SECTION - B

FUNDAMENTALS OF BUSINESS STATISTICS

STATISTICAL REPRESENTATION OF DATA 4

This Module includes:

4.1 Diagrammatic Representation of Data

4.2 Frequency Distribution

4.3 Graphical representation of Frequency Distribution – Histogram, Frequency Polygon Curve, Ogive, Pie-chart

STATISTICAL REPRESENTATION OF DATA

Module Learning Objectives:

After studying this Module, the students will be able to understand –

- Meaning and definitions of Statistics
- Statistical Methods
- Representation of Information through Data
- Different analysis of Data

Introduction

4

The word 'Statistics' has been derived from the Latin word 'Status' which means a political state. It has also its root either to the Italian word 'Statista' or the German word 'Statistik' each one of which means a political state. For several decades, the word 'statistics' was associated solely with the display of facts and figures pertaining to the economic, demographic and political situations prevailing in a country,usually, collected and brought out by the local governments.

Statistics is a tool in the hands of mankind to translate complex facts into simple and understandable statements of facts.

Meaning and definition of Statistics:

Meaning of statistics: The word Statistics is used in two different senses - Plural and singular. In its plural form, it refers to the numerical data collected in a systematic manner with some definite aim or object in view such as the number of persons suffering from malaria in different colonies of Delhi or number of unemployed girls in different states of India and so on. In Singular form, the word statistics means the science of statistics that deals with the principles, devices or statistical methods of collecting, analyzing and interpreting numerical data.

Thus, 'statistics' when used in singular form refers to that branch of knowledge which implies Applied Mathematics. The science of statistics is an old science and it has developed through ages. This science has been defined in different ways by different authors and even the same author has defined it in different ways on different occasions.

It is impossible to enumerate all the definitions given to statistics both as "Numerical Data i.e., Plural Form: and "Statistical Methods, i.e., Singular Form". However, we have given below some selected definitions of both the forms.

Definitions of "Statistics in Plural Form or Numerical Data": Different authors have given different definitions of statistics. Some of the definitions of statistics describing it quantitatively or in plural form are:

"Statistics are the classified facts representing the conditions of the people in a state especially those facts which can be stated in number or in a table of numbers or in any tabular or classified arrangement.

This definition is narrow as it is confined only to the collection of data regarding the people in a state. But the following definition given by Secrist is modern and convincing. It also brings out the major characteristics of statistical data.

"By Statistics we mean the aggregate of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standards of accuracy collected in a systematic manner for a pre-determined purpose and placed in relation to each other"

This definition makes it clear that statistics (in plural form or numerical data) should possess the following characteristics.

I. Statistics means aggregate of facts

II. Statistics is affected by a large number of causes

III. Statistics is always numerically expressed

IV. Statistics should be enumerated or estimated according to reasonable standards of acuracy

V. Statistics should be collected in a systematic manner

VI. Statistics should be collected for a pre-determined purpose

VII. Statistics should be placed in relation to each other.

Statistics as Statistical methods or Statistics in Singular Sense:

We give below the definitions of statistics used in singular sense, i.e., statistics as statistical methods.

Statistical methods provide a set of tools which can be profitably used by different sciences in the manner they deem fit. The term statistics in this context has been defined differently by different authors.

A few definitions are given below:

"Statistics may be called the science of counting"

This definition covers only one aspect, i.e., counting, but the other aspects such as classification, tabulation, etc., have been ignored. As such, the definition is inadequate and incomplete

"Statistics may be defined as the collection, presentation, analysis and interpretation of numerical data"

This definition given by Croxton and Cowden is simple, clear and concise.

According to this definition, there are four stages – collection of data, presentation of data, analysis of data, and interpretation of data. However, one more stage may be added and that is the organization of data. Thus, there are four stages:

1. **Collection and Organization of data**: There are various methods for collecting the data such as census, sampling, primary and secondary data etc.

2. **Presentation of data**: The mass data collected should be presented in a suitable, concise form as the mass data collected is difficult to understand and analyse

3. **Analysis of data**: The mass data collected should be presented in a suitable, concise form for further analysis. Analysis includes condensation, summarisation conclusion, etc., by means of measures of central tendencies, dispersion, skewness, kurtosis, correlation, regression, etc.

4. **Interpretation of data**: The last step is drawing conclusions from the data collected as the figures do not speak for themselves.

   Having briefly discussed some of the definitions of the term statistics and having seen their drawbacks we are now in a position to give a simple and complete definition of the 'Statistics' in the following words:

Statistics (as used in the sense of data) are numerical statements of facts capable of analysis and interpretation and the science of statistics is a study of the principles and methods used in the collection, presentation, analysis and interpretation of numerical data in any sphere of enquiry.

Importance and Scope of Statistics:

I. Statistics and Economics: According to Prof. Alfred Marshall, "Statistics are the straws out of which I like every other economist, have to make bricks." The following are some of the fields of economics where statistics is extensively used.

(a) **Consumption**: Statistical data of consumption enable us to find out the ways in which people in different strata of society spend their incomes.

(b) **Production**: The statistics of production describe the total productivity in the country. This enables us to compare ourselves with other countries of the world.

(c) **Exchange**: In the field of exchange, an economist studies markets, laws of prices which are determined by the forces of demand and supply, cost of production, monopoly, competition, banking etc. A systematic study of all these can be made only with the help of statistics

(d) **Econometrics**: With the help of econometrics, economics has become exact science. Econometrics is the combination of economics, mathematics and statistics.

(e) **Public Finance:** Public finance studies the revenue and expenditure activities of a country. Budget, (a statistical document), fiscal policy, deficit financing, etc., are the concepts of economics which are based on statistics.

(f) **Input-Output Analysis**: The input-output analysis is based on statistical data which explain the relationship between the input and the output. Sampling, Time series, Index numbers, Probability, Correlation and Regression are some other concepts which are used in economic analysis.

II. Statistics and Commerce:

Statistical methods are widely applied in the solution of most of the business and trade activities such as production, financial analysis, costing, manpower, planning, business, market research, distribution and forecasting etc. A shrewd businessman always makes a proper and scientific analysis of the past records in order to predict the future course of the business conditions. Index numbers help in predicting the future course of business and economic events. Statistics or statistical methods help the business establishments in analysing the business activities such as:

(a) **Organization of Business**: Any businessman makes extensive use of statistical data to arrive at the conclusion which guides him in establishing a new firm or business house

(b) **Production**: The production department of an organisation prepares the forecast regarding the production of the commodity with the help of statistical tools.

(c) **Scientific Management and Business Forecasting**: Better and efficient control of a business can be achieved by scientific management with the help of statistical data. "The success of businessman lies on the accuracy of forecast made". The successful businessman is one whose estimates most closely approach accuracy," said Prof. Boddington.

(d) **Purchase:** The price statistics of different markets help the businessman in arriving at the correct decision. Raw material is purchased from those markets only where the prices are low.

III. Statistics and 'Auditing and Accounting': Statistics is widely used in accounting and auditing.

IV. Statistics and Economic Planning: According to Prof. Dickinson, "Economic Planning is making of major decisions – what and how much is to be produced, and to whom it is to be allocated – by the conscious

decisions of a determinate authority on the basis of a comprehensive survey of economy as a whole. "The various documents accompanying preceding and following each of the eight Five Year Plans of India are a standing testimony to the fact that statistics is an indispensable tool in economic planning.

V. **Statistics and Astronomy**: Statistics were first collected by astronomers for the study of the movement of stars and planets. As there are a few things which are common between physical sciences, and statistical methods, astronomers apply statistical methods to go deep in their study. Astronomers generally take a large number of measurements and in most cases there is some difference between several observations. In order to have the best possible measurement they have to make use of the technique of the law of errors in the form of method of least squares.

VI. **Statistics and Meteorology**: Statistics is related to meteorology. To compare the present with the past or to forecast for the future either temperature or humidity of air or barometrical pressures etc., it becomes necessary to average these figures and thus to study their trends and fluctuations. All this cannot be done without the use of statistical methods. Thus, the science of statics helps meteorology in a large number of ways.

VII. **Statistics and Biology**: The development of biological theories has been found to be closely associated with statistical methods. Professor Karl Pearson in his Grammar of Sciences has written, "The whole doctrine of heredity rests on statistical basis".

VIII. **Statistics and Mathematics**: Mathematics and Statistics have been closely in touch with each other ever since the 17th Century when the theory of probability was found to have influence on various statistical methods. Bowley was right when he said, "Acknowledge of Statistics is like knowledge of foreign language or of algebra: it may prove of use at any time under any circumstances".

Thus we observe that**:**

"Science without statistics bear no fruit, statistics without sciences have no Root".

IX. **Statistics and Research**: Statistical techniques are indispensable in research work. Most of advancement in knowledge has taken place because of experiments conducted with the help of statistical methods.

X. **Statistics and natural sciences**: Statistics finds an extensive application in physical sciences, especially in engineering physics, chemistry, geology, mathematics, astronomy, medicine, botany, meteorology, zoology, etc.

XI. **Statistics and Education:** There is an extensive application of statistics in Education. Statistics is necessary for formulation of policies to start new courses, infrastructure required for new courses consideration of facilities available for new courses etc.

XII. **Statistics and Business:** Statistics is an indispensable tool in all aspects of business. When a man enters business he enters the profession of forecasting because success in business is always the result of precision in forecasting and failure in business is very often due to wrong expectations which arise in turn due to faulty reasoning and inaccurate analysis of various causes affecting a particular phenomenon. Boddington observes, "The successful businessman is the one, whose estimate most closely approaches the accuracy".

LIMITATIONS OF STATISTICS

Statistics and its techniques are widely used in every branch of knowledge. W.I. King rightly says:

"Science of statistics is the most useful servant, but only of great value to those who understand its proper use". The scope of statistics is very wide and it has great utility; but these are restricted by its limitations. Following are the important limitations of statistics:

1. **Statistics does not deal with individual item**: King says, "Statistics from the very nature of the subject cannot and never will be able to take into account individual cases". Statistics proves inadequate, where one wants to study individual cases. Thus, it fails to reveal the true position.
2. **Statistics deals with quantitative data**: According to Prof. Horace Secrist, "Some phenomenon cannot be quantitatively measured; honesty, resourcefulness, integrity, goodwill, all important in industry as well as in life, are generally not susceptible to direct statistical measurement".
3. **Statistical laws are true only on averages**. According to W.I. King, "Statistics largely deals with averages and these may be made up of individual items radically different from each other".

   Statistics are the means and not a solution to a problem.
4. **Statistics does not reveal the entire story**: According to Marshall, "Statistics are the straws, out of which, I, like every other economist, have to make bricks". Croxton says: "It must not be assumed that statistical method is the only method or use in research; neither should this method be considered the best way to attack for every problem".
5. **Statistics is liable to be misused**: According to Bowley, "Statistics only furnishes a tool though imperfect, which is dangerous in the hands of those who do not know its use and deficiencies".

   W.I. King states, "Statistics are like clay of which you can make a God or Devil as you please". He remarks, "Science of Statistics is the useful servant, but only of great value to those who understand its proper use".
6. **Statically data should be uniform and homogeneous**

STATISTICAL TOOLS USED IN ECONOMIC ANALYSIS

The following are some of the important statistical techniques which are applied in economic analysis:

(a) Collection of data

(b) Tabulation

(c) Measures of Central Tendency

(d) Measures of Dispersion

(e) Time Series

(f) Probability

(g) Index Numbers

(h) Sampling and its uses

(i) Business Forecasting

(j) Tests of Significance and analysis of variance

(k) Statistical Quality Control

Collection of Data:

Data that is the information collected through censuses and surveys or in a routine manner or other sources is called a raw data. The word data means information (its literary meaning is given as facts). The adjective raw attached to data indicates that the information thus collected and recorded cannot be put to any use immediately and directly.

It has to be converted into more suitable form or processed before it begins to make sense to be utilized gainfully. A raw data is a statistical data in original form before any statistical techniques are used to redefine, process or summarize it.

There are two types of statistical data:

(i) Primary data

(ii) Secondary data

1. Primary Data: It is the data collected by a particular person or organization for his own use from the primary source

2. Secondary data: It is the data collected by some other person or organization for their own use but the investigator also gets it for his use.

Methods of Collecting Primary Data:

The primary data can be collected by the following methods:

1. **Direct personal observation**: In this method, the investigator collects the data personally and, therefore, it gives reliable and correct information.

2. **Indirect oral investigation**: In this method, a third person is contacted who is expected to know the necessary details about the persons for whom the enquiry is meant.

3. **Estimates from the local sources and correspondence**. Here the investigator appoints agents and correspondents to collect the data

4. **Data through questionnaires**. The data can be collected by preparing a questionnaire and getting it filed by the persons concerned.

5. **Investigations through enumerators**. This method is generally employed by the Government for population census, etc.

Methods of Collecting Secondary data:

The secondary data can be collected from the following sources:

1. Information collected through newspapers and periodicals.

2. Information obtained from the publications of trade associations.

3. Information obtained from the research papers published by University departments or research bureau or UGC.

4. Information obtained from the official publications of the central and the state governments dealing with crop statistics, industrial statistics, trade and transport statistics etc.

5. Information obtained from the official publications of the foreign governments for international organizations.

Classification of Data: The process of arranging things in groups or classes according to their common characteristics and affinities is called the classification of data.

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

Thus classification is the process of arranging the available data into various homogenous classes and sub-classes according to some common characteristics or attributes or objectives of investigation.

Requisites of a Good Classification:

The main characteristics of a good classification are:

1. It should be exhaustive
2. It should be unambiguous
3. It should be mutually exclusive
4. It should be stable
5. It should be flexible
6. It should have suitability
7. It should be homogeneous
8. It should be a revealing classification
9. It should be reliable
10. It should be adequate

Advantages of classification of data:

(i) It condenses the data and ignores unnecessary details

(ii) It facilitates comparison of data

(iii) It helps in studying the relationships between several characteristics

(iv) It facilitates further statistical treatments

Types of Classification of Data:

There are four types of classification of data:

(i) Quantitative Classification

(ii) Temporal Classification

(iii) Spatial Classification and

(iv) Qualitative Classification

(i) **Quantitative Classification**: When the basis of classification is according to differences in quantity, the classification is called quantitative

A quantitative classification refers to classification that is based on figures: In other words, it is a classification which is based on such characteristics which are quantifiable such as height, weight, marks obtained by students of a class.

(ii) **Temporal Classification**: When the basis of classification is according to differences in time, the classification is called temporal or chronological classification.

(iii) **Spatial or Geographical Classification**: When the basis of classification is according to geographic location or place, the classification is called spatial or geographical classification.

(iv) **Qualitative Classification**: When the basis of classification is according to characteristics or attributes like social status etc. is called qualitative classification.

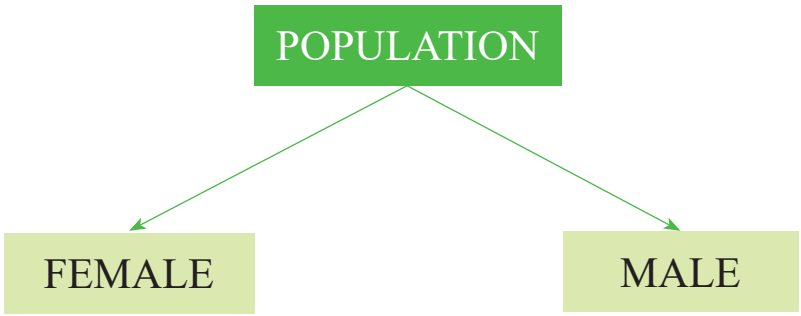
Classification according to attributes is a method in which the data are divided on the basis of qualities.(i.e., married or single; honest or dishonest; beautiful or ugly; on the basis of religion, viz., Hindu, Muslim, Sikh, Christian etc., known as attributes), which cannot be quantified.

Qualitative Classification is of two types:

(i) Simple Classification or Two-Fold Classification

(ii) Manifold Classification

1. **Simple Classification or Two-fold Classification**: If the data are classified only into two categories according to the presence or absence of only one attribute, the classification is known as simple or two-fold classification or Dichotomous For example, the population of India may be divided into males and females; literate and illiterate etc.



Moreover, if the classification is done according to a single attribute it is also known as one way classification.

2. **Manifold Classification**: It is a classification where more than one attributes are involved.

MODE OF PRESENTATION OF DATA

In this section we shall consider the following three modes of presentation of data

(a) Textual Presentation

(b) Tabular Presentation or Tabulation

(c) Diagrammatic presentation

Textual Presentation:

In this method presentation of data is done with the help of a paragraph or a number of paragraphs. The official report of an enquiry commission is usually made by textual presentation. Following are the examples of textual presentation.

Example1: In 1995, out of total 2,000 students in a college, 1,400 were for graduation and the rest for post-graduation (P.G.). Out of 1,400 Graduation students 100 were girls. However, in all there were 600 girls in the college. In 2000, number of graduation students increased to 1,700 out of which 250 were girls, but the number of P.G. students fall to 500 of which only 50 were boys. In 2005, out of 800 girls 650 were for graduation, whereas the total number of graduation students was 2,200. The number of boys and girls in P.G. classes were equal.

Merits and Demerits of Textual Presentation: The merit of this mode of presentation lies in its simplicity and even a layman can present and understand the data by this method. The observations with exact magnitude can be presented with the help of textual presentation. This type of presentation can be taken as the first step towards the other methods of presentation.

Textual presentation, however, is not preferred by a statistician simply because it is dull, monotonous and comparison between different observations is not possible in this method. For manifold classification, this method cannot be recommended.

Tabular presentation or tabulation of data: Tabulation is a scientific process used in setting out the collected data in an understandable form

Tabulation may be defined as logical and systematic arrangement of statistical data in rows and columns. It is designed to simplify the presentation of data for the purposes of analysis and statistical inferences.

Secrist has defined tabulation in the following words:

"Tables are a means of recording in permanent form the analysis that is made through classification and by placing in juxtaposition things that are similar and should be compared".

The above definition clearly points out that tabulation is a process which gives classification of data in a systematic form and is meant for the purpose of making comparative studies.

Professor Bowley refers to tabulation as:

"The intermediate process between the accumulation of data in whatsoever form they are obtained, and the final reasoned account of the result shown by the statistics"

"Tabulation is the process of condensing classified data in the form of a table so that it may be more easily understood and so that any comparison involved may be more readily made".

Thus tabulation is one of the most important and ingenious devices of presenting the data in a condensed and readily comprehensible form. It attempts to furnish the maximum information in the minimum possible space, without sacrificing the quality and usefulness of the data.

Objectives of Tabulation:

The purpose of tabulation is to summarise lots of information in such a simple manner that it can be easily analysed and interpreted.

The main objectives of the Tabulation are:

1. To simplify the complex data.
2. To clarify the objective of investigation
3. Economise space.
4. To facilitate comparison
5. To depict trend and pattern of data
6. To act as reference for future studies.
7. To facilitate statistical analysis.
8. To detect errors and omissions in the data
9. To clarify the characteristics of data.

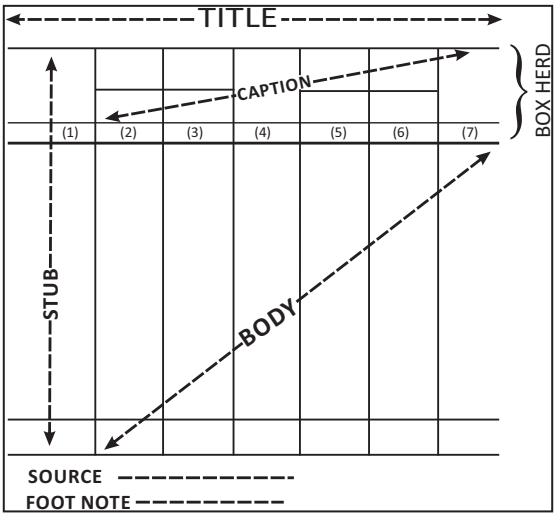
Essential Parts of a Statistical Table:

A good statistical table should invariably have the following parts:

1. **Table Number**: A table should be numbered for identification, especially, when there are a large number of tables in a study. The number may be put at the centre, above the title or at the bottom of the table.

2. **Title of the table**: Every table should have a title. It should be clear, brief and self-explanatory. The title should be set in bold type so as to give prominence.
3. **Date**: The date of preparation of a table should always be written on the table. It enables to recollect the chronological order of the table prepared.
4. **Stubs or Row Designations**: Each row of the table must have a heading. The designations of the rows are called stubs or stub items. Stubs clarify the figures in the rows. As far as possible, the same items should be considered so that they can be included in a single row.
5. **Captions or Column headings**: A table has many columns. Sub-headings of the columns are called captions or headings. They should be well-defined and brief.
6. **Body of the table**: It is the most vital part of the table. It contains the numerical information. It should be made as comprehensive as possible. The actual data should be arranged in such a manner that any figure may be readily located. Different categories of numerical variables should be set out in an ascending order, from left to right in rows and in the same fashion in the columns, from top downwards.
7. **Unit of Measurements**: The unit of measurements should always be stated along with the title, if this is uniform throughout. If different units have been adopted, then they should be stated along the stubs or captions.
8. **Source Notes**. A note at the bottom of the table should always be given to indicate the primary source as well as the secondary source from where the data has been taken, particularly, when there is more than one source.
9. **Foot Notes and References**: It is always placed at the bottom of the table. It is a statement which contains explanation of some specific items, which cannot be understood by the reader from the title, or captions and stubs.

Different Parts of Table



Difference between Textual and Tabular Presentation: The tabulation method is usually preferred to textual presentation as:

(i) It facilitates comparison between rows and columns

(ii) Complicated data can be represented using tabulation

(iii) Without tabulation, statistical analysis of data is not possible.

(iv) It is a must for diagrammatic representation.

Diagrammatic Representation of Data

4.1

The representation of statistical data through charts, diagrams and picture is another attractive and alternative method. Unlike the first two methods of representation of data, diagrammatic representation can be used for both the educated section and uneducated section of the society. Furthermore, any hidden trend presented in the given data can be noticed only in this mode of representation.

However, compared to tabulation, this is less accurate. So if there is a priority for accuracy, we have to recommend tabulation.

In this chapter we shall consider the following three types of diagrams:

I. Line diagram;

II. Bar diagram;

III. Pie chart.

LINE CHART

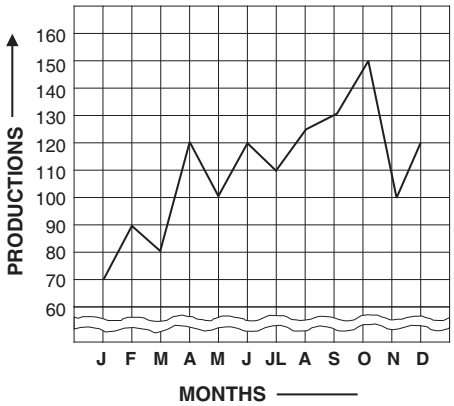
We take a rectangular axes. Along the abscissa, we take the independent variable (x or time) and along the ordinate the dependent variable (y or production related to time). After plotting the points, they are joined by a scale, which represents a line chart. The idea will be clear from the following example.

Example : Represent the following data by line chart.

The monthly production of motor cars in India during 2011-12

Jan	Feb	Mar	April	May	June	July	Aug	Sept	Oct	Nov	Dce
70	90	80	120	100	120	110	125	130	150	100	120

Graph showing production of motor cars.



BAR DIAGRAM

The simplest type of graph is the bar diagram. It is especially useful in comparing qualitative data or quantitative data of discrete type. A bar diagram is a graph on which the data are represented in the form of bars. It consists of a number of bars or rectangles which are of uniform width with equal space between them on the x-axis. The height (for vertical bars) or length (for horizontal bars) of the bar is proportional to the value it represents. It should be seen that the bars are neither too short nor too long. The scale should be clearly indicated and baseline be clearly shown.

Bars may be drawn either horizontally or vertically. A good rule to use in determining the direction is that if the legend describing the bar is to be written under the bars then vertical bars should be used; when it is not a requirement, horizontal ones must be used. In this way, the legends can be read without turning the graph. The descriptive legend should not be written at the ends of the bars or within the bars, since such writing may distort the comparison. Usually the diagram will be more attractive if the bars are wider than the space between them.

The width of bars is not governed by any set rules. It is an arbitrary factor. Regarding the space between two bars, it is conventional to have a space about one half of the width of a bar.

The data capable of representation through bar diagrams, may be in the form of row scores, or total scores, or frequencies, or computed statistics and summarised figures like percentages and averages etc.

The bar diagram is generally used for comparison of quantitative data. It is also used in presenting data involving time factor. When two or more sets of data over a certain period of time are to be compared a group bar diagram is prepared by placing the related data side by side in the shape of bars. The bars may be vertical or horizontal in a bar diagram. If the bars are placed horizontally, it is called a Horizontal Bar Diagram. When the bars are placed vertically, it is called a Vertical Bar Diagram.

There are six types of Bar diagram:

(i) Simple Bar Diagram;

(ii) Multiple or Grouped Bar Diagram;

(iii) Subdivided or Component Bar Diagram;

(iv) Percentage Subdivided Bar Diagram;

(v) Deviation or Bilateral Bar Diagram;

(vi) Broken Bars.

Simple Bar Diagram:

It is used to compare two or more items related to a variable. In this case, the data are presented with the help of bars. These bars are usually arranged according to relative magnitude of bars. The length of a bar is determined by the value or the amount of the variable. A limitation of Simple Bar Diagram is that only one variable can be represented on it.

Multiple or Grouped Bar Diagram:

A multiple or grouped bar diagram is used when a number of items are to be compared in respect of two, three or more values. In this case, the numerical values of major categories are arranged in ascending or descending order so that the categories can be readily distinguished. Different shades or colours are used for each category.

Sub-divided or Component Bar Diagram:

A component bar diagram is one which is formed by dividing a single bar into several component parts.

A single bar represents the aggregate value whereas the component parts represent the component values of the aggregate value. It shows the relationship among the different parts and also between the different parts and the main bar.

Percentage Sub-divided Bar Diagram:

It consists of one or more than one bars where each bar totals 100%. Its construction is similar to the subdivided bar diagram with the only difference is that in the sub-divided bar diagram segments are used in absolute quantities and in the percentage bar diagram the quantities are transformed into percentages.

PIE DIAGRAM OR ANGULAR DIAGRAM

A pie diagram is a circular graph which represents the total value with its components. The area of a circle represents the total value and the different sectors of the circle represent the different parts. The circle is divided into sectors by radii and the areas of the sectors are proportional to the angles at the centre. It is generally used for comparing the relation between various components of a value and between components and the total value. In pie diagram, the data are expressed as percentages.

Each component is expressed as percentage of the total value. A pie diagram is also known as angular diagram.

The name pie diagram is given to a circle diagram because in determining the circumference of a circle we have to take into consideration a quantity known as 'pie' (written as π).

Method of Construction: The surface area of a circle is known to cover 2 radians or 360 degrees.

The data to be represented through a circle diagram may therefore be presented through 360 degrees or parts or sections of a circle. The total frequencies or value is equated to 360° and then the angles corresponding to component parts are calculated (or the component parts are expressed as percentages of the total and then multiplied by 360/100 or 3.6). After determining these angles the required sectors in the circle are drawn. Different shades or colours of designs or different types of hatchings are used to distinguish the various sectors of the circle.

Illustration 1.

120 students of a college were asked to opt for different work experiences. The details of these options are as under.

Areas of work experience	No. of students
Photography	6
Clay modeling	30
Kitchen gardening	48
Doll making	12
Book binding	24

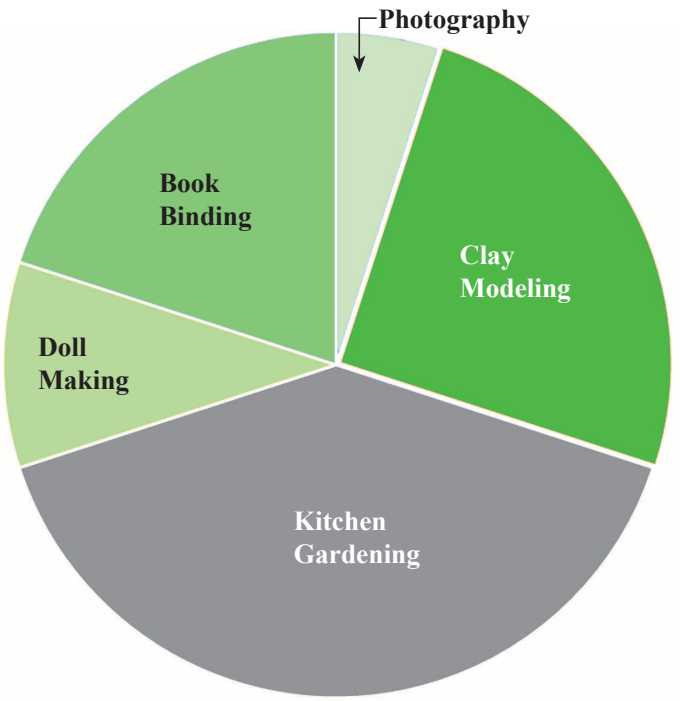
Represent the above data through a pie diagram.

Solution:

The numerical data may be converted into the angle of the circles as given below:

Area of Work Experience	No of Students	Angle subtended at the centre of the circle (degrees)
Photography	6	(6/120) × 360 = 18
Clay Modelling	30	(30/120) × 360 = 90
Kitchen Gardening	48	(48/120) × 360 = 144
Doll Making	12	(12/120) × 360 = 36
Book Binding	24	(24/120) × 360 = 72
Total	**120**	**360**

With the help of above computations the following Pie diagram is constructed.



Frequency Distribution

4.2

TALLY BARS AND FREQUENCY

In order to make the data easily understandable, we tabulate the data in the form of tables or charts.

A table has three columns

(i) Variable

(ii) Tally marks

(iii) Frequency

(i) Variable: Any character which can vary from one individual to another is called a variable or a variate. For example, age, income, height, intelligence, colour etc. are variates. Some variates are measurable and others are not directly measurable. The examples of measurable variates are age, height, temperature, etc., whereas colour and intelligence are the examples of those variates which cannot be measured numerically. Variables or observations having numerics as possible values are called quantitative variables, whereas those with names of places, quality, and things etc., as possible values are called qualitative variables or attributes.

Variables are of two types i) Continuous; ii) Discontinuous or Discrete. Quantities which can take all numerical values within a certain interval are called continuous variables; But those variables which can take only a finite number of values are called discrete variables; For example, number of students in a particular class, number of sections in a school etc.

(ii) Tally: It is a method of keeping count in blocks of five.

For example: | = 1, || = 2, ||| =3, |||| = 4, 𝍸 = 5

Tally Bars: These are the straight bars used in the Tally.

Each item falling in the class interval, a stroke (vertical Bar) is marked against it. This stroke (Vertical Bar)is called the Tally Bar. Usually, after every four strokes (Tally Bar), in a class, the fifth item is marked by a horizontal or slanted line across the Tally Bars already drawn. For example the frequency 5, 6, 7 is represented by 𝍸, 𝍸 |, 𝍸 || respectively.

The above method of presentation of data is known as 'Frequency Distribution'. Marks represent the numerical value of the number of students who secured a particular marks and is called Frequency of that particular value of the variate.

In the first column of the table, we write all marks from lowest to highest. We now look at the first mark or value in the given raw data and put a bar (vertical line) in the second column against it. We then, see the second mark or value in the given raw data and put a bar against it in the second column.

This process is repeated till all the observations in the given raw data are exhausted. The bars drawn in the second column are known as tally marks and to facilitate we record tally marks in bunches of five, the fifth tally marks is drawn across the first four.

For example 𝍸 = 8. We finally count the number of tally marks corresponding to each observation and write in the third column headed by frequency or number of students.

(iii) Frequency: The number of times an observation occurs in the given data is called the frequency of the observation.

Frequency Distribution: A frequency distribution is the arrangement of the given data in the form of a table showing frequency with which each variable occurs. In other words, Frequency distribution of a variable is the ordered set $\{x, f\}$, where f is the frequency. It shows all scores in a set of data together with the frequency of each score.

Types of frequency distributions:

Frequency distributions are of two types:

(i) Discrete Frequency Distribution

(ii) Grouped (or Continuous) Frequency Distribution

Discrete Frequency Distribution

The construction of discrete frequency distribution from the given raw data is done by the method of tally marks as explained earlier.

Construction of Discrete Frequency Distribution Table

The frequency distribution table has three columns headed by

1. Variables (or classes)
2. Tally Mark or Bars
3. Frequency

The table is constructed by the following steps:

Step 1: Prepare three columns, viz., one for the variable (or classes), another for tally marks and the third for the frequency corresponding to the variable (or class).

Step 2: Arrange the given data (or values) from the lowest to the highest in the first column under the heading variable (or classes)

Step 3: Take the first observation in the raw data and put a bar (or vertical line|) in the second column under Tally Marks against it. Then take a second observation and put a tally marks against it, continue this process till all the observations of the given raw data are exhausted. For the sake of convenience, record the tally marks in bunches of five, the fifth bar is placed diagonally crossing the other four (5 is represented by 𝍸 or 𝍸) leave some space between each block of bars.

Step 4: Count the tally marks of column 2 and place this number opposite to the value of the variable in the third column headed by Frequency.

Step 5: Give a suitable title to the frequency distribution table so that it exactly conveys the information contained in the table.

SOME STATISTICAL TERMS

Raw Data or Data: A raw data is a statistical data in original form before any statistical technique is applied to redefine process or summarize it.

Continuous Variable: A continuous variable is capable of assuming any value within a certain range or interval. The height, weight, age and temperature of any person can be expressed not only in integral part but also in fractions of any part. For example, the weight of a boy may be 44.0 kg or 44.6 kg or 44.65kg, similarly, his height may be 56 inches or 56.4 inches and age may be 10 years or 10.5 years. Thus, the height, weight, age or temperature etc. are continuous variables.

Discrete Variable: A discrete variable can assume only integral values and can have exact measurement. In other words, those variables which can take only a finite set of values are called discrete variables. For example: the number of students in a particular class, or the number of sections in a school, etc. are the examples of discrete variables. Discrete variables are also known as discontinuous variables.

Continuous Series: When the continuous variables are arranged in the form of a series, it is called continuous series or exclusive series.

Discrete or Discontinuous Series: When the discrete variables are arranged in the form of a series, it is called a discrete or discontinuous series.

Array: An array is an arrangement of data in order of magnitude either in descending or ascending order.

Descending Order: When data is arranged from the highest value to the lowest value, the array so formed is in descending order.

Ascending order: When the data is arranged from the lowest value to the highest value, the array so formed is in ascending order.

Illustration 2.

If the given data is 17, 7, 11, 5, 13, 9 then

Array in ascending order: 5, 7, 9, 11, 13, 17.

Array in descending order: 17,13,11,9,7, 5.

Range: It is the difference between the largest and the smallest number in the given data

The range of the data given in illustration above is 17 – 5 = 12.

Class, Class-Interval and Class limits. If the observations of a series are divided into groups and the groups are bounded by limits, then each group is called a class. The end values of a class are called class limits. The smaller value of the two limits is called the lower limit and the higher value of the same is called the upper limit of the class. These two class limits are sometimes called the stated class limits.

Class Width: The difference between the lower limit (L) and the upper limit (U) of the class is known as class Width (I).

Thus: I = U – L.

In other words, the range of a class is called its Class Width.

Illustration 3.

The given data is

Marks obtained	Tally Marks	No of Students
1-10	𝍸 |	6
11- 20	|||	3
21-30	𝍸 ||	7
31-40	||	2
Total		**18**

In the above data the classes are: 1-10, 11-20, 21-30, and 31-40.

Class Interval: The range of the marks from 1 to 40 is grouped into four classes or groups viz: 1 – 10, 11 - 20, 21-30, and 31-40. Each group is known as class interval. The interval between one class and its adjacent class being 9. [As 10-1 = 9, 20 – 11 = 9, 30 – 21 = 9, etc.]

Class Limits: In the first class 1 – 10, its lower limit is 1 and upper limit is 10. Similarly, 31 is the lower limit and40 is the upper limit of the class interval 31-40.

Actual Class Limit or Class Boundaries: In the illustration, there is a gap of 1 mark between the limits of any two adjacent classes. This gap may be filled up by extending the two limits of each class by half of the value of the gap. Thus

Lower class boundary = lower class limit – ½ of the gap

Upper class boundary = Upper class limit + ½ of the gap

The class boundary of the class 11 to 20 are

Lower class boundary = 11– ½ of 1 = 11 – 0.5 = 10.5

Upper class boundary = 20+ ½ of 1 = 20+ 0.5 = 20.5

In other words, the class boundaries are the limits up to which the two limits, (actual) of each class maybe extended to fill up the gap that exists between the classes. The class boundaries of each class, so obtained are called the Actual class limits or True class limits.

True lower class limit = Lower class limit - ½ of the gap

True upper class limit = Upper class limit + ½ of the gap

Note: In the case of exclusive series True class limits are the same as class limits

Illustration 4.

Class Interval	Class Boundaries
11 – 20	10.5 – 20.5
21 – 30	20.5 – 30.5
31 – 40	30.5 – 40.5

Class-mark or Mid-point or Mid-value: The central value of the class interval is called the mid-point or mid-value or class mark. It is the arithmetic mean of the lower class and upper class limit of the same class.

Mid-value of Class = (Lower class limit + Upper class limit)/2

Or,

Class mark = (True Upper class limit + True Lower class limit)/2

The class mark of the class 11-20 is (11+ 20)/2= 15.5

Class Magnitude: It is the difference between the upper class boundary and the lower class boundary of the class. In the illustration the class magnitude of the class, 20.5 – 30.5 is (30.5 – 20.5) = 10.

Inclusive and Exclusive Series: In the above illustration, all the marks we considered were integers.

Hence, it was possible for us to choose classes 11 to 20, 21 to 30 etc. there is a gap of 1 between the upper limit of a class and the lower limit of its succeeding class, which has not created any difficulty. But there can be situations where the raw data is not in integers. For example, in the information regarding maximum temperature of the city or time required to solve a statistical problem is recorded in the data, it may contain fractions as well. In such cases, the consecutive classes have to be necessarily continuous. We have the following:

Inclusive Series: When the class-intervals are so fixed that the upper limit of the class is included in that class, it is known as inclusive method of classification, e.g., 0-5, 6-10, 11-15, 16-20.

In the inclusive series, the upper limit and lower limit are included in that class interval. For example, in illustration, the marks 11 and 20 are included in the class 11-20. It is a discontinuous series or inclusive series. In order to make it a continuous one, some adjustment with the class limits is necessary. The class limits are extended to class boundaries by the adjusting adjustment factor, which is equal to half of the difference between the upper limit of the one class and lower limit of the next class. The series so obtained is continuous and is known as exclusive series.

Exclusive or Continuous Series: In this series the upper limit of the class is the lower limit of the other class, the common point of the two classes is included in the higher class. For example, 10-15, 15-20, 20-25… represent a continuous series or the exclusive series. In this series, 15 is included in the class 15-20 and 20 is included in 20-30. Here the class intervals overlap and the upper limit of each class is treated as less than that limit and lower limit of each class actually represents exact value. Thus when the class-intervals are so fixed that the upper limit of one class is the lower limit of the next class, it is known as Exclusive method of classification.

Relative frequency, Percentage frequency and Frequency Density of a class interval

Relative Frequency: Frequency of each class can also be expressed as a fraction of total frequency. These are known as relative frequencies. In other words, a relative frequency is the class frequency expressed as a ratio of the total frequency, i.e.

Relative frequency = Class frequency/Total frequency

Percentage Frequency: Percentage frequency of a class interval may be defined as the ratio of the class frequency to the total frequency expressed as a percentage.

Percentage frequency = (Class frequency/ Total frequency) × 100

Frequency Density: This in defined as the ratio of Frequency of a class to its Width.

Frequency Density = Class Frequency / Class Width

Graphical Representation of Frequency Distribution

4.3

The graphs of frequency distribution are designed to present the characteristic features of a frequency data. They facilitate comparative study of two or more frequency distributions regarding their shape and pattern.

The most commonly used graphs are:

1. Histogram
2. Frequency Polygon
3. Frequency Curve
4. Cumulative Frequency Curve or Ogive.

Illustration 5.

Draw the pie diagram from the following information

Vehicle Sales by type at a Car Dealers shop are given as:

Type	Numbers Sold
Wagon R	72
S-Cross	54
Amaze	27
Punch	18
City	9
Total	**180**

Solution:

The detail computations are as follows:

Type	Nos. Sold	Angle subtended at the centre by the Area (degrees)
Wagon R	72	144
S-Cross	54	108
Amaze	27	54
Punch	18	36
City	9	18
Total	**180**	**360**

Angle subtended at the centre of the circle by differnent Areas (in Degrees) are

Wagon $R = \frac{360}{180} \times 72 = 144$

S-Cross $= \frac{360}{180} \times 54 = 108$

Amaze $= \frac{360}{180} \times 27 = 54$

Punch $= \frac{360}{180} \times 18 = 36$

City $= \frac{360}{180} \times 9 = 18$

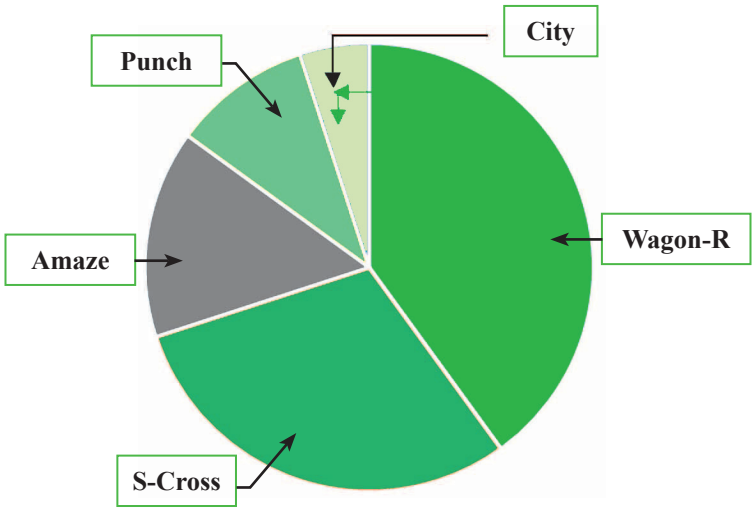


Illustration 6.

Prepare a statistical table from the following information of Daily earnings (in ₹) by Doctors:

88	23	27	28	86	96	94	93	86	99
82	24	24	55	88	99	55	86	82	36
96	39	26	54	87	100	56	84	83	46
102	48	27	26	29	100	59	83	84	48
104	46	30	29	40	101	60	89	46	49
106	33	36	30	40	103	70	90	49	50
104	36	37	40	40	106	72	94	50	60
24	39	49	46	66	107	76	96	46	67
26	78	50	44	43	46	79	99	36	68
29	67	56	99	93	48	80	102	32	51

Solution:

Detailed calculations are as follows:

Class Internal of Earnings (₹)	Tally Marks	No of Doctors	Class Boundaries
20 – 29	𝍸 𝍸 |||	13	19.5 – 29.5
30 – 39	𝍸 𝍸 |	11	29.5 – 39.5
40 – 49	𝍸 𝍸 𝍸 |||	18	39.5 – 49.5
50 – 59	𝍸 𝍸	10	49.5 – 59.5
60 – 69	𝍸 |	6	59.5 – 60.5
70 – 79	𝍸	5	60.5 – 79.5
80 – 89	𝍸 𝍸 ||||	14	79.5 – 89.5
90 – 99	𝍸 𝍸 ||	12	89.5 – 99.5
100 – 109	𝍸 𝍸 |	11	99.5 – 109.5
Total	–	**100**	–

Illustration 7.

The national income of a country for the year 1999-2000 to 2001-2002 at current prices was 8650, 9010 and 9530 crores of rupees respectively and per capita income for these years was 250, 256 and 267 rupees. The corresponding figures of national income and per capita income at 1999-2000 prices for the above years were 8650, 8820 and 8850 crores of rupees and 250, 252 and 248 respectively. Present the above data in a table.

Solution:

National Income and Per capita income of a country for the year 1999-2000 to 2001-2002

Year	National Income		Per Capita Income	
	At Current Prices (` crores)	At 1999-2000 Prices (₹ crores)	At Current Prices	At 1999-2000 Prices
1999-2000	8650	8650	200	250
2000-2001	9010	8820	256	251
2001-2002	9530	8850	267	248

Illustration 8.

Draw a Histogram of the frequency distribution given below:

Class Interval	Frequency	Class Interval	Frequency
58 – 61	2	70 – 73	56
61 – 64	10	73 – 76	16
64 – 67	48	76 – 79	4
67 – 70	64		

Solution:

In this case widths of the classes are of equal magnitude throughout. Width of each class interval is drawn on the X axis i.e. base of each rectangle is proportional to the magnitude of the width of class interval. With width of each class interval as base a rectangle is drawn with height proportional to the corresponding frequency of the class.

Histogram is as follows:

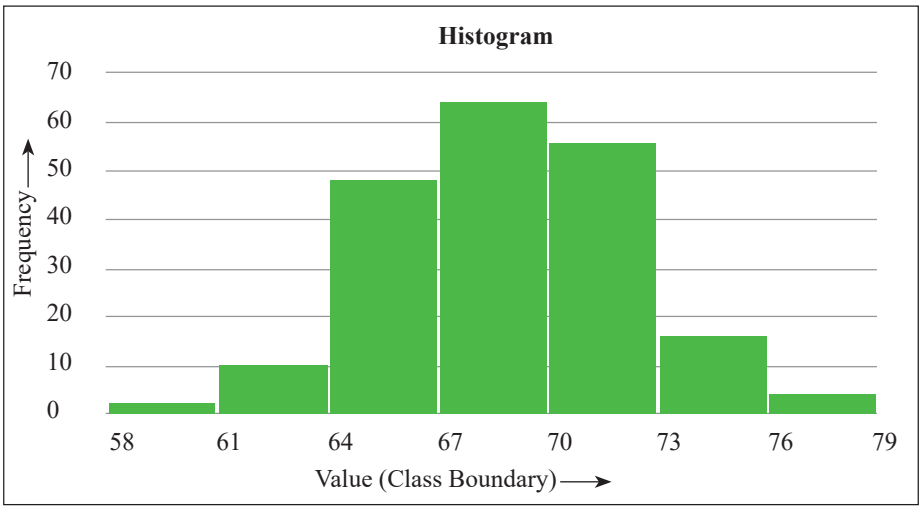


Illustration 9.

Draw a Histogram of the frequency distribution given below:

Class Interval	Frequency	Class Interval	Frequency
10-14	4	40 – 49	14
15 – 19	12	50 – 74	25
20 – 29	20	75 – 99	10
30 – 39	18		

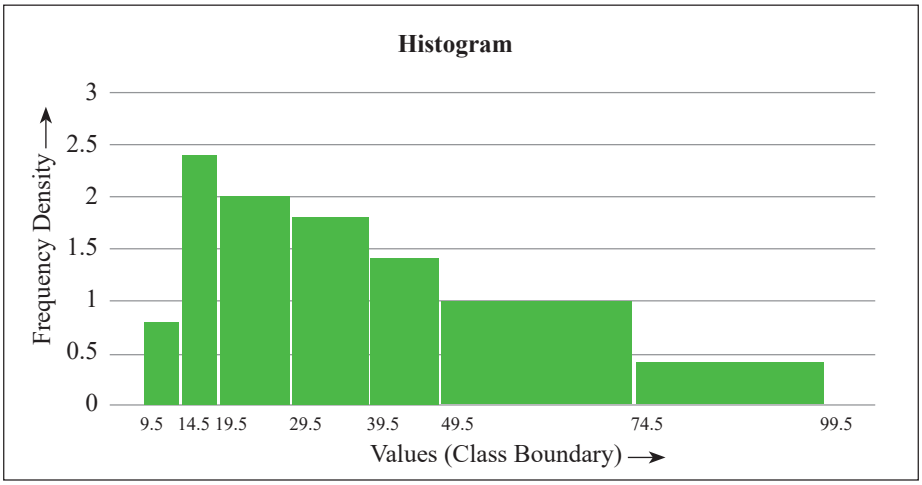
Solution:

In the given data it is to be noted that the widths of the class intervals are unequal and class boundaries are also not provided. There is a gap of 1 between the limits of any two adjacent classes. Since rectangle in a histogram should have area proportional to the class frequencies, the height of each rectangle should be made proportional to the frequency density. Secondly we can eliminate the gap between the limits of two adjacent classes by bringing in the class boundaries instead of class limits. Thus in this case the class boundaries are 9.5 – 14.5, 14.5 – 19.5 and so on.

Calculaltion of Frequency Density

Class Internal Showing Class Limits	Class Internal Showing Class Boundaries	Class Width*	Frequency	Frequency Density
(1)	**(2)**	**(3)**	**(4)**	**(5) = (4) ÷(3)**
10 - 14	9.5 - 14.5	5	4	0.8
15 - 19	14.5 - 19.5	5	12	2.4
20 - 29	19.5 - 29.5	10	20	2
30 - 39	29.5 - 39.5	10	18	1.8
40 - 49	39.5 - 49.5	10	14	1.4
50 - 74	49.5 - 74.5	25	25	1
75 - 99	74.5 - 99.5	25	10	0.4

* Class Width = Upper Class Boundary – Lower Class Boudary



The histogram above was thus drawn by plotting the Values (Class boundary) on X axis and frequency density on Y axis

Illustration 10.

Draw histogram, frequency polygon and ogives (both "less-than" and "more-than" types) for the following frequency distribution:

Wages (₹)	50-59	60-69	70-79	80-89	90-99	100-109	110-119
No.of employees	8	10	16	14	10	5	2

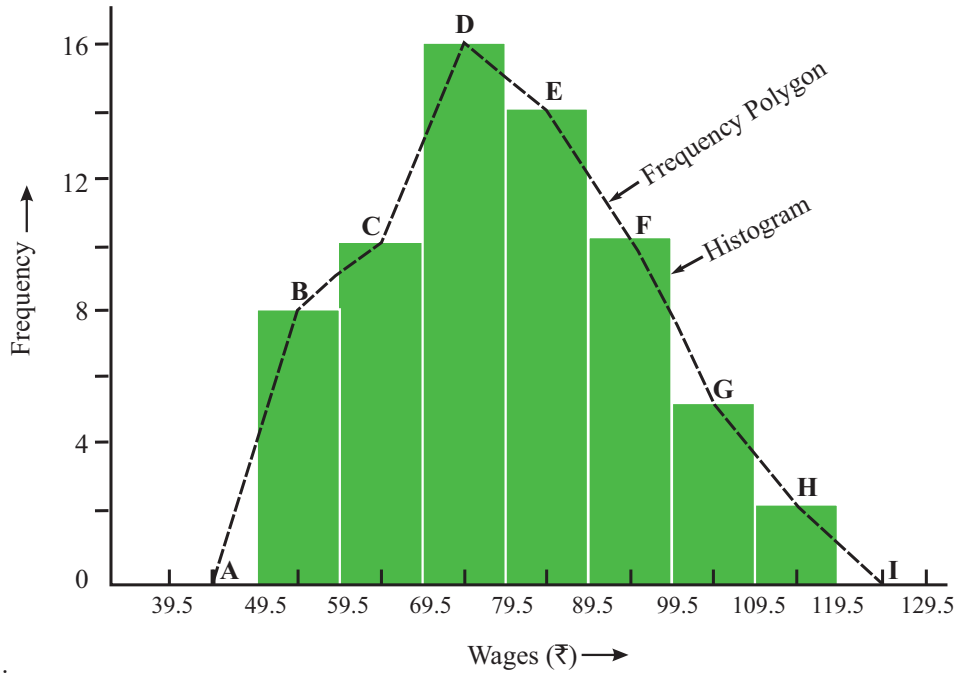
Solution:

[**Note:** Here the class intervals are defined by class limits and so we have to find the class boundaries for drawing the histogram. All the classes have the same width and therefore when drawing the histogram, heights of the rectangles may be represented by the class frequencies. Also for drawing the ogives, we have to calculate both 'less-than' and 'more-than' cumulative frequencies.]

Calculations for Drawing Histogram

Class	Class boundaries	Frequency
50 – 59	49.5 – 59.6	8
60 – 69	59.5 – 69.5	10
70 – 79	69.5 – 79.5	16
80 – 89	79.5 – 89.5	14
90 – 99	89.5 – 99.5	10
100 – 109	99.5 – 109.5	5
110 – 119	109.5 – 119.5	2

Histogram and Frequency Polygon for Wage Distribution



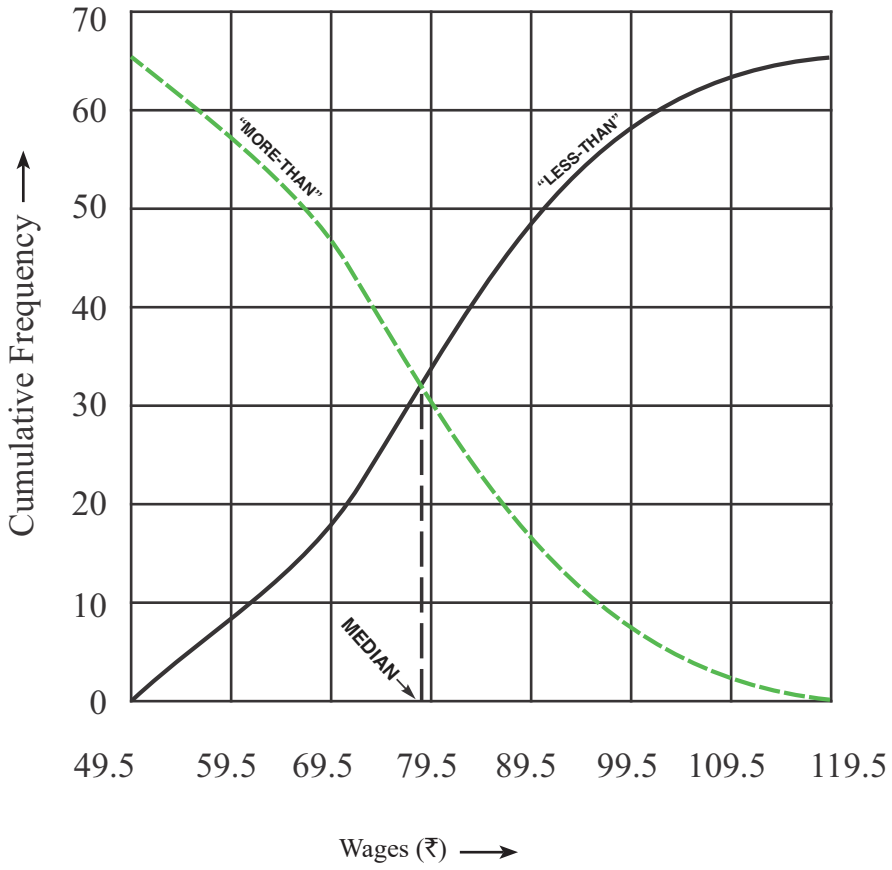
Frequency Polygon is drawn by joining the mid points of the top sides of the frequency bars. Here the points B, C, D, E, F, G and H are the mid points of the frequency bars. Two end points of the Polygon are drawn by considering mid points of two class widths drawn on the two sides of the Histogram. Here A is the mid point of class width considered on the left side of the Histogram and I is that of the width considered on the right side of the Histogram.

Thus for the given data the Frequency Polygon is ABCDEFGHI.

Calculations for Drawing Ogives

Class Boundary	Cumulative Frequency	
	'less-than'	'more-than'
49.5	0	65 = *N*
59.5	8	57
69.5	18	47
79.5	34	31
89.5	48	17
99.5	58	7
109.5	63	2
119.5	65 = *N*	0

Ogives for Wage Distribution



Exercise:

Theoretical Questions

Multiple Choice Questions (MCQ)

1. The word 'Statistics' has been derived from the Latin word 'Status' which means

   (a) A numerical state

   (b) A mathematical state

   (c) A neutral state

   (d) A political state

2. Which one of the following has synonymous words?

   (a) Status, Staistik, Statista

   (b) Staistik, Statista, Stats

   (c) Statistic, Statistia, Stats

   (d) Statistic, Statistia, Status

3. Statistics is a tool in the hands of mankind

   (a) To translate complex facts into simple and understandable statements of facts

   (b) To translate statistical information into mathematical statements of facts

   (c) To translate business information into simple and understandable statements of facts

   (d) To translate complex facts into presentable forms through tables and diagrams

4. The word statistics in plural form means

   (a) The science of statistics that collect data in a systematic manner with some definite aim or object towards further analysis in graphical way

   (b) The numerical data collected in a systematic manner with some definite aim or object in view such as the number of persons unemployed in a country.

   (c) The numerical data collected in a systematic manner for analysing the state of affairs with the help of mathematical applications

   (d) The science of statistics that deals with the principles, devices or statistical methods of collecting, analyzing and interpreting numerical data

5. The word statistics in singular form means

   (a) The science of statistics that collect data in a systematic manner with some definite aim or object towards further analysis in graphical way

   (b) The numerical data collected in a systematic manner with some definite aim or object in view such as the number of persons unemployed in a country.

   (c) The numerical data collected in a systematic manner for analysing the state of affairs with the help of mathematical applications

   (d) The science of statistics that deals with the principles, devices or statistical methods of collecting, analyzing and interpreting numerical data

6. Which one of the following is not a characteristic of Plural form of Statistics?

   (a) Statistics should be collected for a pre-determined purpose

   (b) Statistics should be enumerated or estimated

   (c) Statistics are always graphically expressed

   (d) Statistics are affected by a large number of causes

8. Measurement of skewness is

   (a) Analysis of data

   (b) Presentation of data

   (c) Organisation of data

   (d) Interpretation of data

9. Find the odd one out from the following

   (a) Regression

   (b) Kurtosis

   (c) Sampling

   (d) Central Tendency

10. Raw data is

   (a) Information which can be interpreted to take decision

   (b) Information which can't be put to use directly

   (c) Information which is not amenable to conversion

   (d) Information which are useless

11. There are four person named A, B, C, & D. A is a sales person whereas B, C, D are students. A collected sales figures for his region and B, C, D used these data in order to study sales pattern. Which one of the following is correct?

    (a) B uses secondary data

    (b) A & B both are using primary data

    (c) A, B, C, D all are using secondary data

    (d) B, C, D are using primary data

12. Which one of the following is a method of collecting primary data?

    (a) Information collected through newspapers and periodicals

    (b) Information obtained from the publications of trade associations

    (c) Information collected by Government through Census

    (d) Information gathered from research paper published in research journal

13. Classification of data is

    (a) The process of arranging things in groups or classes according to their common frequencies

    (b) The process of arranging things in groups or classes according to their common characteristics and affinities

    (c) The process of arranging things in groups or classes according to their common differences and tally marks

    (d) The process of arranging things in groups or classes according to their common deviations from respective mean

14. In Statistics classification

    (a) Separates data into different unrelated parts

    (b) Separates data into different dispersed groups

    (c) Separates data into different but related parts

    (d) Separates data into different modal groups

15. Which one of the following is a characteristic of a good classification?

    (a) Classification should be heterogeneous

    (b) Classified groups must have overlapping data

    (c) Classification should be stable

    (d) Classification should not be inclusive

16. When the basis of classification is according to differences in time it is called

    (a) Temporal Classification

    (b) Quantitative Classification

    (c) Spatial Classification

    (d) Qualitative Classification

17. Dichotomous Classification is

    (a) When data is classified according to presence or absence of two attributes

    (b) When data is classified into two groups containing all the attributes

    (c) When data is classified according to presence of two attributes

    (d) When data is classified into two groups according to presence or absence of one attribute

18. Tabulation Condenses classified data so that

    (a) Data may be more easily understood

    (b) Data may be easily presented textually

    (c) Data may be more easily synchronized

    (d) More comprehensive secondary result could be obtained

19. Tabulation is preferred to textual presentation because

    (a) Tabulation never compares between rows and columns

    (b) Tabulation is helpful for diagrammatic representation

    (c) Tabulation always get constructed with simple data

    (d) None of the above

20. With respect to accuracy

    (a) Diagrammatic presentation is preferable to Tabular presentation

    (b) Textual presentation is preferable to diagrammatic presentation

    (c) Tabular presentation is preferable to Diagrammatic presentation

    (d) Textual presentation is preferable to Tabular presentation

State True or False

1. The class boundaries are the limits up to which the two limits, (actual) of each class may be extended to fill up the gap that exists between the classes;
2. In an exclusive series the class limits are extended to class boundaries by the adjusting adjustment factor;
3. In a Continuous Series lower limit of each class actually represents exact value;
4. Ratio of the class frequency to the total frequency expressed as a percentage is called percentage frequency
5. In Histogram class intervals are taken as heights and corresponding frequencies as breadth;
6. Frequency density = Class frequency/ Width of class interval;
7. A discrete variable can assume only integral values and therefore is not capable of exact measurement;
8. Variables or observations those with names of places, attributes, and things etc., as possible values are called descriptive variables;
9. A raw data is a statistical data in original form before any statistical technique is applied to redefine process or summarize it;
10. Any character which can vary from one individual to another is called a variable or a variate

Solution:

- **Multiple Choice Questions (MCQ)**

1.	2.	3.	4.	5.
(d)	(a)	(a)	(b)	(d)
6.	7.	8.	9.	10.
(c)	(a)	(c)	(b)	(a)
11.	12.	13.	14.	15.
(c)	(b)	(c)	(c)	(a)
16.	17.	18.	19.	
(d)	(a)	(b)	(c)	

- **State True or False**

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