

# UNIT I

**Introduction:** Importance of Statistics, Concepts of Statistics, population and a sample; quantitative and qualitative data; Collection of Primary and Secondary data; Classification and Tabulation of data. Construction of Univariate and bivariate frequency distribution; Diagrammatic and Graphical representation of data.

## 1.1 INTRODUCTION TO STATISTICS

**Q1. Define the term statistics in both Plural and Singular Sense.**

(OR)

**Define statistics in singular sense.**

(OR)

**Define the statistics in plural sense.**

*Ans :*

### Meaning

#### Statistics in Singular Sense

In singular sense, statistics acts as a helpful device used for the purpose of collecting, classifying, presenting and interpreting the data. It is otherwise known as 'Analytical statistics'.

Some of the definitions of statistics are presented as follows,

- According to A.L. Bowley**, "Statistics is the science of counting".
- According to A.L. Bowley**, "Statistics may rightly be called the science of averages".
- According to Turtle**, "Statistics is a body of principles and techniques of collecting, classifying, presenting, comparing and interpreting the quantitative data".

Thus, the above definitions specify that statistics in singular sense, is a science which comprises different statistical methods used-for collection, organization, classification, presentation and interpretation of data.

#### Statistics in Plural Sense

In plural sense, the term "Statistics" implies systematical collection of numerical facts like statistics of population, production, price-level, national income and so on.

#### Definition of Statistics in Plural Sense-Numerical Data

1. **According to 'Dr. A.L. Bowley'**, "statistics are numerical statements of facts in any department of enquiry placed in relation to each other".
2. **According to 'L.R. Comor'**, "Statistics are measurements, enumeration or estimates of natural or social phenomena, systematically arranged so as to exhibit their inter-relations".
3. **According to 'Prof. Yule and Prof. Kendall'**, "By statistics we mean quantitative data affected to a marked extent by multiplicity of causes".
4. **According to 'Horace Secrist'**, "By statistics we mean aggregate of facts affected to a marked extent by multiplicity of causes numerically expressed, enumerated or estimated according to a reasonable standard of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other".
5. **According to Webster**, "Statistics are classified facts respecting the conditions of the people in a state - especially those facts which can be stated in numbers or in tables of numbers or in any other tabular or classified arrangement."

6. **According to Achenwall,** "A collection of noteworthy facts concerning state, both historical and descriptive."
7. **According to H. Secrist,** By statistics we mean aggregate of facts affected to a marked extent by a multiplicity of causes, numerically expressed, enumerated or estimated according to a reasonable standard of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other."

**Q2. What are the characteristics of statistics?**

(OR)

**Explain the characteristics of statistics.**

(OR)

**Enumerate the characteristics of a statistics.**

*Aus :* (Imp.)

The following are the characteristics of a statistics :

**1. Statistics are aggregates of facts**

Single and unconnected figures are not statistics, for example, if it is stated that Raj, a student, secured 50 marks, it is not statistics. However, marks secured by the students of the class would constitute statistics. A single figure relating to production income, marks height, etc. cannot be regarded as statistics but aggregates of such facts would be regarded as statistics.

**2. Statistics are affected to a marked extent by multiplicity of causes**

Numerical facts are affected by a multiplicity of factors. For example, the price, of a commodity is affected by number of factory such as supply, demand, imports, exports, money in circulation, competitive products in the market and so on. It is very difficult to study separately the effect of these factors on the price of the commodity. In physical sciences, it is possible to isolate the effect of various factors on a single item, but statistics

are commonly used in social sciences and in social sciences it is very difficult to study the effect of any one factor separately. In statistical methods, the effects of various factors affecting a particular phenomenon are generally studied in a combined form, though attempts are also made to study the effects of different factors separately as well.

**3. Statistics are numerically expressed**

Only numerical data constitutes statistics. Qualitative expressions like good, bad, young, old etc. cannot be regarded as statistics. Statement like the standard of living of people in India has improved or production of petroleum products has increased in India do not constitute statistics. But Statements like "Rice production in India in 2003-04 is expected to be 86.4 million tones as against 72.66 million tons in 2002-03" (Source Economic Survey 2003-2004) is statistics.

**4. Statistics are enumerated or estimated according to reasonable standards of Accuracy**

Numerical information can be either enumerated or estimated. If they are enumerated i.e., actually counted or measured, they are supposed to be exact and accurate. If complete enumeration is not possible because of the large size of high cost, then data is estimated by using sampling technique. Estimated figures cannot be absolutely accurate and precise.

The degree of accuracy depends to a large extent on the particular purpose for which are information is collected and the nature of the particular problem about which the data is collected. There cannot be uniform standard of accuracy for all types of data collection. For example, while calculating the marks of students in an entrance examination or the number of votes received by a candidate in an election, it is necessary that it should be absolutely accurate; but while calculating the number of persons watching a cricket match on Television, the numbers need not be accurate.

**5. Statistics are collected in a systematic manner**

The data should be collected in a very systematic manner. For any socioeconomic survey, a proper schedule depending on the object of enquiry should be prepared and trained investigators should be used to collect the data. An attempt should be made to reduce personal bias to the minimum. Data collected in haphazard manner may give inaccurate results.

**6. Statistics are collected for a predetermined purpose**

The Purpose of collecting the date should be pre-determined. Otherwise, the data collected may not serve any purpose and may become useless. One should not waste time and money in collecting information that is irrelevant for the enquiry. For example, if the purpose of enquiry is to measure the cost of living of low-income group people, we should collect information about the items that are generally consumed by them. Hence for such an index it is useless to collect information on items such as cars, refrigerators, cell-phones etc.

**7. Statistics should be placed in relation to each other**

Statistical information is collected mostly for the purpose of comparison. If the data collected cannot be compared, then much of the purpose of collection will be lost. The information collected should be homogeneous and not heterogeneous in character. Statistical data are often compared period wise or region-wise. For example, the data relating to population of a country for different years or population of different countries in some fixed period will be regarded as statistics. But data relating to the size of shoe of an individual and his intelligence quotient do not constitute statistics.

Statistics should contain the above characteristics. In the absence of such characteristics, numerical data cannot be called Statistics. Hence, "all statistics are numerical "statements of facts, but all numerical statements of facts are not statistics."

**1.1.1 Importance of Statistics**

**Q3. Explain the importance of statistics.**

(OR)

**Explain the importance of statistics in economic analysis and planning.**

**Ans :**

(Imp.)

**1. Importance to the State**

We know that the subject of statistics originated for helping the ancient rulers in the assessment of their military and economic strength. Gradually its scope was enlarged to tackle other problems relating to political activities of the State. In modern era, the role of State has increased and various governments of the world also take care of the welfare of its people. Therefore, these governments require much greater information in the form of numerical figures for the fulfillment of welfare objectives in addition to the efficient running of their administration.

In a democratic form of government, various political groups are also guided by the statistical analysis regarding their popularity in the masses. Thus, it can be said that it is impossible to think about the functioning of modern state in the absence of statistics.

**2. Importance in economics**

Statistics is an indispensable tool for a proper understanding of various economic problems. It also provides important guidelines for the formulation of various economic policies.

Almost every economic problem is capable of being expressed in the form of numerical figures, e.g., the output of agriculture or of industry, volume of exports and imports, prices of commodities, income of the people, distribution of land holding, etc. In each case, the data are affected by a multiplicity of factors. Further, it can be shown that the other conditions prescribed for statistical data are also satisfied. Thus, we can say that the study of various economic problems is essentially the one of a statistical nature."

### 3. Importance in national income accounting

The system of keeping the accounts of income and expenditure of a country is known as national income accounting. These accounts contain information on various macroeconomic variables like national income, expenditure, production, savings, investments, volume of exports and imports, etc. The national income accounts of a country are very useful in having an idea about the broad features of its economy or of a particular region. The preparation of these accounts require data, regarding various variables, at the macro-level. Since such information is very difficult, if not impossible, to obtain, is often estimated by using techniques and principles of statistics.

### 4. Importance in planning

Planning is indispensable for achieving faster rate of growth through the best use of a nation's resources. It also requires a good deal of statistical data on various aspects of the economy. One of the aims of planning could be to achieve a specified rate of growth of the economy. Using statistical techniques, it is possible to assess the amounts of various resources available in the economy and accordingly determine whether the specified rate of growth is sustainable or not.

The statistical analysis of data regarding an economy may reveal certain areas which might require special attention, e.g., a situation of growing unemployment or a situation of rising prices during past few years. Statistical techniques and principles can also guide the Government in adopting suitable policy measures to rectify such situations. In addition to this, these techniques can be used to assess various policies of the Government in the past. Thus, it is rather impossible to think of a situation where planning and evaluation of various policies can be done without the use of statistical techniques. In view of this it is sometimes said that, "Planning without statistics is a ship without rudder and compass Hence statistics is an important tool for the quantification of various planning policies.

### 5. Importance in business

With the increase in size of business of a firm and with the uncertainties of business because of cut throat competition, the need for statistical information and statistical analysis of various business situations has increased tremendously. Prior to this, when the size of business used to be smaller without much complexities, a single person, usually the owner or manager of a firm, used to take all decisions regarding its business. For example, he used to decide, from where the necessary raw materials and other factors of production were to be obtained, how much of output will be produced, where it will be sold, etc. This type of decision making was usually based on experience and expectations of this single man and as such had no scientific basis.

### Q4. Explain briefly about the scope of statistics.

*Ans :*

Statistics is viewed not just as a device for collecting numerical data, but as a means of sound techniques for their handling, analysis and for drawing valid inference from them. From this perspective, the scope (or) subject matter to statistics can be broadly studied under 2 heads namely,

- (A) Statistical Methods and
- (B) Applied Statistics

#### (A) Statistical Methods

Statistical methods are the tools that are in the hands of the statistician. They include all the general principles and techniques that are commonly used in the collection, analysis and interpretation of data. These methods are applicable to all kinds of data. The stage involved in study of any kind of data are:

- (i) Observation and Collection
- (ii) Organization
- (iii) Presentation
- (iv) Analysis and
- (v) Interpretation

**(B) Applied Statistics**

Applied Statistics deals with application of Statistical methods of specific problems or concrete forms. To illustrate, if a software services firm is experiencing attrition (loss of personnel), it may be worthwhile to investigate the reasons for the same. Special techniques can be employed to understand the underlying trends. For example, a correlation analysis could probably indicate that increasing opportunities (growth in the industry) and attrition are highly correlated. All such techniques and the results obtained by employing such techniques form part of applied Statistics.

Applied Statistics can be further classified into different categories. These categories are

- (i) **Descriptive Statistics:** Descriptive Statistics deals with data that is known. They describe the main features of such data. They are 'basic' to any statistical analysis and bring out such characteristics of data which could have escaped attention of the statistician. Tabulation, Averages and trends are examples of such descriptive statistics.
- (ii) **Scientific Statistics:** Scientific statistics deals with formulation of statistical laws. These laws are based on quantitative data and generally hold true. For example, when we refer to the properties of standard normal distribution. Scientific statistics are heavily used for the purpose of business forecasting.
- (iii) **Analytical Statistics:** This includes methods such as correlation, regression, etc that help in establishing functional relationship between variables. Two or more sets of data are compared and analyzed to arrive at the relationship between them. Such relationship could be continuous or at a defined period of time.
- (iv) **Inferential Statistics:** Statistical methods that help us arrive at certain conclusion based on study of sample

data are part of inferential statistics. It must be noted that no guess work is being made. The size of the sample, criteria for inclusion of a person/activity/thing into the sample, etc are all clearly defined. Based on the results of the study on the sample, inferences about the total population can be drawn.

- (v) **Inductive Statistics:** Statistical methods that help in arriving at general consensus based on a study of random observations are part of inductive statistics. Unlike inferential statistics, the sample chosen under inductive statistics may not 'representative'. Hence, the conclusions arrived at need to be double checked.

**Q5. What are the functions of statistics.**

**(OR)**

**Explain the various functions of statistics.**

**Ans :**

**(Imp.)**

The following are the main functions of statistics:

**1. Presents facts in numerical figures**

The first function of statistics is to present a given problem in terms of numerical figures. We know that the numerical presentation helps in having a better understanding of the nature of a problem. Facts expressed in words are not very useful because they are often vague and are likely to be understood differently by different people. For example, the statement that a large proportion of total work force of India is engaged in agriculture, is vague and uncertain. On the other hand, the statement that 70% of the total work force is engaged in agriculture is more specific and easier to grasp. Similarly, the statement that the annual rate of inflation in a country is 10% is more convincing than the statement that prices are rising.

**2. Presents complex facts in a simplified form**

Generally a problem to be investigated is represented by a large mass of numerical

figures which are very difficult to understand and remember. Using various statistical methods, this large mass of data can be presented in a simplified form. This simplification is achieved by the summarization of data so that broad features of the given problem are brought into focus. Various statistical techniques such as presentation of data in the form of diagrams, graphs, frequency distributions and calculation of average, dispersion, correlation, etc., make the given data intelligible and easily understandable.

### 3. Studies relationship between two (or more phenomena)

Statistics can be used to investigate whether two or more phenomena are related. For example, the relationship between income and consumption, demand and supply, etc., can be studied by measuring correlation between relevant variables. Furthermore, a given mathematical relation can also be fitted to the given data by using the technique of regression analysis.

### 4. Provides techniques for the comparison of phenomena

Many a times, the purpose of undertaking a statistical analysis is to compare various phenomena by computing one or more measures like mean, variance, ratios, percentages and various types of coefficients. For example, when we compute the consumer price index for a particular group of workers, then our aim could be to compare this index with that of previous year or to compare it with the consumer price index of a similar group of workers of some other city, etc. Similarly, the inequalities of income in various countries may be computed for the sake of their comparison.

### Enlarges individual experiences

An important function of statistics is that it enlarges human experience in the solution of various problems. In the words of A.L. Bowley, "the proper function of statistics, indeed is to enlarge individual experience." Statistics is like a master key that is used to

solve problems of mankind in every field. It would not be an exaggeration to say that many fields of knowledge would have remained closed to the mankind forever but for the efficient and useful techniques and methodology of the science of statistics.

### 6. Helps in the formulation of policies

Statistical analysis of data is the starting point in the formulation of policies in various economic, business and government activities. For example, using statistical techniques a firm can know the tastes and preferences of the consumers and decide to make its product accordingly. Similarly, the Government policies regarding taxation, prices, investments, unemployment, imports and exports, etc., are also guided by statistical studies in the relevant areas.

### 7. Helps in forecasting

The success of planning by the Government or of a business depends to a large extent upon the accuracy of their forecasts. Statistics provides a scientific basis for making such forecasts. Various techniques used for forecasting are time series analysis, regression analysis, etc.

### 8. Provides techniques for testing of a hypothesis

A hypothesis is a statement about some characteristics of a population (or universe). For example, the statement that average height of students of a college is 66 inches, is a hypothesis. Here students of the college constitute the population. It is possible to test the validity of this statement by the use of statistical techniques.

### 9. Provides techniques for making decisions under uncertainty

Many a times we face an uncertain situation where any one of the many alternatives may be adopted. For example, a person may face a situation of rain or no rain and he wants to decide whether to take his umbrella or not. Similarly, a businessman might face a situation of uncertain investment opportunities in which he can lose or gain.

He may be interested in knowing whether to undertake a particular investment or not. The answer to such problems are provided by the statistical techniques of decision-making under uncertainty.

**Q6. Discuss the limitations of statistics.**

(OR)

**State the various limitations of statistics.**

(OR)

**List out the limitations of statistics.**

*Ans :*

(Imp.)

**1. Statistics deals with numerical facts only**

Broadly speaking there are two types of facts:

- (a) quantitative and
- (b) qualitative.

(a) **Quantitative:** Quantitative facts are capable of being represented in the form of numerical figures and therefore, are also known as numerical facts. These facts can be analysed and interpreted with the help of statistical methods.

(b) **Qualitative:** Qualitative facts, on the other hand, represent only the qualitative characteristics like honesty, intelligence, colour of eyes, beauty, etc. and statistical methods cannot be used to study these types of characteristics. Sometimes, however, it is possible to make an indirect study of such characteristics through their conversion into numerical figures. For example, we may assign a number 0 for a male and 1 for a female, etc.

**2. Statistics deals only with groups and not with individuals**

Statistical studies are undertaken to study the characteristics of a group rather than individuals. These studies are done to compare the general behaviour of the group at different points of time or the behaviour of different groups at a particular point of

time. For example, the economic performance of a country in a year is measured by its national income in that year and by comparing national income of various years, one can know whether performance of the country is improving or not. Further, by comparing national income of different countries, one can know its relative position vis-a-vis other countries.

**3. Statistical results are true only on the average**

Statistical results give the behaviour of the group on the average and these may not hold for an individual of that very group. Thus, the statement that average wages of workers of a certain factory is ₹ 1,500 p.m. does not necessarily mean that each worker is getting this wage. In fact, some of the workers may be getting more while others less than or equal to ₹ 1,500. Further, when value of a variable is estimated by using some explanatory variable, the estimated value represents the value on the average for a particular value of the explanatory variable. In a similar way, all the laws of statistics are true only on the average.

**4. Statistical results are only approximately true**

Most of the statistical studies are based on a sample taken from the population. Under certain circumstances the estimated data are also used. Therefore, conclusions about a population based on such information are to be true only approximately.

**5. Statistical methods constitute only one set of methods to study a problem**

A given problem m can often be studied in many ways. Statistical methods are used to simplify the mass of data and obtain quantitative results by its analysis. However, one should not depend entirely on statistical results. These results must invariably be supplemented by the results of alternative methods of analyzing the problem. It should be kept in mind that statistics is only a means and not an end.

## 6. Statistics are liable to be misused

Statistical data are likely to be misused to draw any type of conclusion. If the attitude of the investigator is biased towards a particular aspect of the problem, he is likely to collect only such data which give more importance to that aspect. The conclusions drawn on the basis of such information are bound to be misleading. Suppose, for example, the attitude of the Government is biased and it wants to compute a price index which should show a smaller rise of prices than the actual one. In such a situation, the Government might use only those price quotations that are obtained from markets having lower prices.

## 7. Statistics must be used only by experts

Statistics, being a technical subject, is very difficult for a common man to understand. Only the experts of statistics can use it correctly and derive right conclusions from the analysis.

## 1.2 CONCEPTS OF STATISTICS

### 1.2.1 Population and a Sample

#### Q7. Discuss the basic concepts in statistics.

*Ans :*

The following are the basic concepts in statistics.

##### (i) Population

The term Population is used to describe persons, objects, animals, elements, or even reactions that display unique characteristics. If a researcher wants to perform a study on adolescents in Kolkata, then her/his population will be all the adolescents in Kolkata. Though, The Population can be finite or infinite (Mohanty and Misra). An Example of a finite population is the number of students in a school who have passed in psychology and an Example of an infinite population would be the number of stars in the sky.

##### (ii) Sample

The term Sample can be described as the group of individuals who participate in the

research. For Example, in a cold drink factory, if the quality inspector wants to find out the quality and taste of the cold drink that is adequate or not, then she/he will not test all the bottles of cold drink. She/He takes a sample and test it. Ideally, this sample is taken randomly. That's how a sample is a smaller group from the population that participates in the research.

The sample must be representative of the population. It is as much as possible similar to the population or has the same elements/characteristics as the population (Mohanty and Misra). Thus, sampling techniques are applicable in research. It helps in the selection of samples. Though, Sampling Techniques can be classified into probability sampling and non-probability sampling.

##### (iii) Parameter

The term Parameter is called as a value that provides information about the population which is examined in the research. It can be explained as "a measure of the population and refers to the indices of a central value, dispersion, correlation and so on of all the individuals of the population" (Mohanty and Misra). For example, if a researcher wants to know the mean weight of newly born infants in India in a given year, this can be called a parameter. It also describes the weight of all the newly born infants in India in the given year. An exact parameter isn't always easy to obtain and all parameters will have a different statistic.

##### (iv) Statistic

As aspects of a population are measured by a parameter and aspects of a sample are measured by Statistics. Accordingly, the researcher will measure the weight of approximately 500 newly born infants (a sample representing all the newborn infants) in a given year and calculate a mean weight. This means weight can be called as a statistic.

##### (v) Variable(s)

Variable signify something that varies. It can also be described as a number or quantity

that will differ or will have different values. Emotional intelligence and self-esteem can be termed as variables in this study. Emotional intelligence can be low or high as can self-esteem be. Both these variables can have differing values. Even gender can be called a variable because it will differ in terms of males or females.

### 1.3 QUANTITATIVE AND QUALITATIVE DATA

**Q8. What is Quantitative Data? State its strength and weakness.**

*Ans :*

#### Meaning

Quantitative data is a set of numbers collected from a group of people and involves statistical analysis. For example if you conduct a satisfaction survey from participants and ask them to rate their experience on a scale of 1 to 5. You can collect the ratings and being numerical in nature, you will use statistical techniques to draw conclusions about participants satisfaction.

#### Strength

- **Specific:** Quantitative data is clear and specific to the survey conducted.
- **High Reliability:** If collected properly, quantitative data is normally accurate and hence highly reliable.
- **Easy communication:** Quantitative data is easy to communicate and elaborate using charts, graphs etc.
- **Existing support:** Many large datasets may be already present that can be analyzed to check the relevance of the survey.

#### Weakness

- **Limited Options** - Respondents are required to choose from limited options.
- **High Complexity** - Qualitative data may need complex procedures to get correct sample.
- **Require Expertise** - Analysis of qualitative data requires certain expertise in statistical analysis.

**Q9. What is Qualitative Data? State its strength and weakness.**

*Ans :*

#### Meaning

Qualitative data is a set of information which can not be measured using numbers. It generally consist of words, subjective narratives. Result of an qualitative data analysis can come in form of highlighting key words, extracting information and concepts elaboration. For example, a study on parents perception about the current education system for their kids. The resulted information collected from them might be in narrative form and you need to deduce the analysis that they are satisfied, un-satisfied or need improvement in certain areas and so on.

#### Strength

- Better understanding - Qualitative data gives a better understanding of the perspectives and needs of participants.
- Provide Explanation - Qualitative data along with quantitative data can explain the result of the survey and can measure the correction of the quantitative data.
- Better Identification of behavior patterns - Qualitative data can provide detailed information which can prove itself useful in identification of behavioral patterns.

#### Weakness

- Lesser reachability - Being subjective in nature, small population is generally covered to represent the large population.
- Time Consuming - Qualitative data is time consuming as large data is to be understood.
- Possibility of Bias - Being subjective analysis; evaluator bias is quite feasible.

### 1.4 COLLECTION OF PRIMARY AND SECONDARY DATA

**Q10. Define :**

- (i) Primary Data
- (ii) Secondary Data

**Ans :**

**(i) Primary Data**

As the name suggests is original, problem (or) project-specific and collected for the specific objectives and needs spelt out by the researcher. The authenticity and relevance is reasonably high. The monetary and resource implications of this are quite high and sometimes a researcher might not have the resources or the time or both to go ahead with this method. In this case, the researcher can look at alternative sources of data which are economical and authentic enough to take the study forward. These include the second category of data sources namely the secondary data.

**(ii) Secondary Data**

As the name implies is that information which is not topical or research specific and has been collected and compiled by some other researcher or investigative body. The said information is recorded and published in a structured format, and thus, is quicker to access and manage. Secondly, in most instances, unless it is a data product, it is not too expensive to collect. As suggested in the opening vignette, the data to track consumer preferences is readily available and the information required is readily available as a data product or as the audit information which the researcher or the organization can procure and use it for arriving at quick decisions.

**Q11. What are the various Sources of Primary Data ?**

**(OR)**

**Discuss various sources for collecting Primary Data.**

**Ans :**

Primary data may be collected either through observation or through direct communication with respondents in one form or another through personal interviews.

There are several ways of collecting primary data.

**(i) Observation Method**

In observation method, the information is sought by way of investigator's own direct observation without asking from the respondent. The main advantage of this method is that it is free from subjective business, as it is free from respondent's willingness. It is, however, an expensive and time consuming method. Moreover, the information provided by this method is very limited and some of the more busy people like executives may not be accessible to direct observation.

**(ii) Interview Method**

(a) In the personal interviews the interviewer asks questions generally in a face to face contact. Through interview method more and reliable information may be obtained. Personal information can be obtained easily under this method. It is, however, a very expensive and time consuming method, especially when large and widely spread geographical sample is taken. Certain types of respondents, such as officials, executives or people of high income groups, may not be easily accessible.

(b) In telephonic interviews contact is made with the respondents through telephone.

**(iii) Questionnaire Method**

In this method a questionnaire is mailed to the person concerned with a request to answer the questions and return the questionnaire. This method is most extensively applied in various researches of human and economic geography.

**(iv) Schedule Method**

This method of data collection is very much like the collection of data through questionnaires, with little difference that lies in the fact that schedules (are being filled in by the enumerators who are specially appointed for this purpose. Enumerators explain the aims and objects of the investigation and also remove the difficulties which any respondent may feel in understanding the implications of a particular question.

This method is very useful in extensive enquiries and can lead to fairly reliable results. It is, however, very expensive and is usually adopted in investigations conducted by governmental agencies or by some organizations. Population census all over the world is conducted through this method.

#### **Q12. What are the advantages and disadvantages of primary data ?**

*Aus :* (Imp.)

##### **Advantages**

Following are the advantages of using primary data :

1. **Accuracy:** It is more accurate and reliable as it relates to the objectives of the study.
2. **Original:** It is original in character as it is based on the first hand information.
3. **More Detailed:** It is more detailed and it is according to the requirements of the study and or defined research problem.
4. **Authentic:** It is more authentic as it is collected by the investigator himself and recognizes the origin of data.
5. **Necessity:** collection of primary data becomes necessary when secondary data is not available for the study under question.

##### **Disadvantages**

Primary data has the following demerits :

1. **More Expensive:** Collection of primary data requires more finance and it is expensive process as well.
2. **More Labour:** Many people are required to collect data from primary sources, hence it labour intensive task.
3. **More Time:** Collection of primary data requires more time as data are personally collected by the investigator.
4. **More Skills:** For collection of primary data investigator must be highly trained, skilled and qualified in that field.

5. **Non Availability:** Sometimes due to various limitations it is not possible to collect primary data. i.e., data regarding military establishment of a country are not possible to collect.

#### **Q13. Discuss the method of collecting data through secondary data.**

*Aus :*

Secondary data are ready made data which are already collected and used for some purpose by some one or more agencies. As such, there is no necessity of employing any specific method for the collection of secondary data as are employed for collecting the primary data.

However, for collecting the secondary data, the investigator may employ either of the following method :

##### **(i) Personal Visit to the Sources**

Under this method, the researcher can personally visit the source of information viz. libraries, museums, government offices, private and semi-government institutions and private persons and requests them to provide him with the necessary information and permit him to note down the required data there form. For this purpose the researcher may be required to spend some time and make a vivid study at the source of the information. This method is comparatively more expensive, time consuming and arduous in nature, but it gives more accurate and dependable results.

##### **(ii) Correspondence with the Authorities**

As per this method, the researcher can make correspondence with the authorities of the relevant source for getting the desired data by post or through some special messengers. For this purpose researcher has to write down letters and reminders to the authorities concerned in the humble language to supply the necessary data for the required study. He also appraises them of the need and importance of his study and seeks their cooperation in the matters. He should also assure them to return the data after use and to keep the information secret if needed. In this method, the data are collected usually in

the form of report, circulars, bulletins, statutes etc. This method of collecting the data is quite inexpensive and comfortable provided the authorities of the source are quite kind to extend their corporation. However, this method of collecting the data may be very time consuming and vexatious.

### (iii) Subscription to the Periodicals

To collect the secondary data under this method, the researcher can subscribe to certain periodicals such as Magazines, Journals, newspapers etc., regularly to obtain the relevant data there from.

This method of obtaining a secondary data is, no doubt, very comfortable but at the same it is very much time taking and expensive. Moreover, the researcher may not get all the information from the periodicals to which he may subscribe. He has to seek for some other sources for this purpose. Thus, this method is inadequate and undependable. However, if the field of enquiry is very confined and small such as a particular institution (or) organization, this method of collecting a secondary data is most suitable. In addition to it researcher can also visit various internet sites to collect the secondary data.

## Q14. What are the Advantages & Disadvantages of Secondary Data ?

(OR)

### State the Merits and Demerits of Secondary Data.

*Ans :*

#### Advantages

##### 1. Economy

Such data is cheaper. The amount of money spent in acquiring secondary data is generally less than that needed to obtain primary data. The various secondary data from libraries can be obtained at no cost.

##### 2. Quickness

Most of these data are 'instant' since they already exist and merely need to be

discovered. Thus, the time in collecting secondary data is largely search time and usually requires few hours or days.

##### 3. Quality

An individual investigator cannot match the quality or size of the firms that obtain much of the existing secondary data. The information is gathered by trained personnel specialised in data collection. Also many organizations may not release their data to individual researcher but may give it to firms.

##### 4. No Need of Measuring Instruments

When information is gathered from secondary sources, there is a problem in designing information gathering instruments as information is already collected by someone else. The only problem is to locate the appropriate source and method of recording desired information.

##### 5. Availability

Secondary data is sometimes available even in those cases where primary investigations are not possible.

##### 6. Bases for Comparison

Secondary data is useful in the case of exploratory researches as they provide increased understanding of the problem.

##### 7. Useful in Exploratory Research

Secondary data act as a basis for comparison after primary data is collected.

##### 8. Generates Feasible Alternatives

Secondary data are very useful in generating viable alternatives to solve problems. The multiplicity of data sources, research approaches and managerial styles usually lead to a number of possibilities which should be examined by the researcher.

#### Disadvantages

It is however difficult to find secondary data that exactly fits into the needs of some specific research investigation. The problems experienced in respect of secondary sources can be in terms of:

**1. Relevance**

The data may not fit into the needs of investigation. There may be difference in the units of measurement; there may be surrogated data; discrepancy of classes and data may pertain to some other period of time.

**2. Accuracy**

It is observed that it is rather difficult to measure the degree of approximations used in the collection of information as well as the competence of the investigator in motivating the persons to supply the desired information.

**3. Existence of Obsolete Information**

Information may be outdated or obsolete.

**4. Non-Disclosure of Research Findings**

All the findings of a research study may not be made public.

**5. Seldom Catering to the Need**

The available data may not suit the current purpose of research, due to incompleteness, generalities and so on.

**6. Other Limitations**

- i) There may be difficulties in the identification of the source.
- ii) Errors may be there in recording or transferring information from secondary sources.
- iii) The facilities or capabilities of the agency that originally collect the data might be questionable.

**Q15. What are the differences between primary data and secondary data ?**

(Imp.)

*Ans :*

S.No.	Criteria	Primary Data	Secondary Data
1.	Accuracy	Primary data is more accurate	Secondary data is less accurate in comparison to primary data.
2.	Cost	Collection of primary data is more expensive.	Collection of secondary data is relatively less expensive. It requires huge funds.
3.	Time	Collection of primary data requires more time.	Secondary data is collected in less time. It does not require so much time.
4.	Source of Information	Primary data is collected by investigator(s) from the respondents directly.	Secondary data are second hand information and these are collected by some other agency.
5.	Original	Primary data is original in nature because it is collected for the first time.	Secondary data is not original in nature because it has been used many times.
6.	Reliability	Primary data is more reliable and suitable.	Secondary data is comparatively less reliability.
7.	Extra Caution	In the use of primary data no special precautions are followed.	Use of secondary data requires extra cautions to be taken into account.
8.	Concern with problem	Primary data is directly related with the problem under study.	Secondary data is not directly related with problem.

## **1.5 CLASSIFICATION AND TABULATION OF DATA**

**Q16. Define classification. What are the objectives of classification?**

*Ans :*

### **Meaning**

The largest amount of data and the greatest amount of detail may not convey the most useful information for decision making. An important aspect of Statistics is to organize and present data so as to convey critical information quickly.

Classification means grouping of a whole into different groups or classes. However, each of these groups should have a common characteristic.

For example, all living things can be classified into plants and animals. Animals can be classified into Amphibians, Reptiles, and Mammals etc.

### **Definition**

**According to Horace Secrets,** "Classification is the process of arranging data into sequences and groups according to their common characteristics, or separating them into different but related parts".

### **Objectives**

1. To condense the mass of data into a form that is easily understandable.
2. To enable the user to get a proper grasp of the significance of the information contained in the data.
3. To sort data in such a manner that irrelevant details are ignored and relevant details only are considered.
4. To present the data in tabular form and enable the user to apply statistical tools and techniques on the same
5. To enable interpretation, analysis and generalization of data
6. To facilitate comparative study of variables.
7. To establish relationships between variable on study of various characteristics of data

8. To bring out similarities and disparities in data collected.
9. To provide an executive summary of the data collected at a glance.

**Q17. What are the different types of classifications?**

*Ans :*

The nature of classification depends upon the purpose and objective of investigation. The following are some very common types of classification:

#### **1. Geographical (or spatial) classification**

When the data are classified according to geographical location or region, it is called a geographical classification.

#### **2. Chronological classification**

When the data are classified on the basis of its time of occurrence, it is called a chronological classification. Various time series such as National Income figures (annual), annual output of wheat, monthly expenditure of a household, daily consumption of milk, etc., are some examples of chronological classification.

#### **3. Conditional classification**

When the data are classified according to certain conditions, other than geographical or chronological, it is called a conditional classification.

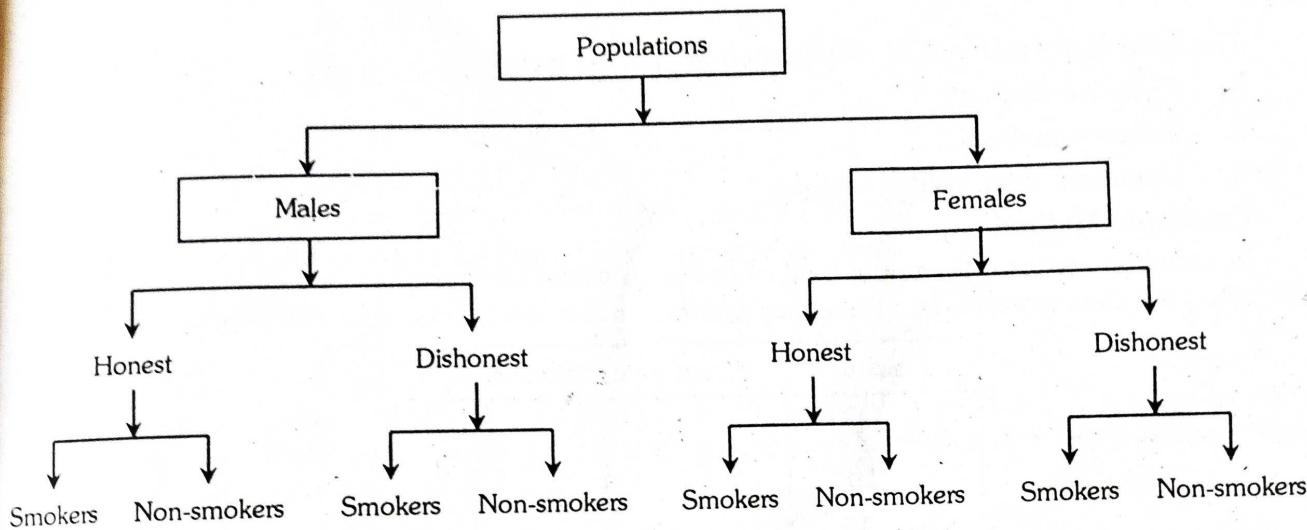
#### **4. Qualitative classification or classification according to attributes**

When the characteristics of the data are non-measurable, it is called a qualitative data. The examples of non-measurable characteristics are sex of a person, marital status, colour, honesty, intelligence, etc.

These characteristics are also known as attributes. When qualitative data are given, various items can be classified into two or more groups according to a characteristic. If the data are classified only into two categories according to the presence or absence of an attribute, the classification is termed as dichotomous or two-fold classification. On the other hand, if the data are classified into

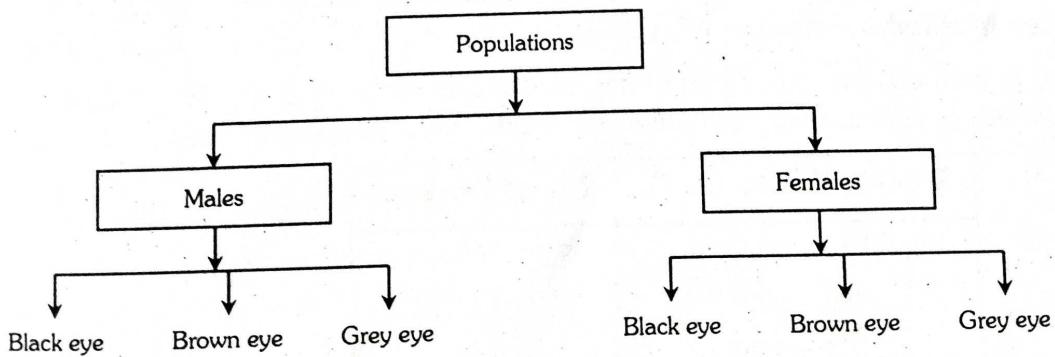
more than two categories according to an attribute, it is called a manifold classification. For example, classification of various students of a college according to the colour of their eyes like black, brown, grey, blue, etc. The conditional classification, given above, is also an example of a manifold classification.

If the classification is done according to a single attribute, it is known as a one-way classification. On the other hand, the classification done according to two or more attributes is known as a two-way or multi-way classification respectively. The example of a three-way classification, where population is dichotomised according to each attribute, sex, honesty and smoking habit, is shown as following:



We note that there will be eight sub-groups of individuals like (male, honest, smokers), (male, honest, non-smokers), etc.

In the classification, given above, the population is dichotomised with respect to each of the three attributes. There may be situations where classification with respect to one attribute is dichotomous while it is manifold with respect to the other. A two-way classification of this type is shown as:



## 5. Quantitative classification or classification according to variables

In case of quantitative data, the characteristic is measurable in terms of real numbers and is termed as variable, e.g., weight, height, income, the number of children in a family, the number of crime cases in a city, life. of an electric bulb of a company, etc. A variable can take a different value corresponding to a different item of the population or universe.

Variables can be of two types:

- (a) **Discrete Variable:** A discrete variable can assume only some specific values in a given interval. For example, the number of children in a family, the number of rooms on each floor of a multistoried building, etc.
- (b) **Continuous Variable:** A continuous variable can assume any value in a given interval. For example, monthly income of a worker can take any value, say, between ₹ 1,000 to 2,500. The income of a worker can be ₹ 1,500.25, etc. Similarly, the life of an electric bulb is a continuous variable that can take any value from 0 to  $\infty$ .

#### **Q18. Explain the different methods of classifications?**

*Ans :*

The three major methods of classification of data are as follows,

1. Exclusive method
2. Inclusive method
3. Open-end class interval method

##### **1. Exclusive Method**

In exclusive method, the upper limit of one class would be the lower limit of the successive class when the class intervals in a frequency distribution are fixed. This is explained below,

Marks	Number of Students
0 – 5	2
5 – 10	4
10 – 15	5
15 – 20	12
20 – 25	16
25 – 30	22
30 – 35	25
35 – 40	27

##### **2. Inclusive Method**

In inclusive method, the upper limit of one class is also included in the similar class itself when the class intervals in a frequency distribution are fixed. This is explained below,

Weekly Wages (in ₹)	No. of Workers
0 – 50	7
51 – 100	11
101 – 150	14
151 – 200	16
201 – 250	18
251 – 300	22
301 – 350	26
351 – 400	29

### 3. Open-end Class Interval Method

An open-end class interval method does not provide any upper limit and lower limit. This is explained below.

Monthly Income (in ₹)	Number of Persons
Below 500	8
501 – 1000	14
1001 – 1500	22
1501 – 2000	27
2001 – 2500	34
2501 – 3000	38
Above 3001	43

### Q19. What is tabulation? What are the objectives of tabulation?

Ans :

(Imp.)

#### Meaning

A Table is a systematic arrangement of statistical data in columns and rows. Rows are horizontal arrangements, whereas columns are vertical. Tabulation is systematic presentation of data in a form suitable for analysis and interpretation.

#### Objectives

1. **Simplify Data:** The main objective of tabulation is to simplify a mass of information into its simplest form, making it easy to understand.
2. **Provide Clarity:** Tabulation provides clarity on data collected, Many of the questions that are related to the objective of the study can be easily answered through tabulation
3. **Facilitate Comparison:** Tabulation facilitates easy comparison. The two competing options that need to be evaluated can be placed side by side and compared.
4. **Trend Analysis:** Tabulation helps to catch the trend by more observation. Tabulated data is well placed to perform statistical calculation of trend and other features of data.
5. **Economy of Space and Time:** Presentation of data in tabular form saves space without any compromise on quality or completeness of data. Not only presentation but usage of data also results in savings of time for the users of such data.
6. **Enables Easy Reference:** Tabulated information with appropriate title and number are easy to remember and convenient to refer at future date. Most Research papers will have a separate Table of contents listing down the various table in which useful statistical data is presented and their page numbers.
7. **Detections of Errors and Omissions:** Particularly in case of numeric data, tabulation serves as a control. If any of the totals of rows or columns are not as they should be, it gets immediately known.
8. **Facilitate Statistical Processing:** It is only after tabulation that data becomes fit for statistical processing we can apply various statistical methods and techniques such as correlation, calculation of various measures of central tendency and dispersion etc, only after data is correctly tabulated.

- 9. Clarity on Characteristics of Data:** A concise tabular form clearly reveals the characteristics of data and highlights its significant characteristics.

**Q20. Explain the Essential parts of a table.**

**Ans :**

**1. Table Number**

Indicates the serial number of the table. Table Numbers help in easy identification and reference of the table in future. Hence, they should be stated in clear and visible terms. It is preferred to write the Table Number at the top of the table, usually in center.

**2. Table Title**

Indicates the subject matter of the table. The title should be as unambiguous as possible. It should be clearly worded and indicate the nature of data contained in the table. The title should ideally indicate then 'what, where and How of the classified data and hint at the period of which it relates. Titles should not be too long, However, if the title is becoming very lengthy, it is advised to have a 'catch-line' with the rest of title stated in the next line. The table title should be very prominent in bold, and at the center, just below the Table Number.

**3. Captions**

It refers to the headings of the columns. Captions usually have a heading and also sub-heads, depending on the type of data being tabulated. A caption should be brief, concise and self-explanatory. Care should be taken that each of the column names does not spill into other column areas. It is preferable to number the captions and the sub-heads for case of reference, particularly when calculations are involved.

**4. Sub**

It refers to the headings of the rows. As with captions, rows can also have further sub-divisions. However, stubs do not need to be very brief. Often, the choice of arrangement is done in such a way that the more descriptive items are taken as stubs. As in case of captions, it is desirable to number the stubs as well.

**5. Body**

It refers to the numerical information entered into the table. The entire statistical data that is to be presented constitutes the 'body'. Tabulation is an effort to present this 'body' in a more meaningful manner.

**6. Head Note**

It is a brief explanatory statement with respect to some or all parts of the table to facilitate better understanding of the nature of information provided in the table.

**7. Foot Note**

Footnotes provide further explanation with regard to the contents of the table. They may include exceptions to data, classifications etc.

**8. Source Data**

In case the table is not generated from primary data collected by the statistical investigator, the source from which data is collected is given at the bottom of the table. The source is stated below the footnotes. Stating the source adds credibility to the data. Usually, we state the 'original source' of data along with other indicators such as page number, table number, etc so that it is easy for anybody to refer to it at any time.

**Q21. What are the differences between classification and tabulation?**

*Ans :*

Sl. No.	Basis for Comparison	Classification	Tabulation
1.	<b>Meaning</b>	Classification is the process of grouping data into different categories, on the basis of nature, behavior, or common characteristics.	Tabulation is a process of summarizing data and presenting it in a compact form, by putting data into statistical table.
2.	<b>Order</b>	After data collection	After classification
3.	<b>Arrangement</b>	Attributes and variables	Columns and rows
4.	<b>Purpose</b>	To analyse data	To present data
5.	<b>Bifurcates data into</b>	Categories and sub-categories	Headings and sub-headings

### 1.6 CONSTRUCTION OF UNIVARIATE AND BIVARIATE FREQUENCY DISTRIBUTION

**Q22. Explain univariate tabulation with an example.**

*Ans :*

Uni-variate analysis is one of the simplest form of statistical analyses which includes only one question. This one question has to be tabulated and this is called as a univariate tabulation. Univariate tables are usually used when the question itself gives the vital important for the analysis. Univariate table can be explained as follows,

**Example**

Find the percentage of the customers opinions regarding the product performance.

Response Category	Number of Responses	Percentage
Excellent	40	26.66
Good	50	33.33
Fair	30	20
Poor	30	20
<b>Total</b>	<b>150</b>	<b>100</b>

33.33% of the responses opinion is good regarding the product and 20% opined to be fair and 20% opined for poor performance of the product.

## **Q23. What do you understand bivariate / multivariate tabulation.**

**Ans :**

### **Bivariate/Multivariate Tabulation**

Even though majority of the researchers use univariate tables based on the responses to only one question yet, they always attempt to get more helpful information by counting the responses to two or more queries either together or individually.

#### **Example**

It might be possible that women are not smokers. But women from a specific high income group can be smokers. In order to analyze this, we need to club the category of 'women' and 'high income group'.

While on the other hand, the univariate tabulation leads in frequency distribution of responses to one question, the bivariate tabulation results in a frequency distribution of responses to two or more questions. The bivariate/multivariate tabulation comprises a table having two or more rows and two or more columns.

In order to include the responses to a third question. The rows and columns needs to be divided. Otherwise, a separate bivariate tables can also be created based on the number of classes, in the third question. Depending on the class of responses to the third question, a separate bivariate table is prepared which includes the responses to the first two question.

An example of bivariate/multivariate tabulation that represents the first alternative is as follows,

M/F	Age	Agree	Disagree	Not Sure	Total
Men	Below 35	-	-	-	-
	Above 35	-	-	-	-
	Total	-	-	-	-
Women	Below 35	-	-	-	-
	Above 35	-	-	-	-
	Total	-	-	-	-

From the above table, it is being observed that the table can be considered as two separate bivariate tables one for women and another for men.

In bivariate/multivariate tabulations, the main problem is in choosing the correct combinations of questions out of the large number of available combinations. It is essential for a researcher to select the right combination clearly. Otherwise it would be a waste of time in creating unnecessary or complex tables which not only creates the chaos among the reader but also provides unnecessary information. The objectives of the research can help in selecting the appropriate combination.

**Q24. Differentiate between univariate and multivariate data.***Ans :*

Differences between univariate and bivariate data.

(Jan.-20)

Sl.No.	Univariate Data	Sl.No.	Bivariate Data
1.	It involving a single variable	1.	It involving two variables
2.	It does not deal with causes or relationships	2.	It deals with causes or relationships
3.	The major purpose of univariate analysis is to describe	3.	The major purpose of bivariate analysis is to explain
4.	Central tendency - mean, mode, median	4.	Analysis of two variables simultaneously
5.	Dispersion - range, variance, max, min, quartiles, standard deviation.	5.	Correlations comparisons, relationships, causes explanations
6.	Bar graph, histogram, pie chart, line graph, box-and-whisker plot	6.	Tables where one variable is contingent on the values of the other variable.

**Q25. What is frequency distribution? Explain the formation of discrete and continuous frequency distribution.***Ans :*

(Imp.)

**Meaning**

Frequency distribution refers to the tabular arrangement of data, when arranged into groups or categories according to conveniently established divisions of the range of the observations. In a frequency distribution, raw data is expressed in distinct groups called 'classes'. The number of observations that fall into each of these classes is known as 'frequency'. Thus, a frequency distribution has two parts, the left part represents the classes and the right part represents the frequencies.

Arrangement of raw data in the form of row or column is called data array. When data is described by a continuous variable, it is called continuous data and when described using discrete variable it is called discrete data.

The following are the two examples of discrete and continuous frequency distributions.

No. of Students	No. of Colleges
100	15
120	25
140	32
160	50
180	27
200	25

Table (1): Discrete Frequency Distribution

Age (In years)	No. of Students
10-15	12
15-20	15
20-25	24
25-30	30
30-35	20
40-45	22

Table (2): Continuous Frequency Distribution

### i) Construction of a Discrete Frequency Distribution

The process of preparing frequency distribution is very simple. For discrete data, place all the possible variables, values in ascending order in one column and then prepare another column for 'Tally marks' (||) to count the number of marks. A block of five marks is prepared and some space is left in between the blocks. The frequency column refers to the number of "Tally marks", a particular class will contain. The construction of a discrete frequency distribution is illustrated as follows.

Consider a sample study in which 40 families were surveyed to find the number of children in each family. The data obtained are,

3	2	2	1	3	4	2	1	3	4
5	0	2	1	2	3	3	2	1	1
2	3	0	3	2	1	4	3	5	5
4	3	6	5	4	3	1	0	6	5

'Tally marks' are used to condense this data into a discrete frequency distribution as shown below.

No. of Children	No. of Families	Frequency
0		3
1		7
2		8
3		10
4		5
5		5
6		2
Total		40

Table: Discrete Frequency Distribution

### ii) Construction of a Continuous Frequency Distribution

Before constructing the frequency distribution for continuous data, it is necessary to identify some of the important terms that are used frequently.

#### ➤ Class Limits

The class limits represent the smallest and the largest values to be included in a class. The lowest and highest boundaries of a class, are known as the lower and upper limits of the class respectively.

For example, in a class among the values ranging between 70-79, 70 is the lower limit and 79 is the upper limit or we can say that there can be no value which is less than 70 and greater than 79.

➤ **Class Intervals**

The class interval represents the width (span or size) of a class. The width may be determined by subtracting the lower limit of one class from the lower limit of the following class (alternatively successive upper limits may be used).

For example, if the two classes are 20 – 30 and 30 – 40, the width of the class interval would be the difference between the two successive lower limits i.e.,  $30 - 20 = 10$  or the difference between the upper limit and lower limit of the same class, i.e.,  $30 - 20 = 10$ .

➤ **Class Midpoint**

Midpoint of a class is defined as the average of the lower and upper limits of a class. The value thus obtained lies between the lower and upper class limits. For example the midpoint of a class 10-20 is,

$$\frac{10 + 20}{2} = \frac{30}{2} = 15$$

➤ **Class Frequency**

The number of observations that fall into each of the class is known as the class frequency. For example, the class frequency for the class interval 20 – 30 is '10'.

### 1.7 DIAGRAMMATIC AND GRAPHICAL REPRESENTATION OF DATA

**Q26. Define diagram and diagrammatical representation. Explain general rules for constructing diagrams.**

*Ans :*

(Imp.)

**Meaning**

(i) **Diagrams**

Diagram is a visual presentation of statistical information. The pictorial presentation helps in proper understanding the data. Diagrams are of different types like pie diagram, rectangles, lines, pictures and maps.

(ii) **Diagrammatic Representation of Data**

Apart from classification and tabulation methods of efficiently presenting the statistical data, the statistical data can also be presented in a convincing, appealing and easily understandable form using the diagrammatic representation (i.e., making use of diagrams).

**General Rules For Constructing Diagrams**

The following general rules should be observed while constructing diagrams:

**1. Title**

Every diagram must be given a suitable title. The title should convey in as few words as possible the main idea that the diagrams intend to portray. However, the brevity should not be secured at the cost of clarity or omission of essential details. The title may be given either at the top of the diagram or below it.

**2. Proportion between Width and Height**

A proper proportion between the height and width of the diagram should be maintained. If either the height and width is too short or too long in proportion, the diagram would give an ugly look. While there are no fixed rules about the dimensions, a convenient standard as suggested by Lutz in the book entitled "Graphic Presentation" may be adopted for general use. It is known as "Root-two" that is, a ratio of 1 (short side) to 1.414 (long side). Modifications may, no doubt, be made to accommodate a diagram in the space available.

**3. Selection of Scale**

The scale showing the values should be in even numbers or in multiples of five or ten, e.g., 25, 50, 75, or 20, 40, 60. Odd values like 1, 3, 5, 7 should be avoided. Again no rigid rules can be laid down about the number of rulings on the amount scale, but ordinarily it should not exceed five. The scale should also specify the size of the unit and what it represents; for example, "million tonnes", "number of persons in thousands", "units produced in lakhs", etc. All lettering should be easily readable without turning the chart up and down.

**4. Footnotes**

In order to clarify certain points about the diagram, footnote may be given at the bottom of the diagram.

**5. Index**

An index illustrating different types of lines or different shades, colours should be given so that the reader can easily make out the meaning of the diagram.

**6. Neatness and Cleanliness**

Diagrams should be absolutely neat and clean.

**7. Simplicity**

Diagrams should be as simple as possible so that the reader can understand their meaning clearly and easily. For the sake of simplicity, it is important that too much material should not be loaded in a single diagram otherwise it may become too confusing and prove worthless. Several simple charts are often better and more effective than one or two complex ones which may present the same material in a confusing way.

**Q27. Explain advantages and disadvantages of diagrams.**

*Ans :*

**Advantages**

1. Diagrams give a clear picture of the data. Data presented with the help of diagrams can be understood and grasp even by a common man in a very short time.
2. Diagrammatical representation of data can be used universally at any place.
3. Diagrammatical representation not only saves time and energy but also is economical.
4. By using diagrammatic representation, data can be condensed.
5. By diagrammatic representation, comparison between data can be made without actually computing the statistical measures.
6. Diagrams are more impressive, attractive and fascinating compared to any other form of representation.

7. More information can be obtained using diagrammatic representation when compared to tabular representation of data.

8. Diagrams have an appealing effect to the user and are more easily remembered than any other representation.

**Disadvantages**

Some of the disadvantages of diagrammatic presentation are,

1. Diagrams do not show exact values.
2. Drawing diagrams is a difficult task.
3. Skills are required to present the data in a diagrammatic form.
4. Diagrammatic presentation of data reveals limited facts.
5. Diagrams are not substitute for tabular presentation.
6. Diagrams give only rough idea. Analyzing the data in detail is not possible through diagrams.
7. Diagrams drawn through false data give false idea.

**Q28. What do you mean by graphical representation of data ? State the rules for graphing. What are the advantages and disadvantages of graphs ?**

*Ans :*

**Meaning**

The graphical representation refers to the way of representing the data with the help of graphs. Graphs play a significant role in representation of data related to time series. The frequency distributions can also be efficiently represented using graphs. The categorical and geographical data can be best represented using a diagrammatic representation. Similarly, graphical representation proves to be the best approach for representing the data related to time series and frequency distribution.

**Rules**

The various rules for constructing graphs are,

**1. Title**

A heading depicting the matter and contents of the data must be provided as title for all the graphical representations. The title must be clear, simple and precise.

**2. Organization of the Graph**

Each graph must be structurally drawn in an attractive manner. The portion of axis must be chosen and drawn neatly. We must represent independent and dependent parameters on X and Y axis respectively.

**3. Scale**

The scale selected must satisfy all the values to be plotted on the graph.

**4. Index**

The index must be provided to show the scale of X and Y axis. The various lines drawn in the graph must be defined clearly in the index.

**5. Source of Data**

The source of data gives the information about the data retrieval and is mentioned at the bottom of the graph.

**6. False Base Line**

The vertical scale must start with zero for representing the given variable effectively. When the fluctuations are maximum and the starting values are very far from zero, a false base line is used to break the vertical scale and to plot the values.

To draw a false base line, the vertical scale is broken into two parts and the values of the dependent variable from zero to the lowest value are omitted by drawing two zig-zag horizontal lines above the base line i.e., X-axis.

**7. Line Designs**

In case of representing several variables on the same graph for comparison, we may differentiate between them by using dotted lines, dash-dotted lines, broken lines, thick lines and thin lines. An index must also be provided to clarify the meanings of the lines.

**Advantages**

Some of the advantages of graphical presentation are,

- (i) Graphs represent the data in a simple manner which is easy to understand by the users.
- (ii) The curves drawn on graphs are helpful in determining the values which are not available.
- (iii) Easy comparisons can be made, if the data is presented in a graphical manner.
- (iv) Graphs are highly attractive.
- (v) Graphs can be used for further data analysis.
- (vi) Graphs are useful in extrapolation and interpolation methods.

**Disadvantages**

There are certain disadvantages of Graphs. They are,

- (i) One should possess skills to understand the graphs.
- (ii) Graphs are not precise when compared to tables.
- (iii) Constructing a graph is a time consuming process.
- (iv) At times, graphs create problems.
- (v) Graphs provide restricted information.

**Q29. List and explain the different ways of representing frequency distribution using graphs.****Ans :**

(Imp.)

The graphical method of representing the frequency distribution is very useful. The frequency graphs provide a clear view, which can be easily understood and remembered by the users as compared to the tabulated data. The frequency distribution can be presented graphically in any of the following ways,

1. Histogram
2. Frequency polygon
3. Smoothed frequency curve
4. Cumulative frequency curves or Ogives.

## 1. Histogram

One of the most commonly used and easily understood methods for graphical representation of frequency distribution data is called 'histogram'. It is also known as 'column diagram', depicting the class frequencies in a frequency distribution by vertical adjacent rectangles.

## 2. Frequency Polygon

A frequency polygon is a graph representing frequency distribution. It has more than four sides and is very helpful in comparing several frequency distributions plotted on the same graph. This feature of frequency polygon is an advantage over histogram.

The frequency polygon can be constructed in two ways. They are,

- (a) Using histograms
- (b) Using midpoints.

(a) **Using Histograms:** A histogram is drawn by using the given data. Identify the midpoints of the upper horizontal side of the rectangles. Join the midpoints by a line. The figure thus obtained is known as 'frequency polygon'. The area of the frequency polygon is approximated by joining the ends of the polygon with the base line and by assuming that both the frequencies of the class before the first class and the class after the last class are zero.

(b) **Using Midpoints:** Calculate the midpoints of the various class intervals. Plot the midpoints and their corresponding frequencies on the graph. Join these points by a straight line. The figure thus formed is called 'frequency polygon'.

## 3. Smoothed Frequency Curve

A smoothed frequency curve is a free hand curve obtained from the various points of the polygon. The area included under the curve is approximately the same as that of polygon.

It is another way of presenting data where only frequency distributions based on samples should be smoothed. The continuous series thus generated is smoothed. The smoothed frequency curve is constructed by first drawing a histogram, plotting the frequencies at the midpoints of class intervals that results in a polygon. Finally, it is smoothed to produce a smoothed frequency curve. Therefore, the total area under the curve is equal to the area under the original histogram or polygon.

## 4. Cumulative Frequency Curves (or) Ogives

The cumulative frequency curve also known as Ogive is a graph of cumulative frequency distribution. These cumulative frequencies are obtained by adding the frequencies and listing them in a tabular form.

The two techniques for constructing an Ogive are,

- (a) More than technique
- (b) Less than technique.

(a) **More than Technique:** Here, more than cumulative frequencies are plotted on Y-axis and the lower limit of the class interval on X-axis. The points are then joined to obtain a smooth free hand curve called the 'More than Ogive'.

(b) **Less than Technique:** In this technique, less than cumulative frequencies are plotted on Y-axis and the upper limit of the class interval on X-axis. The points are then joined so as to form a smooth free hand curve called the 'Less than Ogive'.

**Q30. Differentiate between diagrams and graphs.**

*Ans :*

The differences between Diagrams and Graphs are as follows,

Sl.No.	Diagrams	Sl.No.	Graphs
1.	Diagrams are usually constructed on a plain sheet and they are useful for comparison purpose.	1.	Graphs are constructed on a special sheet known as graph paper and they are used for studying the relationship between the variables
2.	In Diagrammatic presentation, data is presented by using bars, pie charts, circles, cubes etc.	2.	In Graphical presentation, data is presented through lines or points of various types like dots, dashes, dot-dash and so on.
3.	Diagrams give approximate idea about different types of diagrams.	3.	Graphs provide reliable, accurate and authentic information.
4.	Compared to graphs, constructing diagrams is a difficult task.	4.	They are easy to construct compared to diagrams.
5.	Geographical and categorical data can be presented through diagrams.	5.	Data concerning time series and frequency distribution can be presented through graphs.

### 1.7.1 One-Dimensional, Two-Dimensional and Three-Dimensional Diagrams and Graphs

**Q31. State different types of diagrams.**

*Ans :*

In practice, large variety of diagrams is in use. Diagrams are classified on the basis of their length, width and shape. We will discuss the important types of diagrams which are more frequently used. For sake of application and simplicity several types of diagrams are categories under the following heads.

1. One dimensional diagrams
2. Two dimensional diagrams
3. Three dimensional diagrams
4. Pictogram or Ideographs
5. Cartograms or Statistical Maps

**Q32. Explain briefly about one-dimensional diagrams with an examples.**

*Ans :*

This is the most common type of diagrams. They are called one-dimensional diagrams because only length of the bar matters and not the width. For large number of observations lines may be drawn instead of bars to save space.

#### Merits of Bar diagrams

1. They are easily understood.
2. They are simplest and easiest to make
3. They are simplest and easiest in comparing two or more diagrams.

### Types of Bar Diagrams

- a) Simple bar diagram
- b) Subdivided bar diagram
- c) Multiple bar diagram
- d) Percentage bar diagram
- e) Deviation bars

#### a) Simple Bar Diagram

A simple bar diagram is used to represent only one variable. It should be kept in mind that, only length is taken into account and not width. Width should be uniform for all bars and the gap between each bar is normally identical. For example the figures of production, Sales, profits etc for various years can be shown by bar diagrams.

#### Example

Prepare a simple bar diagram for following data related to wheat exports.

Year	Exports (in million tons)
2001	177
2002	219
2003	420
2004	326
2005	202
2006	225

Sol:

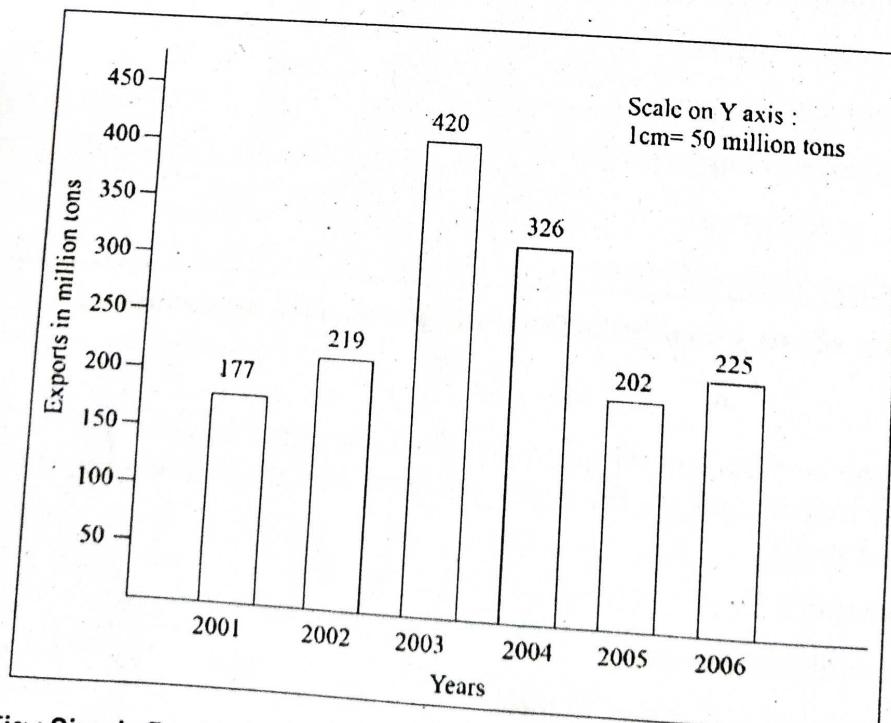


Fig.: Simple Bar Diagram Showing the Wheat Exports in Different Years

### b) Subdivided Bar Diagram

In this diagram, one bar is constructed for total value of the different components of the same variable. Further, it is subdivided proportionately to the various components of that variable.

A bar is represented in the order of magnitude from the largest component at the base of the bar to the smallest at the end of the bar, but the order of various components in each bar is kept in the same order. Different shades or colors are used to distinguish between different components. To explain such differences, the index should be used in the bar diagram. The subdivided bar diagrams can be constructed both on horizontal and vertical bases.

#### Example

The following data shows the production of rice for the period 2010 to 2018. Represent the data by a subdivided bar diagram.

Year	Non-Basmati Rice (in Million metric tons)	Basmati Rice (in Million metric tons)	Total (in Million metric tons)
2010	29	35	64
2011	35	33	68
2012	25	35	60
2013	40	30	70
2014	42	32	74
2015	32	40	72

Sol :

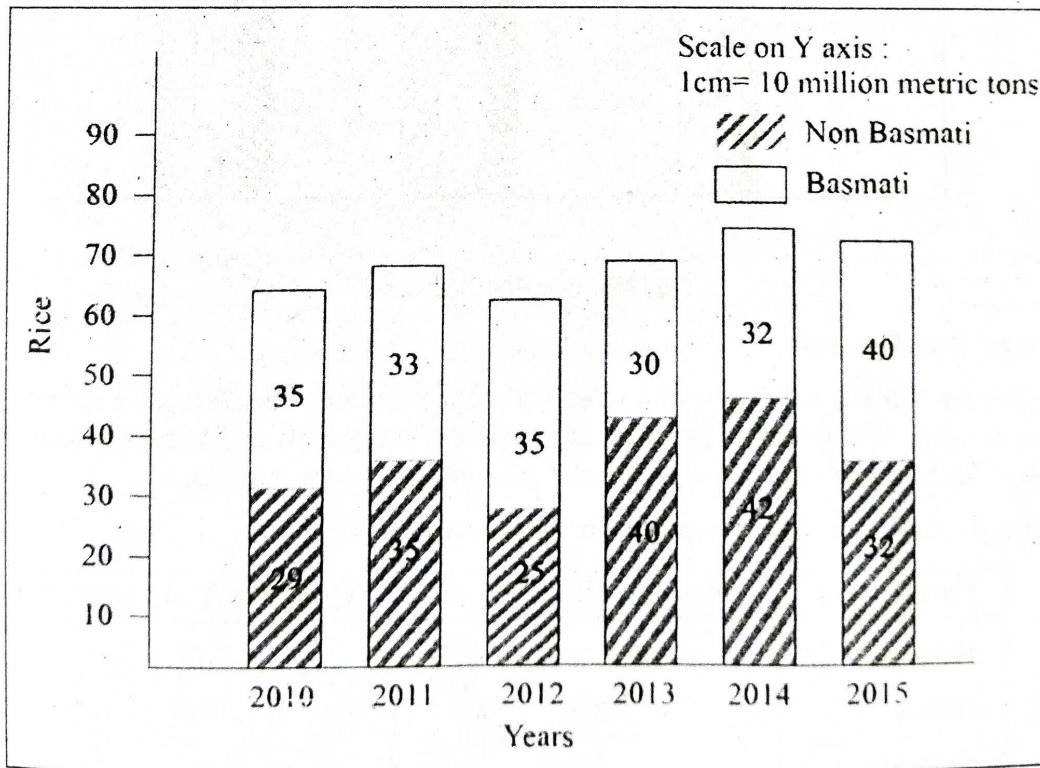


Fig.: Subdivided Bar Diagram Showing the production of Rice (in Different Years)

c) **Multiple Bar Diagram**

Whenever the comparison between two or more related variables is to be made, multiple bar diagram should be preferred. In multiple bar diagrams two or more groups of interrelated data are presented. The technique of drawing such type of diagrams is the same as that of simple bar diagram. The only difference is that since more than one components are represented in each group, so different shades, colors, dots or crossing are used to distinguish between the bars of the same group.

**Example**

Represent the following data by a multiple bar diagram.

Class	Physics	Chemistry	Mathematics
Student A	50	63	57
Student B	55	60	68
Student C	48	60	55

*Sol :*

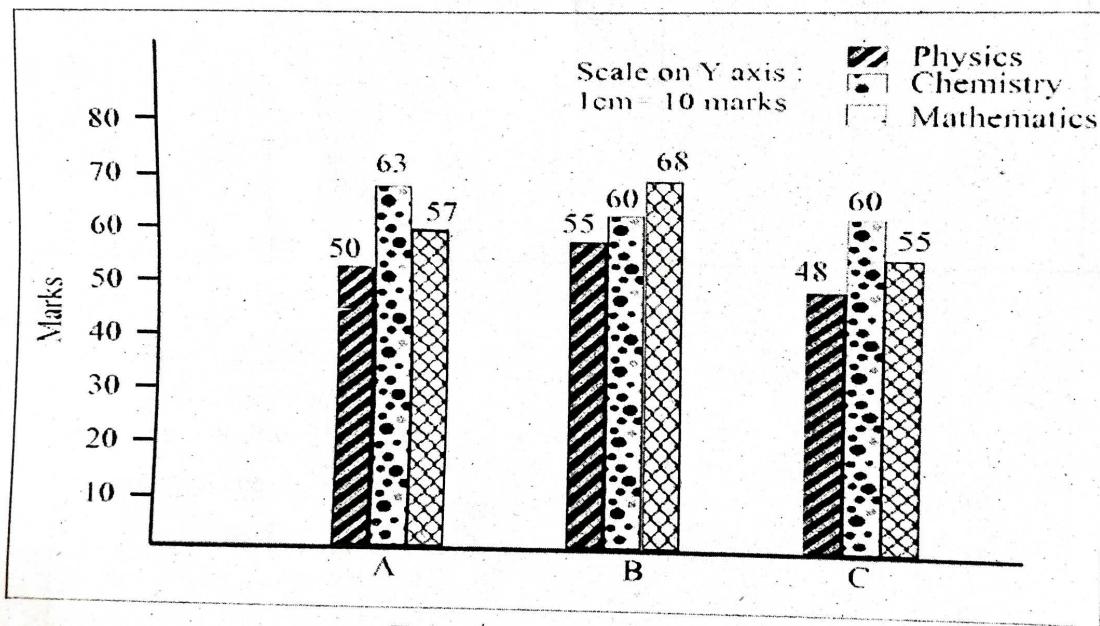


Fig.: Multiple Bar Diagram

d) **Percentage Bar Diagram**

Percentage bars are particularly useful in statistical work which requires the representation of the relative changes in data. When such diagrams are prepared, the length of the bars is kept equal to 100 and segments are cut in these bars to represent the percentages of an average.

**Example:** Draw percentage bar diagram for following data.

Particulars	Cost Per Unit (2010)	Cost Per Unit (2020)
Material	22	35
Lobour	30	40
Delivery	10	20
Total	62	95

*Sol:*

Express the values in terms of percentage for both the years.

Particulars	Cost Per Unit (2010)	% Cost	Cumulative % cost	Cost Per Unit (2020)	% Cost	Cumulative % cost
Material	22	35.48	35.48	35	36.84	36.84
Lobour	30	48.38	83.86	40	42.10	78.94
Delivery	10	16.12	99.98	20	21.05	99.99
<b>Total</b>	<b>62</b>	<b>100</b>		<b>95</b>	<b>100</b>	

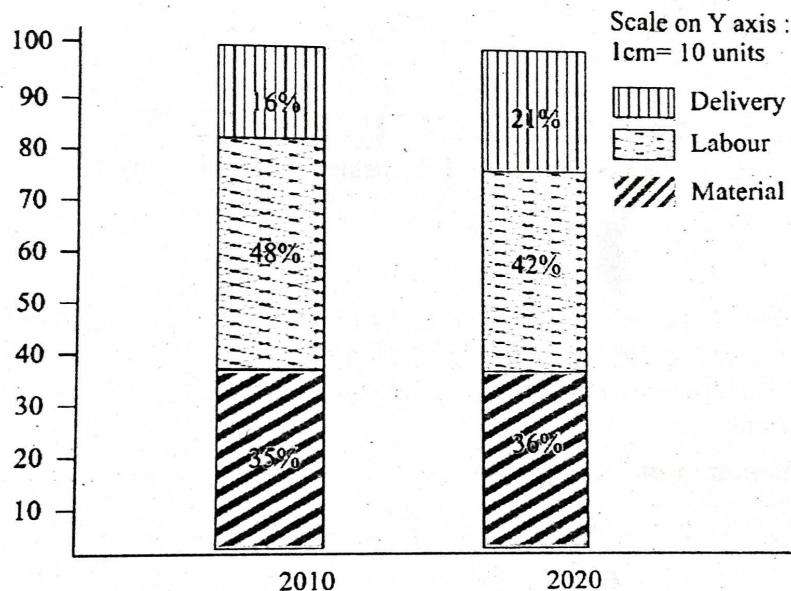


Fig.: Percentage Bar Diagram

#### e) Deviation Bar Diagram

Deviation bars are used for representing net quantities – excess or deficit. i. e. net profit, net loss, net exports or imports etc. For representing net quantities excess or deficit, i.e. net profit, net loss, net exports, net imports, etc., This kind of bars represent both positive and negative values. The values which are positive can be drawn above the base line and negative values can be drawn below it.

#### Example

Draw deviation bar diagram for following data.

Year	Sales	Profits
2010	24%	29%
2011	15%	-10%
2012	23%	-5%

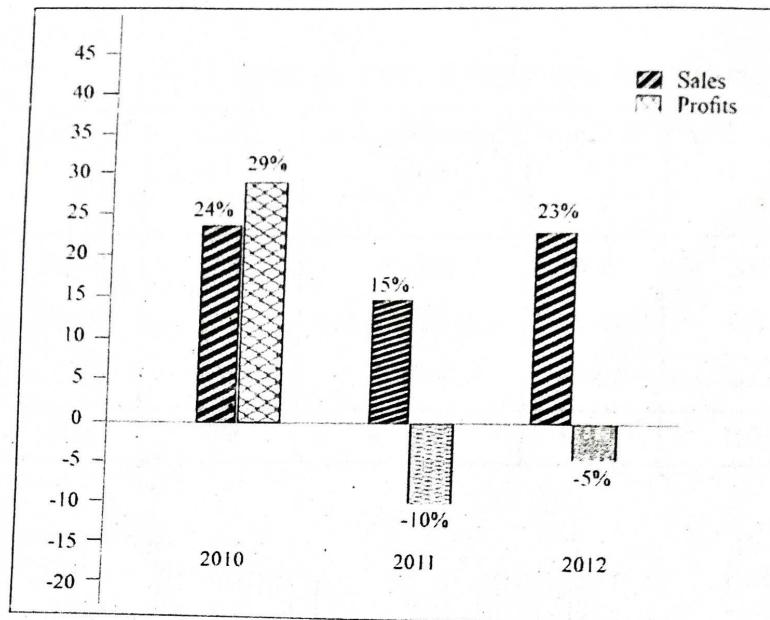


Fig.: Deviation Bar Diagram for sales and profits

**Q33.** Explain briefly about two-dimensional diagrams with an examples.

**Ans :**

**Meaning .**

In one-dimensional diagrams, only length of the bar is considered and comparison of bars are done on the basis of length only. In two-dimensional diagrams the length as well as width of the bars is considered. Thus, the area of the bars represent the given data. Two-dimensional diagrams are also known as surface diagrams or area diagrams.

#### Types of Two Dimensional Diagrams

- (a) Rectangles
- (b) Squares
- (c) Circles

#### a) Rectangles

In rectangle diagram, given numerical figures are represented by areas of the rectangles. We know that area of rectangle = length x width. While constructing such a diagram both length and width are considered. We may represent the figures as they are given or may convert them to percentage and then subdivide the length into various components.

#### Example

Represent the following data of monthly expenditure (in rupees) of two families by suitable diagram.

Expenditure	Family A	Family B
Food	2400	1800
Clothing	1600	1200
Education	1000	800
Electricity	200	200
Miscellaneous	800	500

Sol:

First convert the figures into percentage and take the cumulative sum of percentage.

Expenditure	Family A			Family B		
	Rs.	%	Cumulative %	Rs.	%	Cumulative %
Food	2400	40	40	1800	40	40
Clothing	1600	27	67	1200	27	67
Education	1000	17	84	800	18	85
Electricity	200	3	87	200	4	89
Miscellaneous	3200	13	100	500	11	100
<b>Total</b>	<b>6000</b>	<b>100</b>		<b>4500</b>	<b>100</b>	

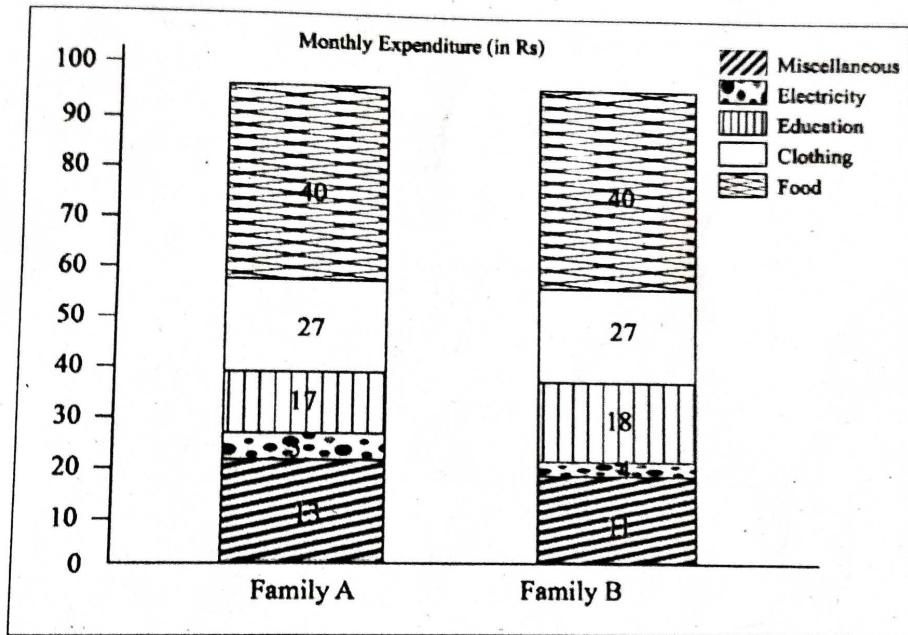


Fig.: Diagram of Rectangles

## ii) Squares

The rectangular method of diagrammatic presentation is difficult to use where the values of items vary widely. The method of drawing a square is simple. Take square roots of the given numerical observations as sides of the corresponding squares and then select a suitable scale to draw the squares.

### Example

Represent the following data of the number of hospitals in a city in 2000-05, 2005-10, 2010-15 and 2015-20 in square diagram.

Year	No. of Hospitals
2000-05	16
2005-10	64
2010-15	400
2015-20	576

*Sol :*

Since there is a big gap between first year and last year, a square diagram is suitable here. To decide the side of a square consider following calculations.

Year	No. of Hospitals	Square Root	Side of square in cms = Square Root/4
2000-05	16	4	1
2005-10	64	8	2
2010-15	400	20	5
2015-20	576	24	6

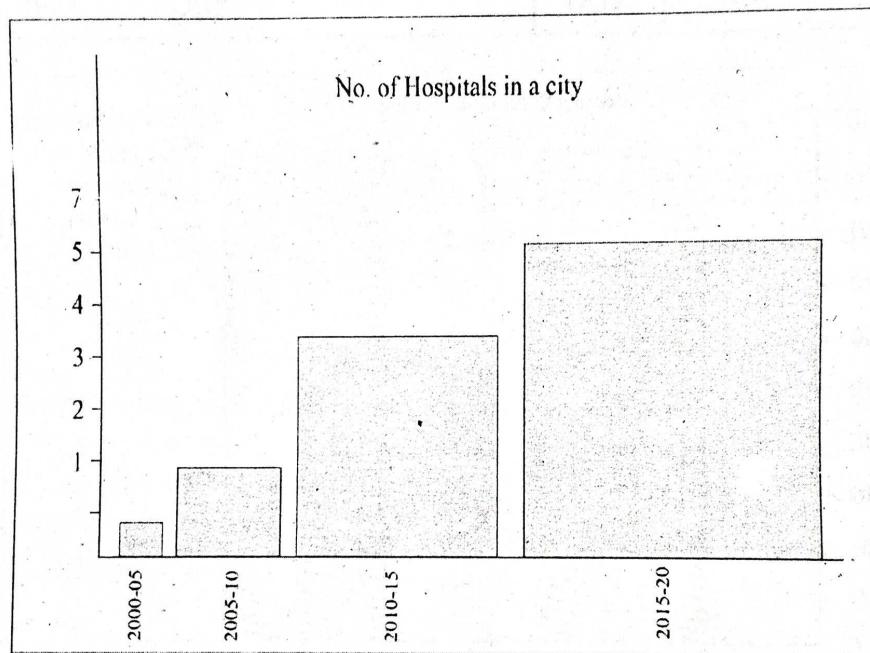


Fig.: Diagram for Squares

### C) Circles

As in square diagram, we took given figures/observations as the areas of the corresponding squares. Similarly, here we take given numerical figures/ observations as areas of the corresponding circles. The area of a circle is proportional to the squares of its radius. The radius of circles can be obtained by dividing the value of pie and taking square root. Circles can be used in all those cases in which squares are used.

Circles are difficult to compare and as such are not popular. When it is necessary to use circles, they should be compared on an area basis rather than on diameter basis.

Area of a circle =  $\pi r^2$  where  $r$  is radius of circle

$$\therefore r^2 = \frac{\text{Area}}{\pi}, \pi = \frac{22}{7}$$

$$\therefore r = \frac{\text{Area}}{\pi}$$

**Example**

Represent the following data of the number of hospitals in a city in 2000-05, 2005-10, 2010-15 and 2015-20 in circle diagram.

Year	No. of hospitals
2000-05	16
2005-10	64
2010-15	400
2015-20	576

**Sol.:**

Year (I)	No. of Hospitals (n) (II)	$n / \left(\frac{22}{7}\right)$ (III)	Square Root of [ $n / \left(\frac{22}{7}\right)$ ] (IV)	Col (IV)/2 (V)
2000-05	16	5.09	2.25	1.125
2005-10	64	20.36	4.51	2.255
2010-15	400	12	11.28	5.64
2015-20	576	183.27	13.54	6.77

[Note: To get smaller value of radius of the circle divide each figure in Col (IV) by 2]

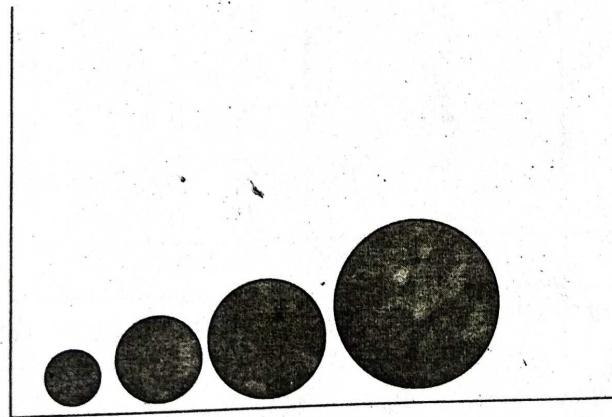


Fig.: for Circles

**Q34. Define piediagram. How to create a pie diagram with an example.**

**Ans :**

A pie Diagram is a type of graph that displays data in a circular graph. The pieces of the graph are proportional to the fraction of the whole in each category. Pie diagrams are very popularly used in practice to show percentage breakdown. While making comparisons, pie diagrams should be used on a percentage basis and not on absolute basis.

**Steps**

1. Take a total of all observations
2. Divide each observation by total and multiply by 100 to get percent. (if instead of percentage, observations are given )
3. Next to know how many degrees for each "pie sector" we need, we will take a full circle of  $360^\circ$  and follow the calculations as below:
4. The central angle of each component =  $(\text{Value of each component}/\text{sum of values of all the components}) \times 360^\circ$
5. Draw a circle of appropriate size with compass and use the protractor to measure the degree of each sector.

In laying out the sectors for pie diagram, it is common practice to begin the largest component sector of pie diagram at 12 O'clock position on the circle. The other component sectors are placed in clockwise direction in descending order of magnitude. Give descriptive label for identification of each sector.

**Example**

Draw the pie diagram for the following data of cost of construction of house.

Bricks	15%
Steel	35%
Cement	20%
Labour	20%
Supervision	10%

$$\frac{\text{Percentage outlay}}{100} \times 360 = \text{Percentage outlay} \times 3.6$$

Sector	Percentage	Angle outlay
Bricks	15	$15 \times 3.6 = 54$
Steel	35	$35 \times 3.6 = 126$
Cement	20	$20 \times 3.6 = 72$
Labour	20	$20 \times 3.6 = 72$
Supervision	10	$10 \times 3.6 = 36$
<b>Total</b>	<b>100</b>	<b>360</b>

(Note: The angles have been arranged in ascending order)

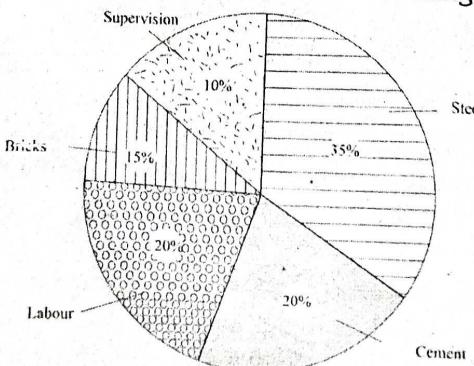


Fig.: Pie Diagram

**Q35. Explain briefly about three-dimensional diagrams with an examples.**

*Ans :*

In three dimensional diagrams three things namely, length, width and height have to be considered. Three-dimensional diagrams, also known as volume diagrams consist of cubes, cylinders, spheres etc. Such diagrams are used where the range of difference between the smallest and the largest value is very large. For example, if two values are in the ratio of 1:1000 and if bar diagram are used to represent them, the shortest bar would be of one-thousandth part of the largest bar. If squares or circles are used then the side of the square or the radius of one circle would be proportionately too large or too small than the other. If cubes are used then their sides would be in the ratio of 1:10. This example makes it clear that three-dimensional diagrams have an important role to play when the gap between the smallest and the largest value is very large.

The disadvantage of three-dimensional data is the side of a cube must be proportionate to the cube root of the magnitude to be represented. It is very difficult for the eye to read precisely such diagrams and hence they are not recommended for statistical presentation.

### **PROBLEMS**

1. Draw a pie diagram for the following data of Sixth Five - Year Plan Public Sector outlays :

<b>Agriculture and Rural Development</b>	<b>12.9%</b>
<b>Irrigation, etc.</b>	<b>12.5%</b>
<b>Energy</b>	<b>27.2%</b>
<b>Industry and Minerals</b>	<b>15.4%</b>
<b>Transport, Communication, etc.</b>	<b>15.9%</b>
<b>Social Services and Others</b>	<b>16.1%</b>

*Sol :*

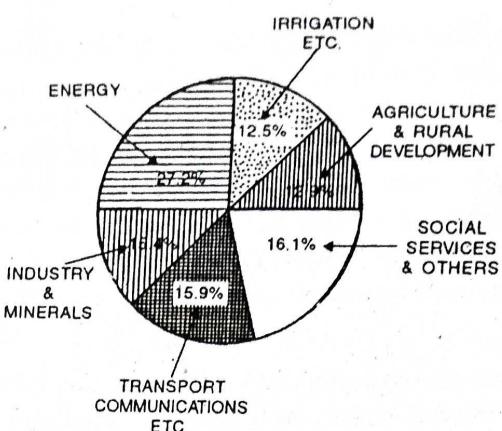
The angle at the centre is given by

$$\frac{\text{Percentage outlay}}{100} \times 360 = \text{Percentage outlay} \times 3.6$$

### **Computation for PIE Diagram**

<b>Sector</b>	<b>Percentage</b>	<b>Angle outlay</b>
Agriculture and Rural Development	12.9	$12.9 \times 3.6 = 46^\circ$
Irrigation, etc.	12.5	$12.5 \times 3.6 = 45^\circ$
Energy	27.2	$27.2 \times 3.6 = 98^\circ$
Industry and Minerals	15.4	$15.4 \times 3.6 = 56^\circ$
Transport, Communication, etc.	15.9	$15.9 \times 3.6 = 57^\circ$
Social Services and others	16.1	$16.1 \times 3.6 = 58^\circ$
<b>Total</b>	<b>100.0</b>	<b>360°</b>

Now a circle shall be drawn suited to the size of the paper and divided into 6 parts according to degrees of angle at the centre. (The angles have been arranged in descending order).



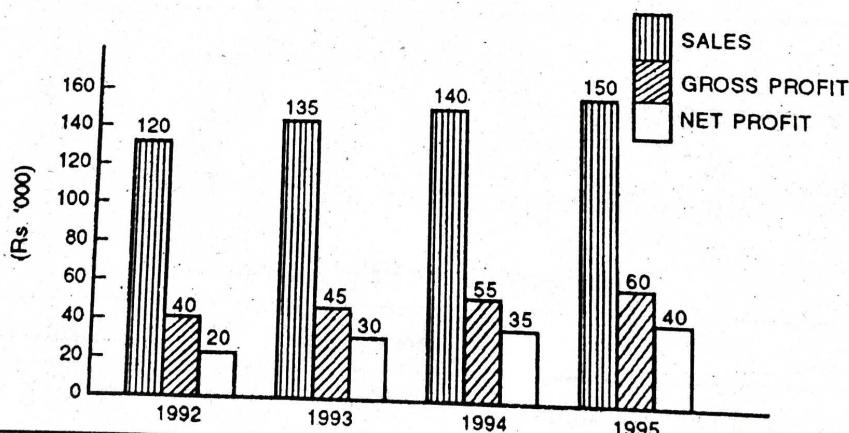
2. Draw a multiple bar diagram from the following data :

Year	Sales ('000 Rs.)	Gross Profit ('000 Rs.)	Net Profit ('000 Rs.)
1992	120	40	20
1993	135	45	30
1994	140	55	35
1995	150	60	40

Sol:

(Imp.)

Sales, Gross Profits & Net Profits (For the year 1992-95)



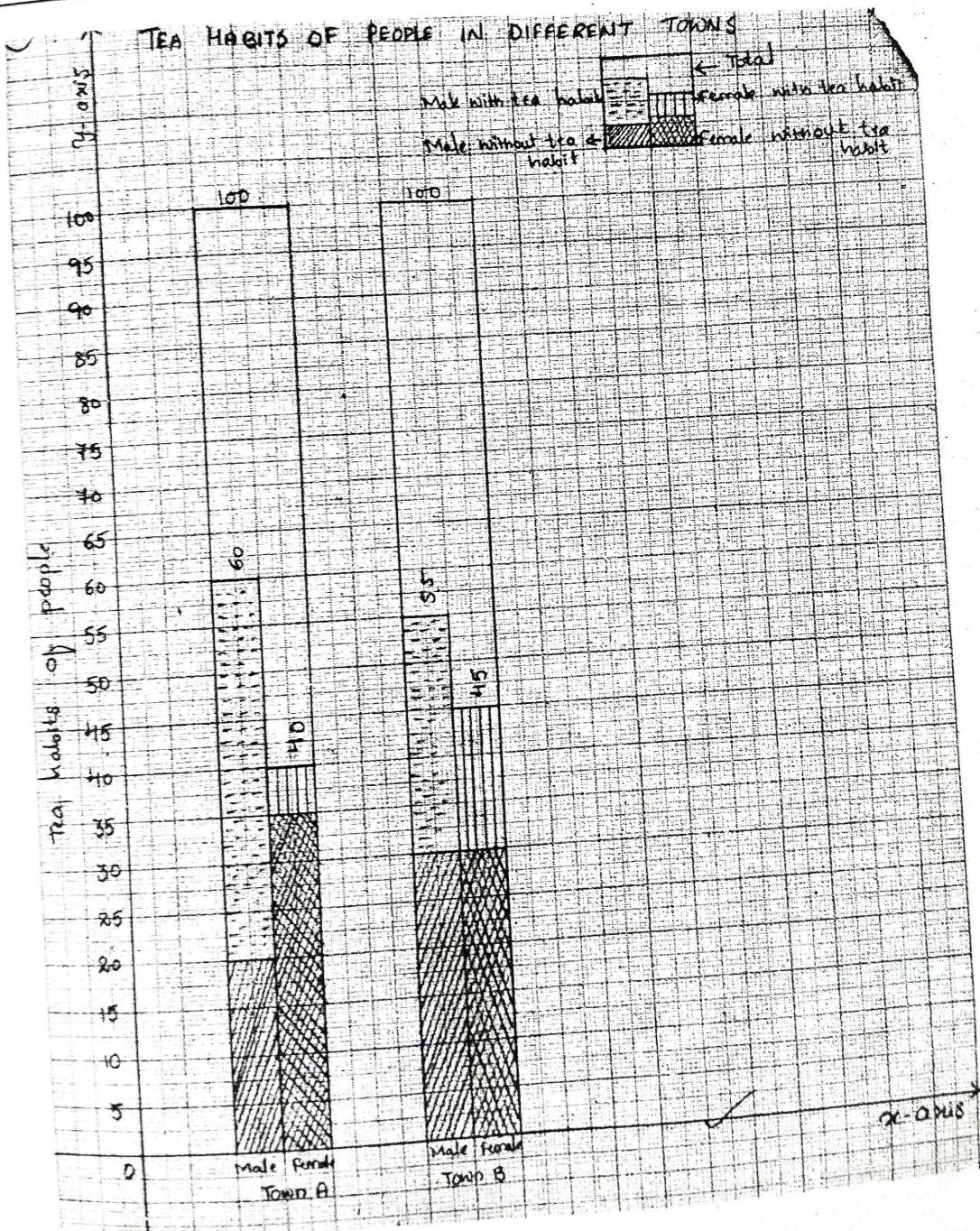
3. Represent the data shown in the following table by using suitable graphical method.

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 with - out tea habit	20	35	55	30	30	60
Total	60	40	100	55	45	100

Sol:

## Representation of Data by Sub - divided Multiple Bars Method

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 with -out tea habit	20	35	55	30	30	60
Total	60	40	100	55	45	100

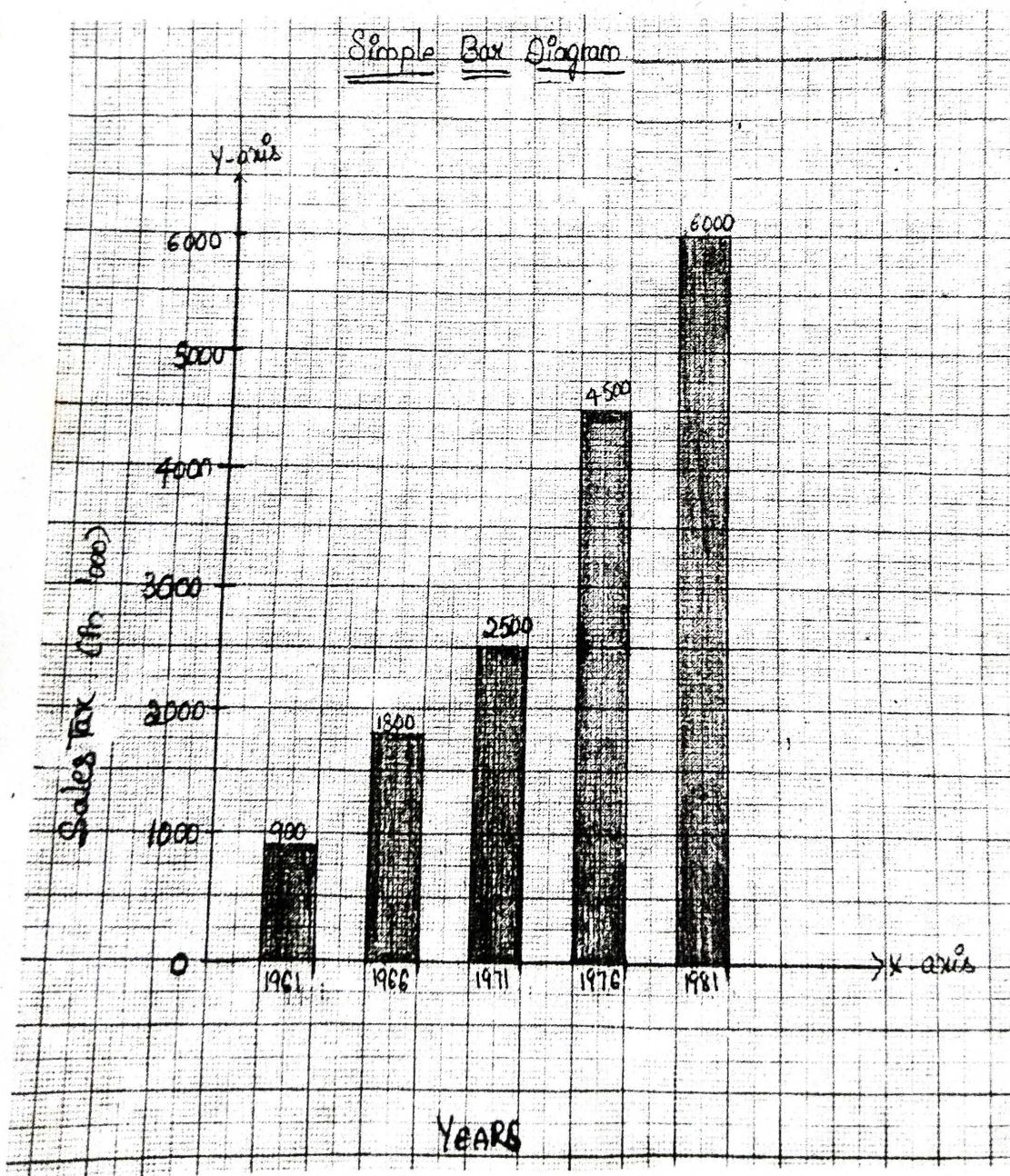


4. Represent the below given data by a simple bar diagram :

Year	Sales Tax (in '000)
1961	900
1966	1800
1971	2500
1976	4500
1981	6000

Sol :

(Imp.)



5. Draw pie diagram to represent the following data of proposed expenditure by a State Government for the year 1999 - 2000.

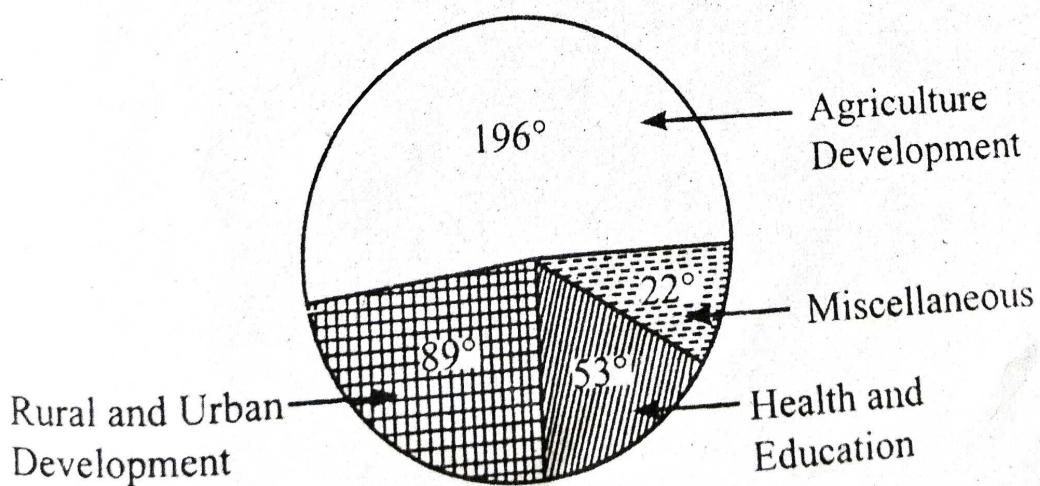
Items	Agriculture Development	Rural and Urban Development	Health and Education	Miscellaneous
Proposed Expenditure (in Million Rupees)	4400	2000	1200	500
(a)	(b)	(c)		

Sol:

Calculations for Pie Chart

Items	Proposed Expenditure (in Million Rupees)	Angle at the Centre
(a)	(b)	(c) = $\frac{(b)}{8100} \times 360^\circ$
Agriculture Development	4400	$\frac{4400}{8100} \times 360^\circ = 196^\circ$
Rural and Urban Development	2000	$\frac{2000}{8100} \times 360^\circ = 89^\circ$
Health and Education	1200	$\frac{1200}{8100} \times 360^\circ = 53^\circ$
Miscellaneous	500	$\frac{500}{8100} \times 360^\circ = 22^\circ$
	8100	360°

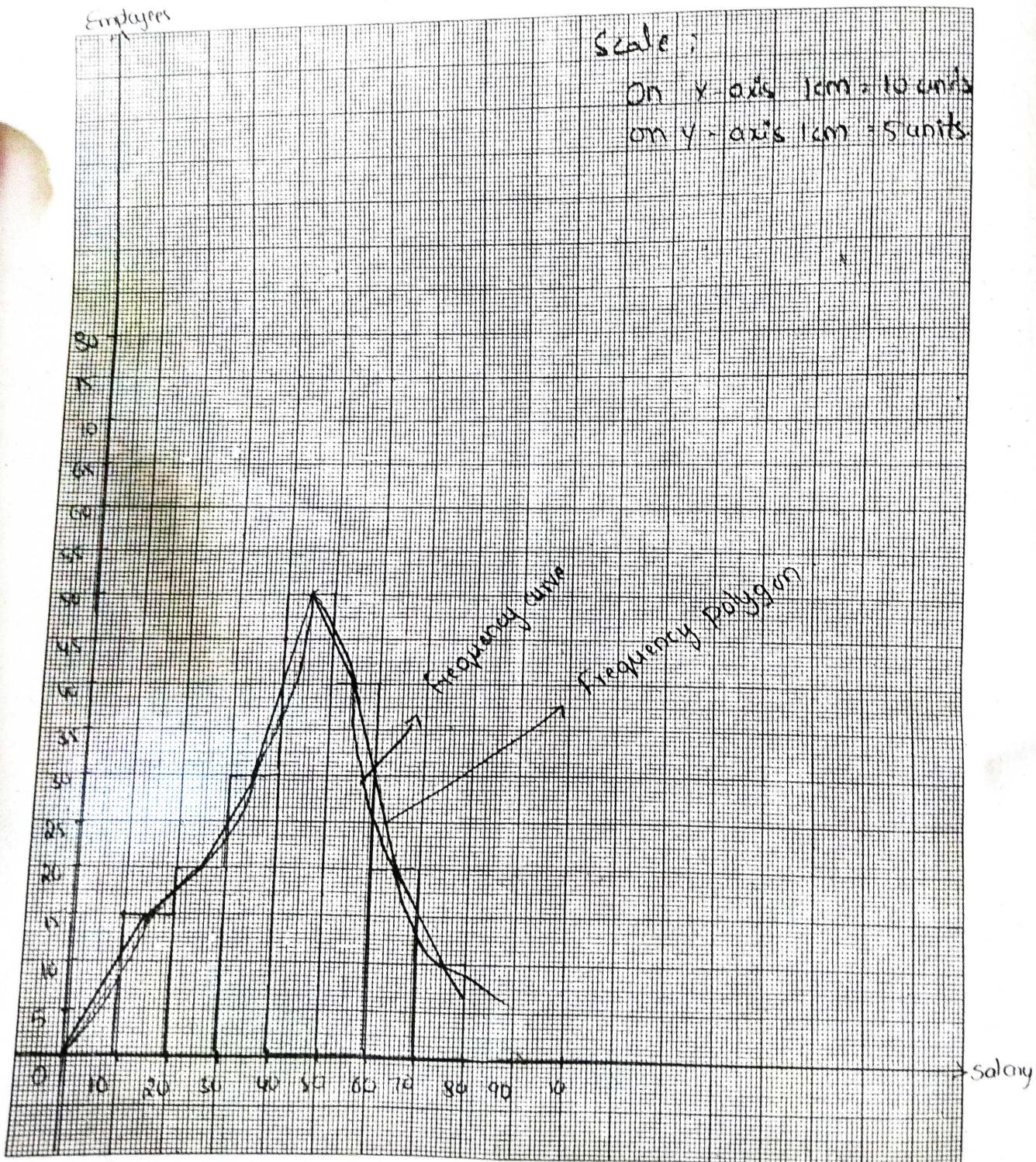
Pie diagram representing proposed expenditure by state government on different items for 1999 - 2000.



6. For the following data, draw Histogram, Frequency Curve and frequency Polygon.

Salary (Rs. in '000)	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Employees	15	20	30	50	40	20	10

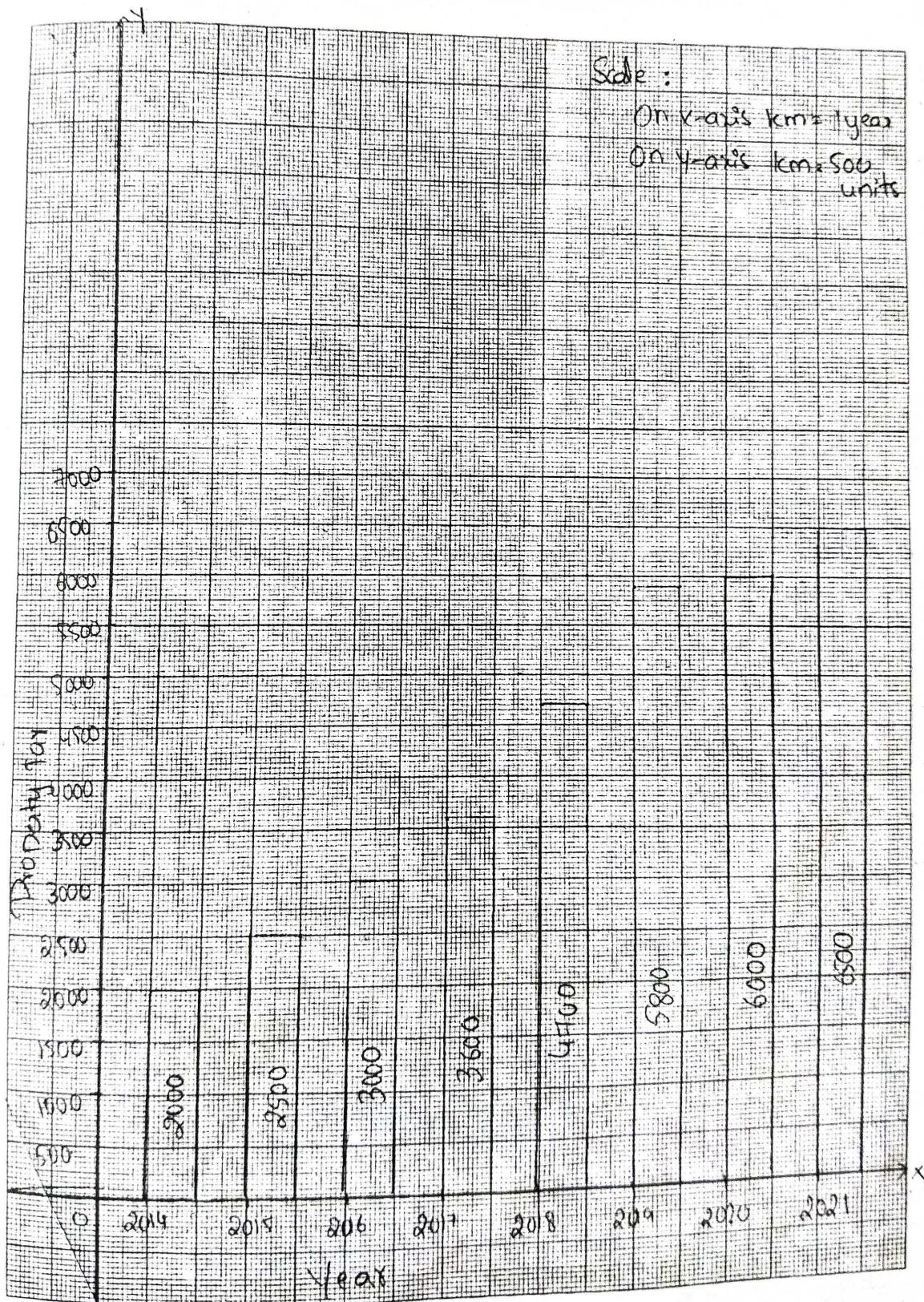
Sol :



7. For the following data, prepare a Bar Diagram

Year	2014	2015	2016	2017	2018	2019	2020	2021
Property Tax (Rs. Crore)	2000	2500	3000	3600	4700	5800	6000	6500

Sol :



## Short Question and Answers

### 1. Define statistics

**Ans :**

Some of the definitions of statistics are presented as follows,

- (a) **According to A.L. Bowley**, "Statistics is the science of counting".
- (b) **According to A.L. Bowley**, "Statistics may rightly be called the science of averages".
- (c) **According to Turtle**, "Statistics is a body of principles and techniques of collecting, classifying, presenting, comparing and interpreting the quantitative data".

Thus, the above definitions specify that statistics in singular sense, is a science which comprises different statistical methods used-for collection, organization, classification, presentation and interpretation of data.

### 2. Functions of Statistics

**Ans :**

The, following are the main functions of statistics:

#### (i) Presents facts in numerical figures

The first function of statistics is to present a given problem in terms of numerical figures. We know that the numerical presentation helps in having a better understanding of the nature of a problem. Facts expressed in words are not very useful because they are often vague and are likely to be understood differently by different people. For example, the statement that a large proportion of total work force of India is engaged in agriculture, is vague and uncertain. On the other hand, the statement that 70% of the total work force is engaged in agriculture is more specific and easier to grasp. Similarly, the statement that the annual rate of inflation in a country is 10% is more convincing than the statement that prices are rising.

#### (ii) Presents complex facts in a simplified form

Generally a problem to be investigated is represented by a large mass of numerical figures which are very difficult to understand and remember. Using various statistical methods, this large mass of data can be presented in a simplified form. This simplification is achieved by the summarization of data so that broad features of the given problem are brought into focus. Various statistical techniques such as presentation of data in the form of diagrams, graphs, frequency distributions and calculation of average, dispersion, correlation, etc., make the given data intelligible and easily understandable.

#### (iii) Studies relationship between two (or more phenomena)

Statistics can be used to investigate whether two or more phenomena are related. For example, the relationship between income and consumption, demand and supply, etc., can be studied by measuring correlation between relevant variables. Furthermore, a given mathematical relation can also be fitted to the given data by using the technique of regression analysis.

### 3. What is Quantitative Data?

**Ans :**

Quantitative data is a set of numbers collected from a group of people and involves statistical analysis. For example if you conduct a satisfaction survey from participants and ask them to rate their experience on a scale of 1 to 5. You can collect the ratings and being numerical in nature, you will use statistical techniques to draw conclusions about participants satisfaction.

### 4. Primary Data

**Ans :**

As the name suggests is original, problem (or) project-specific and collected for the specific

objectives and needs spelt out by the researcher. The authenticity and relevance is reasonably high. The monetary and resource implications of this are quite high and sometimes a researcher might not have the resources or the time or both to go ahead with this method. In this case, the researcher can look at alternative sources of data which are economical and authentic enough to take the study forward. These include the second category of data sources namely the secondary data.

### 5. Secondary Data

*Ans :*

As the name implies is that information which is not topical or research specific and has been collected and compiled by some other researcher or investigative body. The said information is recorded and published in a structured format, and thus, is quicker to access and manage. Secondly, in most instances, unless it is a data product, it is not too expensive to collect. As suggested in the opening vignette, the data to track consumer preferences is readily available and the information required is readily available as a data product or as the audit information which the researcher or the organization can procure and use it for arriving at quick decisions.

### 6. Define classification.

*Ans :*

**Meaning**

The largest amount of data and the greatest amount of detail may not convey the most useful information for decision making. An important aspect of Statistics is to organize and present data so as to convey critical information quickly.

Classification means grouping of a whole into different groups or classes. However, each of these groups should have a common characteristic.

For example, all living things can be classified into plants and animals. Animals can be classified into Amphibians, Reptiles, and Mammals etc.

**Definition**

**According to Horace Secrets**, "Classification is the process of arranging data into sequences and groups according to their common characteristics, or separating them into different but related parts".

### 7. What is frequency distribution?

*Ans :*

Frequency distribution refers to the tabular arrangement of data, when arranged into groups or categories according to conveniently established divisions of the range of the observations. In a frequency distribution, raw data is expressed in distinct groups called 'classes'. The number of observations that fall into each of these classes is known as 'frequency'. Thus, a frequency distribution has two parts, the left part represents the classes and the right part represents the frequencies.

Arrangement of raw data in the form of row or column is called data array. When data is described by a continuous variable, it is called continuous data and when described using discrete variable it is called discrete data.

### 8. Diagrams

*Ans :*

Diagram is a visual presentation of statistical information. The pictorial presentation helps in proper understanding the data . Diagrams are of different types like pie diagram, rectangles, lines, pictures and maps.

**9. Define piediagram.****Ans :**

A pie Diagram is a type of graph that displays data in a circular graph. The pieces of the graph are proportional to the fraction of the whole in each category. Pie diagrams are very popularly used in practice to show percentage breakdown. While making comparisons, pie diagrams should be used on a percentage basis and not on absolute basis.

**10. Population****Ans :**

The term Population is used to describe persons, objects, animals, elements, or even reactions that display unique characteristics. If a researcher wants to perform a study on adolescents in Kolkata, then her/his population will be all the adolescents in Kolkata. Though, The Population can be finite or infinite (Mohanty and Misra). An Example of a finite population is the number of students in a school who have passed in psychology and an Example of an infinite population would be the number of stars in the sky.

## Choose the Correct Answer

1. What is the arrangement of data in rows and columns known as? [ c ]
  - (a) Frequency distribution
  - (b) Cumulative frequency distribution
  - (c) Tabulation
  - (d) Classification
  
2. When the quantitative and qualitative data are arranged according to a single feature, what is the tabulation known as? [ a ]
  - (a) One-way
  - (b) Bivariate
  - (c) Manifold division
  - (d) Dichotomy
  
3. Which function does the tabulation origin spot specify? [ a ]
  - (a) The list of integers
  - (b) The list of maxterms
  - (c) The list of minterms
  - (d) None of the above
  
4. What does the tabulation form exercise? [ c ]
  - (a) Gates
  - (b) Demorgan's postulate
  - (c) Matching process cycle
  - (d) Venn diagram
  
5. What was the first tabulation method known as? [ d ]
  - (a) Quine-McCluskey
  - (b) Cluskey
  - (c) McQuine
  - (d) None of the above
  
6. A frequency distribution is a tabular summary of data showing the [ d ]
  - (a) Fraction of items in several classes
  - (b) Percentage of items in several classes
  - (c) Relative percentage of items in several classes
  - (d) Number of items in several classes
  
7. A frequency distribution is [ c ]
  - (a) A tabular summary of a set of data showing the relative frequency
  - (b) A graphical form of representing data
  - (c) A tabular summary of a set of data showing the frequency of items in each of several nonoverlapping classes
  - (d) A graphical device for presenting qualitative data
  
8. A tabular summary of a set of data showing the fraction of the total number of items in several classes is a [ b ]
  - (a) Frequency distribution
  - (b) Relative frequency distribution
  - (c) Frequency
  - (d) Cumulative frequency distribution
  
9. Qualitative data can be graphically represented by using a(n) [ d ]
  - (a) Histogram
  - (b) Frequency polygon
  - (c) Ogive
  - (d) Bar graph
  
10. The relative frequency of a class is computed by [ d ]
  - (a) Dividing the midpoint of the class by the sample size
  - (b) Dividing the frequency of the class by the midpoint
  - (c) Dividing the sample size by the frequency of the class
  - (d) Dividing the frequency of the class by the sample size

## Fill in the blanks

1. \_\_\_\_\_ are numerical facts in any department of enquiry placed in relation to each other.
2. The \_\_\_\_\_ is an important tool for gathering primary data.
3. \_\_\_\_\_ is the process of arranging the data based on the existing similarities and dissimilarities between them.
4. \_\_\_\_\_ is a visual presentation of statistical information.
5. \_\_\_\_\_ bars are particularly useful in statistical work which requires the representation of the relative changes in data.
6. A pie Diagram is a type of graph that displays data in a \_\_\_\_\_.
7. \_\_\_\_\_ are used to display series of numerical data in a graphical format.
8. \_\_\_\_\_ chart shows changes or contribution of a value in a given period of time or over categories.
9. \_\_\_\_\_ chart is similar to that of column chart but are horizontally oriented.
10. \_\_\_\_\_ chart is frequently used for stock price data and as well as for scientific data.

### ANSWERS

1. Statistics
2. Questionnaire
3. Classification
4. Diagram
5. Percentage
6. Circular graph
7. Charts
8. Area
9. Bar
10. Stock