Chapter - 2.1

Meaning and Definition of Statistics

Modern time is the time of knowledge and scientific activities. Each branch of Knowledge and scientific activities is concerned directly and indirectly with the numerical facts. All the small incidents of life can be assigned the numerical values. Without numbers the accuracy of facts is not completed. Information of social, economic, political and other problems are analysed on the basis of numerical values by the state and thus efforts are made for their solution. Thus, the knowledge based on numerical facts is real and actual. According to the British writer Lord Kelvin, "If the topic about which we are talking can be measured and expressed by assigning numerical values, it means we know something about that topic. But when we can not measure it and can not express in numbers then our knowledge is unsatisfactory and less."

Origin of Statistics:

Statistics has its origin along with the national organizations. The rulers in the old time collected data about the issues concerned with management.

There were many facts related with statistics collected about administration, social system, war, etc. According to the Arthashastra of Kautilya. Tendency of data collection is also found in Rome and other countries also. Statistics was considered as the Science of State efficiency or the Science of kind because the policies of the State used based on data.

Meaning of Statistics:

The English word Statistics is derived from the word state. The word state is knows as 'status' in Letin, as 'stato' in Roman as 'statistik in German and as 'Statista' in Italian. All these words mean the state. The relationship between state and statistics has been very deep. Sometimes this word had been used for such person who was expert in state related work. We find the use of statistics in India in many old books like Arthashastra of Kautilya. The German mathematician Gottfried Achenwall was the father of statistics who, first of all, used the word 'Statistics' in 1749 and enhanced and developed it as the branch of knowledge. In modern time data are widely used so their demand has increased. Introduction of new methods in statistics has saved time, labour and cost.

The english word STATISTICS is used in three sense in Hindi as data, Statistics as a subject and Statistic (value calculated from samples). Generally 'Statistics' word is used for data like, data of death and birth in India, data about crime, data about price level, etc. Data are the collection of numerical values of facts.

Secondly, statistics mean the methods which are used in statistics. These are the methods relating with theories and devices which

are concerned with the collection, analysis and interpretation of data.

The word 'Statistics is used for the plural groups or data of statistics, e.g. census data.

Thirdly, statistics means the values which are calculated from the collected data like sample mean, sample mode, sample standard deviation, etc. These values are known as statistic and used as estimates of the population parameters.

Definition of Statistics :

There are many definitions of statistics. **Quetelet** gave 180 definitions of statistics in 1869. In his book John Griffin writes," it is difficult to define Statistics. Till now there is no definition of Statistics which has universally been accepted. Some authors have defined statistics as below:

According to Bowley, "Statistics is science of counting." At the other place Bowley writes, "Statistics is properly the science of ends."

According to Boddington, "Statistics is the science of investigations and probabilities."

Accoding to Saligman. "Statistics is the science which deals with the methods of collections, classification presentation, comparison and analysis of the collected data" This is considered as a wide definition of Statistics.

It is clear from all the above definitions that like economists, there are many ways of giving definitions of statistics. This difference in giving an idle definition of statistics is due to the fact that it is not an easy work to define Statistics.

Scope of Statistics :

In ancient times the scope of statistics war very limited. Statistics originated as the science of Kings. But in modern time scope of statistics is very wide. Statistical methods are used in every science as a tool. So, it is correct to say that without statistics, science is not complete and without science statistics is not complete.

The subject matter of statistics is of two types :-

(a) Statistical Methods

(b) Applied Statistics

(a) Statistical Methods :

Statistical methods are helpful in collection of data, proper presentation of data and making data comparable and understandable. Proper conclusions are also derived with the help of the statistical methods. Thus, according to Johnson and Jackson," Statistical methods are the processes which are used for the collection, organization, summary, analysis, interpretation and presentation of the numerical facts." It is clear that following functions are included in the statistical methods.

i. Collection of data :

This function of statistical methods explains that how much, from where and how the data are to be collected.

ii. Classification :

Under this function different methods of classification of data are studied. The classification makes data easier to understand and compare.

iii. Tabulation:

In tables, data are properly presented in rows and columns.

iv. Presentation:

Under this function of statistical methods data are presented in the form of diagrams and graphs so that they can be easy to understand, compare and interpret.

v. Analysis:

Measures of central tendency, dispersion, skewness, regression, correlation, etc. are used to

analyse the data. These all are the various statistical methods.

vi. Interpretation:

Under interpretation it is studied that what the messages or informations are given by the various calculated values with the help of various statistical methods.

vii. Forecasting:

The values of variables are forecast on the basis of old and present given information and thus necessary announcement are made.

(b) Applied Statistics:

Statistical methods provide the theoretical knowledge. How are the statistical methods can be used practically is the subject matter of applied statistics. For example, data on population, national income, industrial production, prices, wages, etc. are the applied data. Applied data are related with economics, commerce, sociology, administration, biology, psychology, etc. There are two types of applied statistics:

i. Descriptive Statistics:-

In descriptive statistics we study the collected data of past and present related with any sector of economy or society.

ii. Scientific Applied Statistics:

Under this type of statistics the applied data are collected for the specification of the scientific laws about different subjects. Law of demand, trade cycles, etc. are such examples.

Under applied statistics, statistical methods are used for the study, analysis and solution of the different applied problems. Thus, the scope of statistics is very wide.

Role of Statistics in Economics:

Statistics and economics are deeply related. Statistical data play pivotal in different laws and theories of economics. Here data are just like the foundation stone. According to Marshall-"data are those elements, in which just like each economist, I also make bricks (economic laws)."

The essential role of data both in theoretical and applied economics has been proved. The inductive and deductive methods are based on data which are used to test economic laws. Theory of population, quantity theory money, theory of distribution, etc. which are the important theories of economics, have been propounded with the help of data. These all theories have been verified by the statistical methods. In applied economics statistical data are required to evaluate the progress of development plans and their implementation. Diagrams and graphs are used to analyse the effects of plans on economic variables. According to Bowley, "any student of economics cannot be a perfect one if he or she does not have the full knowledge of statistical methods."

Limitations of Statistics:

Scope of statistics is very wide. At present, the popularity of statistical methods is continuously increasing but there are certain limitations of statistics as described below:

i. Study of only the Numerical Facts:

Statistics studies the numerical facts only. It does not study the qualitative facts. In other words, statistics studies only those problems which are expressed in numbers only, e.g. age, height, production, price and wage. The quality issues like health, mental level, poverty, etc. are not directly studied under statistics.

ii. Study of Groups, not of Individual Units :

In Statistics we study the aggregate characteristics of numerical facts like per capita average income in the country. This per capita average income highlights only the aggregative characteristics and does not tell any thing about the per capita average income of poor, begger, rich persons.

iii. Not a Single Method of Study :

Statistical method is not a single method to study a problem. According to Croxton and Cowden, "It should not be answered that statistical method is the only method to use in study of any investigation. The results obtained by the statistical method shuld not be accepted as true until proved so by the other methods or verified by the other results."

iv. Illusion Oriented Conclusion:

To understand the statistical conclusions very well it is necessary to study their references also otherwise they can be proved wrong. According to Bowley, "students who use data should not be satisfied by assuming the results of investigation as proved but should also study all the related parts of knowledge.

v. Long Run Correctness of Statistical Rules only in the Average Form:-

Statistical rules, as the rules of Physics, are not completely correct and can not be applied always and in all circumstances. These rules are applicable only in the long run and in average form.

vi. To be used by Experts:

The work of collection, analysis and interpretation of data should be done only by experts who have the special knowledge of statistical methods. If the data are dealt by unqualified and untrained persons the conclusions can be derived or wrong conclusions either cannot be obtained. As Yule and Kendall write, "statistical methods in the hands of unqualified persons are like dangerous equipments."

As Conclusion, it is proper to mention that in the present time there is no science where statistics is not used.

Important Points :

- Statistics is a science which relates with significant numbers.
- The subject of statistics can be divided into two parts-

  (i) Statistical Method and (ii) Applied Statistics
- The laws and principles of Economics are derived from statistical data.
- There are limitations of applications of statistics. If these limitation are not kept in mind, the results may be illusive and biased.

Questions For Exercise

Objectives Type Questions:

1. In plural sense STATISTICS means-

(a) Statistics as Science

(b) Data

(c) Statistical Measures

(d) All of the above ( )

2. Statistics is –

(a) Science of Counting

(b) Science of Estimates and Probability

(c) Science of Data Interpretation and analysis

(d) All of the above ( )

3. Who defines statistics as "Statistics is science of counting or ends"-

(a) Bowley (b) Cowden

(c) John Griffin (d) Parsen ( )

Very Short Answer Questions :

1. Write the meaning of Statistics.
2. Write the meaning of Statistics in plural sense.
3. Define data.
4. Write any two limitations of statistics.
5. What are the areas in which applied statistics is used?

6. Who is the father of Statistics?

7. Define Statistics in singular sense.

Short Answer Questions :

1. What do you understand by Statistics ?

2. Explain the scope of Statistics in brief.

3. What are the different statistical methods?

4. Explain any two limitations of Statistics.

5. Explain the relation of Statistics with Economics in brief.

Essay Type Questions :

1. Explain the meaning and scope of Statistics.

2. Explain Statistics in brief. Describe its relation with Economics.

3. Define Statistics. Describe its limitations.

4. Define Statistics. Explain the role of Statistics in Economics.

Answer to objective Type Questions

(1) b (2) c (3) a

Reference books:

1. Kailash Nath Nagar- Sankhyiki ke Mool Tatva.

2. Ranga, Gupta, Goyal, Bhatnagar, Shah, Raghuvanshi- Sankhyiki, 2nd Edition.

Chapter - 2.2

Collection of Data

Statistical investigation is a complicated process. After planning the statistical investigation a proper method is chosen and then the work of data collection is started. Collection of data is the fundamental activity of the science of Statistics. Data are the base of the statistical Science. The success of analysis and interpretation of data and their further activities depends on the correctness of the data collection. The work of data collection for a statistical investigator is just like a building constructor who collects stone, sand, cement, brick, etc. As without the collection of building material the construction of building is not possible, in the same way investigation is not possible without data collection. If the collected data lack the accuracy and sufficiency, the results deried from these data will be doubtful and incorrect. So there is need of precautions in the data collection.

Primary and Secondary Data:

Data are of two types on the basis of the sources of data collection –

(i) Primary Data

(ii) Secondary Data

Primary Data:

Primary data are data which are collected first time newly by the investigator for the own use. In other words if the investigator collects data first time from the beginning to the end are known as primary data. When original data are collected about the playing habits of students after going on the play ground, these are the primary data.

Secondary Data :

There are the data which have already been collected by other persons or institutions and which have been published. The investigator only uses these data. In other words, if the data have already been collected by some investigator and other investigator uses these data for the investigation then for the second investigator the data are known as the secondary data. For example, if the investigator uses the data collected and published by government on agriculture, labour, employment then these data are known as secondary data.

Difference between Primary and Secondary Data :

These differences can be seen as below:

(i) Nature of Data:

Primary data are original and as the raw material for the statistical methods. Secondary data have already gone through the scope of statistics and are like the finished goods.

(ii) Collection of Data:

Primary data are collected either by the

investigator or by the representative of the investigator Secondary data are collected by other people or institutions which have already been published.

(iii) Planning of Collection of Data:

Primary data are newly collected after an independent planning while secondary data are already available. It means the data which are available in publications, reports, etc. are the secondary data.

(iv) Objective of Data:

Primary data are always collected according to the investigation while less money and time are used in the collection of secondary data. Secondary data are set according to the objective of the study.

(v) Money, Time and Effort :

More money, time and effort are required in the collection of secondary data.

Methods of Collection of Primary Data:

Following are the different methods used for collecting the primary data-

(i) Direct Personal Investigation

(ii) Indirect Oral Investigation

(iii) Information through corres- pondents

(iv) Information through schedules to be filled by the informants.

(v) Information through Enumerators

1. Direct Personal Investigation Method:

This method is appropriate for such investigations whose area is limited, has the local nature and in which more emphasis is given to the originality, accuracy and secrecy of data. In this method of primary data collection the investigator himself or herself meets to the persons from whom the data have to be collected in the investigation area and directly comes in the contact of the persons and thus collects data on the basis of inspection and experience. Investigation relating with income-expenditure, condition of labourers, educated unemployment etc. in the limited area are carried out by this method of data collection. Arther Young used this method in the study of agriculture production. This method has the merits of accuracy, uniformity, reliability and flexibility. In this method of data collection the investigator can be biased, there can be more expenditure and some others are the demerits.

2. Indirect Oral Investigation:

In this method of collection of primary data the information is not collected from the persons who are directly related to the problem to be investigated but is collected from the third persons by the way of oral investigation. Persons from which the information is to be collected are not contacted directly. For example, collecting information about way of life of labourers from trade unions or mill-owners instead of from labourors. This method is used when the area is large. Being economical, biaslessness, etc. are the merits of this method of data collection. Getting information indirectly, possibility of getting false information, etc. are the demerits of this method.

3. Information through Corresponds:

In this method the investigator appoints the local people or special corresponds who send or collect the time to time information and send the same to the investigator.

4. Information through Schedules to be filled by the Informants:

In this method the investigator prepares a list of questions or questionnaire and this questionnaire is sent by post to persons from which the information has to be collected.

Along with this the investigator also requests the informants to maintain secrecy and to send the information within the time. This method is appropriate when the area is very large. For the success of this methods the informants should be literate or educated. This method is used to collect data on family budget, polling survey, unemployment, etc. Being original, economic, etc. are the merits and being biasedness, lack of accuracy, etc. are the demerits of this method.

5. Information through Enumerators :

In this method of data collection a list of questions is prepared and given to enumerators who ask the questions and thus the schedules are filled in. Enumerators are trained and know the local language. This method is appropriate when labour and money resources are sufficient. Generally governments use this method for data collection. Census, Socio-economic survey, etc. are the areas in which this method is used. Accuracy, reliability, biasednessless, etc. are the merits of this method. This method is not economic, and there is chance of being delayed due to training.

Schedule and Questionnaire:

Schedules and questionnaires are used in the survey to collect data. Generally schedule is got filled in by the enumerator after asking the questions and the questionnaire is got filled in by the informants.

Merits of a Good Questionnaire :

Following are the merits of a good questionnaire :

(i) Size of questionnaire should be small and the number of questions should be less.

(ii) Questions should be simple and easy to understand.

(iii) Questions should be set in such a manner that their answers can be given correctly. Questions can be objective type.

(iv) Question should not be of the type to hurt the self respect of the informants. The spirit hurting questions should not be asked.

(v) Questions should be of the type so that the informants can give the answer.

(vi) Words used in questions should be correct and relevant.

(vii) Questions must have a clear-cut answer.

(viii) Order of questions should be proper and consistent so that answer can be given with strictly convenience.

(ix) Questions should strictly be related with the investigation.

(x) There should be all alternatives to a questions.

(xi) Open questions should be less in number (for example, give the suggestion to remove corruption).

(xii) Written instructions should be given to maintain secrecy regarding information and the given answers.

(xiii) Such questions should be asked so that their correctness can be tested.

(xiv) There should be clear instructions about filling in the questionnaire.

Difference between Questionnaire and Schedule :

Below mentioned are the differences between questionnaire and schedule.

(i) Questionnaire is filled in by the informants while schedule is filled in by the enumerator by asking the questions.

(ii) Questionnaire is sent to informants by post while schedule is filled in by enumerator who meets the informants.

(iii) There is no direct contact between investigator and the informants in case of

questionnaire while in case of schedule there is direct contact.

(iv) Questionnaire is related with an educated informant while schedule is related both with educated and uneducated informants.

(v) System of questionnaire is economic while system of schedule is related with higher expenditure.

(vi) In questionnaire the level of accuracy is lower but in schedule the level of reliability is higher.

Collection of Secondary Data:

Secondary data are the data which have already been collected by any person or institution or government. Thus secondary data are the data which are reused in the investigation.

There are two sources of secondary data:

(i) Published Source

(ii) Unpublished Source

(i) Published Source : Many investigating agencies, government departments, research institutions and corporations collect and publish the original data. Following are the sources of published data:

1. Government publications
2. Report of committers and commissions
3. Publications of semi-government institutions
4. Publications of commercial institutions
5. Publications of news paper and magazines
6. Publications of research institutions
7. Research work of universities
8. Publications of international institutions
9. Original books of expert persons

(ii) Unpublished Source: Sometimes government or other institution or persons (specially researchers) collect information or data on different subjects but they do not get it published. Such unpublished material is available or can be obtained from the office files, records, registers and diaries of researchers.

Scrutiny and Uses & Secondary Data:

Before using the secondary data, these data should critically be scrutinized and then the data should be edited in an enlarged way. There are many short comings in secondary data so these data should be used with precautions. The investigators should look into the reliability, sufficiency and accuracy of the secondary data before using these data in the study.

Precautions in Use of Secondary Data:

Following points should be noticed while scrutinizing the data on the basis of reliability, appropriatness and sufficiency:

i. First of all it is necessary to know that which investigator collected the secondary data primarily

ii. It is also essential to know that the used method of data collection is useful or not.

iii. Before using the secondary data it should be noticed whether the collected data are proper and useful as per the objective and area of investigation. If there is difference in objective and area these data are not considered to be appropriate.

iv. It should be ensured that the to which time the available data are related and in which circumstance data are collected. Data collected in war circumstances can not be used in the circumstances of peace. If there is difference in the circumstance of primary collection of data and their use, such data can be of less utility.

v. The level of precision in data collection should be kept high so that these data can be more reliable.

vi. Data should be compared if these are collected from different sources so that their reality can be judged.

Method of Statistical Investigation :

All the statistical investigations try to get the information about population. This information can be obtained in two ways which are known as the methods of statistical investigation. These two methods are :

(i) Census method

(ii) Sample method

i. Census method :

When investigator gets information from all the units of population, this is known as the census method or the complete enumeration. Under this method information of each unit of the population is collected. Population census in a country is an example of this method. Under population census information about each house and each person is collected. This information is wide one. Results of this method are pure and reliable. This method is very expensive, more labour using and time consuming.

ii. Sample method:

Under this method some units are selected from the population and conclusions are obtained from these selected units. While purchasing wheat, rice, etc. we do not examine each wheat or rice but a few quantity of wheat or rice is examined and thus on the basis of this examination we take decision whether to purchase wheat or rice or not. This method saves the time and money. This method requires some additional precaution otherwise there is possibility of wrong conclusion.

Sample method is based on three principles, known as three laws: These laws are: (i) Principle of Probability (ii) Statistical Regularity Law, and (iii) Law of Large Inertia.

Methods of Sample Selection :

When statistical investigation is carried out from the sample method, a sufficient number of unit are selected from the population. These are the samples which are unbiased and representative. Following are the methods of sample selection:

1. Purposive Sample :

When samples are selected by the investigator with the help of his own mind and experience, the samples are known as purposive sample.

2. Random Sample :

If samples are selected randomly and thus each unit of population has same probability to be included in the sample, the samples are known as random sample. Following are the methods of taking random samples.

i. Lottery Methods:

Under the lottery method chits of paper are prepared for each unit of the population and then one by one chits of paper are taken away by any unbiased person.

ii. By Rotating Drum:

Under this method of sample selection small pieces of wood or iron are taken and the number of all the units are mentioned one by one on these pieces of wood or iron. These numbered pieces are then put in this drum. Now drum is rotated. An unbiased person now takes away the pieces of wood or iron from the drum in the predetermined number.

iii. Blindfold Method :

Under this method a map is prepared by making as many sections on it as the number of units in the population. Now this map is used for the sample selection by a person having a piece of cloth on the own eyes. The person now throws arrow on the map and in which section of the map the

arrow touches that marked number becomes a sample.

iv. Arrangement in Definite Order :

Under this method all the units of population are arranged in a proper way according to the alphabet. Then desired units are selected from this list on the basis of the equal differences.

v. Using Random Number Tables:

Many Statisticians have prepared random number tables. Tippet's random number table has 10400 random numbers of 4 digits which have been arranged in an order. By giving the order number all selected with the help of the Tippet's table and samples are chosen.

3. Stratified Sample :

Under this method the units of population are divided into different groups on some basis. These groups are known as strata. Then out of these strata desired size of samples are drawn randomly form each stratum.

Other Methods :

There are some others methods of taking samples like multiphase random sampling multistage sampling, quota sampling and convenience sampling.

Important Points:

- On the basis of source data are of two types
  - i. Primary data
  - ii. Secondary data.
- Data collected by investigator directly from beginning to the end are the primary data.
- Data primarily collected by any other person and have been published are secondary data.
- Questionnaire is filled in by the informant.
- Schedule is filled in by the enumerator.
- Published source and unpublished source are the two sources of secondary data.

Question for Exercise

Objective Type Questions :

1. Primary data are:
   - (a) Original data
   - (b) Collected firstly
   - (c) Already not in existence
   - (d) All of the above ( )
2. Secondary data are collected by :
   - (a) Through schedules
   - (b) Through questionnaire
   - (c) Through published and unpublished sources
   - (d) All of the above ( )

Very Short Answer Questions :

1. How many types of data are from the collection point of view ?
2. What are primary data ?
3. What are secondary data ?
4. Why are primary data known as original data ?
5. What is direct personal investigation ?
6. What is meaning of questionnaire ?
7. What is random sample ?

Short Answer Questions :

1. What is difference between published and unpublished sources ?
2. State any there differences between primary and secondary data.
3. What are the primary data ? Give example.
4. What are the secondary data ? Give example.

5. Write any three merits of a good questionnaire.
6. Write any there differences between questionnaire and schedule.
7. Explain the sources of secondary data in brief.

Essay Type Questions :

1. Explain the difference between primary and secondary data and describe the methods of the primary data collection.
2. What do you understand by secondary data ? What are different sources of secondary data?
3. Critically analyse the different methods of the primary data collection.
4. Describe the methods of statistical investigation.

Answer to objective questions:

(1) d (2) c

Reference Books :

1. Kailash Nath Nagar- “Sankhyiki ke Mool Tatva.
2. Ranga, Gupta, Goyal, Bhatnagar, Shah, Raghuvanshi – "Sankhyiki", 2nd Edition, 2007.

Chapter - 2.3

Classification of Data

The collected data are generally complicated and unsystematic. These data are like a huge collection of facts which are not easy to understand and thus it becomes difficult to derive the desired results. So it is necessary to present these data in the systematic order to understand them easily and to derive the proper results. These unsystematic and unrefined data are just like the store of a street hawker who moves in streets and purchases old books and note-books, empty glass bottles, plastic and iron items old news papers, etc. and at last the hawker sells these items to the whole seller. There are many street hawkers who sell such items to whole sellers. These whole sellers then divide these goods/items into different classes according to the quality and type. This process enables the traders to sell these goods in the market according to demand. Thus unsystematic and unrefined data are classified to make them systematic and statistically useful for the analysis. Classification of data enables statisticians to compare them and to get the required conclusions.

Example : A teacher presents the data of marks of 50 students of a class in Economics as below:

22, 20, 30, 26, 31, 48, 25, 14, 19, 24, 11, 45, 27, 25, 06. 40, 13, 03, 29, 11, 24, 47, 02, 09, 45, 31, 20, 12, 15, 41, 49, 01, 27, 24, 31, 07, 02, 49, 14, 19, 17, 44, 47, 26, 09, 02, 42, 35, 26 and 17.

Can we draw any conclusion from these data? Because these data are not systematic and classified, so these data have to be arranged in an order (ascending or descending) and then the data will become comparable and meaningful.

Under classification collected data are divided into different classes according to different properties and qualities. Classification is a process in which the data are arranged in different classes or groups on the basis of equality and uniformity. Classification can be real or imaginary. In this process the uniformity in diversity of data is reflected.

Objectives of Classification :

Different objectives of classification are as described below:

1. Making Simple and Brief:

The objective of classifications is to put the collected data in the simple and short form so that data can easily be understood. So, in above example, it is difficult to get the average and other information If these data are classified as 0-10, 10-20, then they can easily be understood.

2. Explaining Equality and Inequality:

Under classification data are put into different classes according to different characteristics and thus they reflect the equality and inequality.

3. **Help in Comparison :**

The comparison of data becomes easier when they are classified. If the population of two cities or villages is classified as literate and illiterate or married and unmarried or employed and unemployed then both the cities or villages can easily be compared.

4. **Logic based Arrangement :**

Classification is a logic based activity. Under this activity data are presented systematically. For example, division of censes data in the classes of age, sex, caste, religion, state, urban or rural, etc. is a logic based activity.

5. **Basis of Tabulation :**

Tabulation is impossible without classification. Thus classification provides the basis for the tabulation of data. Also statistical analysis is impossible without classification.

Requirements of Idle classification :

The following elements are required for an idle classification:

1. **Clarity** : It should be clear and certain that in which class or group the collected data have to be arranged. Classes must be constructed in such a way that these classes should reflect the simplicity and clarity. Each similar item should be included in one class.

2. **Stability :** The classification should be stable so that data can be compared and the result can be meaningful.

3. **Universality** : The classes are constructed in such a way that any item could not be left without inclusion. Each and every item must be included in any of the classes. If necessary, some sub-classes under main classes can be formed. For example, under the married and unmarried classes, sub-classes of widow, widower divorcee, etc., can be constructed. So classification must be complete and wide.

4. **Appropriate:** Classes should be objective oriented. For example, to know the economic condition or saving behavior of people, the classes should be formed on the basis of income.

5. **Flexibility :** Classification should be flexible so that different classes can be changed, modified and adjusted as per the new requirements.

6. **Uniformity :** The units of each class should be uniform. In each class or group, all the same units should be kept.

Methods of Classification :

There are two types of statistical data- (i) Descriptive or qualitative data, and (ii) Numerical or quantitative data. The qualitative data can not be measured directly. These are measured only by the presence or absence criterion. For example, literacy, marital status, employment, etc. are the qualitative variables. Quantitative data are measured directly or by assigning the numerical values, for example, income, age, weight, height, etc. These facts are known as variables also. variables are always having changing values.

On the basis of the above two types of data, there are two types of classification also – (i) Qualitative classification and (ii) Quantitative classification or classification according to class-intervals.

1. Qualitative Classification :

This classification is also known as the classification according to attributes. When facts are classified on the basis of characteristics, it is known as the classification according to attributes. This classification has two types as below :

(i) When the facts are classified in two classes on the basis of the presence or absence of a characteristic, this is known as simple classification or biclassification. For example, the available data are classified as rural or urban and male or female.

(ii) When facts are divided on the basis of two or more characteristics, this is known as the multiple classification. For example, the classification of census data as male and female and then these both are again classified as literate and illiterate and then further classified as employed and un-employed is the multiple classification.

These both types of classification can be shown as below :

Whole population is classified as:

Male :

i. Literate

ii. Illiterate

i. Literate :

(a) Employed

(b) Unemployed

ii. Illiterate :

(a) Employed

(b) Unemployed

Female

i. Literate

(a) Employed

(b) Unemployed

ii. Illiterate

(a) Employed

(b) Unemployed

2. Quantitative Classification:

This is the classification according to class-intervals. Under this classification data are divided in different classes on the basis of some characteristics, for example, income, production, marks obtained, age, etc.

Example 2 : The marks of 100 students of a class are classified as below in the class intervals of 10-10. Marks are between 00 and 98.

Marks obtained	No. of Students
0-10	3
10-20	7
20-30	8
30-40	11
40-50	30
50-60	20
60-70	13
70-80	4
80-90	3
90-100	1

Following are the concepts related with the quantitative classification:

(i) Class Limits :- Each class has two limits. The first limit is known as the lower limit (L_1) and the second limit is known as the upper limit (L_2). In the class (50-60), L_1 is 50 and L_2 is 60.

(ii) Class- Intervals :- It is the difference of L_2 from L_1, denoted by i. so, i= L_2-L_1. In the above example, i is 10 for each class. The class-interval can be same or different for all classes.

(iii) Mid-Point or Mid-Value : Mid-point is the value in the centre of the class. It is the half of the sum of both the limits of a class.

Thus,

Mid-point = $\frac{L_1+L_2}{2}$

(iv) Class-Frequency :- Class frequency is the number of items in the both limits of a

class. In the above example, number of students is 11 of the class (30-40) which is known as the class frequency. The class-frequency is denoted by 'F'.

Frequency Distribution :

A frequency distribution explains the classification of data on the basis of a measurable variable. This distribution is a table in which the values of a variable are shown in different classes along with their frequencies. Thus, frequency distribution is a systematic arrangement of classes of different values with their frequencies. Clearly, frequency distribution is a function of variable and frequency.

Variables are of two types and are denoted by 'X'. These are.

(i) Discrete variable, and

(ii) Continuous Variable

Discrete variables are the variables which have the definite values. They do not have range and their units can not be divided.

For Example; 0, 1, 2, 3, … as the marks obtained in examination, number of children in the family, number of goals in a football match, etc.

Continuous variable is that variable whose value is not certain or fixed. This variable can take any value in the given class. For example, in the class (10-20) the values which can be included are any between 10.01 and 19.999.

Following table shows both the discrete and continuous frequency distribution for the purpose of comparison.

Discrete Frequency Distribution		**Continuous Frequency Distribution**	
No. of Children	**No. of Family**	**Marks Obtained**	**No. of Students**
0	5	10-20	7
1	18	20-30	12
2	35	30-40	20
3	21	40-50	40
4	13	50-60	16
5	8	60-70	5
Total	**N = 100**	**Total**	**N = 100**

When data are very large in number and spread in irregular way then tally marks are used for arranging the frequencies in different classes. This makes the classification simple and comfortable. For these tally marks a symbol (I) is used. If there are two values in a class then the tally marks used are (II). For five values in a class the symbol used is (𝍸). After this all the tally marks are added which becomes the frequency of a particular class.

Problems of Classification according to Classes:

Under the quantitative classification, following problems are faced and thus these problems must properly be solved.

(i) Number of Classes :

There is not any fixed rule to fix the number of classes. Number of classes should properly be decided so that important characteristics of data can be reflected and important information can be included. Generally, number of classes should be in between 5 to 15.

(ii) Magnitude :

Normally the magnitude depends on two factors- (i) the maximum and minimum values in given data, and (ii) number of classes. It should be tried to keep the same magnitude for all classes as for as possible. The magnitude (i) is calculated as below:

$$i = \frac{Maximum\ value - minimum\ value}{N}$$

$$= \frac{L-S}{N}$$

where, N - number of classes.

(iii) Class Limits:

The class-limits of different classes should be clearly determined and as for as possible these limits should not be in fractions. Each and every unit of the group must be included in any class. Class-limits can be in both forms of inclusive and exclusive. But for simplicity class-limits should be in the exclusive form.

Statistical Series:

We study different series in statistics. In reality, a series is a systematic and logical order of items or facts. Statistical series are the systematic order of arrangement of statistical factors according to some fixed basis. Statistical series are of three types on the basis of their construction. These are:

(a) Individual Series

(b) Discrete Series

(c) Continuous Series

(a) Individual Series :

In these series each item is individually and independently important. Each item is individually measured and thus not included in any group or class. If marks of 40 student in a class or the monthly expenditure of 25 families are shown individually this becomes an individual series.

Example: 30 students of a class obtain marks out of 10 in a monthly test as follows:

8, 2, 9, 3, 5, 8, 6, 1, 0, 5, 5, 4, 2, 9, 8, 8, 4, 5, 3, 7, 7, 2, 3, 5, 9, 3, 4, 6, 1 and 7.

In this series there is no column of frequency.

(b) Discrete Series :

The series in which each unit has its actual or real measurement is known as the discrete series. In this series the given values of a variable are not divisible. Thus discrete series is that one which is obtained by orderly arrangement of the values of the discrete variable. The variable generally does not have the values in fractions.

For example, number of children in a family, number of pages in a book, accidents in a city, marks obtained by students, etc. are the discrete variables.

Example : The marks of 30 students (given in the above example) can be arranged in a discrete series as below:

Marks (X)	No. of Students (F)
0	1
1	2
2	3
3	4
4	3
5	5
6	2
7	3
8	4
9	3
10	0
Total	**N = 30**

In this series X is a discrete variable and F is the frequency.

(c) Continuous Series :

This series is concerned with the continuous variables. Continuous variables do not have fixed value, but they can have any value in the certain class or limit. Values of variables are shown in classes rather than showing individually. This type of series is known as continuous series. Each value is kept in a class and thus the real measure is not clear. There is continuity in the continuous series which means the upper limit of the

preceding class is equal to the lower limit of the next class. The continuous series is used for the continuous variables like age, height, weight, income, production, saving, consumption, etc.

Example :- By using the above data of marks of 30 students, continuous series can be constructed with the magnitude of 2-2 marks as shown below:

Class of Marks (X)	No. of Student (F)
0-2	3
2-4	7
4-6	8
6-8	5
8-10	7
Total	**N=30**

Difference between Discrete and Continuous Series :

The difference between discrete and continuous series can be shown as below:

(i) **On the Basis of Structure :** In discrete series the value of units is given while in continuous series classes are given.

(ii) **On the Basis of Measurement :** The measurement in discrete series is actual and generally not fractional. In continuous series the measurement is not actual but it is artificially constructed in terms of classes.

(iii) **On the Basis of Discontinuity :** There is discontinuity in the discrete series and there can be a fixed difference in item-values. In continuous series there is continuity.

(iv) **On the Basis of the Source of construction :** Discrete series is constructed by discrete variables and continuous series is constructed by continuous variables.

Methods of Classification according to Class-Intervals :

Here, there are two methods :-

(i) Exclusive Method, and

(ii) Inclusive Method

1. Exclusive Methods :

In this method, the upper limit of the previous class is equal to the lower limit of the next class. This method is exclusive method because the value of variable equal to the upper limit of class is not included in that class but included in the next class. If the income class in an institution is Rs. (400-500) per month then the wage of Rs. 500 is not included in this class but included in the next class of Rs. (500-600).

Example : Exclusive class-intervals can easily be understood by the two below given tables:

Table-I

Marks	
0-10	0 but less than 10
10-20	10 but less than 20
20-30	20 but less than 30
30-40	30 but less than 40
40-50	40 but less than 50

Table II

Income (Rs.)	
0-400	0 but less than 400
400-1000	400 but less than 1000
1000-2000	1000 but less than 2000
2000-5000	2000 but less than 5000
5000-10,000	5000 but less than 10,000

2. Inclusive Series :

The class in which the values of a variable between the lower limit and upper limit are

included is known as the inclusive series. In this method the value of a variable equal to the upper limit of the class has to be included in that class. This method of inclusive series is recognized by the fact that the upper limit of the previous class is not equal to the lower limit of the next class and the maximum difference between these two limits is not more than 1.

Example : The inclusive series can be understood with the help of 3 columns of the below table:

I Column	**II Column**	**III Column**
Weight of children (X)		
40-45	0-9	20-29.5
46-50	10-19	30-39.5
51-55	20-29	40-49.5
56-60	30-39	50-59.5
61-65	40-49	60-695

Changing Inclusive Series in Exclusive Series:

Generally inclusive series is used for the discrete variables like number of labourers, marks obtained, etc. but for continuous variables, like income, age, weight, etc. exclusive series is used. If the data are given in inclusive series it should be changed into exclusive series for the simplicity. Under this process the difference of upper limit of the previous class and the lower limit of the next class (d) is halved. This halved difference is added to the upper limits of all the classes to get the real upper limits and subtracted from the lower limits of all the classes to get the real lower limits/ So,

Real $L_1 = L_1 - \frac{d}{2}$

and Real $L_2 = L_2 + \frac{d}{2}$

In the first column of the above table the difference in all the classes is same. So,

$d = 46 - 45 = 1$

Half of difference = $\frac{d}{2} = 0.5$

Thus, $L_1 = 46\text{-}0.5 = 45.5$

$L_2 = 45+0.5 = 45.5$

In this way the data of first column of the above table are changed in the exclusive series.

To calculate statistical averages (arithmetic mean, mode, median, etc.) the frequency distribution must be normal. If the cumulative frequency distribution is given, the same has to be changed into the normal frequency distribution. With the help of examples the following two data series can easily be understood:

(i) Changing normal frequency distribution in the cumulative frequency distribution.

(ii) Changing cumulative frequency distribution in the normal frequency distribution.

(i) Changing normal frequency distribution in cumulative frequency distribution :

The following table shows the normal frequency distribution.

Classes (x)	Frequency (F)
0-5	4
5-10	16
10-15	20
15-20	8
20-25	2
Total	N=50

The above example can be changed in two types of cumulative frequency distributions, namely:

(a) Less than cumulative frequency Distribution.

(b) More than cumulative frequency distribution.

(a) Less than Cumulative frequency distribution :

By using upper limits of all the classes of

the above table, less than cumulative frequency distribution is shown as below:

'Less than' Cumulative Frequency Distribution

Class (x) (Less than)	Cumulative Frequency (CF)
Less than 5	4
Less than 10	20 (4 + 16)
Less than 15	40 (20 + 20)
Less than 20	48 (40 + 8)
Less than 25	50 (48 + 2)

(b) More than Cumulative Frequency Distribution:

The above normal frequency distribution can be changed in 'more than' cumulative frequency distribution by using the lower limits of all classes as below:

'More than' Cumulative Frequency Distribution:

Marks(x) (More than)	**Cumulative Frequency (CF)**
More than 0	50 (46+4)
More than 5	46 (30 + 16)
More than 10	30 (20 + 10)
More than 15	10 (4 +6)
More than 20	4

(ii) Changing Cumulative frequency Distribution in Normal Frequency Distribution :

In this process the classes are formed on the basis of the two nearby class-limits and then the cumulative frequencies are changed in the normal frequencies.

Example : Change the following series in the normal frequency distribution –

Marks(x) (More than)	**Cumulative Frequency (CF)**
Less than 5	4
Less than 10	20
Less than 15	40
Less than 20	48
Less than 25	50

Solution:

Marks obtain (X) (Less than)	**CF**	**Class**	**F**
Less than 5	4	0-5	4
Less than 10	20	5-10	16(20-4)
Less than 15	40	10-15	20 (40-20)
Less than 20	48	15-20	8 (48-40)
Less than 25	50	20-25	2 (50-48)

In the same way the 'more than' cumulative frequency distribution can also be changed in the normal frequency distribution.

Miscellaneous Question :

1. The marks obtained by 50 students in an examination are given as below:

25	30	45	42	52
50	73	60	58	43
76	78	44	63	42
49	44	38	54	67
41	52	65	65	70
37	61	35	40	84
32	50	61	88	54
49	87	35	65	51
50	35			

Change the above data in a frequency distribution by taking 20-30, 30-40, etc. as classes.

Solution : In this question the first class is taken as (20-30) and with 10 as class-interval, other classes are formed as given below.

Class	Tally Marks	Frequency (F)
20-30	II	2
30-40	~~IIII~~ IIII	9
40-50	IIII IIII I	11
50-60	~~IIII~~ ~~IIII~~ III	13
60-70	~~IIII~~ III	8
70-80	IIII	4
80-90	III	3
Total		**N = 50**

2. Put the following frequency distribution in (i) less than and (ii) more than cumulative frequency distribution.

Marks Obtained	No. of students
0-10	21
10-20	39
20-30	32
30-40	58
40-50	37
50-60	13

Solution : (i) Above distribution has been shown in less than cumulative frequency distribution as below:

Marks Obtained	CF
Less than 10	21
Less than 20	60 (21+39)
Less than 30	92 (60+32)
Less than 40	150 (92+58)
Less than 50	187 (150+37)
Less than 60	200 (187+13)

(ii) The 'more than' cumulative frequency distribution is shown as below :

Marks Obtained	CF
More than 0	200
More than 10	179 (200-21)
More than 20	140 (179-39)
More than 30	108 (140+32)
More than 40	50 (108+58)
More than 50	13 (50-37)

3. Change the following inclusive series in the exclusive series.

Class	F
1-10	5
11-20	10
21-30	15
31-40	8
41-50	7

Solution: There is a difference of 1 between the upper limit of previous class and the lower limit of the next class. The half of this difference has to be subtracted from the lower limit of each class and has to be added to the upper limit of each class. This method changes the above series into the exclusive series as below:-

Class	F
0.5-10.5	5
10.5-20.5	10
20.5-30.5	15
30.5-40.5	8

40.5-50.5	7

4. Change the following classes of unequal intervals to classes of equal intervals.

Class	F	Class	F
0-3	5	18-20	24
3-6	8	20-24	12
6-10	12	24-25	15
10-12	14	25-28	10
12-16	16	28-30	6
16-18	16	30-36	5

Solution : In the above table the class-intervals are not equal for all classes. The class-intervals can be changed in equal class-intervals by considering 6 as the difference. Thus new classes become as 0-6, 6-12, …… Accordingly the frequencies will also be adjusted as shown below:

Class	F
0-6	13
6-12	26
12-18	32
18-24	36
24-30	31
30-36	05

5. Change the following series in the normal continuous series.

Mid-value	F
10.5	3
17.5	7
24.5	15
31.5	25
38.5	40
45.5	18
52.5	12
59.5	10

Solution: The mid-value of above series have to be changed in the equal class-intervals. There is difference of 7 between all the mid-value of the series. To change these data into normal continuous series the half of 7 (i.e.3.5) has to be added to the mid-values to get the upper limit and 3.5 has to be subtracted from the mid-values to get the lower limit of the classes. Thus the changed series appears as below:

Mid-Value (m)	Class	F
10.5	10.5±3.5= 7-14	3
17.5	17.5±3.5 = 14-21	7
24.5	24.5 ± 3.5 = 21-28	15
31.5	31.5±3.5=28-35	25
38.5	38.5±3.5=35-42	40
45.5	45.5±3.5=42-49	18
52.5	52.5±3.5=49-56	12
59.5	59.5±3.5=56-63	10

Important Points:-

- The objective of classification of data is to make tham systematic.
- Classification is the process of arrangement of data in different classes on the basis of similarity and uniformity.
- With the help of classification the collected data are presented in brief and simple forms which become helpful in comparison and in making them similar.
- Clarity, stability, flexibility, and uniformity are the necessary elements of an idle classification.

- There are two methods of classification (i) Qualitative Classification and (ii) Quantitative classification.
- In quantitative classification, the terms like class-limits, class-interval, mid-point, class frequency are used.
- The arranged form of values or classes and their frequencies is known as frequency distribution.
- The frequency distribution is of two types (i) discrete frequency distribution and (ii) continuous frequency distribution.
- Number of classes, class-interval and class-limits are the main issues of classification.
- There are three types of statistical series – (i) individual series, (ii) discrete series, and (iii) continuous series.
- According to classes inclusive and exclusive are the two methods of classification.
- The cumulative frequency series are known as (i) less then and (ii) more than series. Cumulative frequency distribution is changed in normal frequency distribution.

Question for Exercise

Objective Type Questions :

1. In an exclusive series,

   (a) Both the class limits are considered

   (b) Lower limit is not considered

   (c) Upper limit is not considered

   (d) Both limits are not considered ( )

2. The frequency of each item-value in individual series is.

   (a) Equal (b) Unequal

   (c) Both above situations possible

   (d) No truth exists ( )

3. Main objective of classification is,

   (a) To show large data group in brief.

   (b) To make data flexible

   (c) To make data stable

   (d) To make data mutually exclusive ( )

4. Name of the following series is,

   X : 1 2 3 4 5

   No. of Students : 20 4 2 3 1

   (a) Individual

   (b) Discrete

   (c) Continuous inclusive

   (d) Continuous exclusive ( )

5. The mid-value of the class of 10 as lower limit and 20 as upper limit is:

   (a) -15 (b) 10 (c) 15 (d) 30 ( )

6. Which is not a quantitative fact ?

   (a) Height (b) Weight

   (c) Unemployment (d) Age ( )

Very Short Answer Questions :

1. Name the two types of qualitative classification.
2. What is the classification on the basis of variables ?
3. What do you mean by variable ?
4. Write names of different series
5. What are the class-limits ?
6. How is the mid-value calculated ?
7. Which limits are used in 'less than' and 'more than' while getting the cumulative frequency ?

Short Answer Questions :

1. What do you understand by classification of

data ?

2. What are main objectives of classification ?
3. Write any four elements of an idle classification.
4. What is frequency distribution?
5. Differenciate the exclusive and inclusive series.
6. How is the normal frequency distribution changed in the cumulative frequency distribution ? Give Example.
7. Do you agree that classified data are better than raw data ?

Essay Type Questions:

1. Explain the exclusive and inclusive methods used in the classification of data.
2. Explain the necessary elements of an idle classification. What are the objectives of classification ?
3. Construct the discrete and continuous series by taking an imaginary example.

Answer to objective Questions :

(1) c (2) d (3) a (4) b (5) c (6) c

Reference Books:

1. S.C. Gupta & V.K. Kapoor: Fundamentals of Mathematical Statistics, Published by sultan chand and sons
2. S.P.Sing: Sankhyki : Sindhant Avam Vyavhar, S. Chand
3. K.N. Nagar: Sankhyki ke Mool Tatva, Minakshi Prakashan.

Chapter - 2.4

Presentation of Data

Collected data are complicated and unsystematic. Generally data are in large quantity. Originally, the collected data are in the large groups which cannot easily be understood nor any proper logical conclusion can be derived. So there is need of making data presentable in the suitable form. It is necessary that the collected data should be simple, brief and understandable so that comparative study, analysis of data and their interpretation can become possible. To understand the collected data easily and to make them useful, there is need of the proper presentation of data. Generally, presentation of data can be of three types.

(a) Textual or Analytical Presentation

(b) Tabulated Presentation or Tabulation

(c) Diagrammatic and Graphical Presentation

(a) Textual or Analytical Presentation :

Under this type of presentation data are described in the text. This is a proper method for the quantitative data. For Example, during the student strike in the city on October 25, 2015 only 12 shops were opened and 155 shops shut down and only two school run and 24 schools closed.

(b) Tabulation :

The collected data are firstly classified then tabulated for the purpose of their comparison and interpretation. Tabulation makes data simple, brief and easy to understand. Thus, tabulation means the presentation of the classified data in tables for making them simple and brief.

Broadly speaking, tabulation is a process in which data are presented in rows and columns. Tabulation enable in deriving the final logical result.

Objective :-

There are three objectives of tabulation:

(i) Presentation of data in the arranged way.

(ii) Presentation of data in the simple and brief way.

(iii) Making the problem simpler and more clear.

Importance of Tabulation:

Tabulation is the important process between classification of data and their interpretation. Under tabulation we arrange data for the statistical use and for taking necessary decisions.

The importance of tabulation can be explained on the basis of following points:

1. Simplicity :

Tabulation helps the understanding the necessary information easily and speedily and thus it ends the complexity of data. The incorrect

information can easily be found out with the help of tabulation.

2. Comparative Study :

With the help of tabulation, the comparison can easily be made because the comparable data are mentioned in the nearby spaces.

3. Demonstration :

Tabulation helps in presenting data in the form of diagrams and graphs with more attraction.

4. Saving of Time and Space :

The large sized facts and data use the less time and space if presented in tables.

5. Statistical Analysis :

The tabulation is used for necessary calculation of statistical values relating with dispersion, skewness, regression, correlation, etc. These measures are the statistical methods which are used to study the socio-economic variables.

Difference between Tabulation and Classification :

Tabulation and classification both are the important functions of the statistical investigation. The scattered and unsystematic data can be made simple, brief and understandable with the help of tabulation and classification. There are some difference in these two which are as below :-

1. First of all the unsystematic data are classified and then put into different categories. Thus tabulation is based on the classification.
2. Under classification the data are put into different classes or categories on the basis of their similar or dissimilar properties. Under tabulation the classified data are arraged into rows and columns.
3. Classification is a statistical method while tabulation is a process of presentation of data.
4. Under classification data are divided in classes or sub-classes while under tabulation data are shown in titles and subtitles.

Construction and Main Parts of a Table :

Before tabulating the data, it is necessary to know that what are the important parts of a good table. The simplest way of tabulating data is to adjust the data in rows and columns. Following are the main parts of a good table :

1. Table number
2. Title of the table
3. Stubs and captions
4. Main Body of the table
5. Drawing lines and keeping spaces vacant
6. Arrangement of items
7. Unit of measurement
8. Footnotes
9. Origin or sources

The brief analysis of the above parts is as below:

1. Table Number :

The table can be identified with the table number. The table number is mention on the top of the table. For Example, table 6.3 shows that it is table 3 of chapter 6.

2. Title :

The title of the table shows the subject matter of the table. This title should be clear, brief and correct and to be written with the or after the table number.

3. Stubs and Captions :

Each row of table should be given a title, which is known as stub. The stubs are mentioned in the extreme left column of the table. Similarity each column of the table are also given the title known as caption.

4. Main Body of the Table :

The facts or data are given in the main body of the table. Main body is the heart of the table. The size of main body depends on the matter and availability of data.

5. Drawing Lines and Keeping Spaces Vacant :

By drawing lines and keeping spaces vacant the tables can be made beautiful. Which line should be bold, which line should be thin and what should be colour of different lines depend on the subject matter of the table.

6. Arrangement of Items :

The table can be made more attractive and useful by arranging the items systematically and orderly. All the comparable cells should be kept near to each other in the table.

7. Unit of Measurement :

If the whole table has the same unit of measurement, the unit of measurement should be mentioned alongwith the title. The different units of measurement should be mentioned alongwith the stubs and captions.

8. Footnotes:

If any relevant information has not been given in the table or any fact needs some special explanation, it should be given in the end of the table.

9. Source :

The source of data and information should also be given at the end of the table so that the data can be shown doubtless.

The main parts of the table has been shown in the below chart:

Table 6.3

Title of Table

(in %)

Stubs/Row	Captions/Columns	Total

Title	Title		
	Literate	Illiterate	
Rural	Body of Table		
Urban			
Total			

Footnotes : .....

Source : .....

Kinds of Tables :

There are different bases on which the tables are constructed . The following chart shows the classification of tables:

1. Objective Based Table

i. Common Objective

ii. Special Objective Table

2. Origin Based Table

i. Primary Table

ii. Derived Table

3. Structure Based Table

i. Simple Table

ii. Complex Table

A. Complex Table :

i. Two-way Table

ii. Three-way Table

iii. Multiple Table

1. Objective Based Table:

The common objective does not have any special objective. This is also known as the reference table. Special objective table is constructed to fulfill a special objective. The size of special objective table is limited in comparison to the common objective table. This is also known as the analysis table where we use mean, percentage, ratio, etc.

2. Origin Based Table :

The original data are shown in the primary table. This is also known as the classification table. In derived table total, percentage, ratio, multiple, etc. are shown along with the original data.

3. Structure Based Table :

When data are expressed on the basis of only one characteristic, this is known as simple table . Distribution of population according to age/sex/state is the example of the simple table.

When data are expressed on the basis of more than one characteristics, table is known as complex table. A complex table can be two-way table, three-way table or multiple table. The classification of population according to age and sex is the example of two-way table and the classification of population according to age, sex and literacy is the example of three-way table.

Example 1 : Construct a blank table showing the classification of man power data according to age, sex and rural urban domicile.

Solution : A three-way table has to be constructed for the above example.

Man Power According to Age, Sex and Rural-Urban Domicile.

Age Class (Year)	Rural			Urban			Total		
	M	F	Total	M	F	Total	M	F	Total
0-20									
20-40									
40-60									
More than 60									
Total									

Foot Note – M – Male, F- Female

Source : Census 2011

Rules for Construction of Statistical Table:

The construction of a sound table depends on the qualification and experience of the investigator. Following are the rules of constructing a good table:

1. Title :

Each table must have a clear, complete and brief title which can reflect the subject, time, and classification of the data.

2. Rows and Columns :

The number of rows and columns should initially be decided according to the objective of and the data to be presented in the table. Columns must be given the serial numbers. The columns titles must be clear along with the unit of the measurement. Numbers of columns should be not unnecessarily large.

3. Comparison :

The data to be compared must be shown near to each other. The derived values like percentage, ratio, multiple, etc. must be kept nearer to the original data.

4. Lines :

The important information should be shown alongwith the bold lines so that the reader can easily be attracted towards these values.

5. Arrangement of Items:

The different items must be arranged in the table according to importance, size, place, time etc. The data of more importance must be shown in the limited spaces in the table.

6. Special Importance:

The special information should be shown by bold letters in the table.

7. Footnotes :

Any information which needs some special attention and that has not been cleared in the table should be mentioned in the footnote at the end of the table.

8. Origin or Source:

The source of data should be mentioned at the end of the table.

9. General Rules:

An investigator must follow all the above rules of constructing the table alongwith some general rules like: size of table should be according to the size of the paper, if information or data are more, more than and table should be constructed, etc but it should be kept in mind that each table should be complete, simple, economic and that can easily be understood.

(c) Diagrammatic and Graphical Presentation :

Data are interestless and thus showing data in the form of diagrams and graphs can be understood easily. It becomes very easy to take conclusion with the help of diagrams and graphs. Thus, presentation of data in terms of diagrams and groups helps in understanding the facts without putting much burden on the mind. When data are presented by diagrams and graphs in newspapers and magazines, they can easily be understood. Here, there are two methods for the analysis: (1) Diagrammatic presentation of data, and (2) Graphical presentation of data.

1. Diagrammatic Presentation of Data :

Diagrammatic presentation means showing data with the help of simple and attractive geometric figures like bar diagram, rectangle, circle diagram, etc. One diagram is equal to many thousand words. Following are the benefits of diagrammatic presentation of data.

A. Attractive and effective

B. Simple and understandable

C. Labour and time saving

D. Helpful in comparison

E. Wide application

F. Source of entertainment and information.

These benefits are emplaning as below:

A. Attractive and Effective :

Diagrams are more attractive and effective so these are much popular. The matter which is not clear in terms of data can be learnt easily with the help of diagrams.

B. Simple and understandable presentation :

The complicated and unsystematic data can simply and easily be studied and understood with the help of diagrams. Diagrams give instant information about the subject. In the same way, the complicated facts related with data can easily be understood with the help of diagrams.

C. Saving of Labour and Time:

Generally more time and labour are required in deriving conclusions from the given data. It is the benefit of diagrams that these can be studied in less time and with less labour. So diagrams same time and labour.

D. Helpful in Comparison :

Different facts can easily be compared with the help of diagrams. The comparison based on diagrams is more effective than based on numerical comparison. If a table gives data of output for 8 years and on the other hand this information is shown in diagrams, then it is easier to compare with the help of diagram.

E. Wide Application :

Diagrams are widely used in every sphere of our life. Diagrams are intensively as well as extensively used to study the issues relating with trade, commerce, advertisement, education, health, etc.

F. Entertainment and Information :

Diagrams are a very good source of

entertainment and information. So there is an easy attraction of people towards diagrams.

Generals Rules for Constructing Diagrams:

Diagrams should be constructed with precautions. This is the job of expertization and qualification and that should be taken into account while constructing diagrams. The following rules should be followed to make the diagrams attractive and effective :

(1) Attractive and Cleanliness:

Diagrams are the visual aids of statistical data presentation. Diagrams are attractive to our eyes and leave permanent effect on our mind. So, diagrams must be clean, interesting and attractive. Diagrams can be made more attractive by using different types of columns, points and lines.

(2) Accuracy :

While making diagrams attractive, the accuracy should be ensured. The inaccurate and wrong diagrams give the misleading and doubtful results.

(3) Suitable Size :

There is no certain rule about size of diagrams. Diagram should neither be very large sized nor very small sized.

(4) Title and Footnotes :

There should be a clear, appropriate and brief title of the diagrams so that the facts and subject matter can be known. The required footnotes should be given below the diagram on the left side.

(5) Selection of Scale :

Selection of scale is very important in the diagrammatic presentation of data. The scale should be determined according to the size of the paper, nature of data and important characteristics of data. The scale should properly be mentioned both on vertical and horizontal axises. When two or more diagrams are compared the scale should be same.

(6) Index :

The necessary indexes are to be given on the right corner above to explain the various used symbols (dots, lines, shades, squares, etc. so that diagrams can easily be understood.

(7) Simplicity :

The diagrams must be simple to understand so that readers should not face any doubtful situation.

(8) Choice of a Suitable Diagram :

There are many types of diagrams used to present the data. Which diagram has to be used depends on the nature of data, range of data, etc. The knowledge, experience and efficiency are also required for the selection of the proper diagram.

Kinds of Diagrams :

The statistical diagrams are mainly of 5 types:

1. One Dimensional Diagram

These are as below:

i. Line Diagram

ii. Simple Bar Diagram

iii. Multiple Bar Diagram

iv. Sub-divided Bar Diagram

2. Two Dimensional Diagram

These are of 3 types:

i. Rectangular Diagram

ii. Square Diagram

iii. Circular Diagram

3. Three Dimensional Diagram

4. Pictograms

5. Cartograms

Here we study the line diagram, simple bar diagram, rectangular diagram and the circular diagram.

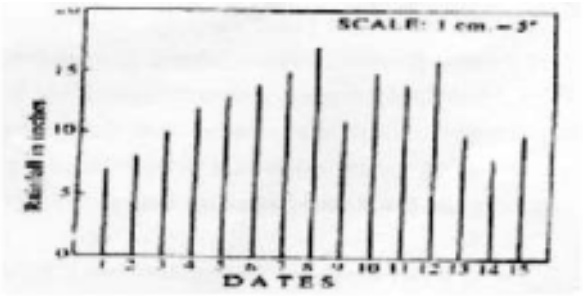
(A) Line Diagram :

Line diagram is a one dimensional diagram. When the number of item-values is large and the difference between small and large values of the data series is less then line diagram is considered to be appropriate. In line diagrams the gap between different lines is kept same and vertical lines are drawn equal to the each value of items. These lines are not bold and thus are less attractive. Thus the given values can easily to compared.

Example : The data about the rainfall of 15 days are given below. Show these data by the appropriate diagram.

Date	Rainfall
1	7
2	8
3	10
4	12
5	13
6	14
7	15
8	17
9	11
10	15
11	14
12	16
13	10
14	18
15	10

Solution : In case of the given data the line diagram is appropriate which is drawn as below :



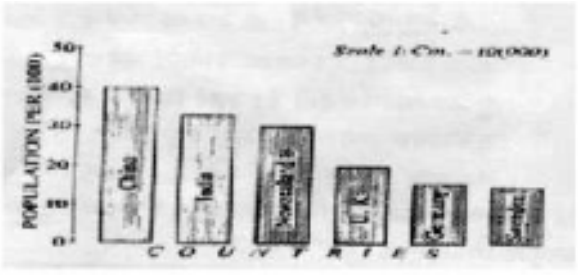
(B) Simple Bar Diagram :

This is also one dimensional diagram. Simple bar diagrams are used if the item values are small. There is a difference between line diagrams and bar diagrams that bar diagrams are constructed with certain width so that they can be more attractive. Simple bar diagrams are drawn with equal width and the height of bar is in the ratio of values of items. In these diagrams the difference among bars is kept equal. Simple bar diagrams can both be vertical and horizontal. These diagrams are more appropriate for the presentation of individual data and the time series data.

Example : The data of birth rate (per thousand) of different countries in a certain time period are given in the below table. Construct a proper diagram.

Country	China	India	Neweland	U.K.	Germany	Sweden
Birth Rate	40	33	30	20	16	15

Solution : Birth Rate (,000) in Different countries scale: 1cm=10,000



(C) Rectangular Diagrams :

These diagrams are two dimensional diagrams. In one dimensional diagrams only one dimension (height/width) is considered in the

construction. The areas of two dimensional diagrams are in the ratio of the item-values so these are also known as surface diagrams or area diagrams.

The rectangular diagrams are used to compare the two or more than two values. There are two types of rectangular diagrams.

(i) Percentage Sub-divided Rectangular Diagram

(ii) Divided Rectangular Diagram

(i) Percentage Sub-divided Rectangular Diagram :

By this diagram, family budgets of different families can be compared. In these diagrams, the total income is assumed to be 100 (%) and the expenditure on different items is changed into percentage. Now, rectangular diagrams are constructed with the equal height of 100. The width of these rectangles is kept in the ratio of total expenditure and then on different items the rectangle is divided into different parts.

Example : Present the below given monthly expenditure (in Rs.) of Two families by the two dimensional diagram.

Item of Expenditure	Family A	Family B
Food	400	500
Cloth	200	500
House Rent	160	200
Fuel	80	100
Misc.	160	400
Total Expenditure	1000	1600

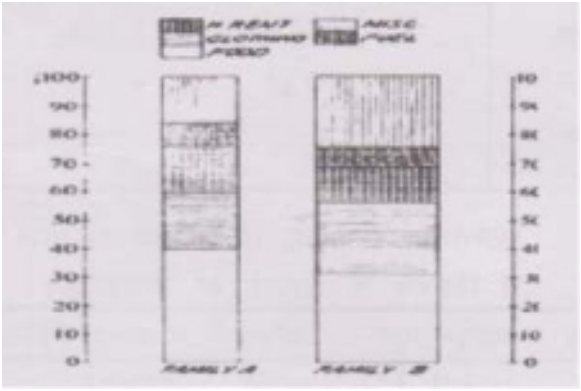
Monthly Expenditure of Two Families

Solution : The expenditure amount can be changed in percentage according to the following table assuming income of Rs. 100 for both families. If the spent amount on various items is less than total expenditure, it may be shown as savings of the families and then construct a rectangular diagram.

Items of Exp.	Family A			Family B		
	Rs.	%	Cum. %	Rs.	%	Cum. %
Food	400	40	40	500	31.3	31.3
Cloth	200	20	60	400	25	56.3
House Rent	160	16	76	200	12.5	68.8
Fuel	80	8	84	100	6.20	75
Misc.	160	16	100	400	25.0	100
Total	1000	100		1600	100	

(ii) Divided Rectangular Diagram:

These diagrams are used to represent the different facts which are related to each other, For example, per unit price of a commodity, quantity of sale and sales receipts are shown with the help of this type of diagram. In these diagrams, the width is taken in the ratio of per unit price and height is taken in the ratio of quantity of sale. For getting the sales value the height has to be multiplied with the width which gives the area of the rectangle. Rest of the procedure is same as followed in the percentage sub-divided rectangular diagram.



(D) Circular or Pie Diagram :

The circular diagram is constructed in the same way as the square diagram. Circular diagrams are also the two dimensional diagrams. Under the construction of circular diagrams, first of all the square roots of the given values are calculated and then radius of circles are obtained in the ratio of square roots. These radius are used for the construction of circles. The circles are constructed on the same base with equal distance. Circular diagrams are sub-divided in sub-parts so that comparison can be made. For the construction of circles the total sum of items is assumed to be equal to 360^o and then angles are calculated for each item. Since the angle in the centre of the circle is of 360^o, so circular diagrams are also known as the angular diagrams.

Example : Present the following data into the angular diagram:

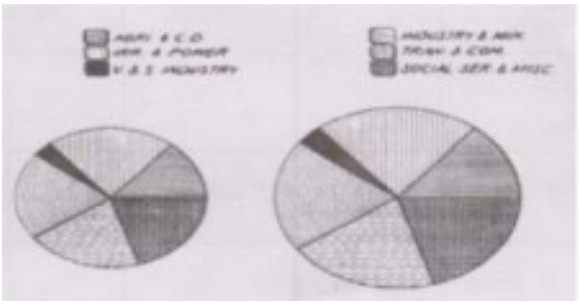
Population of India According to Work-status (2001)

Work-Status	Population (Crore)
Marginal worker	9
Main worker	31
Non-worker	62
Total	102

Solution : For all the given values above the angles are calculated as shown in the following table:

Work-Status	Population	Percentage	Angle (Degree)
Marginal Worker	9	8.8	32^o
Main Worker	31	30.4	109^o
Non-Worker	62	60.8	219^o
Total	**102**	**100.00**	**360^o**

With the help of the above data angular diagrams are constructed as below:



Example : Construct the blank table from the following information:

1. Distribution of population according to age.
2. Classification of population according to age and sex.
3. Classification of population according to age, sex and literary.
4. Classification of population according to age, sex, literacy and distribution in states.

1. Distribution of Population according to Age :

In this case a simple blank table is constructed as below:

Age Group (Years)	No. of Person (Million)
0-20	-
20-40	-
40-50	-
>50	-
Total	**-**

2. Classification of Population according to Age and Sex:

When the given data have two characteristics we construct the two-way table. The classification of population according to age and sex is shown in the following blank table.

Age Group (years)	No. of Person (Million)		
	Male	Female	Total
0-20	_	_	_
20-40	_	_	_
40-50	_	_	_
> 50	_	_	_
Total	_	_	_

3. Classification of Population according to Age, Sex and Literacy:

In the three-way table three characteristics are shown simultaneously. Following blank table shows the classification of population according to age, sex and literacy.

Age Group (Yrs)	No. of Person (Million)						
	Male			Female			Total
	Lit	Illit	Tol	Lit	Illit	Total	
0-20	_	_	_	_	_	_	_
20-40	_	_	_	_	_	_	_
40-50	_	_	_	_	_	_	_
>50	_	_	_	_	_	_	_
Total	_	_	_	_	_	_	_

4. Classification of Population according to Age, Sex, Literacy and Distribution in States:

The Following blank table is a multiple table.

State	Age Group (Yrs)	No. of Person (Million)									
		Male			Female			Total			Total
		Li	Ill	Tot	Li	Ill	Tot	Li	Ill	Tot	
Bihar	0-20	_	_	_	_	_	_	_	_	_	_
	20-40	_	_	_	_	_	_	_	_	_	_
	40-50	_	_	_	_	_	_	_	_	_	_
	>50	_	_	_	_	_	_	_	_	_	_
	Total	_	_	_	_	_	_	_	_	_	_
Rajasthan	0-20	_	_	_	_	_	_	_	_	_	
	20-40	_	_	_	_	_	_	_	_	_	
	40-50	_	_	_	_	_	_	_	_	_	
	>50	_	_	_	_	_	_	_	_	_	
	Total	_	_	_	_	_	_	_	_	_	

Graphs of Frequency Distribution:

Generally data given in grouped frequency distributions can be presented by frequency graphs. Frequency graphs are of the following types.

(a) Histogram

(b) Frequency polygon

(c) Frequency curve

(d) Cumulative frequency curve or ogives.

In constructive the above graphs, the size and classes are measured horizontally and frequencies are measured vertically.

(a) Histogram :

Histograms are used to present data of continuous series. In histograms, rectangles are constructed in the height of the frequencies of the class-intervals. If the class-intervals are inclusive, these are converted into the exclusive class-intervals. With the help of histogram, mode can be determined.

Determination of Mode by Graph :

After construction the histogram, the highest rectangle is considered for the determination of mode and the related class-interval is known as the model class. The upper right corner of this bar is joined with the upper right corner of the preceding bar. In the same way the upper left corner of the highest bar is joined with the upper left corner of the succeeding bar.

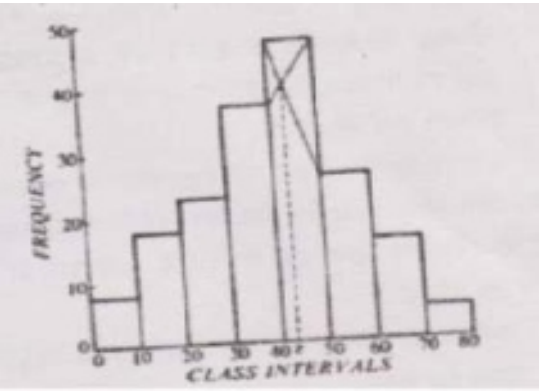
From the intersection point of these two lines a perpendicular is drawn on the horizontal axis and thus the value of mode is determined.

Example: Show the below given data by the histogram and determine the mode.

Class	Frequency
0-10	8
10-20	18
20-30	23
30-40	37
40-50	47
50-60	26
60-70	16
70-80	5

Solution : The histogram has been shown as below:

Determination of Mode by Histogram :



According to the inspection method, (40-50) is the model class in which the value of mode lies. Thus,

$$Z = L_1 + \frac{F_1 - F_0}{2F_1 - F_0 - F_2} \text{ xi}$$

$$= 40 + \frac{47-37}{94-37-26} \text{ x10}$$

$$= 40 + 3.2 = 43.2$$

(b) Frequency Polygon :

Frequency polygon is constructed from the values of mid-points and their frequencies. In this graph the values in discrete or continuous series are measured on the X-axis and the frequencies are measured on the Y-Axis. All the mid-points of all the bars in the histogram are joined and the figure obtained is known as frequency polygon.

(c) Frequency Curve:

The frequency curve is a smoothed figure/curve of the frequency polygon. This curve is constructed as a free hand curve by passing through the nearby mid points of the frequency polygon. It is not necessary that all the points of the frequency curve pass through the all points of the frequency polygon, but it passes through the nearest points on the frequency polygon.

(d) Cumulative Frequency Curve:

Cumulative frequency curve is constructed by measuring the upper limit of class-intervals on the X-axis and cumulative frequencies on the Y-axis. With the help of the cumulative frequency curves the values of quartiles, deciles, octiles and percentiles are determined. There are two types of cumulative frequency curves:

(i) By cumulative frequencies and upper limit of the class-intervals. This curve is known as 'less than' cumulative frequency curve and has the negative stop.

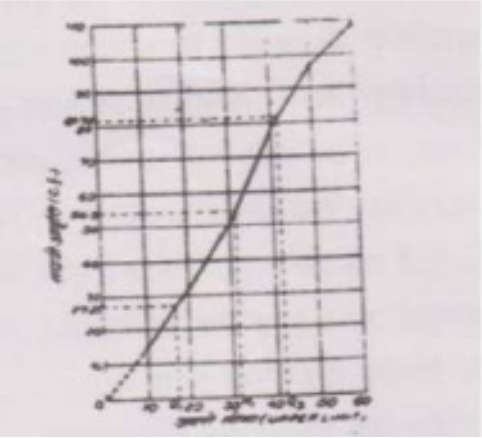
(ii) By cumulative frequency and the lower limit of the class intervals. This curve is known as 'more than' cumulative frequency curve and has the positive slope.

The point at which the two cumulative frequency curve intersect with each other determines the value of median.

Example : Construct a cumulative frequency curve and determine the median from the data given below:

Age(Less than)	10	20	30	40	50	60
No. of Person	15	32	51	78	97	109

Solution : Above data are used to construct the cumulative frequency curve as shown below:



Data in above table are given in the form of the cumulative frequency distribution. So,

Median (M) = Size of($\frac{N}{2}$)th item

= Size of($\frac{109}{2}$)th item

= Size of 54.5th item

By applying the formula, the calculated value of M is 31.3.

It is clear that the graphs are very important in economics. The positional averages can easily be determined with the help of graphs. With the help of diagrams and graphs the message can easily reach to the common people.

Important Points:

- There is need of proper presentation of data for comparative study, analysis and interpretation.
- Data can be presented in there forms:
  (a) Textual Presentation
  (b) Tabulation, and
  (c) Diagrammatic and Graphical Presentation.
- Presenting data into tables for making them simple and brief is known as tabulation. Tabulation is important from the point of view of simplicity, comparative study, presentation, saving of time and place, and statistical analysis.
- Classification is a method of statistical analysis while tabulation is a process of presentation of data.
- Tables are classified according to objective, structure and origin.
- Diagrammatic and graphical presentation of data can be understood simply and easily.
- Data can be represented by two methods:
  (a) Diagrammatic representation
  (b) Graphical representation
- Under graphical representation bar diagrams, rectangular diagrams, circular diagrams, etc. are constructed.
- Data of the grouped frequency distributions can be presented in the form of histogram, frequency polygon, frequency curve and cumulative frequency curves.

Questions for Exercise:

Objective Type Questions :

1. Diagrams used to compare two or more than two related data groups on the basis of quality, time and place are :
   (a) Simple bar diagram
   (b) Multi-bar diagram
   (c) Sub-divided bar diagram
   (d) Rectangular diagram ( )
2. Information required to construct circle is :
   (a) Square (b) Side
   (c) Raddi (d) Circle ( )
3. What is circle ?
   (a) A raddi
   (b) Two dimension

(c) Three dimension

(d) All of the above ( )

4. What will be the angle in the circular diagram to show the ratio of 40% woman literacy in India ?

(a) 60^0 (b) 72^0

(c) 144^o (d) 40^o ( )

5. In which series the mode is determined with the help of histogram ?

(a) Individual Series

(b) Continuous Series

(c) Discrete Series

(d) Exclusive Series ( )

6. Which is not a two dimensional diagram ?

(a) Bar Diagram

(b) Square Diagram

(c) Rectangular Diagram

(d) Circular Diagram ( )

7. The graphical presentation is performed on :

(a) Simple Paper

(b) Graph Paper

(c) Drawing Sheet

(d) Any Paper ( )

Very Short Answer Questions:

1. What is meaning of line diagram ?
2. Which diagram can be used for getting the mode?
3. Explain the tabulation.
4. Write any three bases of data representation.
5. Write any four points about utility of diagrams.
6. Write the name of two dimensional diagrams.
7. Construct a histogram.

Short Answer Questions:

1. State the difference between tabulation and classification.
2. What is difference between diagrams and graphics ?
3. What are the important points to keep noted at the time of graphical representation ?
4. Mention any four points showing the importance of diagrams.
5. What are the bases of classification of tables?

Essay Type Questions:

1. Explain the meaning of tabulation. What are the different parts of a table? What are the points to be noted while doing tabulation ?
2. Construct a blank table for the distribution of population according to education, employment and sex in a city.
3. Explain in brief the different diagrams generally used to present the statistical facts.

Answer to objective questions:

(1) b (2) c (3) b (4) c (5) b (6) a (7) b

Reference Books :

1. S.C. Gupta & V.K. Kapoor : Fundamentals of Mathematical Statistics, Published by Sultan Chand and Sons.
2. S.P. Singh : Sankhyiki : Sidhant avm Vyavahar, S. Chand.
3. Kailash Nath Nagar: Sankhyiki Ke Mool Tatva, Minakshi Prakashan.