I MBA – Semester - I					
Course Code	**STATISTICAL METHODS FOR MANAGERS**	**L**	**T**	**P**	**C**
22MBA115		**3**	**1**	**0**	**4**

Course Educational Objectives (CEO):

CEO1: To provide knowledge on basics of Statistics and data presentation
CEO2: To develop skills for the measurement of right average for the given data
CEO3: To provide knowledge for the measurement of right deviation and coefficient of variation for the given data to take right managerial decision
CEO4: To provide knowledge to find out the relationship between variables and Coefficient of Correlation for the given data to take right managerial decision
CEO5: To develop skills for the application of a right test for the testing Hypothesis

UNIT - I	**Introduction**	Lecture Hrs: 12

Meaning and definition of Statistics,- Nature, Scope, Significance of Statistics, Sources of data – Types of Data – Methods of Data Collection. Classification and Tabulation of data –Types of classification and tables –Rules of Classification and tabulation.

Graphs & Diagrams: Graphs and Diagrams Presentation – Importance and different types of graphs and diagrams.

UNIT - II	**Measures of Central Tendency**	Lecture Hrs: 8

Arithmetic–Weighted Mean. Median, Mode

UNIT - III	**Measures of Dispersion**	Lecture Hrs:8

Range, Quartile Deviation, Mean Deviation. Standard Deviation, Coefficient of Variation

UNIT - IV	**Correlation and Regression**	Lecture Hrs:12

Introduction, Significance and Types of Correlation, Methods of Correlation- Coefficient of Correlation, and Multiple Correlation Analysis.

Regression: Meaning and Purpose of Regression Analysis – Regression Lines and Regression Equations, Multiple Regression Analysis.

UNIT - V	**Testing of Hypothesis**	Lecture Hrs:12

One Sample and Two sample tests for means of small samples (t-Test), F test for two samples.

ANOVA (One- way classification and Two- way Classification), Chi-square test (Goodness of Fit, Independence of attributes).

Course Outcomes :

On successful completion of the course the students will be able to		**Pos related to COs**
CO1	**Demonstrate** knowledge on Sources of data, Types of Data, Methods of Data Collection preparation of Classification and Tabulation of data.	PO1, PO2, PO5
CO2	**Measure** a right deviation and coefficient of variation for the given data to take right managerial decision	PO1,PO2,
CO3	**Apply** correlation and regression techniques for forecasting and decision making	PO1, PO2, PO6,PO7

CO4	**Find out** the relationship between variables and Coefficient of Correlation for the given data to take right managerial decision	PO1, PO2, PO6, PO7
CO5	**Apply** a right Hypothesis test for the given data to take right decision	PO1, PO2, PO6, PO7

Textbooks:

1. Statistical Method, S.P Gupta, Sulthan Chand & Sons, 2017.
2. Statistics for Management, Richard I Levin, David S.Rubin, Pearson, 2008.

References:

1. Business Statistics, Gupta S.C & Ira Gupta, Himalaya Publishing House, Mumbai, 2012.
2. Statistics for Management, P.N.Arora, S.Arora, S.Chand, 2009.
3. Statistics for Management, Levin, Pearson Company, New Delhi, 2013.

Online Learning Resources:

https://archive.nptel.ac.in/courses/110/107/110107114/
https://archive.nptel.ac.in/courses/121/106/121106007/

STATISTICAL METHOD FOR MANGER

Unit 1

Meaning and Definition of Statistics

MEANING

A branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data

Statistics means studying, collecting, analyzing, interpreting, and organizing data. Statistics is a science that helps to gather and analyze numerical data in huge quantities. With the help of statistics, you can measure, control, and communicate uncertainty. It allows you to infer proportions in a whole, derived from a representative sample. In other words, statistics could be described as the feature or characteristic of a sample and is generally used to estimate a population parameter's value.

Definition of Statistics

Statistics is defined as “Science of counting and averages Science of counting and averages.”

- A.L. Bowley,

-

“Statistics is defined as the. Collection, Presentation, Analysis and Interpretation of numerical data.”

- Croxton and Cowden

“By statistics we mean quantitative data affected to a marked extent by multiplicity of causes.”

- Yule and Kendall defined statistics

What are the Features or Characteristics of Statistics?

The following are the main characteristics of statistics:

1. Statistics are aggregate of facts
2. Statistics must be numerically expressed
3. Statistics should be collected for a pre-determined purpose
4. Statistics should be collected in a systematic manner
5. Statistics should be capable of being placed in relation to each other

6. Statistics are affected to a great extent by a multiplicity of causes
7. Statistics are enumerated or estimated according to reasonable standards of accuracy
8. A reasonable standard of accuracy should be maintained in the collection of statistics

1. Statistics are aggregate of facts

A single numerical figure is not statistics. For example, the height of an individual, the price of a certain commodity does not form statistics as are unrelated and incomparable.

But the aggregate of such figures of births, accident suctions, imports, exports, etc. constitute statistics as these figures can study in relation to time, place, and frequency of occurred.

2. Statistics must be numerically expressed

Qualitatively expressed statements such as "**Nepal is a developing country**", "The cement production in Nepal is increasing", "Nepal and China, etc. do not constitute statistics.

But the quantitative statement In 1990, the production of wheat was 20 thousand metric tonnes compared to 15 thousand metric tonnes in 1985" constitutes statistics.

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

The objective or the purpose of the inquiry should be clearly stated before collecting the data. The data collected without any pre-determined purpose objective may not be useful for the inquiry. So it is of at most important to define in clear and concrete terms the objective or the purpose of the inquiry.

4. Statistics should be collected in a systematic manner

A suitable plan should be prepared for data collection and the work should be carried out in a systematic manner. If the data are collected haphazardly, the conclusions may be wrong or miss leading.

5. Statistics should be capable of being placed in relation to each other

The numerical data should be comparable, as statistics are collected mostly for the purpose of comparisons. The numerical data collected constitutes statistics if they are comparable. In order to make valid comparisons, the data should be homogeneous and uniform.

6. Statistics are affected to a great extent by a multiplicity of causes

Numerical figures should be affected by a large number of causes.

7. Statistics are enumerated or estimated according to reasonable standards of accuracy

Data may be collected either by actual counting and measurement or by estimation. The figures obtained by counting and measurement will be exact and accurate but the figures estimated can not be as accurate as those obtained by and measurement.

The degree of accuracy of the estimated values largely depends on the nature and purpose of the inquiry.

8. A reasonable standard of accuracy should be maintained in the collection of statistics:

Statistics deals with a large number of data. Instead of counting each and every item, Statisticians take a sample and apply the result thus obtained from sample to the whole group.

The degree of accuracy of the sample largely depends upon the nature and object of the inquiry. If a reasonable standard of accuracy is not maintained, numbers may give misleading results.

Importance and Scope of Statistics:

The fact those in the modern world statistical methods are universally applicable. It is in itself enough to show how important the science of statistics is.

1. **Statistics are aggregate of facts**:

A single age of 20 or 30 years is not statistics, a series of ages are. Similarly, a single figure relating to production, sales, birth, death etc., would not be statistics although aggregates of such figures would be statistics because of their comparability and relationship.

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

A number of causes affect statistics in a particular field of enquiry, e.g., in production statistics are affected by climate, soil, fertility, availability of raw materials and methods of quick transport.

3. **Statistics are numerically expressed, enumerated or estimated**:

The subject of statistics is concerned essentially with facts expressed in numerical form—with their quantitative details but not qualitative descriptions. Therefore, facts indicated by terms such as ‘good’, ‘poor’ are not statistics unless a numerical equivalent is assigned to each expression. Also this may either be enumerated or estimated, where actual enumeration is either not possible or is very difficult.

4. **Statistics are numerated or estimated according to reasonable standard of accuracy:**

Personal bias and prejudices of the enumeration should not enter into the counting or estimation

of figures, otherwise conclusions from the figures would not be accurate. The figures should be counted or estimated according to reasonable standards of accuracy. Absolute accuracy is neither necessary nor sometimes possible in social sciences. But whatever standard of accuracy is once adopted, should be used throughout the process of collection or estimation.

5. **Statistics should be collected in a systematic manner for a predetermined purpose**:

The statistical methods to be applied on the purpose of enquiry since figures are always collected with some purpose. If there is no predetermined purpose, all the efforts in collecting the figures may prove to be wasteful. The purpose of a series of ages of husbands and wives may be to find whether young husbands have young wives and the old husbands have old wives.

6. **Statistics should be capable of being placed in relation to each other**:

The collected figure should be comparable and well-connected in the same department of inquiry. Ages of husbands are to be compared only with the corresponding ages of wives, and not with, say, heights of trees.

Scope of Statistics:

Statistics is applied in every sphere of human activity – social as well as physical A– like Biology, Commerce, Education, Planning, Business Management, Information Technology, etc.

1. Statistics and Economics:

Statistical data and techniques are immensely useful in solving many economic problems such as fluctuation in wages, prices, production, distribution of income and wealth and so on.

2. Statistics and Firms:

Statistics is widely used in many firms to find whether the product is conforming to specifications or not.

3. Statistics and Commerce:

Statistics are life blood of successful commerce. Market survey plays an important role to exhibit the present conditions and to forecast the likely changes in future.

4. Statistics and Education:

Statistics is necessary for the formulation of policies to start new course, according to the changing environment. There are many educational institutions owned by public and private engaged in research and development work to test the past knowledge and evolve new knowledge. These are possible only through statistics.

5. Statistics and Planning:

Statistics is indispensable in planning. In the modern world, which can be termed as the "world of planning", almost all the organisations in the government are seeking the help of planning for efficient working, for the formulation of policy decisions and execution of the same. In order to achieve the above goals, various advanced statistical techniques are used for processing, analyzing and interpreting data. In India, statistics play an important role in planning, both at the central and state government levels, but the quality of data highly unscientific.

6. Statistics and Medicine:

In Medical sciences, statistical tools are widely used. In order to test the efficiency of a new drug or to compare the efficiency of two drugs or two medicines, t – test for the two samples is used. More and more applications of statistics are at present used in clinical investigation.

7. Statistics and Modern applications:

Recent developments in the fields of computer and information technology have enabled statistics to integrate their models and thus make statistics a part of decision making procedures of many organisations. There are many software packages available for solving simulation problems.

Sources of data

The foundation of data analysis in statistics lies in the collection of data. Data is nothing but unorganized facts and figures which are collected for a certain purpose, like an analysis. The medium through which data is collected is termed as a source of data.

Sources of data are of two types; they are as follows –

- **Statistical Data**

  This type of data source refers to the collection of data that are used for official purposes, such as population census, official surveys, etc.
- **Non-Statistical Data**

  This type of data source refers to the collection of data that are used for various administrative purposes, mainly in the private sector.

Different Sources of Data

Sources of data can also be classified based on its collection methods, which are –

- **Internal Sources of Data**

  In several cases for a certain analysis, data is collected from records, archives, and various other sources within the organization itself. Such sources of data are termed internal sources of data.

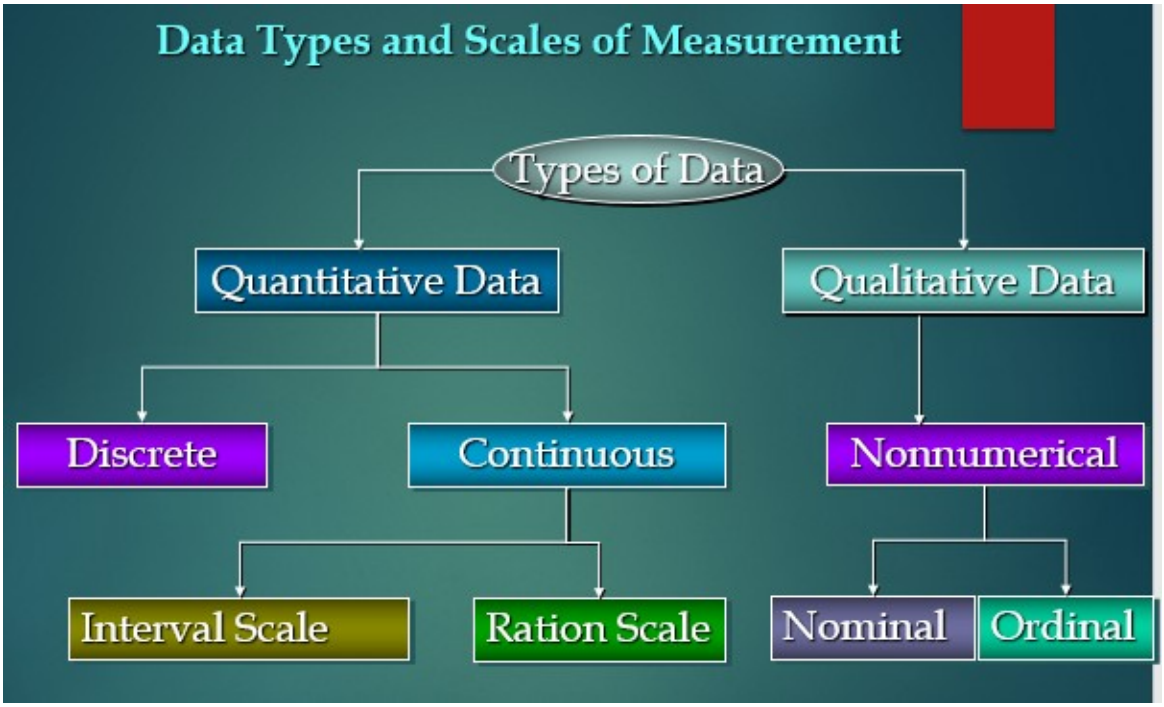
- **Example:** A school is performing an analysis to figure out the highest marks achieved in class 8 science subjects for the last 10 years.

- **External Sources of Data**

  Data may also be collected from various sources outside the organization for analytical purposes. Such sources of data collection are known as external sources of data.

- **Example:** As a patient, you are analyzing the price charts of your nearby hospitals for the treatment of ulcers.

Types of Data



Data is defined as a systematic record corresponding to a specific quantity. Basically, data can be summarised as a set of facts and figures which can be used to serve a specific usage or purpose. For instance, data can be used as a survey or an analysis. Data in a systematic and organized form is referred to as information. In addition to this, the source of data primary or secondary is also an essential factor.

Qualitative Data

Qualitative data is used to represent some characteristics or attributes of the data. The facts and figures depicted by the qualitative data cannot be computed. These properties reflect observable attributes. These are non-numerical in nature. The qualitative data characteristics are exploratory on a larger end than being conclusive in nature. For instance, data on attributes such as honesty, loyalty, wisdom, and creativity for a set of persons defined can be considered as qualitative data.

Examples:

- Attitudes of people to a political system.
- Music and art
- Intelligence
- Beauty of a person

Nominal Data

Nominal data is a sub-category belonging to one of the types of qualitative information. Also known as the nominal scale, it is used to label the variables without providing the numerical value for them. Nominal data attributes can't either be ordered or measured. The nominal data can be both qualitative and quantitative in nature. For instance, some of the nominal data attributes are letters, symbols or gender, etc.

The examination of the nominal data is based on the usage of the grouping method. This method is based on the principle of the grouping of data into different categories. This is followed by the calculation of the frequency or the percentage of the data.

Examples:

- Gender (Women, Men)
- Eye color (Blue, Green, Brown)
- Hair color (Blonde, Brown, Brunette, Red, etc.)
- Marital status (Married, Single)
- Religion (Muslim, Hindu, Christian)

Ordinal Data

Ordinal data/variable is the specific type of data that follows a natural order. The difference between the data values is not determined in the case of nominal data. For instance, ordinal data variable is mostly found in surveys, economics, questionnaires, and finance operations.

The examination of the nominal data is based on the usage of visualization tools. The visualization of this data is done using the bar chart. The ordinal data can be expressed in the form of tables which have each row corresponding to the distinct category.

Examples:

- Feedback is recorded in the form of ratings from 1-10.
- Education level: elementary school, high school, college.
- Economic status: low, medium, and high.

- Letter grades: A, B, C, and etc.
- Customer level of satisfaction: very satisfied, satisfied, neutral, dissatisfied, very dissatisfied.

Quantitative Data

Quantitative data can be measured and is not just observable. The measurement of data is numerically recorded and represented. Calculations and interpretations can then be performed on the obtained results. Numerical data is indicated by quantitative data. For instance, data can be recorded about how many users found a product satisfactory in terms of the collected rating, and therefore, an overall product review can be generated.

Examples:

- Daily temperature
- Price
- Weights
- Income

Discrete Data

Discrete data refers to the data values which can only attain certain specific values. Discrete data can't attain a range of values. Discrete data can be represented using bar charts. For instance, ratings of a product made by the users can only be in discrete numbers.

Examples:

- The number of students in a class,
- The number of chips in a bag,
- The number of stars in the sky

Continuous Data

Continuous Data can contain values between a certain range that is within the highest and lowest values. The corresponding difference between the highest and lowest value of these intervals can be termed as the range of data. Continuous data can be tabulated in what is called a frequency distribution. The frequency distribution table can be computed for the range type of data. It can also be depicted using histograms. For example, the heights of the students in the class can be largely varying in nature, therefore, they can be divided into ranges to summarise the data.

Examples:

- Height and weight of a student,
- Daily temperature recordings of a place
- Wind speed measurement

Methods of Data collection

What is Data Collection?

In Statistics, data collection is a process of gathering information from all the relevant sources to find a solution to the research problem. It helps to evaluate the outcome of the problem. The data collection methods allow a person to conclude an answer to the relevant question. Most of the organizations use data collection methods to make assumptions about future probabilities and trends. Once the data is collected, it is necessary to undergo the data organization process.

The main sources of the data collections methods are “Data”. Data can be classified into two types, namely primary data and secondary data. The primary importance of data collection in any research or business process is that it helps to determine many important things about the company, particularly the performance. So, the data collection process plays an important role in all the streams. Depending on the type of data, the data collection method is divided into two categories namely,

- Primary Data Collection methods
- Secondary Data Collection methods

Primary Data Collection Methods

Primary data or raw data is a type of information that is obtained directly from the first-hand source through experiments, surveys or observations. The primary data collection method is further classified into two types. They are

- Quantitative Data Collection Methods
- Qualitative Data Collection Methods

Let us discuss the different methods performed to collect the data under these two data collection methods.

Quantitative Data Collection Methods

It is based on mathematical calculations using various formats like close-ended questions, correlation and regression methods, mean, median or mode measures. This method is cheaper than qualitative data collection methods and it can be applied in a short duration of time.

Qualitative Data Collection Methods

It does not involve any mathematical calculations. This method is closely associated with elements that are not quantifiable. This qualitative data collection method includes interviews,

questionnaires, observations, case studies, etc. There are several methods to collect this type of data. They are

Observation Method

Observation method is used when the study relates to behavioural science. This method is planned systematically. It is subject to many controls and checks. The different types of observations are:

- Structured and unstructured observation
- Controlled and uncontrolled observation
- Participant, non-participant and disguised observation

Interview Method

The method of collecting data in terms of oral or verbal responses. It is achieved in two ways, such as

- Personal Interview – In this method, a person known as an interviewer is required to ask questions face to face to the other person. The personal interview can be structured or unstructured, direct investigation, focused conversation, etc.
- Telephonic Interview – In this method, an interviewer obtains information by contacting people on the telephone to ask the questions or views orally.

Questionnaire Method

In this method, the set of questions are mailed to the respondent. They should read, reply and subsequently return the questionnaire. The questions are printed in the definite order on the form. A good survey should have the following features:

- Short and simple
- Should follow a logical sequence
- Provide adequate space for answers
- Avoid technical terms
- Should have good physical appearance such as colour, quality of the paper to attract the attention of the respondent

Schedules

This method is similar to the questionnaire method with a slight difference. The enumerations are specially appointed for the purpose of filling the schedules. It explains the aims and objects

of the investigation and may remove misunderstandings, if any have come up. Enumerators should be trained to perform their job with hard work and patience.

Secondary Data Collection Methods

Secondary data is data collected by someone other than the actual user. It means that the information is already available, and someone analyses it. The secondary data includes magazines, newspapers, books, journals, etc. It may be either published data or unpublished data. Published data are available in various resources including

- Government publications
- Public records
- Historical and statistical documents
- Business documents
- Technical and trade journals

Unpublished data includes

- Diaries
- Letters
- Unpublished biographies, etc.

CLASSIFICATION OF DATA

Connor defined classification as:

"The process of arranging things in groups or classes according to their resemblances and affinities and gives expression to the unity of attributes that may subsist amongst a diversity of individuals".

The raw data, collected in real situations and arranged haphazardly, do not give a clear picture. Thus to locate similarities and reduce mental strain we resort to classification. Classification condenses the data by dropping out unnecessary details. It facilitates comparison between different sets of data clearly showing the different points of agreement and disagreement. It enables us to study the relationship between several characteristics and make further statistical treatment like tabulation, etc. During population census, people in the country are classified according to sex (males/ females), marital status (married/unmarried), place of residence (rural/urban), Age (0–5 years, 6– 10 years, 11–15 years, etc.), profession (agriculture, production, commerce, transport, doctor, others), residence in states (West Bengal, Bihar, Mumbai, Delhi, etc.), etc.

Rules of Classification

The primary rules of classification are given below:

(i) There should not be any ambiguity in the definition of classes. It will eliminate all doubts while including a particular item in a class.

(ii) All the classes should preferably have equal width or length. Only in some special cases, we use classes of unequal width.

(iii) The class-limits (integral or fractional) should be selected in such a way that no value of the item in the raw data coincides with the value of the limit.

(iv) The number of classes should preferably be between 10 and 20, i.e., neither too large nor too small.

(v) The classes should be exhaustive, i.e., each value of the raw data should be included in them.

(vi) The classes should be mutually exclusive and non-overlapping, i.e., each item of the raw data should fit only in one class.

(vii) The classification must be suitable for the object of inquiry.

(viii) The classification should be flexible and items included in each class must be homogeneous.

(ix) Width of class-interval is determined by first fixing the no. of class-intervals and then dividing the total range by that number.

Types of Classification

There are four types of classification, viz., (i) qualitative; (ii) quantitative; (iii) temporal and (iv) spatial.

I Qualitative classification: It is done according to attributes or non-measurable characteristics; like social status, sex, nationality, occupation, etc. For example, the population of the whole country can be classified into four categories as married, unmarried, widowed and divorced. When only one attribute, e.g., sex, is used for classification, it is called simple classification. When more than one attributes, e.g., deafness, sex and religion, are used for classification, it is called manifold classification.

II Quantitative classification: It is done according to numerical size like weights in kg or heights in cm. Here we classify the data by assigning arbitrary limits known as class-limits. The quantitative phenomenon under study is called a variable. For example, the population of the whole country may be classified according to different variables like age, income, wage, price, etc. Hence this classification is often called 'classification by variables'.

(a) Variable: A variable in statistics means any measurable characteristic or quantity which can assume a range of numerical values within certain limits, e.g., income, height, age, weight, wage, price, etc. A variable can be classified as either discrete or continuous.

(1) Discrete variable: A variable which can take up only exact values and not any fractional values, is called a 'discrete' variable. Number of workmen in a factory, members of a family, students in a class, number of births in a certain year, number of telephone calls in a month, etc., are examples of discrete-variable.

(2) Continuous variable: A variable which can take up any numerical value (integral/fractional) within a certain range is called a 'continuous' variable. Height, weight, rainfall, time, temperature, etc., are examples of continuous variables. Age of students in a school is a

continuous variable as it can be measured to the nearest fraction of time, i.e., years, months, days, etc.

(iii) Temporal classification: It is done according to time, e.g., index numbers arranged over a period of time, population of a country for several decades, exports and imports of India for different five year plans, etc.

(iv) Spatial classification: It is done with respect to space or places, e.g., production of cereals in quintals in various states, population of a country according to states, etc.

Tabular presentation, or, Tabulation

Tabulation may be defined as the systematic presentation of numerical data in rows or/and columns according to certain characteristics. It expresses the data in concise and attractive form which can be easily understood and used to compare numerical figures.

Objectives of Tabulation:

(1) To simplify complex data

- It reduces the bulk of information, i.e., it reduces raw data in a simplified and meaningful form so that it can be easily interpreted by a common man in less time.

(2) To bring out essential features of data

- It brings out the chief/main characteristics of data.
- It presents facts clearly and precisely without textual explanation.

(3) To facilitate comparison

- The representation of data in rows and columns is helpful in simultaneous detailed comparison on the basis of several parameters.

(4) To facilitate statistical analysis

- Tables serve as the best source of organised data for statistical analysis.
- The task of computing average, dispersion, correlation, etc., becomes easier if data is presented in the form of a table.

(5) To save space

- A table presents facts in a better way than the textual form.
- It saves space without sacrificing the quality and quantity of data.

Parts of tabulation

An ideal statistical table should contain the following items:

(i) Table number: A number must be allotted to the table for identification, particularly when there are many tables in a study.

(ii) Title: The title should explain what is contained in the table. It should be clear, brief and set in bold type on top of the table. It should also indicate the time and place to which the data refer.

(iii) Date: The date of preparation of the table should be given.

(iv) Stubs or, Row designations: Each row of the table should be given a brief heading. Such designations of rows are called “stubs”, or, “stub items” and the entire column is called “stub column”.

(v) Column headings or, Captions: Column designation is given on top of each column to explain to what the figures in the column refer. It should be clear and precise. This is called a “caption”, or, “heading”. Columns should be numbered if there are four, or, more columns

(vi) Body of the table: The data should be arranged in such a way that any figure can be located easily. Various types of numerical variables should be arranged in an ascending order, i.e., from left to right in rows and from top to bottom in columns. Column and row totals should be given.

(vii) Unit of measurement: If the unit of measurement is uniform throughout the table, it is stated at the top right-hand corner of the table along with the title. If different rows and columns contain figures in different units, the units may be stated along with “stubs”, or, “captions”. Very large figures may be rounded up but the method of rounding should be explained.

(viii) Source: At the bottom of the table a note should be added indicating the primary and secondary sources from which data have been collected.

(ix) Footnotes and references: If any item has not been explained properly, a separate explanatory note should be added at the bottom of the table.

Rules of tabulation

(i) A good table must contain all the essential parts, such as, Table number, Title, Head note, Caption, Stub, Body, Foot note and source note.

(ii) A good table should be simple to understand. It should also be compact, complete and self-explanatory.

(iii) A good table should be of proper size. There should be proper space for rows and columns. One table should not be overloaded with details. Sometimes it is difficult to present entire data in a single table. In that case, data are to be divided into more number of tables.

(iv) A good table must have an attractive get up. It should be prepared in such a manner that a scholar can understand the problem without any strain.

(v) Rows and columns of a table must be numbered.

(vi) In all tables the captions and stubs should be arranged in some systematic manner. The manner of presentation may be alphabetically or chronologically depending upon the requirement.

(vii) The unit of measurement should be mentioned in the head note.

(viii) The figures should be rounded off to the nearest hundred, or thousand or lakh. It helps in avoiding unnecessary details.

(ix) Percentages and ratios should be computed. Percentage of the value for item to the total must be given in parenthesis just below the value.

(x) In case of non-availability of information, one should write N.A. or indicate it by dash (-).

The advantages of a tabular presentation over the textual presentation are:

i. It is concise;
ii. There is no repetition of explanatory matter;
iii. Comparisons can be made easily;
iv. The important features can be highlighted; and
v. Errors in the data can be detected.

5. DIAGRAMATIC AND GRAPHICAL REPRESENTATION

5.1 Introduction:

In the previous chapter, we have discussed the techniques of classification and tabulation that help in summarising the collected data and presenting them in a systematic manner. However, these forms of presentation do not always prove to be interesting to the common man. One of the most convincing and appealing ways in which statistical results may be presented is through diagrams and graphs. Just one diagram is enough to represent a given data more effectively than thousand words.

Moreover even a layman who has nothing to do with numbers can also understands diagrams. Evidence of this can be found in newspapers, magazines, journals, advertisement, etc. An attempt is made in this chapter to illustrate some of the major types of diagrams and graphs frequently used in presenting statistical data.

5.2 Diagrams:

A diagram is a visual form for presentation of statistical data, highlighting their basic facts and relationship. If we draw diagrams on the basis of the data collected they will easily be understood and appreciated by all. It is readily intelligible and save a considerable amount of time and energy.

5.3 Significance of Diagrams and Graphs:

Diagrams and graphs are extremely useful because of the following reasons.

1. They are attractive and impressive.
2. They make data simple and intelligible.
3. They make comparison possible
4. They save time and labour.
5. They have universal utility.
6. They give more information.
7. They have a great memorizing effect.

5.4 General rules for constructing diagrams:

The construction of diagrams is an art, which can be acquired through practice. However, observance of some general guidelines can help in making them more attractive and effective. The diagrammatic presentation of statistical facts will be advantageous provided the following rules are observed in drawing diagrams.

1. A diagram should be neatly drawn and attractive.
2. The measurements of geometrical figures used in diagram should be accurate and proportional.
3. The size of the diagrams should match the size of the paper.
4. Every diagram must have a suitable but short heading.
5. The scale should be mentioned in the diagram.
6. Diagrams should be neatly as well as accurately drawn with the help of drawing instruments.
7. Index must be given for identification so that the reader can easily make out the meaning of the diagram.
8. Footnote must be given at the bottom of the diagram.
9. Economy in cost and energy should be exercised in drawing diagram.

5.5 Types of diagrams:

In practice, a very large variety of diagrams are in use and new ones are constantly being added. For the sake of convenience and simplicity, they may be divided under the following heads:

1. One-dimensional diagrams
2. Two-dimensional diagrams
3. Three-dimensional diagrams
4. Pictograms and Cartograms

5.5.1 One-dimensional diagrams:

In such diagrams, only one-dimensional measurement, i.e height is used and the width is not considered. These diagrams are in the form of bar or line charts and can be classified as

1. Line Diagram
2. Simple Diagram
3. Multiple Bar Diagram
4. Sub-divided Bar Diagram
5. Percentage Bar Diagram

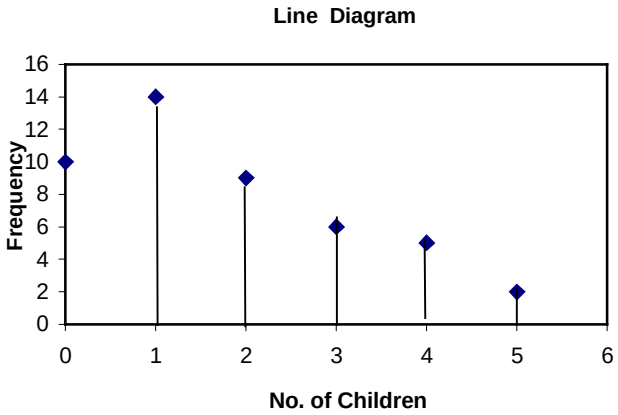
Line Diagram:

Line diagram is used in case where there are many items to be shown and there is not much of difference in their values. Such diagram is prepared by drawing a vertical line for each item according to the scale. The distance between lines is kept uniform. Line diagram makes comparison easy, but it is less attractive.

Example 1:

Show the following data by a line chart:

No. of children	0	1	2	3	4	5
Frequency	10	14	9	6	4	2



Simple Bar Diagram:

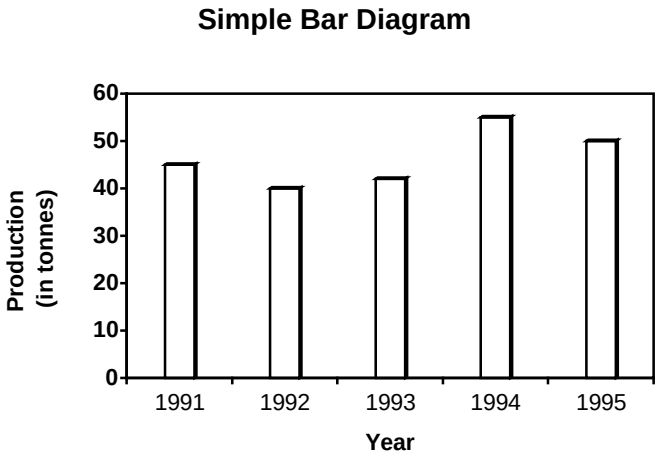
Simple bar diagram can be drawn either on horizontal or vertical base, but bars on horizontal base more common. Bars must be uniform width and intervening space between bars must be equal.While constructing a simple bar diagram, the scale is determined on the basis of the highest value in the series.

To make the diagram attractive, the bars can be coloured. Bar diagram are used in business and economics. However, an important limitation of such diagrams is that they can present only one classification or one category of data. For example, while presenting the population for the last five decades, one can only depict the total population in the simple bar diagrams, and not its sex-wise distribution.

Example 2:
Represent the following data by a bar diagram.

Year	Production (in tones)
1991	45
1992	40
1993	42
1994	55
1995	50

Solution:



Multiple Bar Diagram:

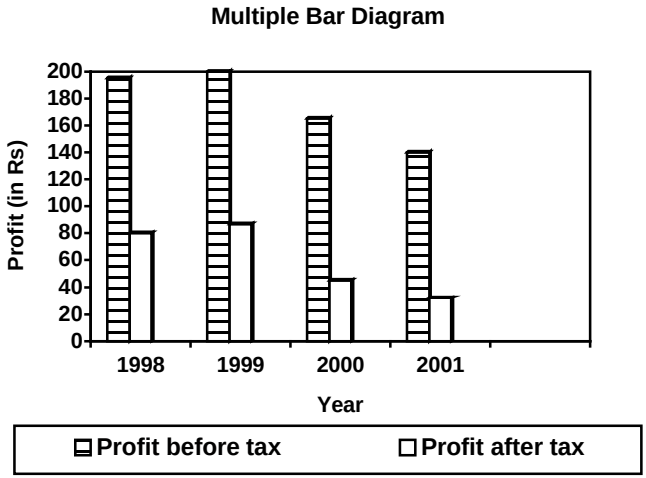
Multiple bar diagram is used for comparing two or more sets of statistical data. Bars are constructed side by side to represent the set of values for comparison. In order to distinguish bars, they may be either differently coloured or there should be different types of crossings or dotting, etc. An index is also prepared to identify the meaning of different colours or dottings.

Example 3:

Draw a multiple bar diagram for the following data.

Year	Profit before tax (in lakhs of rupees)	Profit after tax (in lakhs of rupees)
1998	195	80
1999	200	87
2000	165	45
2001	140	32

Solution:



Sub-divided Bar Diagram:

In a sub-divided bar diagram, the bar is sub-divided into various parts in proportion to the values given in the data and the whole bar represent the total. Such diagrams are also called Component Bar diagrams. The sub divisions are distinguished by different colours or crossings or dottings.

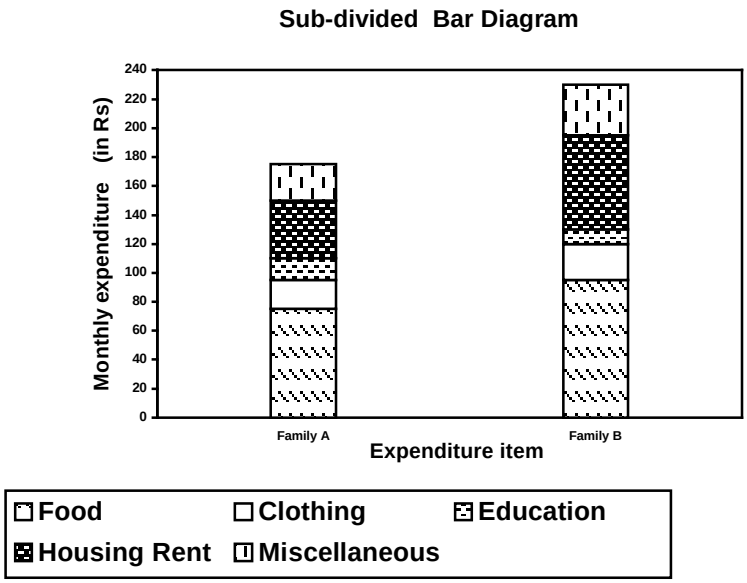
The main defect of such a diagram is that all the parts do not have a common base to enable one to compare accurately the various components of the data.

Example 4:

Represent the following data by a sub-divided bar diagram.

Expenditure items	Monthly expenditure (in Rs.)	
	Family A	Family B
Food	75	95
Clothing	20	25
Education	15	10
Housing Rent	40	65
Miscellaneous	25	35

Solution:



Percentage bar diagram:

This is another form of component bar diagram. Here the components are not the actual values but percentages of the whole. The main difference between the sub-divided bar diagram and percentage bar diagram is that in the former the bars are of different heights since their totals may be different whereas in the latter the bars are of equal height since each bar represents 100 percent. In the case of data having sub-division, percentage bar diagram will be more appealing than sub-divided bar diagram.

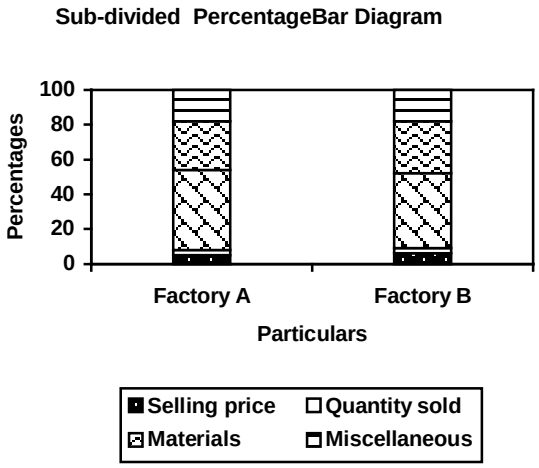
Example 5:
Represent the following data by a percentage bar diagram.

Particular	Factory A	Factory B
Selling Price	400	650
Quantity Sold	240	365
Wages	3500	5000
Materials	2100	3500
Miscellaneous	1400	2100

Solution:
Convert the given values into percentages as follows:

Particulars	Factory A		Factory B	
	Rs.	%	Rs.	%
Selling Price	400	5	650	6
Quantity Sold	240	3	365	3
Wages	3500	46	5000	43
Materials	2100	28	3500	30
Miscellaneous	1400	18	2100	18
Total	7640	100	11615	100

Solution:



5.5.2 Two-dimensional Diagrams:

In one-dimensional diagrams, only length 9 is taken into account. But in two-dimensional diagrams the area represent the data and so the length and breadth have both to be taken into account. Such diagrams are also called area diagrams or surface diagrams. The important types of area diagrams are:

1. Rectangles 2. Squares 3. Pie-diagrams

Rectangles:

Rectangles are used to represent the relative magnitude of two or more values. The area of the rectangles are kept in proportion to the values. Rectangles are placed side by side for comparison. When two sets of figures are to be represented by rectangles, either of the two methods may be adopted.

We may represent the figures as they are given or may convert them to percentages and then subdivide the length into various components. Thus the percentage sub-divided rectangular diagram is more popular than sub-divided rectangular since it enables comparison to be made on a percentage basis.

Example 6:

Represent the following data by sub-divided percentage rectangular diagram.

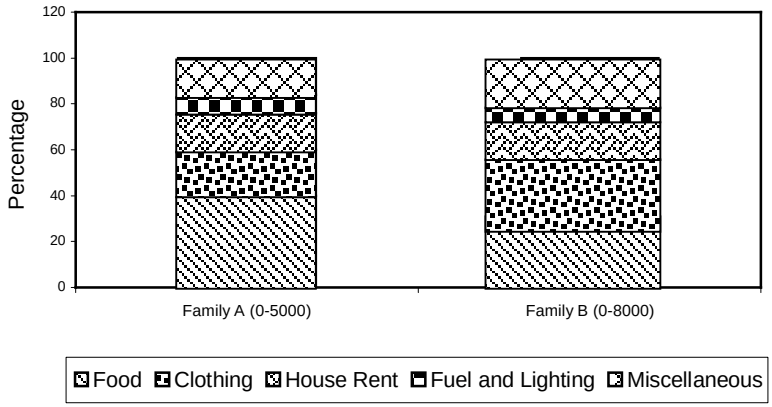
Items of Expenditure	Family A (Income Rs.5000)	Family B (income Rs.8000)
Food	2000	2500
Clothing	1000	2000
House Rent	800	1000
Fuel and lighting	400	500
Miscellaneous	800	2000
Total	5000	8000

Solution:

The items of expenditure will be converted into percentage as shown below:

Items of Expenditure	Family A		Family B	
	Rs.	Y	Rs.	Y
Food	2000	40	2500	31
Clothing	1000	20	2000	25
House Rent	800	16	1000	13
Fuel and Lighting	400	8	500	6
Miscellaneous	800	16	2000	25
Total	5000	100	8000	100

SUBDIVIDED PERCENTAGE RECTANGULAR DIAGRAM



Squares:

The rectangular method of diagrammatic presentation is difficult to use where the values of items vary widely. The method of drawing a square diagram is very simple. One has to take the square root of the values of various item that are to be shown in the diagrams and then select a suitable scale to draw the squares.

Example 7:

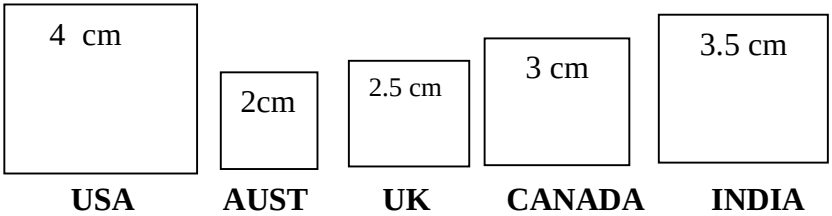
Yield of rice in Kgs. per acre of five countries are

Country	U.S.A	Australia	U.K	Canada	India
Yield of rice in Kgs per acre	6400	1600	2500	3600	4900

Represent the above data by Square diagram.

Solution: To draw the square diagram we calculate as follows:

Country	Yield	Square root	Side of the square in cm
U.S.A	6400	80	4
Australia	1600	40	2
U.K.	2500	50	2.5
Canada	3600	60	3
India	4900	70	3.5



Pie Diagram or Circular Diagram:

Another way of preparing a two-dimensional diagram is in the form of circles. In such diagrams, both the total and the component parts or sectors can be shown. The area of a circle is proportional to the square of its radius.

While making comparisons, pie diagrams should be used on a percentage basis and not on an absolute basis. In constructing a pie diagram the first step is to prepare the data so that various components values can be transposed into corresponding degrees on the circle.

The second step is to draw a circle of appropriate size with a compass. The size of the radius depends upon the available space and other factors of presentation. The third step is to measure points on the circle and representing the size of each sector with the help of a protractor.

Example 8:

Draw a Pie diagram for the following data of production of sugar in quintals of various countries.

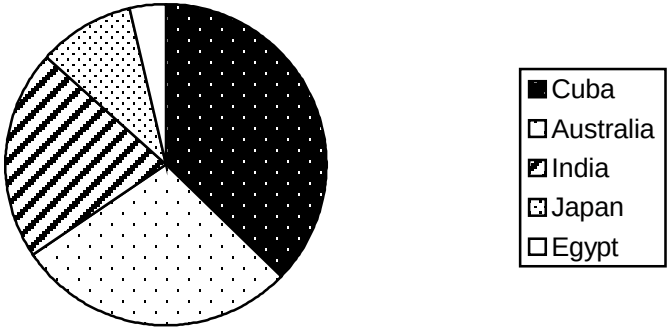
Country	Production of Sugar (in quintals)
Cuba	62
Australia	47
India	35
Japan	16
Egypt	6

Solution:

The values are expressed in terms of degree as follows.

Country	Production of Sugar	
	In Quintals	In Degrees
Cuba	62	134
Australia	47	102
India	35	76
Japan	16	35
Egypt	6	13
Total	166	360

Pie Diagram



5.5.3 Three-dimensional diagrams:

Three-dimensional diagrams, also known as volume diagram, consist of cubes, cylinders, spheres, etc. In such diagrams three things, namely length, width and height have to be taken into account. Of all the figures, making of cubes is easy. Side of a cube is drawn in proportion to the cube root of the magnitude of data.

Cubes of figures can be ascertained with the help of logarithms. The logarithm of the figures can be divided by 3 and the antilog of that value will be the cube-root.

Example 9:

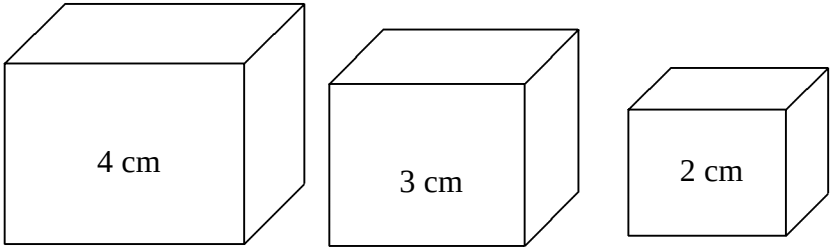
Represent the following data by volume diagram.

Category	Number of Students
Under graduate	64000
Post graduate	27000
Professionals	8000

Solution:

The sides of cubes can be determined as follows

Category	Number of students	Cube root	Side of cube
Undergraduate	64000	40	4 cm
Postgraduate	27000	30	3 cm
Professional	8000	20	2 cm



Undergraduate **Postgraduate** **professional**

5.5.4 Pictograms and Cartograms:

Pictograms are not abstract presentation such as lines or bars but really depict the kind of data we are dealing with. Pictures are attractive and easy to comprehend and as such this method is particularly useful in presenting statistics to the layman. When Pictograms are used, data are represented through a pictorial symbol that is carefully selected.

Cartograms or statistical maps are used to give quantitative information as a geographical basis. They are used to represent

spatial distributions. The quantities on the map can be shown in many ways such as through shades or colours or dots or placing pictogram in each geographical unit.

5.6 Graphs:

A graph is a visual form of presentation of statistical data. A graph is more attractive than a table of figure. Even a common man can understand the message of data from the graph. Comparisons can be made between two or more phenomena very easily with the help of a graph.

However here we shall discuss only some important types of graphs which are more popular and they are

1.Histogram 2. Frequency Polygon
3.Frequency Curve 4. Ogive 5. Lorenz Curve

5.6.1 Histogram:

A histogram is a bar chart or graph showing the frequency of occurrence of each value of the variable being analysed. In histogram, data are plotted as a series of rectangles. Class intervals are shown on the ' X-axis' and the frequencies on the ' Y-axis' .

The height of each rectangle represents the frequency of the class interval. Each rectangle is formed with the other so as to give a continuous picture. Such a graph is also called staircase or block diagram.

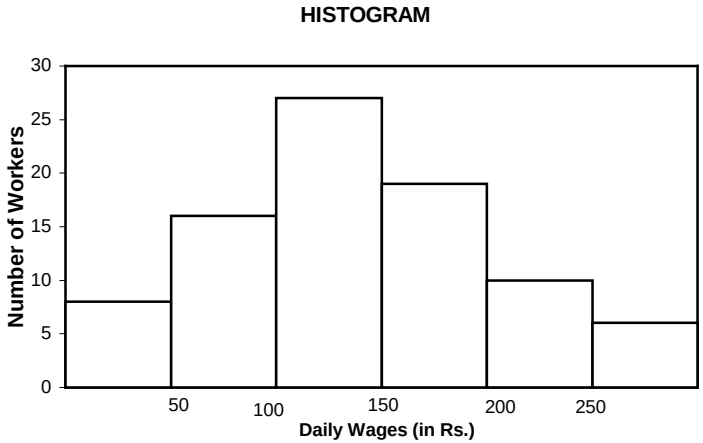
However, we cannot construct a histogram for distribution with open-end classes. It is also quite misleading if the distribution has unequal intervals and suitable adjustments in frequencies are not made.

Example 10:

Draw a histogram for the following data.

Daily Wages	Number of Workers
0-50	8
50-100	16
100-150	27
150-200	19
200-250	10
250-300	6

Solution:



Example 11:

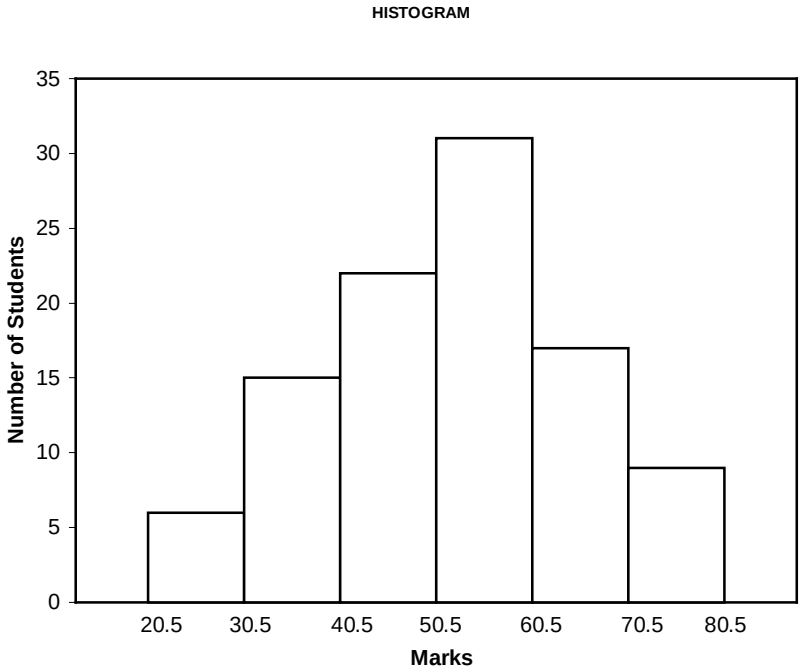
For the following data, draw a histogram.

Marks	Number of Students
21-30	6
31-40	15
41-50	22
51-60	31
61-70	17
71-80	9

Solution:

For drawing a histogram, the frequency distribution should be continuous. If it is not continuous, then first make it continuous as follows.

Marks	Number of Students
20.5-30.5	6
30.5-40.5	15
40.5-50.5	22
50.5-60.5	31
60.5-70.5	17
70.5-80.5	9



Example 12:

Draw a histogram for the following data.

Profits (in lakhs)	Number of Companies
0-10	4
10-20	12
20-30	24
30-50	32
50-80	18
80-90	9
90-100	3

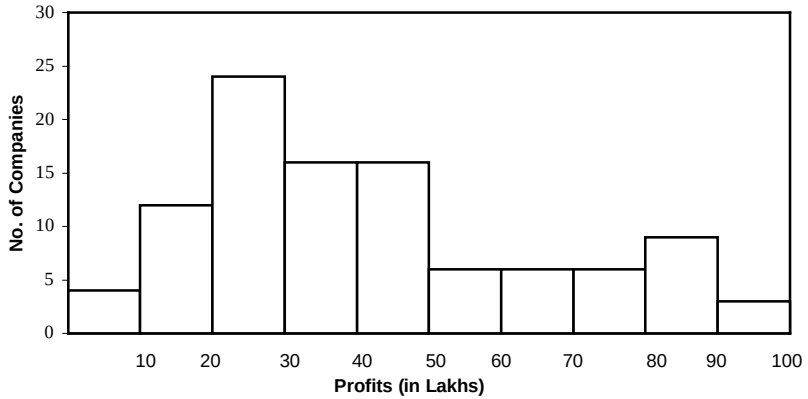
Solution:

When the class intervals are unequal, a correction for unequal class intervals must be made. The frequencies are adjusted as follows: The frequency of the class 30-50 shall be divided by two since the class interval is in double. Similarly the class interval 50-80 can be divided by 3. Then draw the histogram.

Now we rewrite the frequency table as follows.

Profits (in lakhs)	Number of Companies
0-10	4
10-20	12
20-30	24
30-40	16
40-50	16
50-60	6
60-70	6
70-80	6
80-90	9
90-100	3

HISTOGRAM

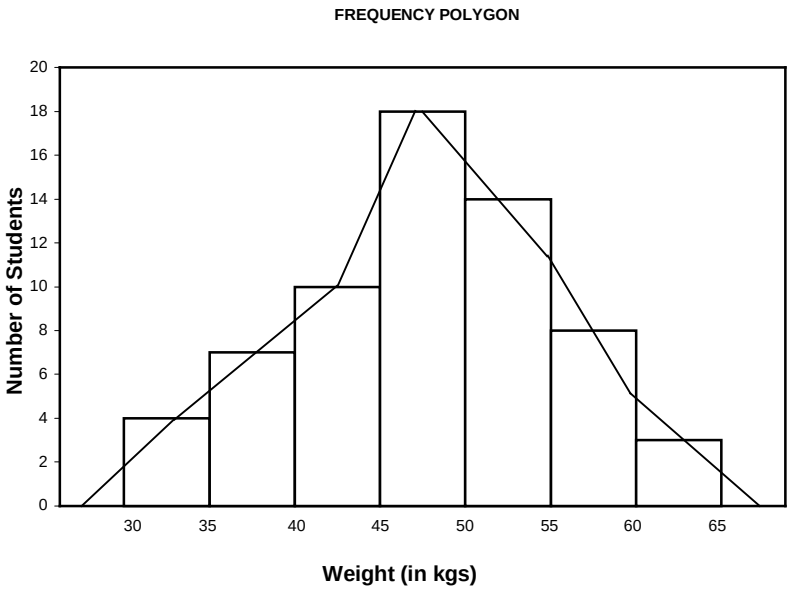


5.6.2 Frequency Polygon:

If we mark the midpoints of the top horizontal sides of the rectangles in a histogram and join them by a straight line, the figure so formed is called a Frequency Polygon. This is done under the assumption that the frequencies in a class interval are evenly distributed throughout the class. The area of the polygon is equal to the area of the histogram, because the area left outside is just equal to the area included in it.

Example 13:
Draw a frequency polygon for the following data.

Weight (in kg)	Number of Students
30-35	4
35-40	7
40-45	10
45-50	18
50-55	14
55-60	8
60-65	3



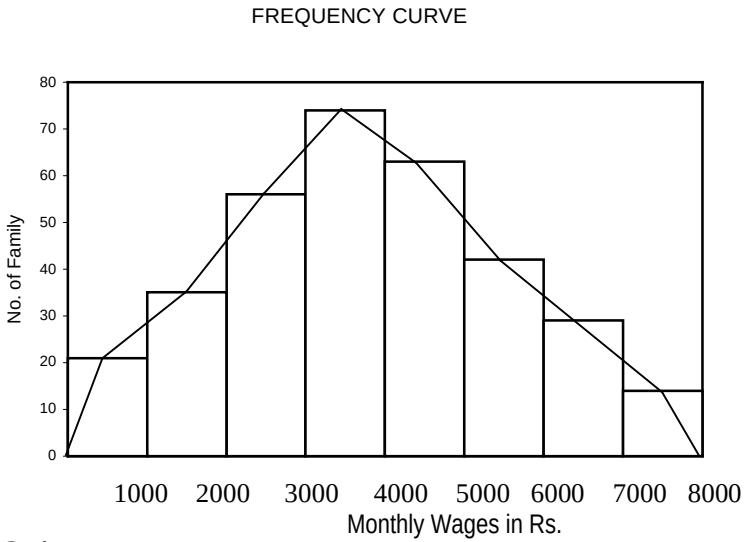
5.6.3 Frequency Curve:

If the middle point of the upper boundaries of the rectangles of a histogram is corrected by a smooth freehand curve, then that diagram is called frequency curve. The curve should begin and end at the base line.

Example 14:
Draw a frequency curve for the following data.

Monthly Wages (in Rs.)	No. of family
0-1000	21
1000-2000	35
2000-3000	56
3000-4000	74
4000-5000	63
5000-6000	40
6000-7000	29
7000-8000	14

Solution:



5.6.4 Ogives:

For a set of observations, we know how to construct a frequency distribution. In some cases we may require the number of observations less than a given value or more than a given value. This is obtained by a accumulating (adding) the frequencies upto

(or above) the give value. This accumulated frequency is called cumulative frequency.

These cumulative frequencies are then listed in a table is called cumulative frequency table. The curve table is obtained by plotting cumulative frequencies is called a cumulative frequency curve or an ogive.

There are two methods of constructing ogive namely:

1. The ‘ less than ogive’ method
2. The ‘ more than ogive’ method.

In less than ogive method we start with the upper limits of the classes and go adding the frequencies. When these frequencies are plotted, we get a rising curve. In more than ogive method, we start with the lower limits of the classes and from the total frequencies we subtract the frequency of each class. When these frequencies are plotted we get a declining curve.

Example 15:

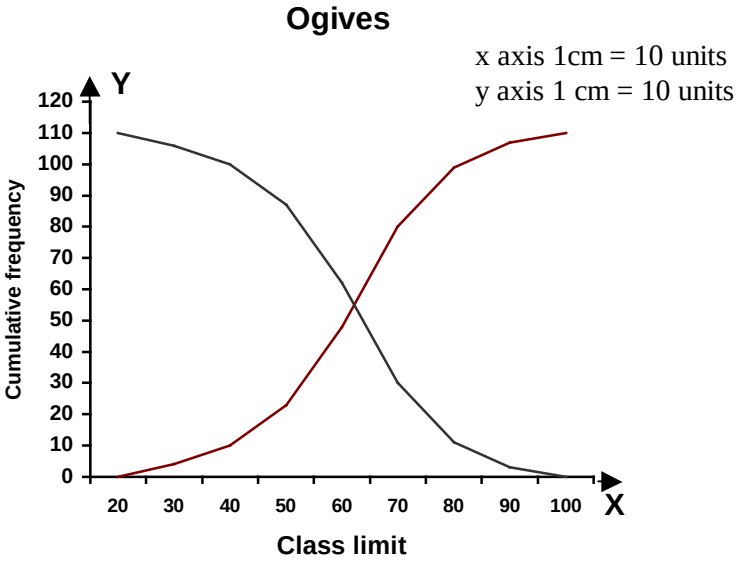
Draw the Ogives for the following data.

Class interval	Frequency
20-30	4
30-40	6
40-50	13
50-60	25
60-70	32
70-80	19
80-90	8
90-100	3

Solution:

Class limit	Less than ogive	More than ogive
20	0	110
30	4	106
40	10	100
50	23	87
60	48	62
70	80	30
80	99	11

90	107	3
100	110	0



5.6.5 Lorenz Curve:

Lorenz curve is a graphical method of studying dispersion. It was introduced by Max.O.Lorenz, a great Economist and a statistician, to study the distribution of wealth and income. It is also used to study the variability in the distribution of profits, wages, revenue, etc.

It is specially used to study the degree of inequality in the distribution of income and wealth between countries or between different periods. It is a percentage of cumulative values of one variable in combined with the percentage of cumulative values in other variable and then Lorenz curve is drawn.

The curve starts from the origin (0,0) and ends at (100,100). If the wealth, revenue, land etc are equally distributed among the people of the country, then the Lorenz curve will be the diagonal of the square. But this is highly impossible.

The deviation of the Lorenz curve from the diagonal, shows how the wealth, revenue, land etc are not equally distributed among people.

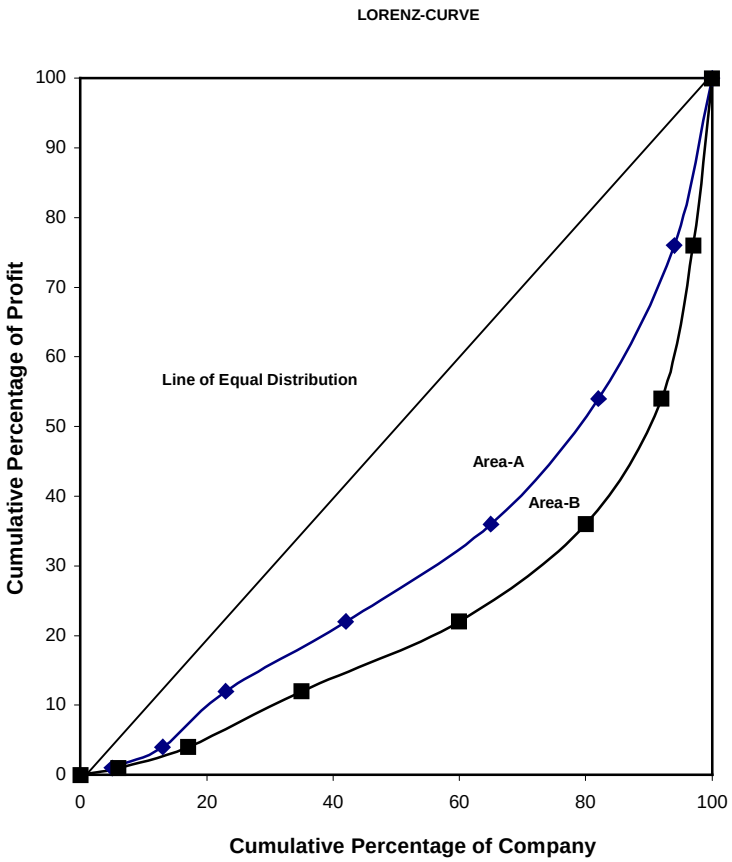
Example 16:

In the following table, profit earned is given from the number of companies belonging to two areas A and B. Draw in the same diagram their Lorenz curves and interpret them.

Profit earned (in thousands)	Number of Companies	
	Area A	Area B
5	7	13
26	12	25
65	14	43
89	28	57
110	33	45
155	25	28
180	18	13
200	8	6

Solution:

Profits			Area A			Area B		
In Rs.	Cumulative profit	Cumulative percentage	No. of companies	Cumulative number	Cumulative percentage	No. of companies	Cumulative number	Cumulative percentage
5	5	1	7	7	5	13	13	6
26	31	4	12	19	13	25	38	17
65	96	12	14	33	23	43	81	35
89	185	22	28	61	42	57	138	60
110	295	36	33	94	65	45	183	80
155	450	54	25	119	82	28	211	92
180	630	76	18	137	94	13	224	97
200	830	100	8	145	100	6	230	100



Exercise – 5

I Choose the best answer:

1. Which of the following is one dimensional diagram.
   (a) Bar diagram (b) Pie diagram (c) Cylinder
   (d) Histogram
2. Percentage bar diagram has
   (a) data expressed in percentages
   (b) equal width
   (c) equal interval
   (d) equal width and equal interval

3. Frequency curve
   (a) begins at the origin (b) passes through the origin
   (c) begins at the horizontal line.
   (d) begins and ends at the base line.
4. With the help of histogram we can draw
   (a) frequency polygon (b) frequency curve
   (c) frequency distribution
   (d) all the above
5. Ogives for more than type and less than type distribution intersect at
   (a) mean (b) median
   (c) mode (d) origin

II Fill in the blanks:

1. Sub-divided bar diagram are also called _____ diagram.
2. In rectangular diagram, comparison is based on _____ of the rectangles.
3. Squares are _____ dimensional diagrams.
4. Ogives for more than type and less than type distribution intersects at _____.
5. _____ Curve is graphical method of studying dispersion.

III. Answer the following:

1. What is diagram?
2. How diagrams are useful in representing statistical data
3. What are the significance of diagrams?
4. What are the rules for making a diagram?
5. What are the various types of diagrams
6. Write short notes on (a) Bar diagram
   (b) Sub divided bar diagram.
7. What is a pie diagram?
8. Write short notes on
   a) Histogram b) Frequency Polygon
   c) Frequency curve d) Ogive
9. What are less than ogive and more than ogive? What purpose do they serve?
10. What is Lorenz curve? Mention its important.

11. Represent the following data by a bar diagram.

Year	Profit (in thousands)
1995	2
1996	6
1997	11
1998	15
1999	20
2000	27

12. Represent the following data by a multiple bar diagram.

Factory	Workers	
	Male	Female
A	125	100
B	210	165
C	276	212

13. Represent the following data by means of percentage sub-divided bar diagram.

Food crops	Area A (in 000,000 acres)	Area B (in 000,000 acres)
Rice	18	10
Wheat	12	14
Barley	10	8
Maize	7	6
Others	12	15

14. Draw a Pie diagram to exhibit the causes of death in the country.

Causes of Death	Numbers
Diarrhoea and enteritis	60
Prematurity and atrophy	170
Bronchitis and pneumonia	90

15. Draw a histogram and frequency polygon for the following data.

Weights (in kg)	Number of men
40-45	8
45-50	14
50-55	21
55-60	18
60-65	10

16. Draw a frequency curve for the following data.

Marks	No. of students
0-20	7
20-40	15
40-60	28
60-80	17
80-100	5

17. The frequency distribution of wages in a certain factory is as follows:

Wages	Number of workers
0- 500	10
500-1000	19
1000-1500	28
1500-2000	15
2000-2500	6

18. The following table given the weekly family income in two different region. Draw the Lorenz curve and compare the two regions of incomes.

Income	No. of families	
	Region A	Region B
1000	12	5
1250	18	10
1500	29	17
1750	42	23
2000	20	15
2500	11	8
3000	6	3

IV. Suggested Activities:

1. Give relevant diagrammatic representations for the activities listed in the previous lessons.
2. Get the previous monthly expenditure of your family and interpret it into bar diagram and pie diagram. Based on the data, propose a budget for the next month and interpreted into bar and pie diagram.

   Compare the two months expenditure through diagrams

Answers

I. 1. (a) 2. (a) 3.(d) 4. (d) 5.(b)

II.

1. Component bar
2. Area
3. Two
4. Median
5. Lorenz