

BUSINESS STATISTICS

Unit-I

INTRODUCTION TO STATISTICS

1. What is singular & plural sense of statistics?

Ans:- The singular sense of statistics is defined as collection, presentation, analysis and interpretation of data. The plural sense of statistics is: It is defined as aggregates of facts, affected to marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other.

2. What are the characteristics of statistics?

Ans:- The characteristics of statistics are:

- a) **Statistics are aggregates of facts**: Single and isolated figures are not statistics, a set of data is termed as statistics for which analysis can be done. For example, 60kgs. Weight of a student is not a statistics, aggregate of certain facts i.e., 50, 65, 70kgs. Weight of students in a class or profits of a firm over different times etc., are liable to statistics.
- b) **Statistics are numerically expressed**: All statistics are numerical statement of facts expressed in numbers so that counting or measurement of data can be made possible. It means that the data or the fact to constitute statistics must be capable of being expressed in some quantitative form as weights of persons are 60, 70, 100 and 90kgs or profits of company Rs.10,000, Rs.20,000 etc., Thus these data must contain numerical figures so that those may be called as numerical statement of facts.
- c) **Statistics are affected to marked extent by multiplicity of causes**: Generally speaking facts and figures are affected to a considerable extent by a number of forces operating together. Ways and means are also being devised in statistics for segregating the effect of various forces on an event.
- d) **Statistics are enumerated or estimated according to reasonable standards of accuracy**: Facts and figures about any phenomenon can be derived in two ways i.e., by actual counting and measurement or by estimate. Estimates cannot be as precise and accurate as actual counts or measurements.
- e) **Statistics are collected in a systematic manner for a predetermined purpose**: Another characteristic of statistics is that the data should be collected in a systematic manner. The data collected in a haphazard manner will lead to difficulties in the process of analysis, and wrong conclusions. A proper plan should be made and trained investigators should be used to collect data so that they may collect statistics. If it is not done, in such cases reliability of data gets decreased. So to get correct results the data must be collected in a precise manner.
- f) **Statistics are placed in relation to each other**: If numerical facts are to be called statistics, they should be comparable. Statistical data are often compared period-wise or region-wise. Valid comparisons can be made only if the data are homogeneous, i.e., relate to the same phenomenon or subject and only likes are compared with likes.
- g) **Statistics are collected for a predetermined purpose**: The purpose of collecting data must be decided in advance. The purpose should be specific and well-defined. A general statement of purpose is not enough. For example, if the objective is to collect data on prices, it would not serve any useful purpose unless one

knows whether he wants to collect data on wholesale or retail prices and what are the relevant commodities in view.

3) What are the functions of statistics?

Ans:- The important functions of statistics are:-

- 1) Systematic collection and presentation of facts:- Statistics ensures that data is collected in systematic manner using different methods. The collected data is presented in a systematic manner in the form of tables and graphs.
- 2) Simplification of mass figures:- Statistics not only present facts in a definite form but also helps in simplifying mass data into a few significant figures. This helps in overall analysis and interpretation of data.
- 3) Facilitates comparison:- Statistics enables impartial and fair comparison of data across a wide range of competing alternatives. For example, inflation rate compares the current prices with prices of previous year and indicates the trend of raising prices.
- 4) Helps in formulation and testing of hypothesis:- Statistical methods are extremely helpful in formulating and testing of hypothesis and develop new theories. It is used in determining the result of medicine on a patient, the impact of newly introduced government policies on general public, etc.,
- 5) Helps in prediction:- Statistical methods provide helpful means to forecast future events. If an investor wishes to purchase shares, he can use statistics to find out whether share price is likely to increase or decrease. The businessman use statistics to conduct market research before entering into a particular business.

4) What are the limitations of statistics?

Ans:- Although statistics is very widely used in all spheres of human activity, it has its own limitations.

- 1) Statistics does not study individual measurements:- Since Statistics deals with aggregates of facts, the study of individual measurements lies outside the scope of statistics. The marks obtained by one student of a class or his height are not statistics, their averages i.e., average marks of students of a class or average height of students in a class is called statistics.
- 2) Statistics does not study qualitative phenomenon:- Statistics are numerical statements of facts. They cannot be used directly for the study of qualitative characteristics like honesty, efficiency, illiterates, etc., but these qualitative characteristics can be measured in other ways. For example, the intelligence of boys on the basis of the marks obtained by them in an examination.
- 3) Statistical laws are true only on averages:- The laws of statistics are not universally true like laws of physics, but they are true only on average. These averages can be affected by a vital role in every fields of human activity. Statistics has important role in determining the existing position of per capita income, unemployment, population growth rate, housing , medical facilities etc., in a country. It is difficult to study the effects of each of these factors separately.
- 4) Statistics can be misused:- The greatest limitation of statistics is that it is liable to be misused. If statistical conclusions are based on incomplete information, one may arise at fallacious conclusions.

5) Explain the importance of statistics?

Ans:- Statistics is viewed not only as a mere device for collecting numerical data but also as a means of developing sound techniques for their analysis and drawing valid inferences. It plays vital role indifferent

fields. Now Statistics holds a central position in almost every field like industry, commerce, trade, Physics, Chemistry, Economics, Mathematics, Biology, Botany, Psychology, Astronomy, etc., so application of statistics is very wide. Now we discuss some important fields in which Statistics is commonly applied.

a) **Statistics in state**:- Statistics is essential for a country. Different policies of the government are based on Statistics. Statistical data are now widely used in taking all administrative decisions. Suppose if the government wants to revise the pay scales of employees in view of an increase in the living Index, Statistical methods will be used to determine the rise in the cost of living. Preparation of federal and provincial government budgets mainly depend upon Statistics because it helps in estimating the expected expenditures and revenue from different sources. So Statistics are the eyes of administration of the state.

b) **Statistics in business**:- A successful businessman must be very quick and accurate in decision making. He knows that what his customers wants, he should therefore, know what to produce and sell and in what quantities.

Statistics help businessman to plan production according to the taste of the customers, the quality of the products can also be checked more efficiently by using statistical methods. So all the activities of the businessman based on statistical information. He can make correct decision about the location of business, marketing of the products, financial resources etc... Following are the typical applications areas:

- ***Quality Management***: Statistical Quality Control (SQC) helps in separating the assignable causes from the chance causes. SQC techniques also help in reducing and stabilising process variation that may reduce cost significantly.
- ***Process Capability***: Statistics plays a crucial role to find out whether the enterprise is capable of meeting customer specification or not.
- ***Finance*** : Comparative statistical analysis reveals the performance of the company in the past years and emerging trends in the future.
- ***Cash Forecasting***: Financial analyst has to forecast cash requirements to meet short term obligations capital, payment to suppliers etc., Statistical analysis helps in working out realistic cash forecasting.
- ***Materials Management Inventory level***: By using appropriate statistical methods one can maintain optimum inventory levels of raw materials and finished products and thus savings are made which impact on profit.
- Similarly, Statistics is used in many other ways in business: Quality assessment for incoming and outgoing items, Marketing, Demand projections.

c) **Statistics and Mathematics**:- Statistics play a central role in almost all natural and social sciences. The methods of natural sciences are most reliable but conclusions drawn from them are only probable, because they are based on incomplete evidence.

Statistics helps in describing these measurements more precisely. Statistics is branch of applied mathematics. The large number of Statistical methods like probability averages, dispersion, estimation etc., used in mathematics and different techniques of pure mathematics like integration, differentiation and algebra are used in Statistics.

d) **Statistics and Economics**:- In Economics, all problems are concerned with statistics. The modern states have followed the 'socialistic pattern of society'. It means equal distribution of income and wealth. This is possible if the data pertaining to income and wealth is collected. This will help us in assessing whether inequalities are on the increase or decline. Moreover, Statistics is the most significant tool for the purpose of research in

Economics. The relationship between supply and demand is studied by Statistical methods, the imports and exports, the inflation rate, the per capita income are the problems which require good knowledge of statistics.

e) **Statistics and Banking**:- Statistics plays an important role in banking. The banks make use of Statistics for a number of purposes. The banks work on the principle that all the people who deposit their money with the banks do not withdraw it at the same time. The bank earns profits out of these deposits by lending to others on interest. The bankers use statistical approaches based on probability to estimate the numbers of depositors and their claims for a certain day.

f) **Statistics and Accounting and Auditing**: Accounting is impossible without exactness. But for decision making purpose, so much precision is not essential the decision may be taken on the basis of approximation, known as Statistics. The correction of the values of current assets is made on the basis of the purchasing power of money or the current value of it.

In auditing sampling techniques are commonly used. An auditor determines the sample size of the book to be audited on the basis of error.

f) **Statistics and Physical and Natural Sciences**:- In Physical Sciences like Astronomy, Geology and Physics, the statistical methods are used. But in the recent times, the importance of statistics has further increased due to added importance of natural sciences like Medicine, Meteorology, Zoology, Botany, etc., Prof. Karl Pearson used statistical methods in Biology. He showed that doctrine of evolution and heredity is based on statistical knowledge.

6) What are different methods of collecting primary data?

Ans:- The systematic, planned and meaningful way of collecting information is known as collection of data. The investigation of the data may be collected either through primary data or through secondary data.

Primary data refers to the data collected for the first time and is original in character. It is collected by enumerator for his investigation and statistical methods are applied on such data for the purpose of investigation. The following are the methods of collecting primary data:

1) **Direct personal Interviews**:- Under this method, the information will be collected by the investigator through personal interview from the informants. The reliability of collected data depends upon the training and attitude of the investigator and supporting attitude of the respondent. This method is most suitable for a type of investigation where the investigation is confidential and the process of investigation is so complex that it requires personal attention of the investigator.

The best example of collecting data using this method is Population census.

Merits:

- i) It is possible to collect original, accurate and exact data.
- ii) The doubts of informants can be checked and clarified.
- iii) Informants' doubts can be clarified in the language most suitable to him.

Demerits:

- i) Skilled enumerators are required to collect data by this method.
- ii) This method consumes resources like time and money.

2) **Indirect Oral Investigation**:- This method is used when the informants are unwilling to provide the required information directly. It is generally adopted by police and C.B.I for the collection of information regarding crimes. In the investigation of crime, they collect data from third party or witness or head of an institution, who is supposed to be in touch with the person under investigation.

Merits:

- i) If the area of investigation is very large, then this method of data collection is most suitable or when the respondents are reluctant to give the information directly.
- ii) If a person is not interested to reveal his habits of drinking, smoking, gambling, etc., by applying this method, the information can be collected from third party.

Demerits:

- i) In the absence of direct contact between investigator and informant, it may happen that many important points remain unnoticed.
- ii) The information may be biased as it is provided by the third party.
- iii) The information collected from different persons may not be same and comparable.

3) **Information through local agencies or correspondents**:- In this method, local agents or correspondents are appointed in different parts of the area under investigation. These agents send the required information at regular intervals of time. This method is often adopted by newspapers, news channels, etc.,

Merits:

- i) It is economical in terms of time, money and labor.
- ii) When periodic information is required at regular intervals and area of investigation is large, this method is very useful.

Demerits:

- i) The information lacks originality.
- ii) The bias of the correspondent affects the information.

4) **Mailed questionnaire Method**:- This method is the most popular method for collecting primary data. A questionnaire is a list of questions pertaining to the enquiry. In this method, a set of questions is sent through post or by mail to the informants. They are expected to answer the questions in the space meant for the purpose on the questionnaire itself and send them back to the investigator. It is very useful when the informants are educated and when the area of investigation is very wide. This method is extensively employed in various economic and business surveys.

Merits:

- i) This method is very economical particularly when the universe is large and spread geographically over a vast area.
- ii) Collected information is free from the bias of the enumerators.
- iii) Respondents can take their own time to answer the questions. So they give well thought out answers.
- iv) Respondents that are at remote places and are not easily approachable can also be reached conveniently.
- v) Large samples can be covered and thus the results can be more dependable and reliable.

Demerits:

- i) Questionnaires can be circulated only among the respondents who are educated and cooperative.
- ii) It is capable only to educated informants.
- iii) Some of the informants may send incomplete questionnaires.
- iv) There is possibility of ambiguous replies or omission of replies to certain questions. Interpretation of omissions is difficult.

- v) This method is likely to be slowest of all, because the respondents take their own time to return the filled in questionnaires.

Before sending them to the respondents, it is advisable to conduct a 'Pilot Survey' for pre-testing it. Pilot survey is in fact the replica and rehearsal of the main survey. From the experience gained in this sort of survey, changes can be made in the questionnaire for the final collection of data. The pre-testing is necessary particularly in case of a big enquiry.

5) **Schedule**:- This method of data collection is similar to that of the questionnaire. The schedule is also a proforma containing a set of questions. The difference between the questionnaire and the schedule is that the schedule is being filled in by the enumerators who are specially appointed for the purpose. The enumerators go to the informants along with the questionnaire, get replies to the questions contained in the schedules. Data collection through schedules requires enumerators for filling up schedules and as such they should be very carefully selected. They should be trained to perform their job well. In drafting the schedule, all points stated for a good questionnaire, must be observed.

Merits:

- i) It can be adopted in those cases where informants are illiterate.
- ii) The problem of non-response is avoided as the enumerators go personally to obtain the information.

Demerits:

- i) This method is very expensive as enumerators are generally paid persons.
- ii) If the investigator is not good in interviewing, most of the information collected by him may be unreliable.

7) What are the features of a good questionnaire?

Ans:- In order to make questionnaire more effective, it must be very carefully drafted. The following are the qualities of a good questionnaire:

- 1) It should be short and simple.
- 2) Questions should proceed in logical sequence starting with easy questions and then moving on to more difficult ones. Personal questions should generally be avoided or may be left to the end.
- 3) Questions may be dichotomous(yes or no type), or multiple choice. Open ended questions are difficult to analyse and should be avoided to the extent possible.
- 4) In order to ensure the reliability of respondent there should be some control questions. They introduce a cross-check to see whether the information collected is correct or not.
- 5) Adequate space for answers should be provided in the questionnaire itself. There should always be provision for indications of uncertainty.
- 6) Layout and design of the questionnaire should also be attractive so that it may attract the attention of the respondents.

8) Mention the differences between questionnaire and schedule?

Ans:- Even though schedule is like a questionnaire, we can note the following differences:

Questionnaire	Schedule
1. Questionnaire is mailed to respondents.	1. Schedules are taken to respondents by enumerators.
2. Questionnaires can be used only with educated respondents.	2. Schedules can be used with uneducated respondents.

3. Responses are filled in by respondents.	3. Responses are filled in by enumerators.
4. Questionnaires do not have adequate flexibility.	4. Enumerators can use tact and intelligence to elicit the required information.
5. Suitable if large areas are to be simultaneously covered.	5. Best suited for survey of small, focused areas
6. The percentage of people responding is low.	6. The percentage of people responding is high.
7. There is minimal influence on the respondent	7. Enumerators can influence the respondent
8. Confidentiality can be ensured.	8. Confidentiality is not promised.
9. It is a cost effective method	9. It is relatively expensive
10. The most important factor is the design of the questionnaire.	10. The most important factor is the personality of the enumerator.

9) What are the differences between Primary and Secondary data?

Ans:- The difference between primary and secondary data is one of degrees only. Data, which is primary in the hands of one, becomes secondary in the hands of another.

Primary Data	Secondary Data
1. Primary data is collected by the person conducting the statistical enquiry.	1. Secondary data is data collected by person agencies other than the one conducting the enquiry.
2. It is raw and original.	2. It is processed and hence, not original.
3. It is most relevant to the study being conducted	3. It may or may not be directly relevant to the study.
4. It involves huge costs, time and effort of the investigation.	4. It is relatively less expensive and takes time and effort
5. There is no need for extra precautions in using the data	5. It should be used with great care.

10) Define classification, objectives of classification and types of classification?

Ans:- Classification means arranging the mass of data into different classes or groups on the basis of their similarities and resemblances.

Objectives of classification:-

- 1) It helps in presenting the mass of data in a concise and simple form.
- 2) It divides the mass of data on the basis of similarities and resemblances so as to enable comparison.
- 3) It is a person of presenting raw data in a systematic manner enabling us to draw meaningful conclusion.
- 4) It provides a basis for tabulation and analysis of data.
- 5) It provides us a meaningful pattern in the data and enables us to identify the possible characteristics in the data.

Types of Classification:-

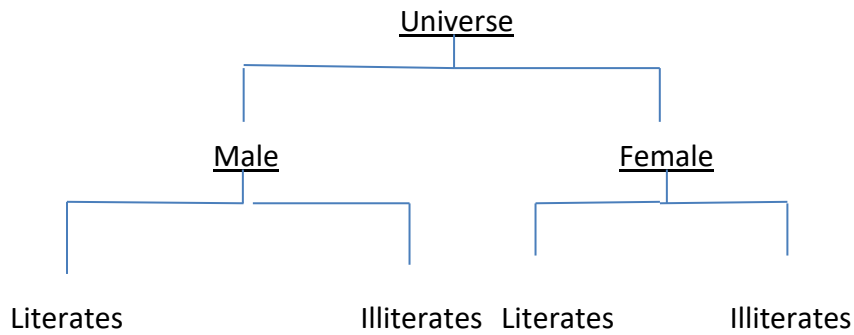
Classification is broadly divided into the following four types.

1) Quantitative classification: Classification based on the quantitative measurement of the data is called quantitative classification. It is a classification based on characteristics which are capable of quantitative measurement. For example, a frequency table is prepared based on quantitative classification.

Marks	0-20	20-40	40-60	60-80	80-100
No. of students	5	15	25	10	5

2) Qualitative Classification: When data is classified on the basis of attributes or qualities which are descriptive in nature, It is called qualitative classification. For example Universe is divided into male and female.

Example:



3) Geographical classification: In geographical classification data is classified on the basis of place. For example, The production of sugarcane can be classified geographically as:

State	U.P	Bihar	Tamilnadu	Maharashtra	A.P
Production of sugarcane	357	1127	7825	1208	18967

4) Chronological classification: When the data is classified on the basis of time then it is called chronological classification. Time Series data is considered as chronological classification. For example, The population of India in different census years are:

Year	1921	1931	1941	1951	1961	1971	1981	1991
Population(in millions)	251	278	318	361	439	548	685	872

11) Define tabulation? Mention the parts of a table? What are the objectives of tabulation? What are types of table?

Ans:- Arrangement of data in the form of rows and columns is called table and a systematic arrangement of data in a form suitable for analysis and interpretation is called tabulation.

Parts of the table:-

- i) **Table Number**: It indicates the serial number of the table. Table number helps an easy identification and reference of the table in future.
- ii) **Table Title**: Table title indicates the subject matter of the table. It should clearly indicate the nature of the data contained in the table.
- iii) **Captions**: It refers to the heading of the columns. A caption should be brief, concise and self explanatory.
- iv) **Stub**: Refers to the heading of the rows.
- v) **Body of the table**: Refers to the numerical information entered into the table.
- vi) **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.
- vii) **Foot Note**: Footnote provide further explanation with regard to the content of the table.
- viii) **Source Note**: It shows the source from which data is collected. It is stated below the foot note.

General form of the table:-

Table No:

Table Title:

Head Note:

STUB	CAPTIONS			TOTAL
	Cap-1	Cap-2	Cap-3	
St-1	BODY OF THE TABLE			
St-2				
St-3				
TOTAL				

Foot Note:

Source Note:

Objectives of a table:-

Tabular presentation serves the following objectives:

- 1) **Systematic Presentation of the data:** Tabulation helps in presenting the data in an orderly manner. The mass of the data is presented in a concise and simple form by means of statistical tables.
- 2) **Facilitates Comparison of data:** The presentation of complete and unorganized data in the form of tables facilitates the comparison of the various aspects of the data.
- 3) **Identification of the desired values:** The desired values can be identified without much difficulty by arranging it in rows and columns.
- 4) **Provides a basis for analysis:** Presentation of data in tabular form provides a basis for analysis of such data. These tables are useful as aids in analysis.
- 5) **Exhibits trend of data:** By presenting data in a condensed form at one place, exhibits the trend of data.

Types of Tabulation:-

- 1) **Simple table or One-way table:** The table based on single variable or characteristics is known as simple or one-way table.

Year	1991	1992	1993	1994	1995
Population(in crores)	7	10	14	18	20

- 2) **Complex Tables:-**

- i) **Two-Way Table:** A table based on two characteristics is called two-way table.

Year	Imports	Exports
2001	585	345
2002	427	543
2003	387	567
2004	231	768
2005	567	566

- ii) **Three-Way Table:** A table based on three characteristics is known as three-way table.

Year	No.of Students				Total
	Male		Female		
	Unmarried	Married	Unmarried	Married	
1982-83	950	50	475	25	1500
1983-84	975	55	490	30	1550
1984-85	1000	55	510	35	1600
1985-86	1035	60	510	30	1635
1986-87	1010	50	520	35	1615
1987-88	1080	50	515	40	1685
Total	6050	320	3020	195	9585

iii) **Main-Fold Tables**: A table consisting of more than three characteristics is called main-fold tables.

iv) **General & Special purpose tables**: The table consisting of wide range of information relating to a specified subject is called general tables. The table showing specific point relating to data and are helpful in statistical analysis is called special purpose tables.

12) Explain distrust of statistics?

Ans:- Distrust of Statistics means lack of confidence in statistical methods and statements. “An ounce of truth will produce tons of statistics”. It is generally stated by people that “There are three types of lies-lies, damned lies and statistics. It can prove anything, so some people do not have confidence in data, statistical statements and statistical methods.

Statistics as a whole require proper usage and the results obtained through statistical analysis require intelligent interpretation”. The following points must be kept in view if the arguments on statistics should be true: (i) Source of information, (ii) Reliability of information, (iii) Their true sense and significance, (iv) Utility of data in relation to problem in hand, (v) Amount of data in relation to problem in hand, (vi) Manner of summarization. The science of statistics is a most useful servant, but only of great value to those who understand its proper use.

In the end, we say that if the data is collected in a proper way and it is uniform, complete, stable, representative and free from personal bias, then it will give the correct picture of the phenomenon.

13) What are the census and sampling methods?

Ans:- Before a study is made, the investigator has to decide the method of data collection. Two methods are generally used to collect the data which are:

(i) **Census Method or Census Enquiry**: When the whole area or population or persons is contacted, the method is known as census method. It is very rarely used, specially in case of social phenomena. In India, census of population is conducted after every 10th year or a decade.

(ii) **Sample Method or Sample Enquiry**: Sample means a smaller representative of huge population which is selected for study. Various methods of sampling are: Random Sampling and Non- Random Sampling.

(a) **Random Sampling**:- It is also called as Probability sampling. In this sampling every item in the population has equal chance of being selected in the sample. The sampling operation is controlled so objectively that the items will be chosen strictly at random.

(b) Non- Random Sampling:- It is also called non-probability sampling. In this method every item in the population do not have equal chance of being selected in the sample. The selection process is, at least, partially subjective.

14) Define Statistical Investigation?

Ans:- An investigation or enquiry or survey means search for knowledge with the help of statistical methods. It is a technique which require special knowledge and skill. It is comprehensive process which lays the structure for decision making. It provides answers to various management problems that the decision makers face. It gives a set of recommendations that help business organizations develop and implement most viable strategies.

Statistical enquires have always required considerable skill on the part of the statistician, rooted in a broad knowledge of the subject-matter area and combined with considerable ingenuity in overcoming practical difficulties.

15) What is statistical errors? Explain.

Ans:- To appreciate the need for sample surveys, it is necessary to understand clearly the role of sampling and non-sampling errors in complete enumeration and sample surveys. The error arising due to drawing inferences about the population on the basis of few observations is termed sampling error. Clearly, the sampling error in this sense is non-existent in complete enumeration survey, since the whole population is surveyed. However, the error mainly arising at the stage of ascertainment and processing of data, which are termed non-sampling errors, are common both in complete enumeration and sample surveys.

Sampling Errors:

The results derived from a sample study may not be exactly equal to the true value in the population. The reason is that estimate is based on a part and not on the whole. Hence sampling gives rise to certain errors known as sampling errors.

These errors would not be present in a complete enumeration survey. However, the errors can be controlled. The modern sampling theory helps- in designing the survey in such a manner that the sampling errors can be made small. The different types of sampling errors are:

I) **Biased Errors**: These error arise from any bias in selection, estimation, etc. For example, if in place of simple random sampling, deliberate sampling has been used in a particular case some bias is introduced in the result and hence such errors are called biased sampling errors.

II) **Unbiased Errors**: These errors arise due to chance differences between the members of population included in the sample and those not included. An error in statistics is the difference between the value of a statistic and that of the corresponding parameter.

UNIT-II

DIAGRAMS AND GRAPHS

1) Explain about diagrams and graphs?

Ans) One of the most convincing and appealing ways in which statistical results may be presented is through diagrams and graphs. A picture is worth a thousand words. The impression created by a picture has much greater impact than any amount of detailed explanation. They make complex data simple and easily understandable. Evidence of this can be found in newspapers, magazines, journals, advertisements, etc., There are numerous ways in which statistical data may be displayed pictorially such as different types of diagrams, graphs and maps.

2) Explain about significance of diagrams and graphs?

Ans) Significance of diagrams and graphs are:

1. Diagrams give a very clear picture of data. Even a layman can understand it very easily and in a short time.
2. We can make comparison between different samples very easily.
3. This technique is used almost in all the subjects and various other fields.
4. Tabulated data has not much impression as compared to Diagrams.
5. This technique can be used for numerical type of statistical analysis.
6. It does not save only time and energy but also is economical. Not much money is needed to prepare even good diagrams.
7. Diagrams which we see leave their lasting impression much more than other data techniques.
8. Data can be condensed with diagrams.

3) Explain about different types of diagrams?

Ans:- A large variety of diagrammatic devices are used in practice to present statistical data. They are:

- One-dimensional diagrams
- Two-dimensional diagrams
- Three-dimensional diagrams
- Pictograms
- Cartograms

One-dimensional diagrams:- The construction of a diagram on the basis of one dimension is called one-dimensional diagrams. The dimension could be height in case of vertical diagrams and length in case of horizontal diagrams. These diagrams can be line or bar diagrams.

Line Diagrams: This is the simplest of all the diagrams. It consists in drawing vertical lines, each line being equal to frequency. Line diagram facilitate comparisons though they are not attractive or appealing to the eye.

Bar Diagrams: These diagrams are one of the easiest and most commonly used devices for presenting statistical data. They consist a group of equidistant rectangles, one for each group or category of data in which values are represented by length or height of the rectangle and the width of the rectangle being immaterial. The following points may be borne in mind to draw bar diagram.

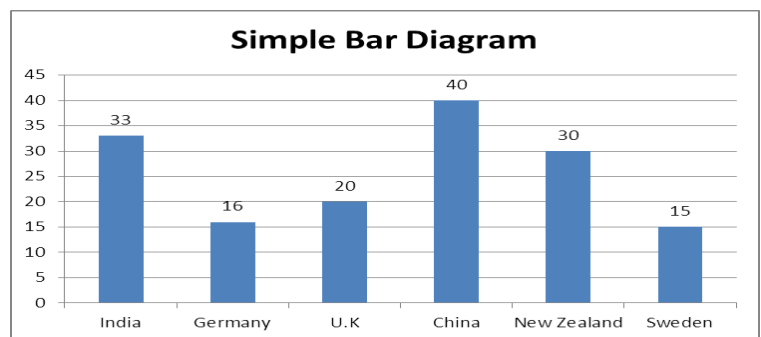
- All the bars drawn in a single study should be of uniform width.
- Uniform spacing should be given between different bars to make the diagrams look more attractive and elegant.
- The height of the rectangles or bars are taken proportional to magnitude of observations.
- All the bars should be constructed on the same base line.
- Bars can be drawn vertically or horizontally, but vertical bars are generally used because they will be attractive.

Types of bar diagrams: The following are the various types of bar diagrams.

A. Simple Bar diagram

Simple Bar diagram is the simplest of the bar diagrams and is used frequently for the comparative study of two or more items or values of a single variable. It consists of rectangular bars in which the magnitude can be represented length wise or height wise according to the construction of diagram i.e., either horizontally or vertically. For example, the figures of sales, production, population, etc., for various years, places, etc.,
Example:- Following table gives the birth rate per thousand of different countries over a certain period:

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

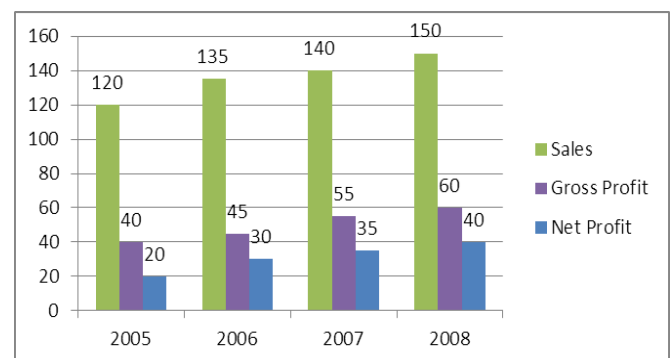


B. Multiple Bar Diagram

It is used to represent two or more sets of interrelated data. The technique of representing the magnitude is same as that of simple Bar diagram. The only difference is that the interrelated data of each set is represented with group of rectangles adjoining each other.

Example:- The Sales, Gross Profit, Net Profit of a company in different years is represented as:

Year	Sales	Gross Profit	Net Profit
2005	120	40	20
2006	135	45	30
2007	140	55	35
2008	150	60	40

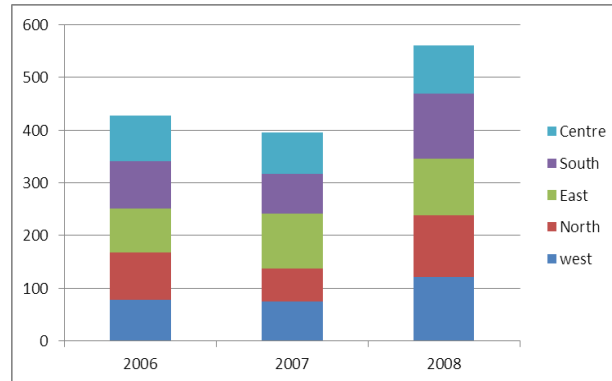


C. Sub-Divided or Component Bar Diagram

This diagram is also used to represent two or more sets of data. Each set of data is represented by drawing a bar with a height of total magnitude of group of related data. Then the bar is divided into different components, each component occupies a part of the bar proportional to its share in the total. To distinguish between the components different shades or colours are used and an index is provided to explain the differences. Sub-divided bar diagram is not used when the components is more than 10 or 12, otherwise the diagram will be overload.

Example:- The region rainfall indices during the year 2006 to 2008 are given below:

Year	West	North	East	South	Centre
2006	78.4	88.9	83.7	89.9	86.5
2007	75.6	62.5	103.6	75.5	77.4
2008	121.2	116.5	107.6	123.9	90.3

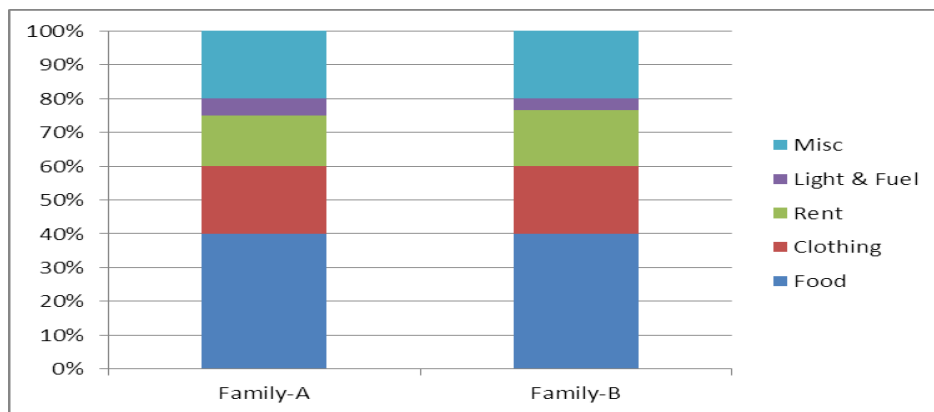


D. Percentage Bar Diagram

This is similar to component bar diagram where the height of all the bars will be 100 and the components inside the bar will be divided according to respective percentage value of its share. To distinguish between the components we use different shades and colors.

Example:- The following data relates to the monthly expenditure of two families A and B:

Items	Family-A	Family-B	% of F-A	% of F-B
Food	16000	12000	40	40
Clothing	8000	6000	20	20
Rent	6000	5000	15	17
Light & Fuel	2000	1000	5	3
Misc	8000	6000	20	20



Two-Dimensional Diagrams:- In one-dimensional diagrams only length or width of the bar is taken into account but in two-dimensional diagrams both length and width is taken account. Thus the area of the bar

represents the data. Two-dimensional diagrams are also known as surface diagrams or area diagrams. Some of them are Rectangles, squares and circles.

A. Rectangular Diagrams

In this diagram a bar is constructed basing on the area of rectangle which is equal to product of length and width and figures are represented as different components. The figures can be represented as given values or by converting it as percentages.

B. Square diagrams

In this method each variable value is represented with one square. The area of square is constructed by its side and is calculated by taking square-root of the given values and expressing them in proportional to the square roots of various quantities. To construct this we require lot of space because each variable value is represented with a square. Also the size of the square increases as the area increases. Thus it requires much time and space to represent diagram.

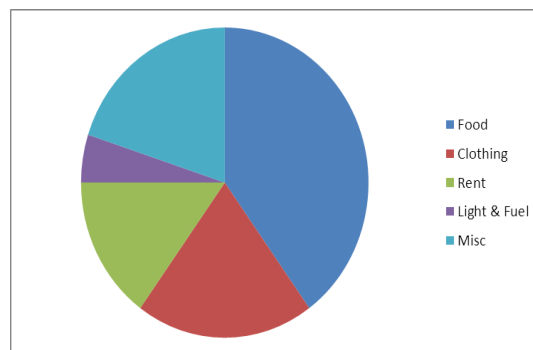
C. Circles

Another way of representing two-dimensional diagrams is circles. Each variable value is represented by circle. Circle is constructed by calculating radii. Basing on the radii value the size of the circle is defined. So, this also take much space and time to represent the data. So, they designed another method using circles which is called Pie-diagram.

PIE-DIAGRAM:- In this method only one circle is constructed with any radii. Each variable value is represented as a sector basing on its angle ($\text{angle} = \frac{\text{Variable Value}}{\text{Total Value}} \times 360$). To differentiate between the components we use different shades or colors.

Example:- Following is the information of Monthly expenditure of a family:

Items	Family-A	Angle
Food	160	144
Clothing	80	72
Rent	60	54
Light & Fuel	20	18
Misc	80	72
	400	



Three-Dimensional Diagrams:- These are also known as volume diagrams. They are Cubes, Cylinders, Spheres, etc., In these diagrams 3 things are taken into account namely length, width and height. These diagrams are used when the gap between first and last value is very large.

Pictograms:- Pictographs are represented using pictorial symbols. They are not abstract presentations such as lines or bars. Pictures are attractive and easy to comprehend and as such this method is particularly useful in presenting statistics to layman.

Cartograms:- These are used to give quantitative information on a geographical basis. They are thus used to represent spatial distributions. The quantities on the map can be shown in many ways, such as through shades or colours, dots, by placing pictograms in each geographical unit along with their numerical figure.

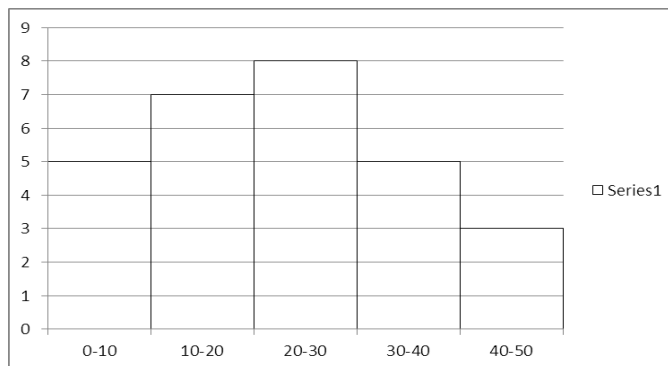
4) Explain different types of graphs?

Ans) A large variety of graphs are used to represent different statistical data. Some important type of graphs are Graphs of Frequency distributions. They are

Histogram

A histogram is a set of vertical bars whose area are proportional to the frequencies represented. It is constructed by taking variable on X-axis and frequencies on Y-axis. For each variable value a frequency is represented as a rectangular bar with a width corresponding to the class-Interval. Each bar will be adjoining each other. Using Histogram we can determine the value of Mode.

Marks	No.Of Students
0-10	5
10-20	7
20-30	8
30-40	5
40-50	3

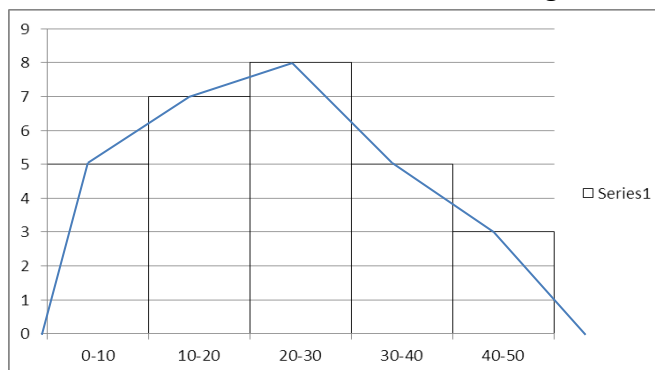


Frequency Polygon

A frequency polygon can be constructed in two ways. One using Histogram and the other without histogram.

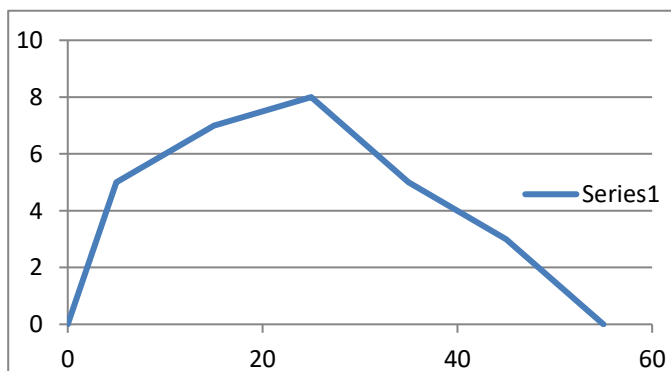
Polygon with Histogram:- A histogram is constructed and on that a curve is represented by marking the mid-values on the top of each bar. Join all the points and extend both the ends of the curve to axis by considering the mid-values of classes before and after the first and last classes that are given.

Marks	No.Of Students	Mid-values
0-10	5	5
10-20	7	15
20-30	8	25
30-40	5	35
40-50	3	45



Polygon without histogram:- A Polygon is constructed by taking mid-values on X-axis and frequencies on Y-axis. Mark a point by representing each frequency against each mid-value and join the points. Extend both the ends of the curve to axis by considering the mid-values of classes before and after the first and last classes that are given.

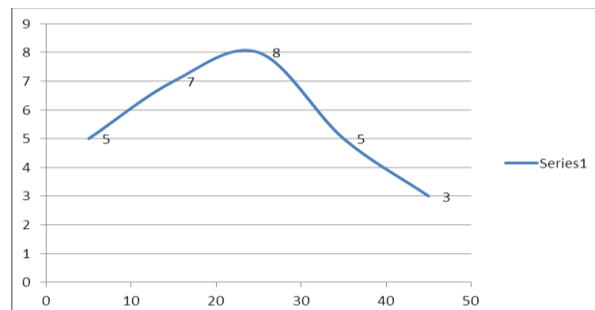
Marks	No.Of Students	Mid-values
0-10	5	5
10-20	7	15
20-30	8	25
30-40	5	35
40-50	3	45



Frequency Curve

A frequency curve is constructed by taking mid-values on X-axis and frequencies on Y-axis. Mark a point by representing each frequency against each mid-value and join the points. The curve so obtained is frequency curve.

Marks	No.Of Students	Mid-values
0-10	5	5
10-20	7	15
20-30	8	25
30-40	5	35
40-50	3	45



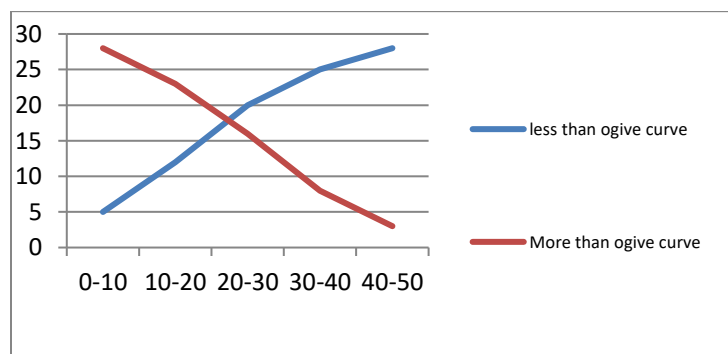
O-Give Curves

These curves are constructed on cumulative frequency distributions by taking Class-Intervals on X-axis and cumulative frequencies on Y-axis. So, they are of two types. Using this curve we can find Median.

Less than Ogive Curve:- It is constructed on less than cumulative frequencies. The L.C.F values are marked against upper limits of the classes. A curve is obtained which will be in upward direction moving from left to right.

More than Ogive Curve:- It is constructed on More than cumulative frequencies. The M.C.F values are marked against lower limits of the classes. A curve is obtained which will be in downward direction moving from left to right.

Marks	No.Of Students	L.C.F	G.C.F
0-10	5	5	28
10-20	7	12	23
20-30	8	20	16
30-40	5	25	8
40-50	3	28	3



T=THEORY QUESTIONS OF OTHER UNITS

SKEWNESS

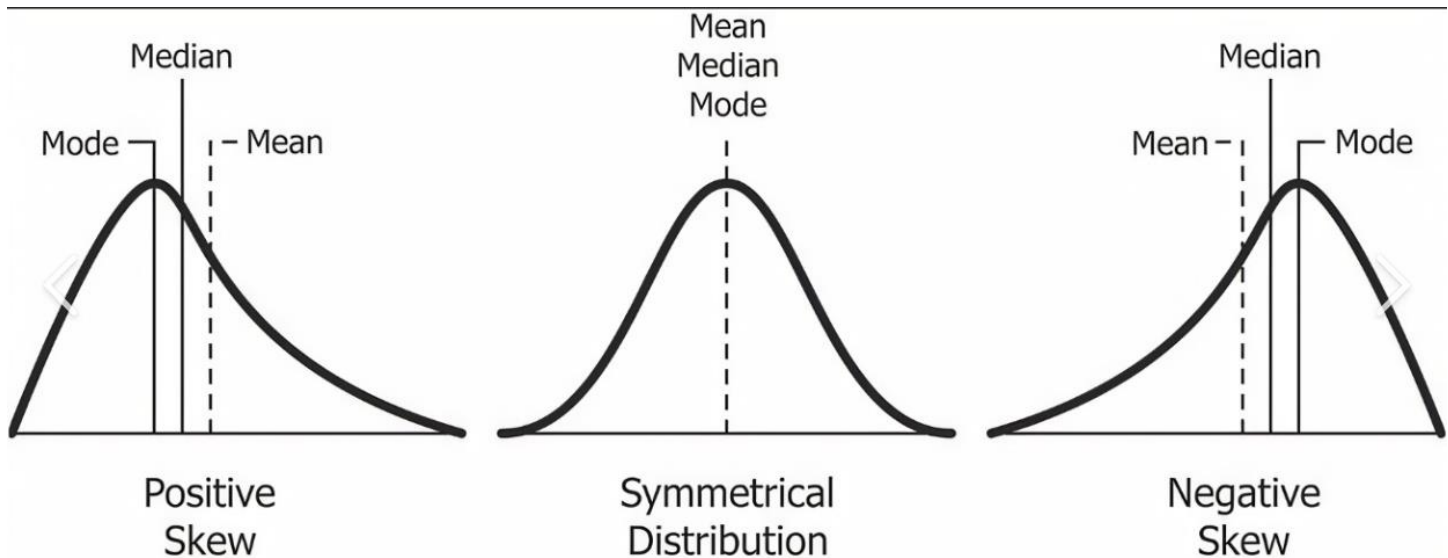
Skewness means lack of symmetry in a frequency distribution. In a non-symmetrical distribution the values of mean, median and mode are not equal. The spread of frequencies is not same on both the sides of central point of the curve.

A frequency distribution is a symmetrical distribution if the value of Mean, Median, Mode coincide i.e., the spread of frequencies is same on both the sides of the central point of the curve.[see figure-2]

In a non-symmetrical frequency distribution the frequency curve will be having a longer tail on one side of the curve i.e., the spread of frequencies will be more on one side [right or left] of the curve.

If the longer tail of the frequency curve of the distribution lies to the right of the central point, it is said to be **positively skewed distribution**. In such a case $\text{Mean} > \text{Median} > \text{Mode}$. [see figure-1]

If the longer tail of the frequency curve of the distribution lies to the left of the central point, it is said to be **negatively skewed distribution**. In such a case $\text{Mean} < \text{Median} < \text{Mode}$. [see figure-3]



KURTOSIS

As skewness helps us to know whether a distribution is symmetric or not. Kurtosis explains whether the distribution is normal or not. Kurtosis tells us whether the distribution when plotted on a graph would give us a normal curve (or) a curve that is more flat than a normal curve (or) a curve that is more peaked than a normal curve.

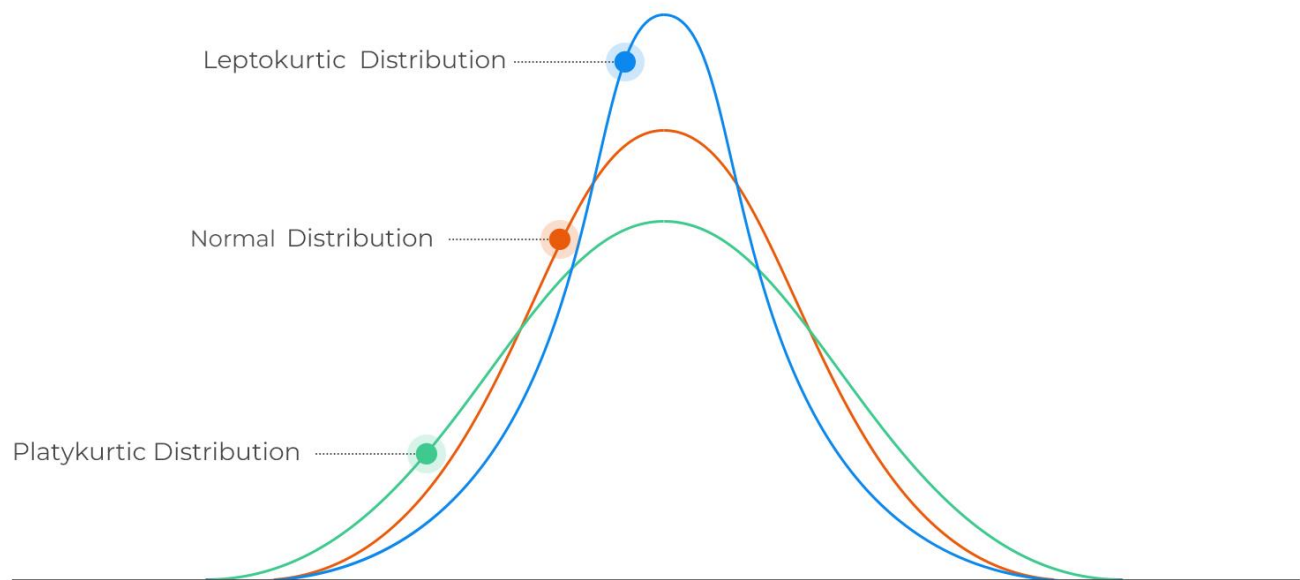
“A measure of kurtosis indicates the degree to which a curve of a frequency distribution is peaked or flat topped”.

If a distribution is more peaked than the normal curve, it is called Leptokurtic distribution. If the distribution is more flat than the normal curve, it is called Platykurtic curve. A normal curve is called Mesokurtic curve.

The kurtosis of any univariate normal distribution is 3. Distribution with kurtosis less than 3 is said to be platykurtic distribution, distribution with kurtosis greater than 3 is said to be Leptokurtic distribution.



Kurtosis



UNIT-V

CORRELATION

1) Define Correlation?

A) The degree of relationship between two variables, where with the changes in the values of one variable, the values of other variable also changes is called correlation. For example, relationship between age of husband, age of wife where with the increase in the age of husband, age of wife also increases.

“ Correlation analysis deals with the association between two or more variables”.--- Simpson & Kafka

“ If two or more quantities vary in sympathy so that movements in one tend to be accompanied by corresponding movements in the other then they are said to be correlated”.--- L.R.Conner

“ Correlation is an analysis of the co-variation between two or more variables”.---A.M. Tuttle

“ Correlation analysis attempts to determine the degree of relationship between variables”.--- Ya Lun Chou

2) Mention the uses of correlation?

A) The Utility of the study of correlation is clear from the following points:

(i) The correlation coefficient helps us in measuring the extent of relationship between two or more than two variables.

(ii) It is through correlation that we can predict about the future.

(iii) If the value of a variable is given, we can know the value of another variable . It is, of course, done with the help of regression analysis.

(iv) Correlation contributes to economic behavior. It helps us in knowing the important variables on which others depend.

(v) The technique of ratio of variation and regression analysis depends totally on the findings of coefficient of correlation.

(vi) In the field of commerce and industry, the technique of correlation coefficient helps to make estimates like sales, price or costs.

(vii) The predictions made on the basis of correlation analysis are considered to be nearer to reality and hence reliable.

3) Mention Types of correlation?

A) Correlation is classified in several different ways. Some of them are as follows:

(1) Positive & Negative Correlation

When the values of two variables move in the same direction i.e., with the increase in the values of one variable the other variable also increases then it is called **Positive Correlation**

For example, Height and weight of persons, As the height increases weight also increase and vice versa, i.e., both the variables move in the same direction.

When the values of two variables move in the opposite direction i.e., with the increase in the values of one variable the other variable decreases and with the decrease in the values of one variable the other variable increases, then it is called **Negative Correlation**.

For example, Price and demand--- As the price of the product increases, demand decreases and as the price decreases, demand increases i.e., both the variables are moving in the opposite direction.

(2) Simple & Multiple correlation

When we find the relationship between two variables then it is **Simple Correlation**.

For example, the relationship between profit and loss is simple correlation.

When we find the relationship between more than two variables then it is **Multiple Correlation**.

For example, the stock market index is based on relationship between many variables like agricultural production, amount of money supply in the economy, level of inflation, etc taken together is multiple correlation.

(3) Partial and Total Correlation

When there are more than two variables under study but the relationship between only two variables is studied assuming that the other variable are constant such correlation is called **Partial correlation**.

When there are more than two variables under study and the relationship between all the variables are studied taken together at one time such correlation is called **Total correlation**.

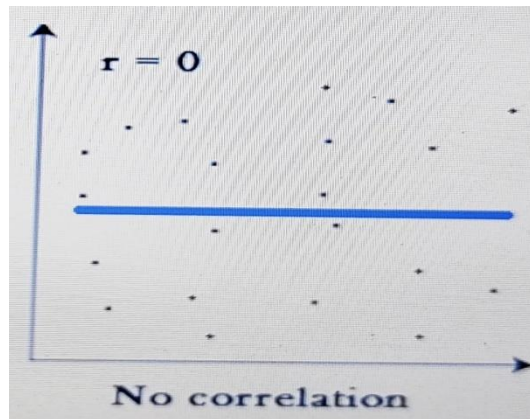
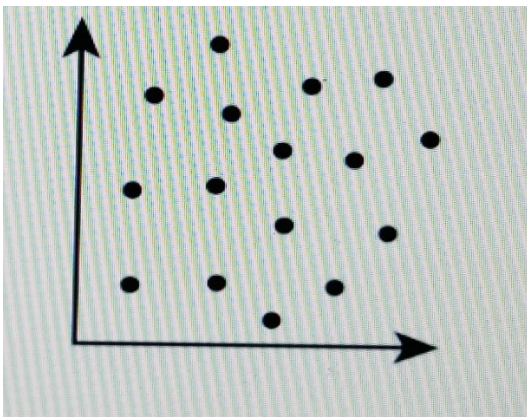
(4) Linear and Non-Linear Correlation

If the relationship between two variables is expressed as a linear equation of the form $y=mx+c$, then it is said to be **Linear Correlation**.

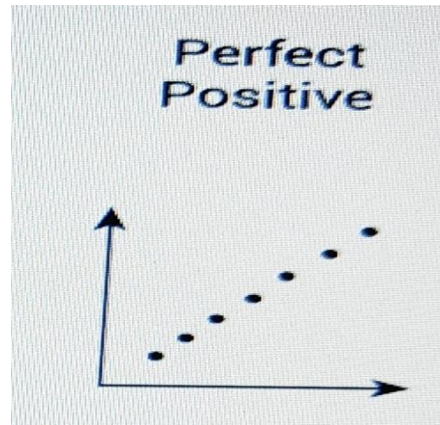
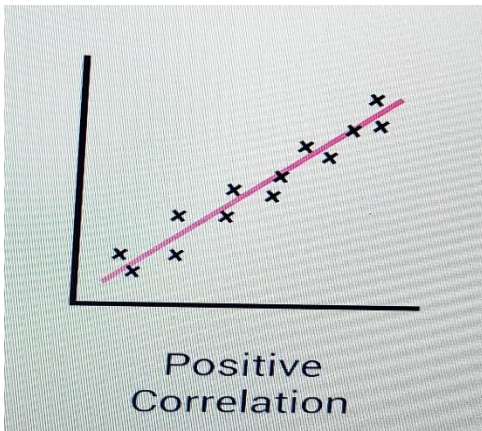
If the relationship between two variables is not expressed as a linear equation then it is said to be **Non-Linear Correlation**.

(4) Explain Scatter diagram?

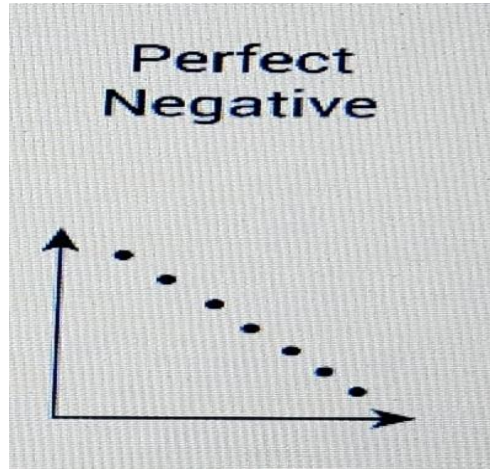
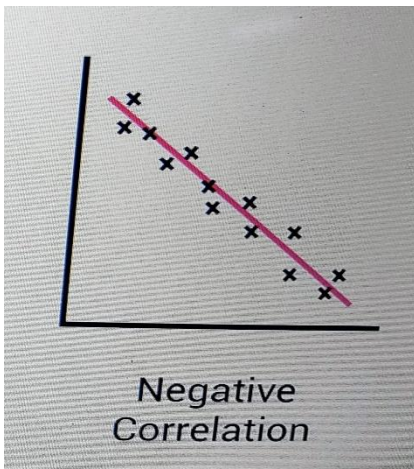
(A) Scatter diagram is one of the simplest method of diagrammatic representation of a bivariate distribution. In this method, the given data are plotted on a graph paper in the form of dots for each pair of X and Y values. If the points are closer there is correlation between the variables, if the points scatter there is no relation between the variables.



If the points move in the same direction i.e., if they move in upward direction from left to right, then there is a ***positive correlation*** between them. If all the points cluster in a single straight line in upward trend then there is a ***perfect positive correlation*** between them.



If the points move in the opposite direction i.e., if they move in downward direction from left to right, then there is a **negative correlation** between them. If all the points cluster in a single straight line in downward trend then there is a **perfect negative correlation** between them.



Merits:-

- i) It is very easy to plot and easy to understand.
- ii) It is not affected by extreme values.
- iii) The shape of the scatter diagram reveals whether correlation is positive or negative, linear or non-linear.
- iv) By drawing a line of best fit we can estimate the value of one variable, given the value of another variable.

De-Merits:-

- i) It is subjective to some extent.
- ii) It does not give an exact measure of the correlation.

iii) It is not suitable if the number of observations is very large.

(5) Explain Graphic method?

(A) In this method, the individual values of the variables are plotted on a graph paper. There will be two curves, one for X variable and another for Y variable. By looking at the direction the two curves, we can say whether there is any correlation between the variables or not.

Merits:-

- i) It is simple to plot and easy to understand.
- ii) The shape of two curves reveal whether correlation is positive or negative, linear or non-linear.
- iii) It is useful if data is given over a period of time and exact correlation is not required.

De-Merits:-

- i) It does not give an exact measure of the correlation.
- ii) It is not suitable if the number of observations is very small.

