GOVERNMENT COLLEGE(AUTONOMOUS)RAJAHMUNDRY

I B.Sc MAJOR & MINOR (Semester-II)

DESCRIPTIVE STATITICS

Syllabus:

Unit – 1: Statistical Description of Data

Origin, history and definitions of Statistics. Importance, Scope and limitations Statistics. Function of Statistics – Collection, Presentation, Analysis and Interpretation. Collection of data - primary and secondary data and its methods. Classification of data. Presentation of data – Textual, Tabular – essential parts.

Unit – 2: Measurement Scales

Measurement Scales - Frequency distribution and types of frequency distributions. Diagrammatic representation of data – Historiagram, Bar, Multiple bar and Pie with simple problems. Graphical representation of data: Histogram, frequency polygon and Ogives with simple problems.

Unit – 3: Measures of Central Tendency (MCT)

Arithmetic Mean – properties, methods. Median, Mode, Geometric Mean (GM), Harmonic Mean (HM). Calculation of mean, median, mode, GM and HM for grouped and ungrouped data.

Unit – 4: Measures of Dispersion

Concept and problems – Range, Quartile Deviation, Mean Deviation and Standard Deviation, Variance. Central and Non – Central moments and their interrelationship. Sheppard's correction for moments. Skewness and kurtosis.

Unit – 5: Elementary Probability

Basic Concepts of Probability, random experiments, trial, outcome, sample space, event, mutually exclusive and exhaustive events, equally likely and favorable outcomes. Mathematical, Statistical, axiomatic definitions of probability. Conditional Probability and independence of events, Addition and multiplication theorems of probability for 2 and for n events and simple problems. Boole's inequality, Bayes theorem and its applications in real life problems.

Descriptive Statistics (Semester-II)

MODEL QUESTION PAPER

Time: 2 1/2 hrs. Max Marks: 50

SECTION-A

Answer any FIVE questions. 5X4= 20M

1	Explain about limitation of statistics.
2	Write about classification of data.
3	Analyze about nominal and ordinal measurement of scale.
4	Explain about properties of A.M
5	Show that Karl Pearson coefficient of skewness lies between ± 3.
6	State and prove multiplication theorem for 2events.
7	Define (a) sample space (b) exhaustive events (c) exclusive events (d) favorable outcomes.
8.	Explain about kurtosis and types.

SECTION – B

Answer Any THREE questions. 3X10=30M

9	Explain about scope of Statistics.
10	Illustrate about graphical representation of data.
11	Describe briefly about measures of central tendency.
12	Explain in detailed about measures of dispersion
13	Explain about Baye's theorem.
14	State and prove additional theorem for n events.

Introduction and Definition of Statistics

1.1. ORIGIN, HISTORY AND DEVELOPMENT OF STATISTICS

In old days statistics was regarded as "Science of State Craft" and used in various administrative activities of the state. It is used in maintaining the records of population, wealth, land, revenue, military forces etc. of the state by the rulers. In India, the King Chandra Gupta Maurya before 300 B.C. an effective system of registration of births and deaths witnessed by Kautilya's Arthshastra.

The word 'Statistics' seems to derived from the Latin word 'Status' or the Italian word "Statista" or the German word "Statistik". The meaning of all these words refer a political state.

inference.

In present days, the collection and analysing the data are two important aspects to draw valid conclusions and to make policy decisions in almost all the fields of human endeavour. Thus, the subject statistics is regarded as the king of sciences applied in all spheres of life.

1.2. DEFINITION OF STATISTICS

Statistics may be defined as collection of data, classification of data, tabulation of the data, analysis of the data and interpretation of the data.

In ancient times statistics was confined only to the affairs of state, but in modern times the utility of statistics has widened considerably. Now statistics applies in almost every sphere of human activity. Hence a number definitions came into existence and mainly defined in two ways.

(1) In Singular Sense

(2) In Plural Sense.

These two definitions of statistics are discussed in detailed below.

1.3. IMPORTANCE OF STATISTICS

In ancient times statistics is used to collect the information on public affairs only, but in modern times it was gradually extended to all most all spheres of life. Statistics is more useful to all scientific experiments where observations are to be recorded and compared so that valid conclusions could be drawn. Various statistical methods are used to test the significance of any hypothesis related to Science, Technology, Engineering, Industry, research any many other fields. Statistics has wide number of applications in the fields of economics, commerce and social sciences. The subject statistics has been extensively used in biology, medical lab technology, meteorology and many research oriented fields.

In present world statistics become indispensable in all phases of human endeavour. It is hardly possible to enumerate a single department of human activity where statistics does not creep in. Statistics is more useful wherever a huge quantity of data needs significant analysis and where estimates are to be made for some uncertain events on the basis of available information.

1.4. SCOPE OF STATISTICS IN VARIOUS FIELDS

Now we discuss the scope of statistics in various fields in detail

In Economics

Many number of statistical methods and techniques re used to solve a variety of economic problems such as sales, prices, wages, production, income, expenditure, national income, analysis of time series, demand analysis etc. Various sampling techniques are used in collection of economic data most effectively. The statistical techniques of forecasting are very much useful for predicting various economic variables

in future. This is one of the very important technique of economic planning. Wide number of applications of mathematics and statistics in the subject economics led to the development of new disciplines like Econometrics and Economic Statistics.

In Planning

Statistics is indispensable in planning. Almost all organisations in the Government are using appropriate statistical methods and techniques for planning and formulating policy decisions. Efficient planning is necessary for all countries in the world for their economic and overall development for which the statistical data relating to production, consumption, birth rates, death rates, income, investments, etc. of important sectors is to be collected and to be analysed continuously from time to time.

In Business

Every decision in business is made with the help of statistical data and statistical methods. Almost all business activities depend on statistical techniques for studying, analysing and forecasting demand of the product and behaviour of the consumer mot effectively. The success of a businessman depends upon the accuracy and precision of forecasting the economic variables. If the predictions are wrong, then it leads to heavy loss. Hence the businessman is supposed to collect relevant statistical data (quantitative facts) about his future plan of production in advance. In business, various statistical methods are used for effective use of raw materials, projection of capital requirements, forecasting sales, estimating consumer's preferences various phases of business research and management.

In Industry

Statistics is immensely used to ensure the quality of the product in any large scale industry. It is invitable to examine the quality of the manufactured product of a company by using techniques of statistical quality control SQC. The statistical tools of SQC *viz.*, Control charts, Sampling inspection plans, etc., are more useful to examine whether the manufacturing product is confirming to specifications or not in industry.

In Mathematics

"Statistics is a branch of Applied Mathematics which specialises in data." **—Connor**

Statistics is intimately related and essentially depend on mathematics. In general statistical techniques are derived from theory of probability and applications advanced mathematics. Increasing role of mathematics in statistics has led to the development of new branch of statistics called Mathematical Statistics. Some of prominent mathematicians who contributed to statistics are Gauss, De-Moivre, James Bernoulli, Laplace, Pascal, R.A. Fisher, Bayes.

In Medical Science

Statistical tools are essential in medical science to observe the facts relating to causes and occurrence of diseases by collecting and analysing the related information. This results in the better usage of drugs and medicines. The efficacy of a manufactured drug or medicine or injection is tested by using the important statistical technique of test of significance.

In Astronomy

In astronomy, the Gaussian theory of Normal Law of errors for studying the movements of stars and planets is developed by using the principle of least squares, an important technique of statistics.

In Biology

Sir Francis Galton was first studied the relation between statistical methods and biological theories in his work 'Regression'. Galton was rightly expressed that without knowledge and application of statistical theory, it is impossible to discuss the national mortality. Now-a-days, the study of biology requires many of statistical techniques like Testing of hypothesis, Sampling theory, Design of experiments, etc., for selecting sample size, correctly choosing samples, analysing the biological experiments and testing the validity and significance of the results. This leads to a new subject called Biostatistics.

In Psychology and Education

Statistics has many number of applications in education and psychology. In psychology, statistics is mostly used to solve the problems like examining common problems most likely to affect the patients, to determine which kind of treatment is more effective etc. Statistics in education is used for analysing efficiency and progress of the students and also useful for improvement of teachers performance. The factor analysis, a technique of statistics is used to determine the reliability and validity of a test in psychology and education, this leads to existence of new subject called "Phychometry".

In Banking

Statistics play very important role in banking. Theory of estimation, statistical tests of significance, various sampling methods etc., are used in banking sector for implementation of their deposit schemes, increasing number of depositors, finding the best mode of operations in the banks, analysing credit-deposit ratio, bank clearings, etc.

In Management

Statistical methods are significantly used for planning business activities more effective by any management. The management personnel rely more on statistical methods and techniques for their development of business in various aspects like long-term planning, to develop new products, to streamline production, to maintain good relationship with customers, to reduce wastage, to assure the quality of the product, to maintain healthy atmosphere between the management and workers, to improve the sales of the product, observe when manufacturing process is to be shifted and many more. Thus utility of statistics is part and parcel of the management. hence, managerial sciences are mostly related with the subject statistics.

In State Affairs

In ancient days, statistics is used by rulers to assess their military and economic growth. Now the welfare of the state depends on more reliable collection of statistical data on social and economic conditions of the people from time to time. The collection and analysis of statistics on regular basis become an essential aid in planning and effective administration of the state.

1.5. LIMITATIONS OF STATISTICS

Even though statistics has wide number of applications in various fields, but it has some limitations which restricts its scope and utility. The following are some important limitations of statistics.

1. Statistics does not study individuals: Statistics deals with an aggregate of facts and does not give any specific importance to the individual items separately. For example, the individual figures of agricultural production or individual marks of a student are meaningless. The aggregates for a group of people only considered as statistics.

2. Statistics does not study qualitative phenomena: Statistics is only applied to the quantitative aspects of the problem, where as it cannot be expressed for qualitative characteristics like intelligence, honesty, beauty, poverty, etc. But some statistical techniques are applied for expressing quantitatively. For example beauty can be studied for a group by taking beautiful persons in the group.

3. Statistical laws are not exact: Statistical laws are only approximations and not exact since some probability laws are applied which are not certain. For example, if ten teams are participated in a game, statistics may be expressed that the winning chance of first team is probably 20%.

4. Statistical results are true only on an average: In general, statistical results reveal the average bahaviour or general trend. Some times the average or trend indicated by statistics is applied to individual cases, which may not be proper. For example, the average income of the group of families is ₹ 50,000, but if take this figure for a family, it may be only ₹ 5,000. Hence it is true only on an average.

5. Statistics does not reveal the entire story: Statistics provide only simple aggregate facts represent to a study but it does not reveal the entire story of the study. Some times statistics may not represent some parts of study group properly. Hence a good statistician should consider all relevant factors to interpret the good results.

6. Statistical relations do not necessarily bring out the cause and effect relationship between phenomena: Statistical data and its methods can only reveal the association between certain sets of data which may not be expressed the cause and effect.

7. Statistics is collected with a given purpose and cannot be indiscriminately applied to any situation: The statistical data collected for some purpose and utilising for another study without taking proper precautions lead to fallacious conclusions. Hence the data collected for a given purpose cannot be indiscriminately applied to any other situation. A proper care and thorough scrutiny is needed for using the secondary data.

8. Statistics is liable to be misused: The most important limitation of statistics is that it must be used by experts. Statistical methods are the most dangerous tools in the hands of inexperts. The use of statistical methods and techniques by untrained and inexperienced persons lead to fallacious conclusions.

Collection of Data

2.1. DATA COLLECTION

Introduction

Statistical investigation is based on systematic collection of data. The reliability of conclusions drawn from the sample data depends to a great extent on the quality of the data. The systematic planned and meaningful way of collecting information is known as collection of data. The methods of collection of data depends on various aspects such as objective, scope and nature of the problems under study. The data can be collected from two main sources which will be discussed below.

Primary and Secondary Data

Statistical data can be collected generally in the following two ways :

1. Primary data
2. Secondary data.

2.2. PRIMARY DATA (AU 2016, 2017)

Primary data are those statistical data which are collected for the first time are original in nature. Primary data are collected originally by the authorities who are required to collect them. The source from which primary data are collected is called primary source. Primary data is collected by field workers, investigators and enumerators. In India, the sources of primary data are the Census of India published by the Government, the Reserve Bank of India Bulletin published by RBI etc.

The primary data may be collected by any one of the methods :

(*i*) Direct personal interview
(*ii*) Indirect personal interview
(*iii*) Mailed questionnaire method
(*iv*) Information from local agents and correspondents.

1. Direct personal interview

According to this method, the investigator personally approaches each respondent and gathers first hand information. The reliability of data depends upon the training and attitude of the investigator and supporting attitude of the respondent.

Merits

1. In this method, the data obtained in original, accurate and exact.
2. This method leads to obtain more reliable information since investigator can clear the doubts and misunderstandings of the respondents.
3. Supplementary information can also be collected about the respondent's personal characteristics and environment. This helps in interpreting results.

Demerits

1. The method is not suitable when the number of respondents are very large.
2. The method is costly, time-consuming.
3. Skilled investigators are required to collect the data.
4. The success of the survey depends on personal qualities of investigator.

2. Indirect personal interview

This method is used when the respondents are reluctant to provide information directly. When the field of investigation is very large, the information about a large number of respondents can indirectly be obtained from one person who may lead the community or head of the organisation. It is generally used by C.B.I. and police for the collection of information.

Merits

1. If the area of investigation is very large, then this method is suitable.
2. Personally, if the respondent may not give the information to the investigator, then one may collect the information from the third person.

Demerits

1. In the absence of direct contact between investigator and respondent, important information may be lost.
2. The information given by the third person may be biased.
3. The information collected from the different persons may not be same and comparable.

3. Mail questionnaire method

In this method, a set of questions are prepared and sent by a mail to the respondents. The respondents are supposed to fill the schedule and mail them back to investigating agency. It is very useful when the respondents are educated and the area of investigation is very wide.

Merits

1. It is useful when the area is large.
2. It is useful when all the respondents are educated and aware.
3. The information collected by this method is free from the bias of investigators.

Demerits

1. It is applicable only to educated respondents.
2. Some of the respondents may not return the questionnaire.
3. Some of the respondents may send incomplete questionnaires.

4. 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 generally used by newspapers.

Merits

1. This method is quite ideal when information is needed from a wide area.
2. It is economic in terms of time and money.

Demerits

1. The information may not be reliable.
2. The data may be affected by the bias of the investigator.

2.3. SECONDARY DATA

(AU 2018)

The secondary data is one which is collected by some earlier agency but is used and analysed by any other for its own use. There are several sources of secondary data which are discussed below.

Sources of secondary data

The sources of secondary data can be broadly classified in two categories :

1. Published sources.
2. Unpublished sources.

1. Published sources

There are a number of national and international agencies which collect statistical data relating to business, trade, labour, prices, consumption, production, agriculture, industry, income, health, population and a number of socio-economic characteristics and publish their findings in statistical reports on a regular basis, *i.e.*, monthly, quarterly, annually etc. The following are some of important published sources of secondary data.

1. Government publications

The following are various government organisations which collect and publish statistical data on various fields.

1. Central Statistical Organisation (CSO).
2. National Sample Survey Organisation (NSSO).
3. Office of the Registrar General and Census Commissioner of India, New Delhi.
4. Directorate of Economics and Statistics.
5. Labour Bureau, Ministry of Labour.

2. International publications

Various foreign governments and international agencies like UNO, World Bank, International Monetary Fund (IMF) regularly publish reports on the data collected by them on various aspects.

3. Semi-official publications

Various local bodies such as District Boards, Municipal Corporations, Banking Organisations, etc., publish periodicals providing information about vital events, socio-economic characters, etc.

4. Private publications

The following private publications may also be used as secondary sources of the data.

1. Publications of professional bodies like ISI (Indian Statistical Institute), CSIR, ICAR, NCERT, etc.
2. Annual reports of private banks.
3. Information published in new papers, books, magazines, etc.
4. Reports prepared by research scholars of the university.

2. Unpublished sources

The information taken from the sources like diaries, letters, unpublished biographics and autobiographics, etc. are called unpublished sources. Unpublished data may also be available with scholars, trade associations and individuals.

Precautions for using secondary data

The following precautions must take before using the secondary data.

1. The organisation must check whether the data is reliable and suitable for the Statistical Survey.
2. The investigating team should check whether the data is sufficient for present investigation.

2.4. DESIGNING A QUESTIONNAIRE AND A SCHEDULE (AU 2016, 2017)

Questionnaire

Collection of the data through questionnaires is the most popular method for collecting primary data. A questionnaire is well prepared list of questions regarding the enquiry of the survey. In this method a questionnaire is sent to various respondents, they answer the questions and return the questionnaires. This method is extensively employed in various economic and business surveys.

Merits

1. This method is very economical when the universe is large and the area is wide.
2. The respondents may furnish the answers well which leads to more accurate results.
3. The data may be collected conveniently from the rural and remote areas.
4. The data is more reliable.

Demerits

1. Some times the respondents may not return the questionnaire.
2. Some questionnaires may not filled up properly, hence incompleteness causes less efficiency.
3. This method cannot be used for illiterates.
4. Once the questionnaires are sent to the respondents, then investigating agency cannot change or modify the questions.
5. The method is not flexible. In case of inadequate or incomplete answers it is difficult to obtain supplementary information.
6. This method is likely to be most time consuming, since the respondents can take their feasible and sufficient time to return the questionnaire.

Features of a good questionnaire

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

1. The size of the questionnaire should be as small as possible.
2. It should be simple, clear and unambiguous.
3. The questions should be brief.
4. The questions should be arranged in a logical order.
5. Questions may be dichotomous (*i.e.* yes or no type) or multiple choice and not of lengthy answers.
6. The case of sensitive and personal nature of questions should be avoided.
7. Questions should not be open-end. Preferably, appropriate answer choices should be given.

Schedule

In this method, a team of enumerators is selected and a special training will be given to them. Now the enumerators fill up the schedule. The difference between schedule and questionnaire methods is that schedules are being filled by enumerators whereas questionnaires are to be filled by respondents. A special team of enumerators is appointed for recording the answers given by respondents. The enumerators explain clearly the objective of survey, the definitions of basic concepts and rules to the respondents and record their responses. Census is usually conducted by using this method in the world.

Merits

1. This method can be for illiterate population.
2. The data collected by this method is more accurate and reliable.
3. In this method, information given by respondents can be checked on the spot by cross questioning.

4. Non-response in this method is very little.
5. The identity of the respondent is known in this method whereas it is not clear in the case of mail questionnaire method.

Demerits

1. It is the most expensive method among all methods of collecting primary data.
2. It is more time consuming.
3. The success of the method mainly depends on the efficiency and skill of the enumerators.
4. The success of the method completely relies on preparation of schedule.

Classification and Tabulation of Data

3.1. CLASSIFICATION OF DATA

Introduction

Already we have learnt about the various methods of collecting primary and secondary data. In any statistical investigation, if the data is collected we cannot interpret and draw the conclusions. To draw meaningful conclusions on the basis of collected data, it is essential to present in summarised and simple form. Classification of data means segregation of various class categories or group heads in which the whole data shall be distributed.

Definition

Classification of data means arranging the collected data into different classes or groups on the basis of their similarities and which have close resemblance. This helps in separating the characters have similarities and dissimilarities. A group or a class has to be determined on the basis of the nature of data and the purpose for which it is going to be used. For example, the data on households can be classified on the basis of age, religion, education, income, occupation, etc.

Objectives of classification of data

The objectives of classification of data may be briefly summarised as follows :

1. It helps in presenting the original data in a concise and simple form.
2. Classification enables comparison between the variables since the data is classified on the basis of similarities and dissimilarities.
3. The study of the relationship between the variables is possible through the classification.

4. It provides a basis for tabulation and analysis of the data.
5. It is a process of presenting raw data in a systematic manner enabling to draw meaningful conclusions.

3.2. METHODS OF CLASSIFICATION

There are four methods of classifications :

1. Temporal (or) Chronological classification
2. Spatial (or) Geographical classification
3. Classification according to attributes (Qualitative)
4. Classification according to variables (Quantitative)

1. Temporal (or) chronological classification

If the data is classified with respect to different time periods is known as temporal (or) chronological classification. A time series data is best example of this classification.

Example. The profits of a company in different years are given below [in accordance with time (year)].

Year	*Profits of a company (in lakhs of ₹)*
2015	235
2016	261
2017	250
2018	282
2019	274
2020	320
2021	342

2. Spatial (or) Geographical Classification

If the data is classified on the basis of geographical region is known as spatial or geographical classification.

Example. The population in different states of a country is given below.

States	*Population (in crores)*
Andhra Pradesh	8.46
Uttar Pradesh	19.98
Maharashtra	11.24
Bihar	10.41
West Bengal	9.13

3. Classification according to attributes (Qualitative)

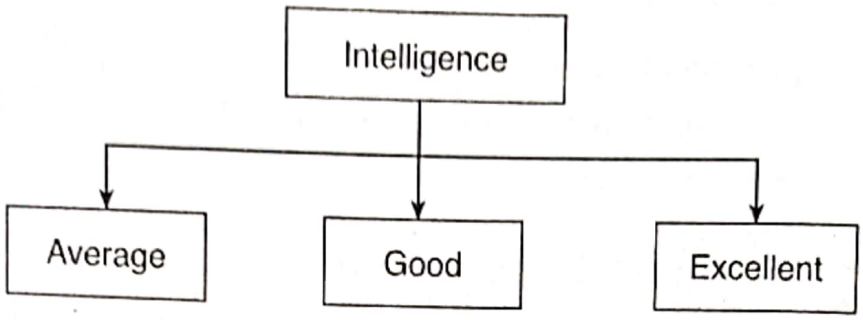
An attribute is a qualitative characteristic which can be measured quantitatively only as presence or absence of the characteristic. For example, intelligence, beauty, sex, etc. One cannot quantify these characteristics. In this case classification is to be done by differentiating presence or absence of the attribute. For example, the sex cannot be

measured quantitatively, then we can classify into how many male persons and female persons are present.

There are two types of classification of attributes.

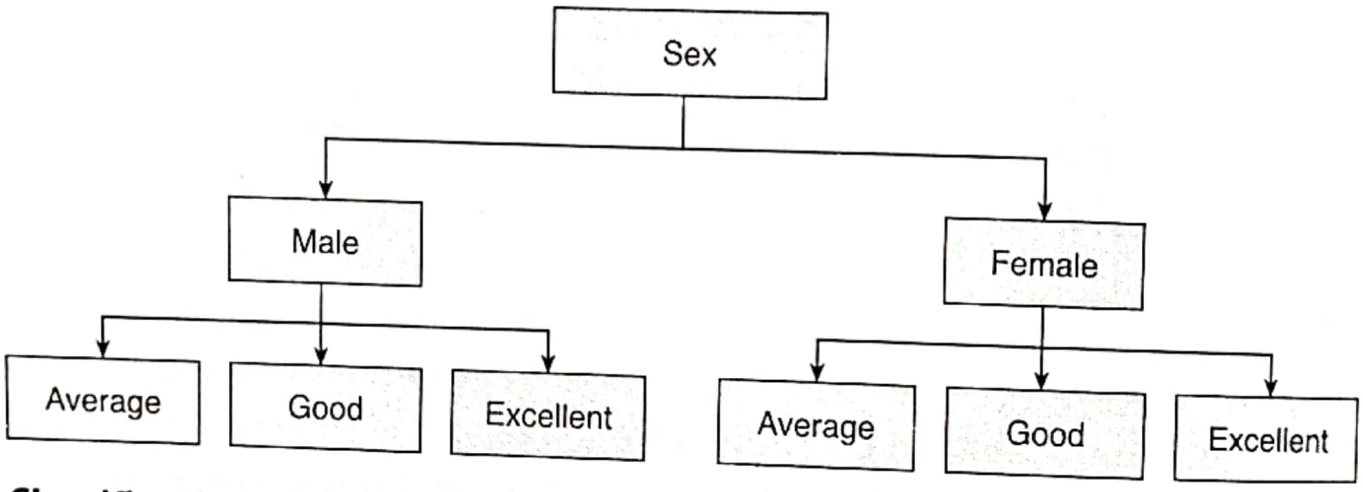
1. Simple Classification. In simple classification, the data is classified on the basis of only one attribute.

For example, intelligence is one attribute, then the classification may be represented in the following diagram.



2. Manifold Classification. In this classification, the data is classified on the basis of more than one attribute.

For example, two attributes intelligence and sex of a group of persons may be classified as shown in the following diagram.



4. Classification according to variables (Quantitative)

Variables are quantifiable characteristics of data which can be measured numerically. Examples are age, height, weight, marks, wages, etc. In this form of classification, the data is shown in the form of a frequency distribution. Based on the number of variables used, there are three categories of frequency distribution.

1. Unvariate frequency distribution.
2. Bivariate frequency distribution.
3. Multivariate frequency distribution.

1. Univariate frequency distribution

The frequency distribution is formed with one variable is called a univariate frequency distribution.

For example, the students in a class may be classified on the basis of variable marks obtained by them. If there are 30 students have obtained the marks between 50 to 100. It can be classified in the following way.

First divide the marks (variable) into 5 classes of equal width (need not be similar), say, then class intervals are 50—60, 60—70, 70—80, 80—90 and 90—100.

Secondly define a class of 50-60 contains the students secured ≥ 50 but < 60 *i.e.* 50 included and 60 not included, where as 60 included in the next class interval etc.

Thirdly, use tally mark to put the student in the class interval and then count the number of students in the class interval called frequency.

If the marks of 30 students are :

52	72	58	55	75	53
54	64	70	74	56	51
50	80	85	92	68	96
76	95	65	78	98	72
88	77	99	62	73	79

Marks as Class interval	*Tally marks*	*No. of Students (frequency)*
50—60	𝍸 III	8
60—70	IIII	4
70—80	𝍸 𝍸	10
80—90	III	3
90—100	𝍸	5

2. Bivariate frequency distribution

The frequency distribution is formed with two variables is called bivariate frequency distribution. For example, if a frequency distribution shows two variables; marks in statistics (first variable) in different age groups (second variable), then it can be treated as bivariate frequency distribution.

Firstly divide the marks in statistics into the class intervals 0—10, 10—20, etc. and age group in to 18—20, 20—22, etc.

Secondly define class interval 0—10 consists ≥ 0 and < 10 (includes 0 and not includes 10) age groups 18—20 consists 18 years and above, less than 20 years (18 years to be included, 20 not included) and so on. This can be represented as for some number of students (say) as follows :

Marks in Statistics	*Age Group of Students*			
	18—20	*20—22*	*22—24*	*24—26*
0—10	II	I	III	II
10—20	—	II	—	—
20—30	—	III	𝍸	III
30—40	III	III	𝍸 I	III
40—50	II	II	II	IIII

(Represented by tally marks).

3. Multivariate frequency distribution

The frequency distribution is formed with more than two variables is called multivariate frequency distribution.

For example, the students in a class may be classified on the basis of marks, age and sex.

First divide into class intervals or groupings.

Secondly, a clear cut definitions *i.e.* marks in a subject of the students 0—10 for which ≥ 0 and < 10 similarly age 18—20 ≥ 18 and < 20, sex as grouping the males as one group and females as another group. It can be shown for some number of students (say) as follows :

Marks	*Age Group*					
	18—20		*20—22*		*22—24*	
	M	*F*	*M*	*F*	*M*	*F*
0—10	—	I	II	—	III	II
10—20	II	—	I	—	II	III
20—30	III	IIII	II	III	𝍸	II
30—40	III	𝍸	IIII	𝍸 I	𝍸 II	𝍸 𝍸

(Represented by tally marks)

3.3. PRESENTATION OF DATA

Once the data is collected, the next important part is presentation of such collected data. There are many number of ways of presenting data. A statistical data may be presented in paragraphs, diagrams, graphs, tables etc. in many ways. But in general, the data in Statistics can be presented in the following three ways.

1. Textual presentation 2. Tabular presentation 3. Graphical presentation

The first two presentations textual and tabular discussed here and graphical presentation is discussed in unit-2.

3.4. TEXTUAL PRESENTATION OF DATA

If the data is presented in a paragraph *i.e.,* simply mentioned as mere text, such presentation of data is known as Textual presentation of data. This is used when the data is small. If the data is large, this presentation is time consuming and most difficult.

Examples:

1. 200 people died in a train accident in a city.
2. Only 48 persons has been awarded Bharat Ratna award.

3.5. TABULATION OF DATA OR TABULAR PRESENTATION OF DATA

Definition

The tabulation of data may be defined as the logical and systematic organisation of statistical data in rows and columns. It is designed to simplify presentation and facilitate analysis.

Objectives of tabulation of data

The main objectives of tabulation are :

1. Systematic presentation of data. Generally the collected data is in fragmented form. The mass data is presented in concise and simple manner by means of statistical tables. Thus, tabulation helps in presenting the data in a systematic way.

2. Facilities comparison of data. Comparison of data in the raw form is difficult. The presentation of complete and unorganised data in the form of tables facilitates the comparison of the various aspects of the data.

3. Identification of the desired values. In tabulation, the data presented in an organised manner in rows and columns. Hence the desired values can be identified without much difficulty.

4. Provides a basis for analysis. A systematic presentation of the data in tabular form provides a basis for analysing data statistically. For example in the calculation of variances, means, ANOVA etc.

5. Exhibits trend of data. By presenting the data in tabular form, we can identify some of the pattern of change in the variable under study.

6. Economy of space. The economy of space is achieved without sacrificing the quality and usefulness of the data.

7. Detection of errors. It is only after tabulation that some vital omissions are detected.

8. For reference. The tabulated information is convenient to refer and to identify at any future time.

Distinguish between classification and tabulation

In classification, the data is divided on the basis of similarity and resemblance, whereas tabulation is the process of recording the classified facts in rows and columns. Tabulation begins where classification ends. In fact, classification provides a basis for tabular presentation. After classifying the data into various classes, they should be shown in the tabular form.

Components (or) Essential Parts of a Table

The following are main components of preparing a good table.

1. Table Number: Numbering in the table helps for easy identification and reference of the contents. The number may be placed at the top of the table either in the centre above the title or in the side of the title.

2. Title: A title is to be placed at the top of the table. A good title clearly explain the nature of the data contained. A title should always prominently lettered to easily

understand the content. In general, a title should state what, where and how a classified data presented. The title also should include the time period to which it relates. If necessary, a sub-title should also to be given and arranged below to the title.

3. Captions: Captions are simply headings or designations of vertical columns in the table. A caption may be either a main heading or a sub-heading and also given as a footnote. The caption should be brief, concise and self-explanatory. The captions are represented in the middle of the columns.

4. Stubs: The stubs are headings represented in the horizontal rows in the table. Usually these are represented by long descriptions with relevant importance. A single stub may be given for two or more columns with some headings.

5. Body: The body of the table contains the numerical information. The arrangement of numerical values is in accordance with the descriptions of rows and columns. The body is arranged generally from left to right in the rows and from top to bottom in the columns.

6. Prefatory notes and Footnotes: If a note is given below the title is known as prefator note. Actually this note is a part of the title but represented for the clarity of the title. If any kind of note is given at the foot of the table is known as footnote.

Footnotes should be given if any kind of explanation is necessary for the content or the values of the data. A footnote is denoted by the symbols +, @, *, $ etc. If the figures are not available, then it is indicated by "N.A." and not at all considerable values, then which are indicated by '—'.

7. Source Note: If the source is specified from which the information has been obtained is known as source note. It is usually presented at end of the table either left hand side or right hand side below the table. It includes the name of the author, title, volume, page, publisher's name and year of publication.

Format of the Table

TITLE

(Prefatory Note)

Stub	Caption						Total (Rows)
	Sub-head		Sub-head		Sub-head		
	Column head	Column head	Column head	Column head	Column head	Column head	
	Body						
Total (Columns)							

Source Note :

Footnote :

Kinds of tables

There are different types of statistical tables depending on the objectives and uses of the data. They can be classified in the following ways :

1. Simple and complex tables
2. General purpose or reference tables
3. Special purpose or summary tables.

1. Simple and complex tables

These types of tables are prepared to show the important characteristics of the collected data.

(*a*) Simple tables

In a simple table the data are classified with respect to a single characteristic and it is also known as one-way table.

Example. The data of students studied in different years of a college represented in a simple table as follows.

Year	*No. of Students*
1998-1999	1500
1999-2000	1800
2000-2001	1900
2001-2002	2000
2002-2003	2500

(*b*) Complex tables

If the data are grouped into different classes with respect to two or more characteristics simultaneously, then the table is known as complex table.

If the data are classified with two characteristics, the table is called two-way, the data is classified with three characteristics, then the table is called three way table.

Example. The number of students in a college according to sex and marital status during 1985-86 to 1988-89 is a complex table, it can be shown as follows.

Year	*No. of Students*			
	Male		*Female*	
	Married	*Unmarried*	*Married*	*Unmarried*
1985-86	35	40	42	46
1986-87	56	20	38	30
1987-88	30	82	46	54
1988-99	61	28	92	68

2. General purpose or reference tables

This type of tables are prepared to store information and they contain wide range of information relating to a specified subject. Such tables are complex tables and are

generally found as appendices to various reports. These tables should be prepared in a systematic manner.

Examples. The tables appended to the census reports, CSO publications, RBI bulletins etc.

3. Special purpose or summary tables

These tables are constructed with a specific point and are very useful for the purpose of comparison. These tables are also called text tables. Generally these tables indicate rates, percentages, averages, etc.

Example. The number sales in two companies of same product are displayed in the following table is an example of summary table.

Year	*Average Sales in Company A (in lakhs of Rs.)*	*Average Sales in Company B (in lakhs of Rs.)*
2008	152	148
2009	158	150
2010	170	165
2011	182	190
2012	196	224

Diagrammatic and Graphical Representation of Data

3.1. DIAGRAMMATIC AND GRAPHICAL REPRESENTATION OF DATA

The easiest method of representing the statistical data is the use of diagrams and graphs. The diagrams and graphs are simple geometrical figures like lines, squares, rectangles, bars, circles, cubes, pictures, charts, maps, etc. The diagrammatic and graphical representations are more attractive, simple to understand and readily comprehensible. Even a layman can easily understand the diagrammatic representation whereas graphical representation needs some knowledge in the subject.

3.2. DIAGRAMMATIC REPRESENTATION OF DATA

Diagrammatic representation is a technique of presenting a numerical data through simple diagrams like bars, circles, maps, pictorials, cartograms etc. This kind representation is most attractive and appealing way to represent statistical data. Even a layman can easily understand this representation. This is only a representation brings out hidden facts and relationships.

There are many types of diagrammatic representation, some of the important methods are :

1. Bar diagram
2. Pie diagram
3. Pictograms and Cartograms.

3.3. BAR DIAGRAM

A bar diagram is diagrammatic representation of a numerical data presents with rectangular bars with heights are proportional to values of the variable under a statistical study. The bars can be plotted vertically or horizontally. The heights or the

lengths of the bars indicate value of the variable of some population characteristic. In the construction of bar diagram width of the bar has no relationship with the measurement and it is only to make the diagram look more elegant and attractive. The bar diagram is useful to compare the different groups of data of some population characteristic. It is more useful to observe changes in data *i.e.*, trends of the variable over time.

The following are different types of bar diagrams :

1. Simple bar diagrams
2. Multiple bar diagrams
3. Subdivided bar diagrams
4. Percentage bar diagrams

3.4. CONSTRUCTION OF BAR DIAGRAM

To construct the bar diagram, first we represent different time periods or places or items etc. on X-axis. The values of the variable are represented as vertical or horizontal bars on Y-axis against different time periods, places etc. This diagrammatic representation of bars is called bar diagram.

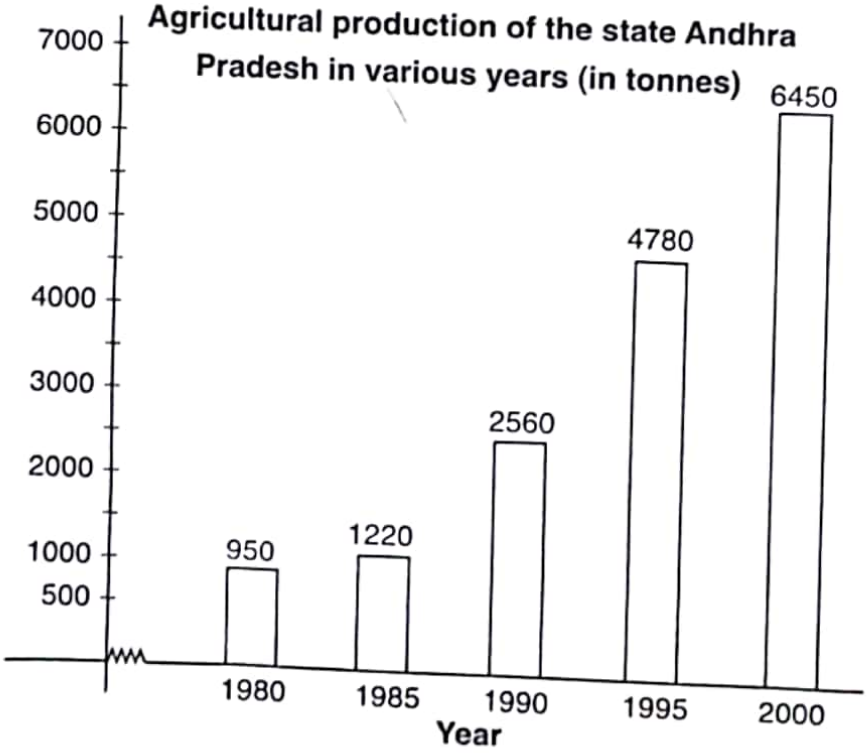
3.5. SIMPLE BAR DIAGRAM

A simple bar diagram represents the data of only one variable. The variable may be sales, profits, population, production, income, agricultural production etc. The construction of bar diagram with one variable is known as simple bar diagram.

PROBLEM 1. The following data relates to agricultural production in tonnes of the state Andhra Pradesh in tonnes for various years. Represent the data by suitable bar diagram.

Year	*1980*	*1985*	*1990*	*1995*	*2000*
Agricultural Production (in tonnes)	950	1220	2560	4780	6450

SOLUTION



3.6. MULTIPLE BAR DIAGRAM

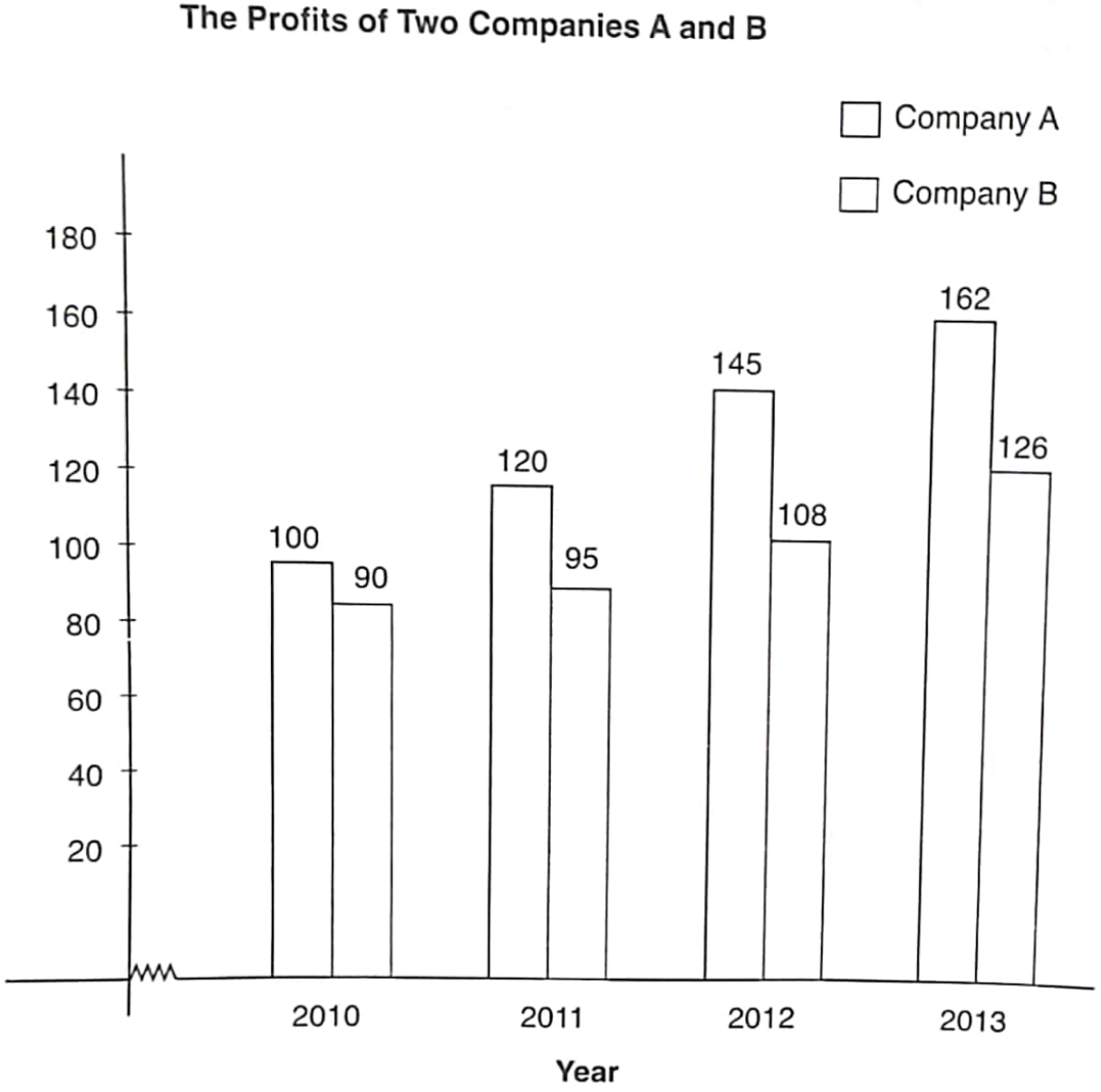
If the comparison is made between two or more sets of related data, then multiple bar diagram is a good choice of representation. The multiple bar diagram represents the data of two or more variables in the study. The construction of bar diagram with two or more variables is known as multiple bar diagram. In multiple bar diagram, different bars for a period are placed together leaving some gap between each set of bars.

PROBLEM 1. The profits of two companies A and B are given below in the following table :

Year	*Profits in ('000 Rupees)*	
	Company A	*Company B*
2010	100	90
2011	120	95
2012	145	108
2013	162	126

Represent the data with suitable bar diagram.

SOLUTION. The given data can be suitably represented by a multiple bar diagram., The diagram is shown below :



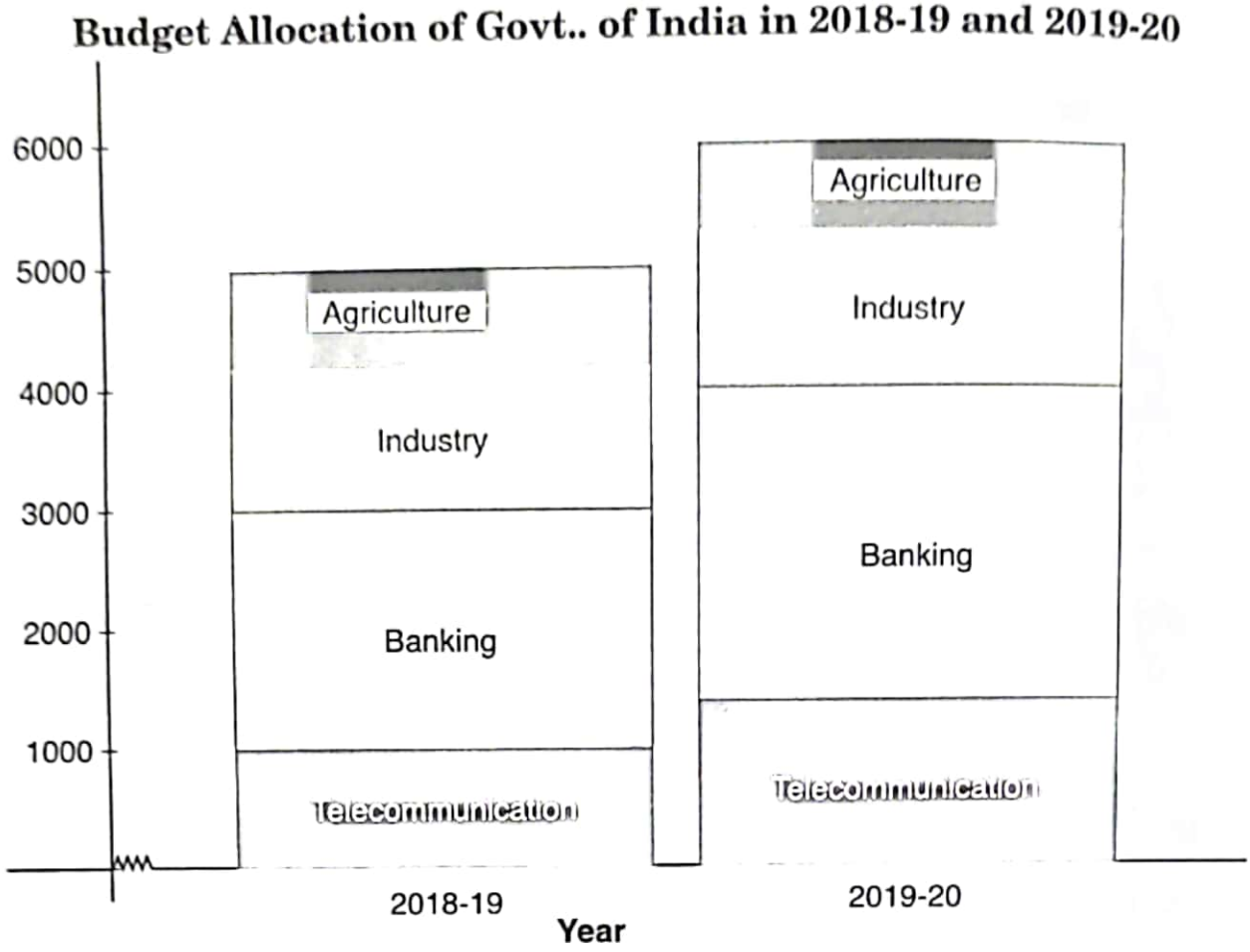
3.7. SUB-DIVIDED BAR DIAGRAM

A sub-divided bar diagram is used if the total magnitude of the given variable is divided into various parts or components or sub-classes. First of all a simple bar diagram is drawn for the given data by considering the total magnitude. Then it is divided into various segments, each segment representing a given component of the total. If all the components or sub-classes of a given variable are represented in a single bar for each class, then such kind of representation of bar diagram is known as **Sub-divided bar diagram**.

PROBLEM 1. Draw sub-divided bar diagram to the following data relates to a special allocation of budget of Govt. of India in crores of rupees for various categories in the years 2018-19 and 2019-20.

Category	*Budget allocation (in crores of rupees)*	
	2018-19	*2019-20*
Agriculture	800	700
Industry	1200	1300
Banking	2000	2600
Telecommunication	1000	1400

SOLUTION



3.8. PERCENTAGE BAR DIAGRAM

If sub-divided bar diagram is represented graphically in the form of percentages, then such bar diagram is known as percentage bar diagram. The most important purpose of percentage bar diagrams is to compare relative changes of various categories in the data. The total for each bar is taken as 100. The vale of each component is expressed as percentage of respective totals. Hence all the bars will be of the same height in percentage bar diagram. These diagrams are more useful for comparing two or more sets of data.

PROBLEM 1. The expenditure details of two families are given below in the following table. Draw percentage bar diagram.

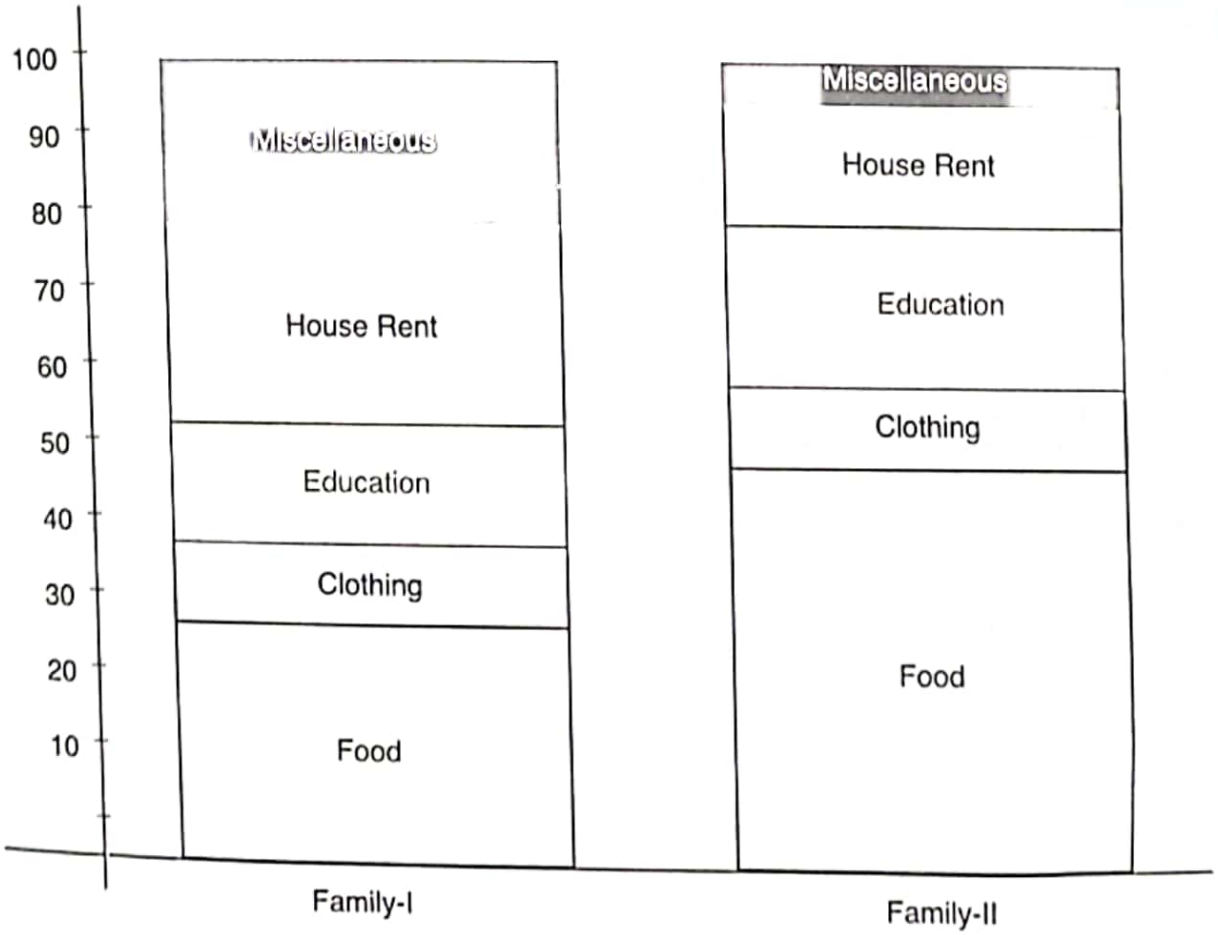
Category	*Expenditure (in Rs.)*	
	Family I	*Family II*
1. Food	3,000	6,000
2. Clothing	1,000	1,200
3. Education	1,500	2,400
4. House rent	2,500	1,800
5. Miscellaneous	2,000	600

SOLUTION. First we have to convert expenditure of various categories into percentages of the total expenditure, which are converted in the following table :

Category	Family I			Family II		
	Expenditure	Percent of expenditure to Total	Cumulative percentage	Expenditure	Percent of expenditure to Total	Cumulative percentage
1. Food	3,000	30	30	6,000	50	50
2. Clothing	1,000	10	40	1,200	10	60
3. Education	1,500	15	55	2,400	20	80
4. House rent	2,500	25	80	1,800	15	95
5. Miscellaneous	2,000	20	100	600	5	100
Total	10,000	100		12,000	100	

SOLUTION

Expenditure Details of two families



3.9. PIE DIAGRAM

Pie diagram is a diagrammatic representation of different categories of a variable in a circular form. The representation of all categories in different sectors of the circle in accordance with degree of the category is called pie diagram.

3.10. CONSTRUCTION OF PIE DIAGRAM

First of all we express the values of different categories of the variable as a percentage of total. Then we convert percentages of categories into degrees as represented as the total of 360°,. The degree of any category can be easily obtained by using the following formula :

$$\text{Degree of a category} = \frac{\text{Value of the category}}{\text{Total value}} \times 360^\circ$$

Now represent all the categories in different sectors of the circle in accordance with the degree of respective categories. This kind of diagrammatic representation in circular form is known as pie diagram.

PROBLEM 1. The revenue sources of Government of India during a particular period is given below :

Source	*Revenue (in thousands of crores)*
Customs	300
Excise	1,380
Income Tax	2,400
Corporate Tax	1,920
Miscellaneous	1,200

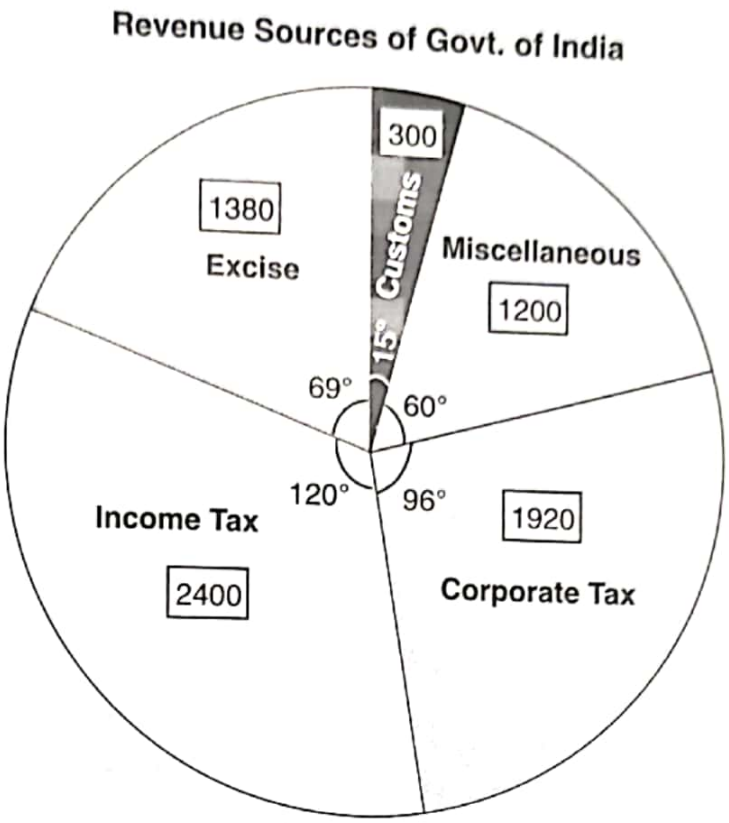
Draw pie chart to represent this data.

SOLUTION

First we have to express degrees of each source to a total 360°. This is calculated in the following table.

Source	*Revenue*	*Degrees of the Source*
Customs	300	$\frac{30}{7,200} \times 360^\circ = 150^\circ$
Excise	1380	$\frac{1,380}{7,200} \times 360^\circ = 69^\circ$
Income Tax	2,400	$\frac{2,400}{7,200} \times 360^\circ = 120^\circ$
Corporate Tax	1,920	$\frac{1,920}{7,200} \times 360^\circ = 96^\circ$
Miscellaneous	1,200	$\frac{1,200}{7,200} \times 360^\circ = 60^\circ$
Total	7,200	360°

$$\text{Degree of the source} = \frac{\text{Value of the source}}{\text{Total value}} \times 360^\circ$$



3.11. GRAPHICAL REPRESENTATION F DATA

Graphical representation is one of the most useful way of analysing numerical data by means of a diagram. In the graphical representation, statistical data is represented in the form of lines or curves and it enables us in studying cause and effect relationship between the variables. This representation is more useful in studying both time series analysis and frequency distribution.

The following are some of important types of graphical representation.

1. Histogram
2. Frequency polygon
3. O-give curve

3.12. HISTOGRAM

Histogram is drawn only for a continuous frequency distribution. If the given grouped frequency distribution is not continuous, then first it is to be converted into continuous distribution. To draw the histogram for a continuous grouped frequency distribution, first we mark class intervals on a suitable scale on X-axis. Then on each class interval draw rectangles with heights proportional to the frequency of corresponding class interval so that the area of rectangle is proportional to the frequency of the class on Y-axis. If class are of unequal width, then the height of the rectangle will be proportional to the ratio of the frequencies to the width of the classes. Now the diagram obtained with continuous rectangles is called histogram of the given distribution.

If the given frequency distribution is not continuous, then we use the following formula to convert into a continuous grouped frequency distribution.

Upper class boundary of new class Interval

$$= \text{Upper class limit of old class interval} + \frac{d}{2}$$

Lower class boundary of new class interval

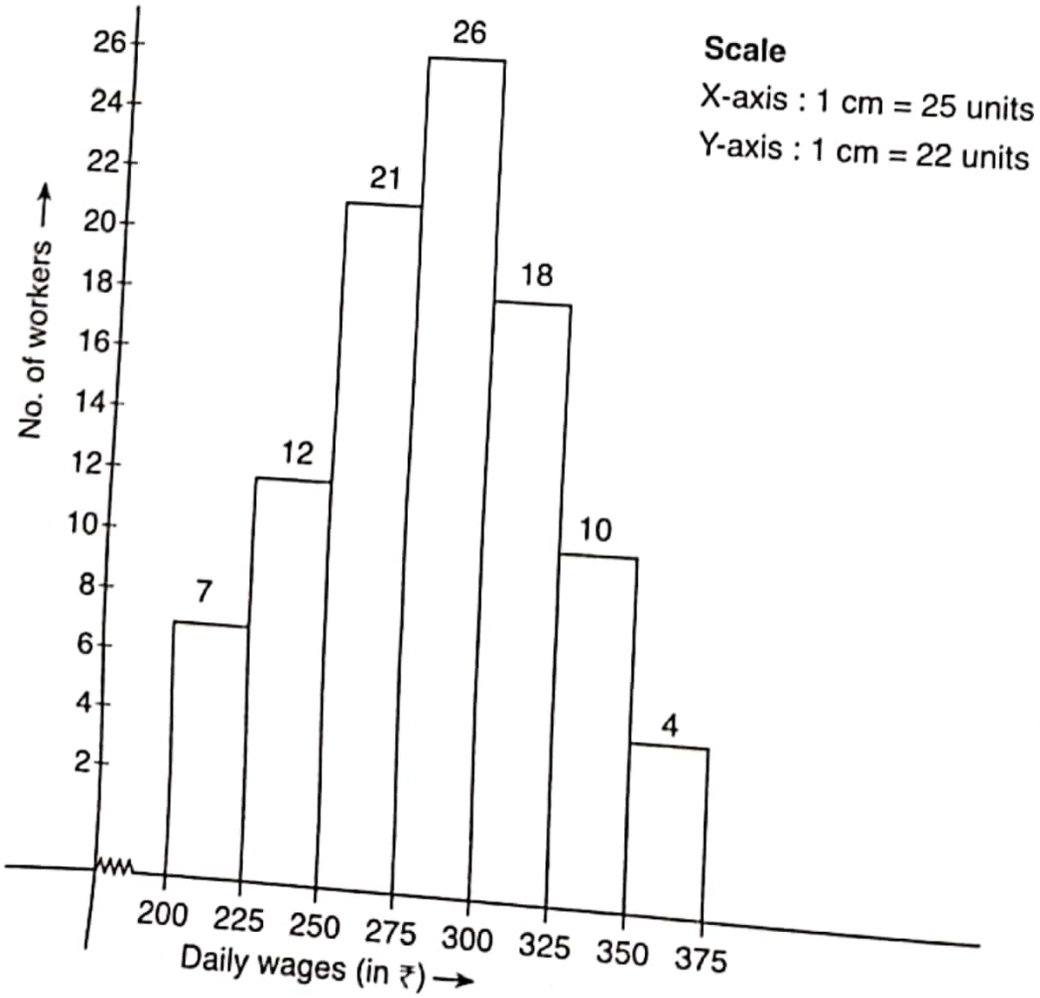
$$= \text{Lower class limit of old class interval} - \frac{d}{2}$$

where 'd' is gap between upper limit of any class and lower limit of the succeeding class.

PROBLEM 1. The daily wages of workers in a locality are given below. Construct histogram to represent the data.

Daily wage	*No. of workers*
200-225	7
225-250	12
250-275	21
275-300	26
300-325	18
325-350	10
350-375	4

SOLUTION. Since the data is a continuous grouped frequency distribution directly we can draw a histogram by plotting daily wages on X-axis and number of workers on Y-axis representing heights of the rectangles with number of workers.



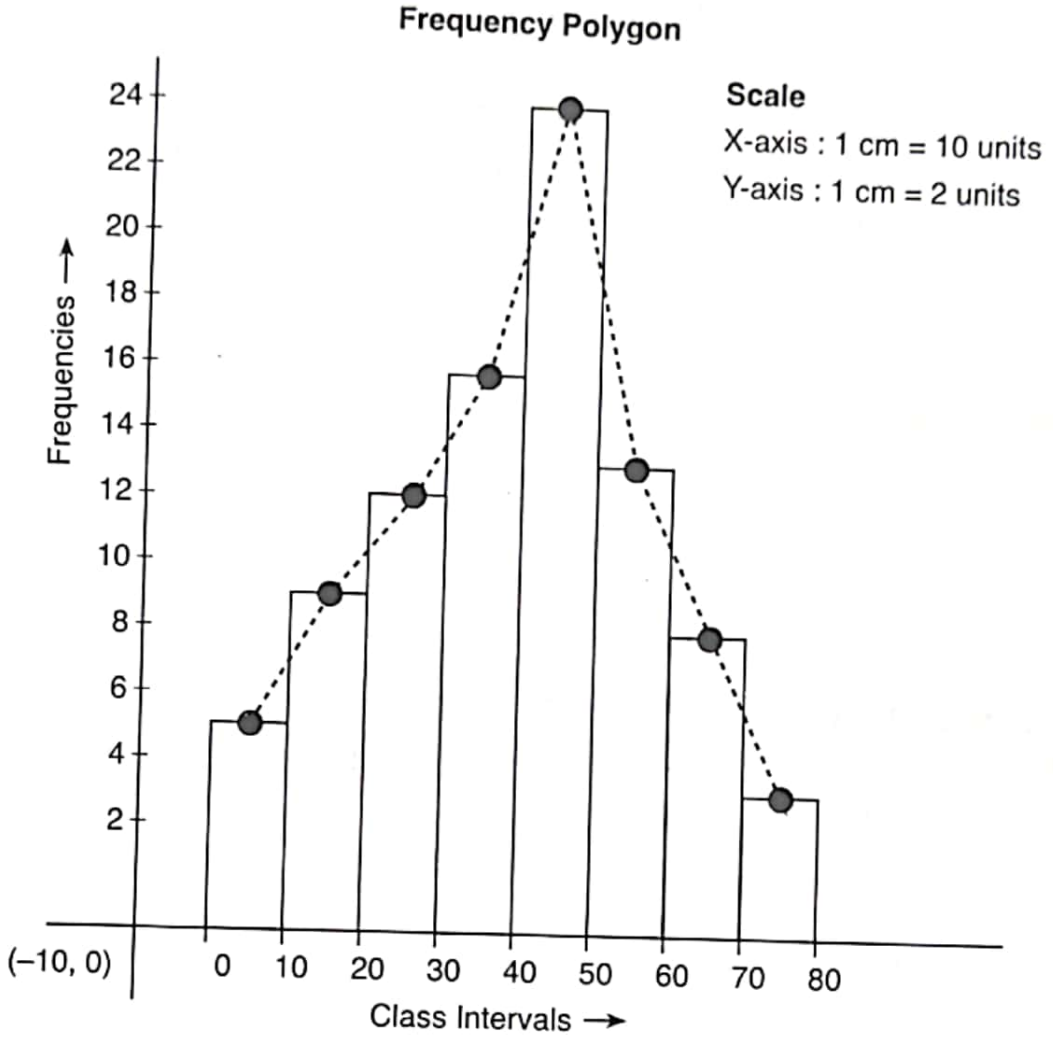
3.13. FREQUENCY POLYGON

To construct a frequency polygon, first we draw a histogram. For a continuous grouped frequency distribution, the histogram is drawn by considering class intervals on X-axis and corresponding frequencies on Y-axis. Now consider the values of the variable at mid values of the class intervals. For equal class intervals the frequency polygon can be obtained by joining the middle points of the upper sides of the adjacent rectangles of the histogram by means straight lines. If the class intervals are of small width, then the frequency polygon can be approximated by a smooth curve.

PROBLEM 1. Draw frequency polygon to the following data :

Class Interval	*Frequency*
0-10	5
10-20	9
20-30	12
30-40	16
40-50	24
50-60	15
60-70	8
70-80	3

SOLUTION. The given data is a continuous grouped frequency distribution. Therefore first we draw histogram and then by joining middle points of the upper sides we can obtain a frequency polygon.



3.14. OGIVE CURVE

Ogive curve is a graphical representation of cumulative frequencies of a distribution. In the construction of ogive curve, consider the class intervals on X-axis and plot corresponding cumulative frequencies on Y-axis against upper limit or lower limits of the respective class interval. The curve obtained by joining these points by means of drawing smooth free hand curve is called ogive curve or cumulative frequency curve.

There are two types of ogive curves, viz.,

1. less than ogive
2. More than ogive

Less than Ogive

Plot less than cumulative frequencies against the upper limits of the corresponding classes. The smooth free hand curve obtained by joining these points is called less than ogive curve.

More than Ogive

Plot more than cumulative frequencies against the lower limits of corresponding classes. The smooth free hand curve obtained by joining these points is called more than ogive curve.

3.15. OBTAINING MEDIAN THROUGH OGIVE CURVE

First draw less than ogive and more than ogive curves for the given frequency distribution. to locate median graphically, mark a point corresponding to $\frac{N}{2}$ along Y-axis. From this point draw a line parallel to X-axis which is meeting ogive. Draw a line perpendicular to X-axis from this point, which gives the value of median.

Alternative Procedure

Draw less than ogive and more than ogive curves for the given frequency distribution. We can observe these two ogive curves intersect at a particular point.

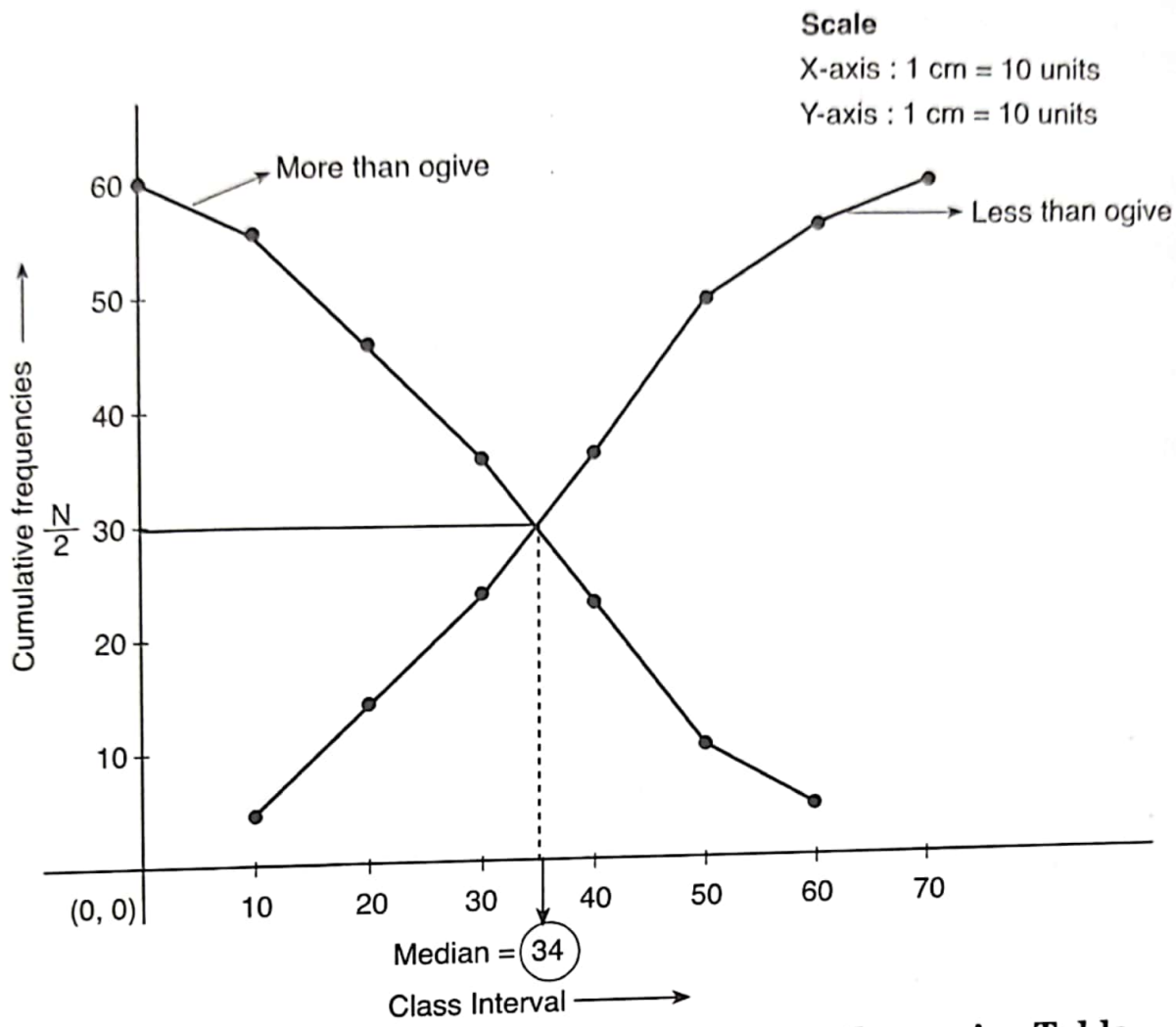
From the point of intersection of these two ogive curves, draw a perpendicular line to the X-axis. The value of the point where this perpendicular line meets the X-axis gives the median.

PROBLEM 1. Draw ogive curve to the following data and obtain median from the ogive curve.

Class Interval	*Frequency*
0-10	4
10-20	8
20-30	11
30-40	15
40-50	12
50-60	6
60-70	4

SOLUTION

C.I.	*f*	*Less than c.f.*	*More than c.f.*
0-10	4	4	60
10-20	8	12	56
20-30	11	23	48
30-40	15	38	37
40-50	12	50	22
50-60	6	56	10
60-70	4	60	4
	N = 60		



Less than ogive Table

Upper limits of C.I.	*Less than c.f.*
10	4
20	12
30	23
40	38
50	50
60	56
70	60

More than ogive Table

Lower limits of C.I.	*More than c.f.*
0	60
10	56
20	48
30	37
40	22
50	10
60	4

$\therefore$ Median obtained through the ogive curve is 34 approximately.

Scales of Measurement

4.1. SCALES OF MEASUREMENT

In general, the statistical data cannot be obtained or measured in similar form or in same pattern. The data obtained in statistical enquiry in various forms are called different scales of measurement. There are four scales of measurement which are used in the collection of statistical data of any statistical enquiry.

1. Nominal Scale
2. Ordinal scale
3. Interval Scale
4. Ratio Scale

4.2. NOMINAL SCALE

Nominal scale is used for labeling the variables into different classifications and does not involve a quantitative value. This scale is used for collecting the qualitative data. In this scale, numbers associated with the variables serve as tags or labels to classify or identify the objects in the statistical study. The order of the variable is not important in this scale. For example, the variable of the study is sex, then it can be given two labels, 'M' for male, 'F' for female (or Male-1, Female-2). For another example, the place of residence, it can be given that '1' for urban, '2' for semi-urban, '3' for rural etc.

4.3. ORDINAL SCALE

the ordinal scale describes order of the variables in a statistical investigation. This scale is used to represent the qualitative or categorical data. Ordinal represents order. In the nominal scale, the variables can be categorised only with labels or tags, whereas in the ordinal scale, it can be grouped and ranked.

Examples :

1. Ratings of different companies.

   1 — for 10 pints
   2 — between 7-9 pints
   3 — between 4-6 points
   4 — between 1-3 points

2. Satisfaction of services in a hotel

Not satisfied	—	1
Satisfied	—	2
Fully satisfied	—	3

3. Performance of the students

Below average	—	1
Average	—	2
Above Average	—	3
Good	—	4
Excellent	—	5

4.4. INTERVAL SCALE

The interval scale contains properties of nominal and ordinal scales of measurement. Interval indicates difference between two objects of variable. In addition to the nominal and ordinal scales, the interval scale measures the difference between the variables. the difference can be added or subtracted from each other, but cannot be multiplied or divided. In interval scale, zero is not absolute, is arbitrary means it represents presence of the variable (example time). the interval scale is mostly used together information or feedback based on satisfaction levels, agreement etc. Interval scale is used for question type surveys with choice of options in numbers so that certain statistical methods are applied. The measures of central tendency like mean, median and mode can be calculated in this scale.

Examples :

1. In banking system, if a question asked "I do not like queue system", the choice of answers provided are :

Strongly disagree	—	1
Disagree	—	2
Neutral	—	3
Agree	—	4
Strongly Agree	—	5

This kind of scale is called attitude scale.

2. The temperature scales are also best examples of interval scale.

4.5. RATIO SCALE

The ratio scale is defined as one of the best variable measurement scale contains all properties of nominal, ordinal and interval scales of measurement along with the unique feature absolute zero. In the ratio scale, zero means total absence of variables. The ratio scale does not have negative values. In this scale, the variables can be orderly added, subtracted and also multiplied and divided. The measures of central tendency like mean, median, mode, harmonic mean, geometric mean and dispersion measures, coefficient of variation also can be calculated for the data obtained in ratio scale.

Frequency Distribution and Its Types

5.1. FREQUENCY DISTRIBUTION

A frequency distribution is an arrangement of number of observations with similar related values of a variable in the order of magnitude with the corresponding frequencies in the form of a statistical table. According to Croxton and Cowdon "Frequency distribution is a statistical table which shows the set of all distinct values of the variable arranged in order of magnitude, either individually or in groups, with their corresponding frequencies side by side."

Examples: 1. The following frequency distribution represents the profits (in crores of rupees) of 20 companies in India.

Profit of a company in crores of ₹	*No. of companies (frequency)*
50	2
60	3
70	5
80	6
90	3
100	1

2. The frequency distribution of marks of 100 students in an examination is given below.

Marks in the examination (C.I.)	*No. of students (frequency)*
30-40	4
40-50	11
50-60	20

60-70	24
70-80	26
80-90	10
90-100	5

5.2. BASIC TERMS OF A FREQUENCY DISTRIBUTION

Now we shall discuss some basic terms to form a frequency distribution.

Tally Mark

A vertical bar (|) is put against the number when it occurs is called a tally mark. The fifth occurrence of value of the variable is represented by putting a cross tally (/) on the first four tallies. This procedure of counting continued till the end (last value of the data). The tally mark facilitates counting the frequency of value of a variable in a systematic manner. The tally marks are used only in preliminary work and not appear in the final representation of a frequency distribution.

Frequency

The number of times a value of the variable occurs in the data is called the frequency of that value of the variable. The frequency is denoted by f.

Example. If 26 students got 68 marks in an examination ,then frequency of the value 68 is 26. *i.e.,* 68 occurs 26 times, *i.e.,* $f = 26$.

Class Internal (C.I.)

If the data is large, then the number of observations can be classified into several groups according to the size of value. Each of these groups defined by an interval is known as Class Interval (C.I.). The class interval is usually specified by two extreme values called class limits. the smaller one is termed as lower limit and the larger one is termed as upper limit of the class interval.

Example. If 200 students appeared for an examination and marks obtained between 50 to 100, then the class intervals can be classified into 50-60, 60-70, 70-80, 80-90 and 90-100.

Class Boundaries

If d is the gap between upper class limit of one class and lower class limit of the next class, then the class boundaries are calculated by

$$\text{Lower class boundary} = \text{lower class limit} - \frac{d}{2}$$

$$\text{Upper class boundary} = \text{Upper class limit} + \frac{d}{2}$$

Example. If class intervals are :

10 – 19

20 – 29

30 – 39 etc.

then $d = 20 - 19 = 1$

Now class boundaries of the C.I. 10 – 19 are :

$$\text{Upper class boundary} = \text{Upper class limit} + \frac{d}{2}$$

$$= 19 + \frac{1}{2} = 19.5$$

$$\text{Lower class boundary} = \text{Low class limit} - \frac{d}{2}$$

$$= 10 - \frac{1}{2} = 9.5$$

$\therefore$ New class interval is 9.5 – 19.5

Like in this way, we can calculate all class intervals.

Mid Value of Class Interval

The value exactly at the middle of the class interval is called mid value of the class interval.

$$\text{Mid value of the C.I.} = \frac{(\text{Lower class limit} + \text{Upper class limit})}{2}$$

Example. If the class interval is 10-20, then mid value of the class interval 10-20 is $\frac{10 + 20}{2} = 15$.

Width of the Class Interval

The difference between upper and lower class boundaries of a class interval is known as width of the class interval. It is denoted by 'h'

h = Upper class boundary – Lower class boundary

Example. If the class intervals are 0-10, 10-20, 20-30, 30-40, etc., then the width of the class interval h is calculated by

$h = 30 - 20 = 10$ (by considering C.I. 20-30)

5.3. TYPES OF FREQUENCY DISTRIBUTION

The frequency distribution may be formed into two ways depending on the nature of data. Hence there are two types of frequency distributions, *viz.*,

1. Discrete (or) Ungrouped Frequency Distribution
2. Continuous Frequency Distribution

1. Discrete (or) Ungrouped Frequency Distribution

If the frequency refers to a given discrete value, then the corresponding frequency distribution is known as Discrete frequency distribution or Ungrouped frequency distribution.

Examples:

1. The number of companies in a country.
2. The number of plots in a field.
3. The following discrete frequency distribution shows number of children in 50 families in a locality.

Number of children	*Frequency (Number of families)*
0	2
1	13
2	30
3	5

2. Continuous Frequency Distribution

If the frequency refers to a continuous variable, then the corresponding frequency distribution is known as continuous frequency distribution. There are some variables which take any fractional value, in such cases, it is good practice to consider the values of the variable in class intervals. This kind of formation gives us continuous frequency distribution.

Examples:

1. The heights of group of 70 persons are given in the following continuous frequency distribution

C.I. (Heights of persons) (in feet)	*Frequency (No. of persons)*
4-5	8
5-6	20
6-7	25
7-8	15
8-9	2

2. The ages of 30 policy holders in a company are given in the following continuous frequency distribution.

Age (C.I.)	*No. of policies (Frequency)*
20-30	18
30-40	5
40-50	3
50-60	2
60-70	2

Note: The class interval 20-30 in the continuous frequency distribution indicates, this includes age 20 and not includes age 30. Similarly C.I. 30-40 includes age 30 and not includes age 40. etc. This is known as exclusive type classification (since upper limits are excluded).

5.4. FORMING OF A FREQUENCY DISTRIBUTION

If the raw data is collected from large number of individuals regarding a single characteristic. Then the simple way of understanding such data is construction in the form of frequency distribution only. To prepare the frequency distribution, first we use tally mark for every observation. This determines frequency of each value of the variable. Next we have to form the class intervals with suitable width if necessary by observing discrete or continuous variables. A statistical table with class interval (if so) and frequencies forms a frequency distribution.