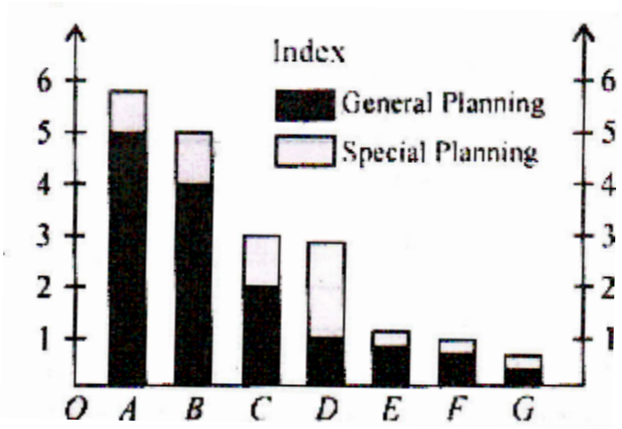
Y – axis = Rs. 2000

(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.

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,

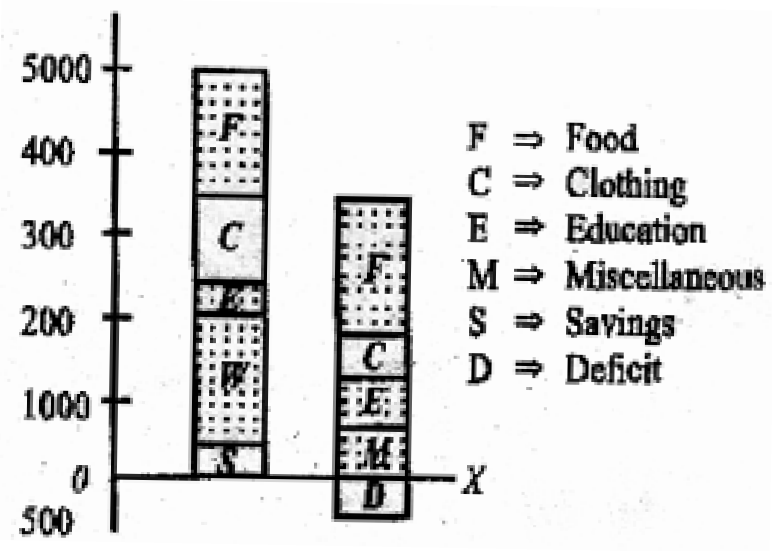
F : Cooperatives,

G : Miscellaneous.

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:



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.

Pie diagram is very useful in drawing comparison among the various components or between a part and the whole.

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.

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

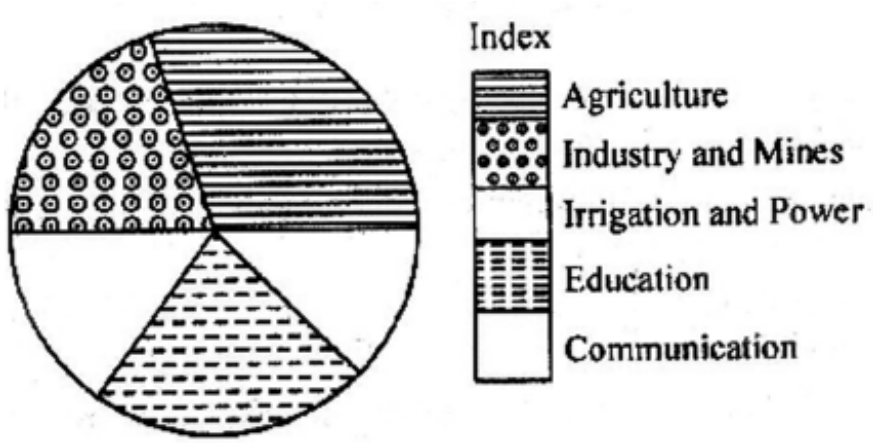
$\therefore$ " " Rs. 12,000=9 x12=108^0

" " 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



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.

*** Solution:** Let, the total of 72,000 votes be represented by 360^0.

The angle corresponding to 1 vote = $\left(\frac{360^0}{72{,}000}\right)=\left(\frac{1^0}{200}\right)$

The angle corresponding to 24,000 votes got by the

1st candidate = $\left(\frac{1}{200}\times 24{,}000\right)$ = 120^0

The angle corresponding to 20,000 votes got by the

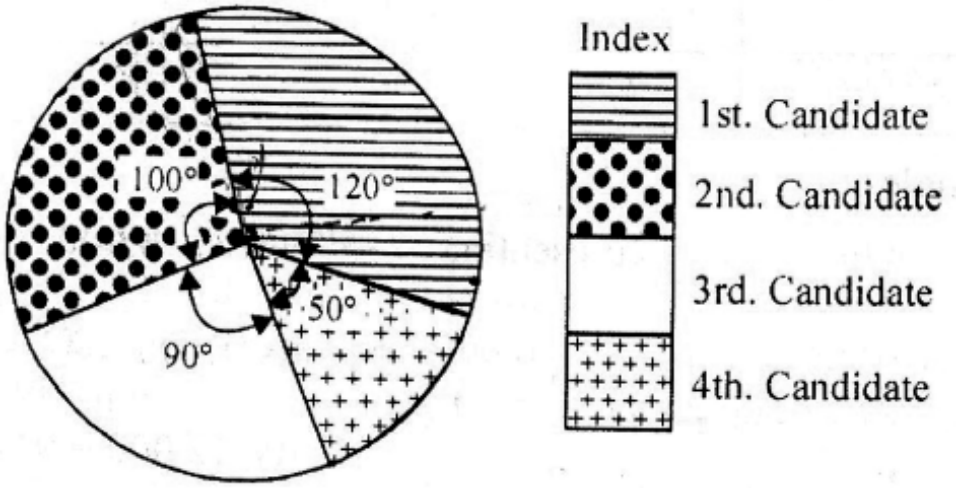
2nd candidate = $\left(\frac{1}{200}\times 20{,}000\right)$ = 100^0

The angle corresponding to 18,000 votes got by the

3rd candidate = $\left(\frac{1}{200}\text{x}18{,}000\right)$ = 90^0

The angle corresponding to 10,000 votes got by the

4th candidate = $\left(\frac{1}{200}\text{x}10{,}000\right)$ = 50^0.



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

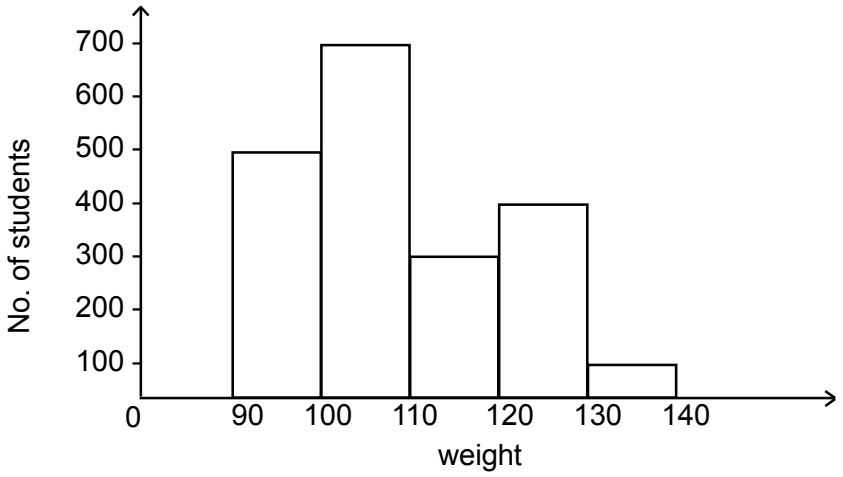
(iii) Cumulative frequency curve or Ogive.

(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.



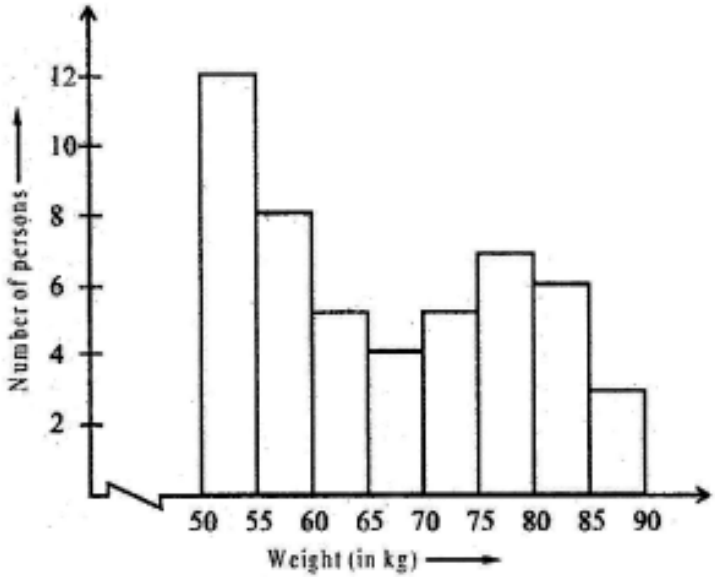
Example 9: The following is the distribution of weights (in kg) of 50 persons:

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.

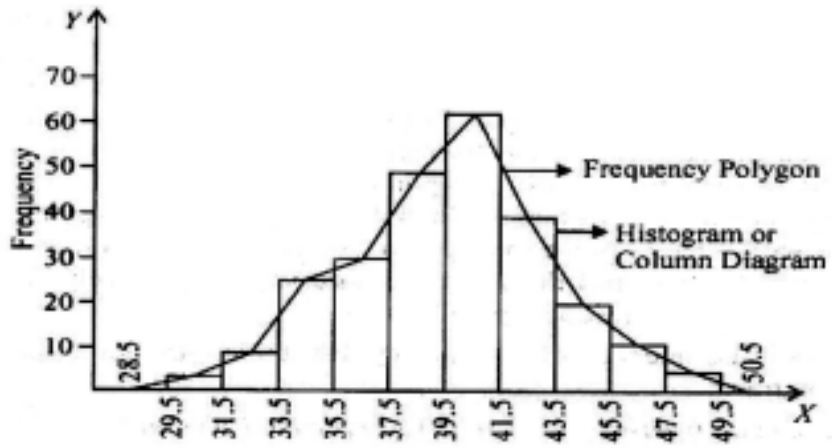
Note: In case the class intervals of a frequency distribution are not all equal then frequency polygon is usually not drawn.

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:

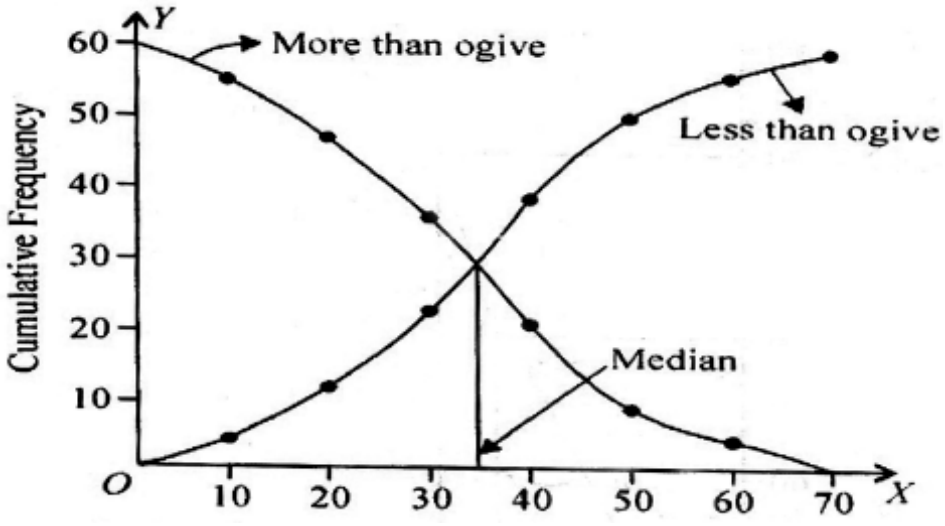
Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of students	4	8	11	15	12	6	3

**** 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

From less-than cumulative frequency column we find that the cumulative frequency of a particular class implies the number of observations included upto the upper limit of that class. For example, the cumulative frequency of the class 30-40 is 38. This means that there are 38 observations less than 40. Again, from the more than cumulative frequency column we find that the cumulative frequency of a class implies the number of observations of the given distribution greater than or equal to the lower limit of that class interval. Thus, the cumulative frequency 59 of the first class (shown by 0-10) tells us the fact that the number of observations of the given distribution greater than or equal to 0 is 59. In case of less than ogive, upper class limits of the class intervals are marked on the X-axis and then points are plotted by taking the corresponding cumulative class frequencies on the Y-axis. The points are joined by a smooth curve with free hand. Then the lower end point of the curve is joined with the lower limit of the first class (class interval). By looking at the

graph of the more-than ogive drawn below students can easily guess the procedure of drawing a more than ogive.



In order to get the median of the distribution, a perpendicular is to be drawn on the X-axis from the point of intersection of the less-than ogive and the more-than ogive. The distance between the origin and the foot of the perpendicular upon the X-axis is the median. In case of the given frequency distribution we get the median to be approximately equal to 35.6 marks.

Note: The number of observations less than 0 is zero. Hence the lower end point of the less-than ogive is at the origin. Again, the number of observations greater than 70 is also zero. Hence the lower end point of the more-than ogive is at the point corresponding to 70 on the X-axis.

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.

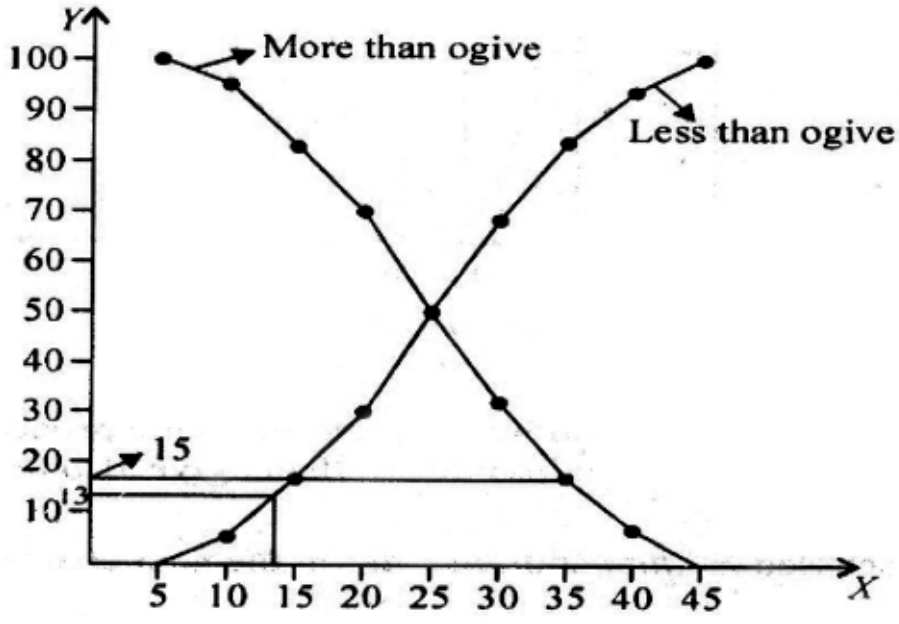
Sales (Rs. in lakhs)	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45
No. of Companies	5	12	13	20	18	15	10	7

*** Solution:** The class intervals of the given frequency is as per the exclusive method of classification and hence the lower and the upper limits of the class intervals may be taken as the actual lower and upper limits.

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.





CHECK YOUR PROGRESS

Q 4: What will be the most suitable diagrams to represent the following sets of data?

(i) Plan outlays for Assam in different five year plans;

(ii) Birth rates and death rates in different states of the North Eastern region according to 2001 census

(iii) Sectoral composition of India's GDP in a given year.

Q 5: Fill in the blanks with suitable word/words:

(i) Frequency polygon can be drawn with the help of _____.

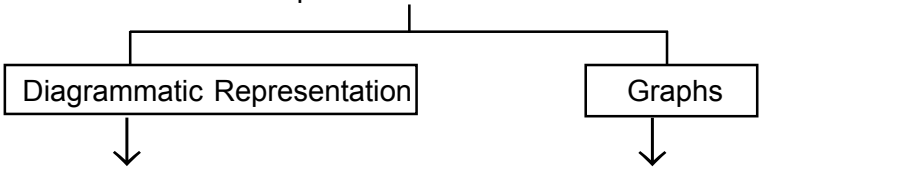
(ii) The ogives of less than type and more than type distributions intersect at_____.

(iii) Pie-chart is always_____.



1.7 LET US SUM UP

- 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



1.8 FURTHER READINGS

1) Business Mathematics, Qazi Zameeruddin, V K Khanna, and S K Bhambri, Vikas Publishing House Pvt Ltd, Second Reprint, 1995.
2) Agarwal, D. R. (2006). Business Statistics. Delhi, Vrinda Publications.
3) Gupta S.C. (1994). Fundamentals of statistics. New Delhi, Himalayan Publishing House.
4) Rajagopalan, S. P. & Sattanthan R. (2009).Business Statistics and Operations Research. New Delhi, Tata McGraw-Hill
5) Sharma, J. K. (2007). Business Statistics. New Delhi. Pearson Education Ltd.
6) Verma, A. P. (2007). Business Statistics. Guwahat, Asian Books Private Limited.



1.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Frequency distribution is a tabular presentation of data into classes in terms of class intervals.

Ans to Q No 2: (i) Discrete (ii) Continuous.

Ans to Q No 3: (i) F, (ii) F, (iii) T, (iv) T, (v) T, (vi) F, (vii) T, (viii) T, (ix) T, (x) F.]

Ans to Q No 4: (i) Simple bar diagram, (ii) Multiple bar diagram, (iii) Pie diagram,

Ans to Q No 5: (i) Histogram, (ii) Median, (iii) Circular.



1.10 MODEL QUESTIONS

Q 1: Define frequency distribution. Explain the construction of a frequency distribution.

Q 2: Marks obtained by 50 students are given below. Prepare a frequency distribution table by taking class intervals each of length 10 by inclusive or exclusive method:

34	54	10	21	51	52	12	43	48	36
48	22	39	26	34	19	10	17	47	28
13	30	30	60	15	59	07	18	40	49
40	51	55	32	41	22	30	35	53	25
14	18	19	40	43	04	17	45	25	43

Q 3: Age at death of 50 persons of a town are given below:

36	48	50	45	49	31	50	48	43	42
37	32	40	39	41	47	45	39	43	47
38	39	37	40	32	52	56	31	54	36
51	46	41	55	58	31	42	53	32	44
53	36	60	59	41	53	58	36	38	60

(a) Arrange the data in frequency distribution having 10 class intervals.

(b) Obtain the percentage frequency in each class interval.

Q 4: Construct a frequency distribution table for the following data. Answer the following questions from the table:

(i) P.C. of students securing more than 30 marks,

(ii) P.C. of students securing less than 30 marks,

Marks secured: 5, 27, 23, 32, 49, 19, 50, 38, 37, 25, 30, 29, 42, 37, 18, 12, 13, 9, 18, 27, 50, 32, 47, 48, 29, 30, 32, 37, 27, 41.

Hint: Here range = 50 - 5 = 45. Length of class intervals = = 4.5 i.e., 5. The class intervals are: 5 - 10, 10 - 15,……50 - 55. No. of students getting marks more than 30 = (5+4+2+3+2) - 2 = 14. [Ans. (i) 46.67%, (ii) 46.67%]

Q 5: What do you mean by classification of data? What are the objectives of classification?

Hint: For objectives of classification see advantages of classification. You are to write down the objectives as: (i) to present the raw-data in a concise and simple form, (ii) to facilitate comparison by dividing the raw-data on the basis of their similarities and resemblances, etc.

Q 6: Discuss the different types of classification.

Q 7: Distinguish between classification according to attributes and classification according to variables.

Q 8: Discuss the importance of classification in statistical analysis.

Q 9: What is tabulation? What are its objectives?

Hint: For objectives see advantages of tabulation. You are to write down the objectives as:

(i) To simplify complicated data, (ii) To present quantitative data in a concise and condensed form, (iii) To facilitate comparison of data, (iv) To provide basis for analysis and interpretation of data.

Q 10: In a sample study about the tea habits in two villages the following data were obtained:

Village A: 70% persons were males. 80% were tea drinks and 62% were male tea drinkers.

Village B. 55% persons were males. 35% were tea drinkers and 52% were male tea drinkers. Tabulate the above supplying the missing link.

Q 11: Prepare a table from the data given below:

In 1987, out of a total of 1,750 workers of a factory 1200 were members of a trade union. The number of women employees was 200, of which 175 did not belong to the trade union.

In 1988, the number of union members increased to 1580, of which 1290 were men.

On the other hand, the number of non-union workers fell down to 208, of which 180 were men.

Q 12: What are the different types of diagrams commonly used for the presentation of statistical data?

Q 13: What is graphical representation of data? What are the various kinds of such graphs?

Q 14: Write short notes on (a) Histogram, (b) Frequency polygon, (c) Pie-diagram.

Q 15: The profits and losses of a business concern (Rs. in lakhs) are given below:

Year	1999	2000	2001	2002	2003
Profit (Rs. in lakhs)	36	-	22	38	-
Loss (Rs. in lakhs)	-	10	-	-	15

Represent the above data by a bar-chart.

Q 16: The following table shows total cost (in Rs.) and its component parts in two consecutive years:

	1990	1991
Direct material	4,000	5,500
Direct labour	5,000	6,000
Direct expenses	1,200	1,500
Overheads	2,500	2,000

Represent the data by a suitable diagram.

Hint: The data may be presented by a sub-divided bar-diagram.

Q 17: Represent the following data by means of a pie-diagram. The data relate to the expenditures of a state during the Second Five Year Plan (Rs. in crores) on different heads:

Industry	127.00
Irrigation	92.50
Agriculture	100.0
Transport and communication	92.50
Others	480.00

Q 18: A table of weekly wages of some tea-garden workers is given below. Draw a histogram for these data.

Salary (Rs.)	200-210	210-220	220-230	230-240	240-250
No. of employees	8	15	21	35	62

*** ***** ***