STATISTICAL REPRESENTATION OF DATA

The word 'Statistics' has been derived from the Latin word 'Status' which means a political state. It has also its root either to the Italian word 'Statista' or the German word 'Statistik' each one of which means a political state. For several decades, the word 'statistics' was associated solely with the display of facts and figures pertaining to the economic, demographic and political situations prevailing in a country, usually, collected and brought out by the local governments.

Statistics is a tool in the hands of mankind to translate complex facts into simple and understandable statements of facts.

Meaning and definition of Statistics:

Meaning of statistics: The word Statistics is used in two different senses - Plural and singular. In its plural form, it refers to the numerical data collected in a systematic manner with some definite aim or object in view such as the number of persons suffering from malaria in different colonies of Delhi or number of unemployed girls in different states of India and so on. In Singular form, the word statistics means the science of statistics that deals with the principles, devices or statistical methods of collecting, analyzing and interpreting numerical data.

Thus, 'statistics' when used in singular refers to that branch of knowledge which implies Applied Mathematics.

The science of statistics is an old science and it has developed through ages. This science has been defined in different ways by different authors and even the same author has defined it in different ways on different occasions.

It is impossible to enumerate all the definitions given to statistics both as "Numerical Data i.e., Plural Form: and "Statistical Methods, i.e., Singular Form". However, we have give below some selected definitions of both the forms.

Definitions of "Statistics in Plural Form or Numerical Data": Different authors have given different definitions of statistics. Some of the definitions of statistics describing it quantitatively or in plural form are:

"Statistics are the classified facts representing the conditions of the people in a state especially those facts which can be stated in number or in a table of numbers or in any tabular or classified arrangement.

This definition is narrow as it is confined only to the collection of the people in a state. But the following definition given by Secrist is modern and convincing. It also brings out the major characteristics of statistical data.

"By Statistics we mean the aggregate of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standards of accuracy collected in a systematic manner for a pre-determined purpose and placed in relation to each other"

This definition makes it clear that statistics (in plural form or numerical data) should possess the following characteristics.

I. Statistics are aggregate of facts

II. Statistics are affected by a large number of causes

III. Statistics are always numerically expressed

IV. Statistics should be enumerated or estimated

V. Statistics should be collected in a systematic manner

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

VII. Statistics should be placed in relation to each other.

Statistics as Statistical methods or Statistics in Singular Sense:

We give below the definitions of statistics used in singular sense, i.e., statistics as statistical methods.

Statistical methods provide a set of tools which can be profitably used by different sciences in the manner they deam fit. The term statistics in this context has been defined differently by different authors. A few definitions are given below:

"Statistics may be called the science of counting"

This definition covers only one aspect, i.e., counting, but the other aspects such as classification, tabulation, etc., have been ignored. As such, the definition is inadequate and incomplete

"Statistics may be defined as the collection, presentation, analysis and interpretation of numerical data"

This definition given by Croxton and Cowden is simple, clear and concise.

According to this definition, there are four stages – collection of data, and presentation of data, analysis of data, and interpretation of data. However, one more stage may be added and that is the organization of data. Thus, there are four stages:

1. **Collection and Organization of data**: Three are various methods for collecting the data such as census, sampling, primary and secondary data etc.

2. **Presentation of data**: The mass data collected should be presented in a suitable, concise form as the mass data collected is difficult to understand and analyse

3. **Analysis of data**: The mass data collected should be presented in a suitable, concise form for further analysis. Analysis includes condensation, summarisation conclusion, etc., through the means of measures of central tendencies, dispersion, skewness, kurtosis, correlation, regression, etc.

4. **Interpretation of data**: The last step is the drawing conclusions from the data collected as the figures do not speak for themselves.

Having briefly discussed some of the definitions of the term statistics and having seen their drawbacks we are now in a position to give a simple and complete definition of the 'Statistics' in the following words:

Statistics (as used in the sense of data) are numerical statements of facts capable of analysis and interpretation and the science of statistics is a study of the principles and methods used in the collection, presentation, analysis and interpretation of numerical data in any sphere of enquiry.

Importance and Scope of Statistics:

I. **Statistics and Economics**: According to Prof. Alfred Marshall, "Statistics are the straws out of which I like every other economist, have to make bricks." The following are some of the fields of economics where statistics is extensively used.

(a) **Consumption**: Statistical data of consumption enable us to find out the ways in which people in different strata of society spend their incomes.

(b) **Production**: The statistics of production describe the total productivity in the country. This enables us to compare ourselves with other countries of the world.

(c) **Exchange**: In the field of exchange, an economist studies markets, laws of prices are determined by the forces of demand and supply, cost of production, monopoly, competition, banking etc. A systematic study of all these can be made only with the help of statistics

(d) **Econometrics**: With the help of econometrics, economics has become exact science. Econometrics is the combination of economics, mathematics and statistics.

(e) **Public Finance:** Public finance studies the revenue and expenditure activities of a country. Budget, (a statistical document), fiscal policy, deficit financing, etc., are the concepts of economics which are based on statistics.

(f) **Input-Output Analysis**: The input-output analysis is based on statistical data which explain the relationship between the input and the output. Sampling, Time series, Index numbers, Probability, Correlation and Regression are some other concepts which are used in economic analysis.

II. **Statistics and Commerce**:

Statistical methods are widely applied in the solution of most of the business and trade activities such as production, financial analysis, costing, manpower, planning, business, market research, distribution and forecasting etc. A shrewd businessman always makes a proper and scientific analysis of the past records in order to predict the future course of the business conditions. Index numbers help in predicting the future course of business and economic events. Statistics or statistical methods help the business establishments in analysing the business activities such as:

(a) **Organization of Business**: Businessman makes extensive use of statistical data to arrive at the conclusion which guides him in establishing a new firm or business house

(b) **Production**: The production department of an organisation prepares the forecast regarding the production of the commodity with the help of statistical tools.

(c) **Scientific Management and Business Forecasting**: Better and efficient control of a business can be achieved by scientific management with the help of statistical data. "The success of businessman lies on the accuracy of forecast made". The successful businessman is one who estimate most closely approaches accuracy," said Prof. Boddington.

(d) **Purchase:** The price statistics of different markets help the businessman in arriving at the correct decisions. Raw material is purchased from those markets only where the prices are low.

III. **Statistics and 'Auditing and Accounting'**: Statistics is widely used in accounting and auditing.

IV. **Statistics and Economic Planning**: According to Prof. Dickinson, "Economic Planning is making of major decisions – what and how much is to be produced, and to whom it is to be allocated – by the conscious decisions of a determinate authority on the basis of a comprehensive survey of economy as a whole. "The various documents accompanying preceding and following each of the eight Five Year Plans of India are a standing testimony to the fact that statistics is an indispensable tool in economic planning.

V. **Statistics and Astronomy**: Statistics were first collected by astronomers for the study of the movement of stars and planets. As there are a few things which are common between physical sciences, and statistical methods, astronomers apply statistical methods to go deep in their study. Astronomers generally take a large number of measurements and in most cases there is some difference between these several observations. In order to have the best possible measurement they have to make use of the technique of the law of errors in the form of method of least squares.

VI. **Statistics and Meteorology**: Statistics is related to meteorology. To compare the present with the past or to forecast for the future either temperature or humidity of air or barometrical pressures etc., it becomes necessary to average these figures and thus to study their trends and fluctuations. All this cannot be done without the use of statistical methods. Thus, the science of statics helps meteorology in a large number of ways.

VII. **Statistics and Biology**: The development of biological theories has been found to be closely associated with statistical methods. Professor Karl Pearson in his Grammar of Sciences has written, "The whole doctrine of heredity rests on statistical basis".

VIII. **Statistics and Mathematics**: Mathematics and Statistics have been closely in touch with each other ever since the 17th Century when the theory of probability was found to have influence on various statistical methods. Bowley was right when he said, "Acknowledge of Statistics is like knowledge of foreign language or of algebra: it may prove of use at any time under any circumstances".

Thus we observe that**:**

"Science without statistics bear no fruit, statistics without sciences have no Root".

IX. **Statistics and Research**: Statistical techniques are indispensable in research work. Most of advancement in knowledge has taken place because of experiments conducted with the help of statistical methods.

X. **Statistics and natural sciences**: Statistics finds an extensive application in physical sciences, especially in engineering physics, chemistry, geology, mathematics, astronomy, medicine, botany, meteorology, zoology, etc.

XI. **Statistics and Education:** There is an extensive application of statistics in Education. Statistics is necessary for formulation of policies to start new courses, infrastructure required for new courses consideration of facilities available for new courses etc.

XII. **Statistics and Business:** Statistics is an indispensable tool in all aspects of business. When a man enters business he enters the profession of forecasting because success in business is always the result of precision in forecasting and failure in business is very often due to wrong expectations. Which arise in turn due to faulty reasoning and inaccurate analysis of various cause affecting a particular phenomenon Boddington observes, "The successful businessman is the one, whose estimate most closely approaches the accuracy".

LIMITATIONS OF STATISTICS:

Statistics and its techniques are widely used in every branch of knowledge. W.I. King rightly says: "Science of statistics is the most useful servant, but only of great value to those who understand its proper use". The scope of statistics is very wide and it has great utility; but these are restricted by its limitations. Following are the important limitations of statistics:

1. **Statistics does not deal with individual item**: King says, "Statistics from the very nature of the subject cannot and never will be able to take into account individual cases". Statistics proves inadequate, where one wants to study individual cases. Thus, it fails to reveal the true position.

2. **Statistics deals with quantitative data**: According to Prof. Horace Secrist, "Some phenomenon cannot be quantitatively measured; honesty, resourcefulness, integrity, goodwill, all important in industry as well as in life, are generally not susceptible to direct statistical measurement".
3. **Statistical laws are true only on averages**. According to W.I. King, "Statistics largely deals with averages and these may be made up of individual items radically different from each other". Statistics are the means and not a solution to a problem.
4. **Statistics does not reveal the entire story**: According to Marshall, "Statistics are the straws, out of which, I like every other economist to have to make bricks. "Croxton says: "It must not be assumed that statistical method is the only method or use in research; neither should this method be considered the best attack for every problem".
5. **Statistics is liable to be misused**: According to Bowley, "Statistics only furnishes a tool though imperfect, which is dangerous in the hands of those who do not know its use and deficiencies".
   W.I. King states, "Statistics are like clay of which you can make a God or Devil as you please". He remarks, "Science of Statistics is the useful servant, but only of great value to those who understand its proper use".
6. **Statical data should be uniform and homogeneous**

STATISTICAL TOOLS USED IN ANALYSIS:

The following are some of the important statistical techniques which are applied in economic analysis:

(a) Collection of data

(b) Tabulation

(c) Measures of Central Tendency

(d) Measures of Dispersion

(e) Time Series

(f) Probability

(g) Index Numbers

(h) Sampling and its uses

(i) Business Forecasting

(j) Tests of Significance and analysis of variance

(k) Statistical Quality Control

Collection of Data:

Data the information collected through censuses and surveys or in a routine manner or other sources is called a raw data. The word data means information (its literary meaning is given as facts). The adjective raw attached to data indicates that the information thus collected and recorded cannot be put to any use immediately and directly. It has to be converted into more suitable form or processed before it begins to make sense to be utilized gainfully. A raw data is a statistical data in original form before any statistical techniques are used to redefine, process or summarize it.

There are two types of statistical data:

(i) Primary data

(ii) Secondary data

1. Primary Data: It is the data collected by a particular person or organization for his own use from the primary source
2. Secondary data: It is the data collected by some other person or organization for their own use but the investigator also gets it for his use.

In other words, the primary data are those data which are collected by you to meet your own specific purpose, whereas the secondary data are those data which are collected by somebody else.

A data can be primary for one person and secondary for the other.

Methods of Collecting Primary Data:

The primary data can be collected by the following methods:

1. **Direct personal observation**: In this method, the investigator collects the data personally and, therefore, it gives reliable and correct information.
2. **Indirect oral investigation**: In this method, a third person is contacted who is expected to know the necessary details about the persons for whom the enquiry is meant.
3. **Estimates from the local sources and correspondence**. Here the investigator appoints agents and correspondents to collect the data
4. **Data through questionnaires**. The data can be collected by preparing a questionnaire and getting it filed by the persons concerned.
5. **Investigations through enumerators**. This method I generally employed by the Government for population census, etc.

Methods of Collecting Secondary data:

The secondary data can be collected from the following sources:

1. Information collected through newspapers and periodicals.
2. Information obtained from the publications of trade associations.
3. Information obtained from the research papers published by University departments or research bureaus or UGC.
4. Information obtained from the official publications of the central, state, and the local governments dealing with crop statistics, industrial statistics, trade and transport statistics etc.
5. Information obtained from the official publications of the foreign governments for international organizations. Like World Bank, ILO, IMF, etc.

Classification of Data: The process of arranging things in groups or classes according to their common characteristics and affinities is called the classification of data.

"Classification is the process of arranging data into sequences and groups according to their common characteristics or separating them into different but related parts – Secrist.

Thus classification is the process of arranging the available data into various homogenous classes and sub-classes according to some common characteristics or attribute or objective of investigation.

Requisites of a Good Classification:

The main characteristics of a good classification are:

1. It should be exhaustive
2. It should be unambiguous
3. It should be mutually exclusive
4. It should be stable
5. It should be flexible
6. It should have suitability
7. It should be homogeneous
8. It should be a revealing classification

9. It should be reliable
10. It should be adequate.

Advantages of classification of data:

(i) It condenses the data and ignores unnecessary details

(ii) It facilitates comparison of data

(iii) It helps in studying the relationships between several characteristics

(iv) It facilitates further statistical treatments

Types of Classification of Data:

There are four types of classification of data:

(i) Quantitative Classification

(ii) Temporal Classification

(iii) Spatial Classification and

(iv) Qualitative Classification

(i) **Quantitative Classification**: When the basis of classification is according to differences in quantity, the classification is called quantitative

A quantitative classification refers to classification that is based on figures: In other words, it is a classification which is based on such characteristics which are capable of quantitative measurement such as height, weight, number of marks obtained by students of a class.

(ii) **Temporal Classification**: When the basis of classification is according to differences in time, the classification is called temporal or chronological classification

(iii) **Spatial or Geographical Classification**: When the basis of classification is according to geographical location or place, the classification is called spatial or geographical

(iv) **Qualitative Classification**: When the basis of classification is according to characteristics or attributes like social status etc. is called qualitative classification.

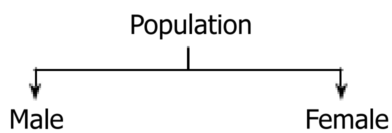
Classification according to attributes is a method in which the data are divided on the basis of qualities. (i.e., married or single; honest or dishonest; beautiful or ugly; on the basis of religion, viz., Hindu, Muslim, Sikh, Christian etc., known as attributes), which cannot be measured quantitatively.

Classification of this nature is of two types:

(i) Simple Classification or Two-Fold Classification

(ii) Manifold Classification

1. **Simple Classification or Two-fold Classification**: If the data are classified only into two categories according to the presence or absence of only one attribute, the classification is known as simple or two-fold classification or Dichotomous For example, the population of India may be divided into males and females; literate and illiterate etc.



Moreover, if the classification is done according to a single attribute it is also known as one way classification.

2. **Manifold Classification**: It is a classification where more than one attributes are involved.

MODE OF PRESENTATION OF DATA:

In this section we shall consider the following three modes of presentation of data

(a) Textual Presentation

(b) Tabular Presentation or Tabulation

(c) Diagrammatic presentation

Textual Presentation:

This method comprises presenting data with the help of a paragraph or a number of paragraphs. The official report of an enquiry commission is usually made by textual presentation. Following are the examples of textual presentation.

Example1: In 1995, out of total of 2,000 students in a college, 1,400 were for graduation and the rest for post-graduation (P.G.) out of 1,400 Graduate students 100 were girls, however, in all there were 600 girls in the college. In 2000, number of graduate students increased to 1,700 out of which 250 were girls, but the number of P.G. students fall to 500 of which only 50 were boys. In 2005, out of 800 girls 650 were for graduation, where as the total number of graduates was 2,200. The number of boys and girls in P.G. classes was equal.

Merits and Demerits of Textual Presentation: The merit of this mode of presentation lies in its simplicity and even a layman can present and understand the data by this method. The observations with exact magnitude can be presented with the help of textual presentation. This type of presentation can be taken as the first step towards the other methods of presentation.

Textual presentation, however, is not preferred by a statistician simply because it is dull, monotonous and comparison between different observations is not possible in this method. For manifold classification, this method cannot be recommended.

Tabular presentation or tabulation of data: Tabulation is a scientific process used in setting out the collected data in an understandable form

Tabulation may be defined as logical and systematic arrangement of statistical data in rows and columns. It is designed to simplify the presentation of data for the purposes of analysis and statistical inferences.

Secrist has defined tabulation in the following words:

"Tables are a means of recording in permanent form the analysis that is made through classification and by placing in juxtaposition things that are similar and should be compared".

The above definition clearly points out that tabulation is a process which gives classification of data in a systematic form and is meant for the purpose of making comparative studies.

Professor Bowley refers to tabulation as:

"The intermediate process between the accumulation of data in whatsoever form they are obtained, and the final reasoned account of the result shown by the statistics".

"Tabulation is the process of condensing classified data in the form of a table so that it may be more easily understood and so that any comparison involved may be more readily made".

Thus tabulation is one of the most important and ingenious devices of presenting the data in a condensed and readily comprehensible form. It attempts to furnish the maximum information in the minimum possible space, without sacrificing the quality and usefulness of the data.

Objectives of Tabulation:

The purpose of tabulation is to summarise lots of information in such a simple manner that it can be easily analysed and interpreted.

The main objectives of the Tabulation are:

1. To simplify the complex data.
2. To clarify the objective of investigation
3. Economise space.

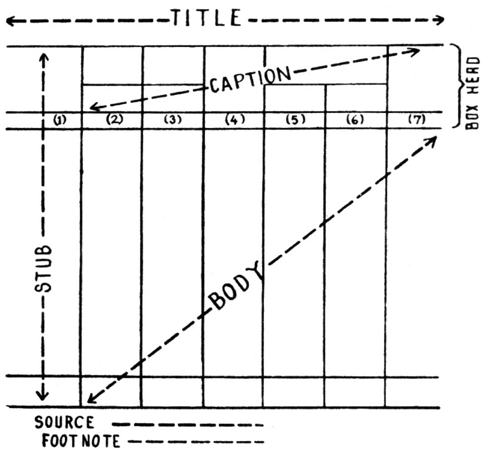
4. To facilitate comparison
5. To depict trend and pattern of data
6. To help reference for future studies.
7. To facilitate statistical analysis.
8. To detect errors and omissions in the data
9. To clarify the characteristics of data.

Essential Parts of a Statistical Table:

A good statistical table should invariably has the following parts:

1. **Table Number**: A table should be numbered for identification, especially, when there are a large number of tables in a study. The number maybe put at the centre, above the title or at the bottom of the table.
2. **Title of the table**: Every table should have a title. It should be clear, brief and self explanatory. The title should be set in bold type so as to give it prominence.
3. **Date**: The date of preparation of a table should always be written on the table. It enables to recollect the chronological order of the table prepared.
4. **Stubs or Row Designations**: Each row of the table must have a heading. The designations of the rows are called stubs or stub items. Stubs clarify the figures in the rows. As far as possible, the items should be considered so that they can be included in a single row.
5. **Captions or Column headings**: A table has many columns. Sub-headings of the columns are called captions or headings. They should be well-defined and brief.
6. **Body of the table**: It is the most vital part of the table. It contains the numerical information. It should be made as comprehensive as possible. The actual data should be arranged in such a manner that any figure may be readily located. Different categories of numerical variables should be set out in an ascending order, from left to right in rows and in the same fashion in the columns, from top downwards.
7. **Unit of Measurements**: The unit of measurements should always be stated along with the title, if this is uniform throughout. If different units have been adopted, then they should be stated along the stubs or captions.
8. **Source Notes**. A note at the bottom of the table should always be given to indicate the primary source as well as the secondary source from where the data has been taken, particularly, when there is more than one source.
9. **Foot Notes and References**: It is always placed at the bottom of the table. It is a statement which contains explanation of some specific items, which cannot be understood by the reader from the title, or captions and stubs.

Different Parts of Table



Difference between Textual and Tabular Presentation: The tabulation method is usually preferred to textual presentation as:

(i) It facilitates comparison between rows and columns

(ii) Complicated data can be represented using tabulation

(iii) Without tabulation, statistical analysis of data is not possible.

(iv) It is a must for diagrammatic representation.

DIAGRAMMATIC REPRESENTATION OF DATA

The representation of statistical data through charts, diagrams and picture is another attractive and alternative method. Unlike the first two methods of representation of data, diagrammatic representation can be used for both the educated section and uneducated section of the society. Furthermore, any hidden trend presented in the given data can be noticed only in this mode of representation. However, compared to tabulation, this is less accurate. So if there is a priority for accuracy, we have to recommend tabulation.

In this chapter we shall consider the following three types of diagrams:

I. Line diagram chart;

II. Bar diagram;

III. Pie chart.

LINE CHART :

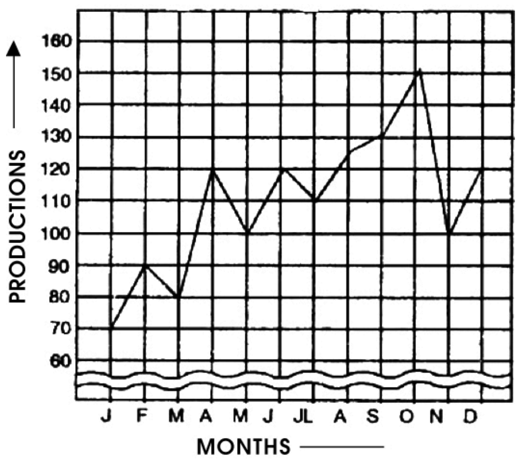
We take a rectangular axes. Along the abscissa, we take the independent variable (x or time) and along the ordinate the dependent variable (y or production related to time). After plotting the points, they are joined by a scale, which represents a line chart. The idea will be clear from the following example.

Example : Represent the following data by line chart.

The monthly production of motor cars in India during 2011-12

Jan	Feb	Mar	April	May	June	July	Aug	Sept	Oct	Nov	Dce
70	90	80	120	100	120	110	125	130	150	100	120

Graph showing production of motor cars.



BAR DIAGRAM:

The simplest type of graph is the bar diagram. It is especially useful in comparing qualitative data or quantitative data of discrete type. A bar diagram is a graph on which the data are represented in the form of bars. It consists of a number of bars or rectangles which are of uniform width with equal space between them on the x-axis. The length of the bar is proportional to the value it represents. It should be seen that the bars are neither too short nor too long. The scale should be clearly indicated and base line be clearly shown.

Bars may be drawn either horizontally or vertically. A good rule to use in determining the direction is that if the legend describing the bar can be written under the bars when drawn vertically, vertical bars should be used; when it cannot be, horizontal ones must be used. In this way, the legends can be read without turning the graph. The descriptive legend should not be written at the ends of the bars or within the bars, since such writing may distort the comparison. Usually the diagram will be more attractive if the bars are wider than the space between them.

The width of bars is not governed by any set rules. It is an arbitrary factor. Regarding the space between two bars, it is conventional to have a space about one half of the width of a bar.

The data capable of representation through bar diagrams, may be in the form of row scores, or total scores, or frequencies, or computed statistics and summarised figures like percentages and averages etc.

The bar diagram is generally used for comparison of quantitative data. It is also used in presenting data involving time factor. When two or more sets of data over a certain period of time are to be compared a group bar diagram is prepared by placing the related data side by side in the shape of bars. The bars may be vertical or horizontal in a bar diagram. If the bars are placed horizontally, it is called a Horizontal Bar Diagram. When the bars are placed vertically, it is called a Vertical Bar Diagram.

There are six types of Bar diagram:

(i) Simple Bar Diagram;

(ii) Multiple or Grouped Bar Diagram;

(iii) subdivided or Component Bar Diagram;

(iv) Percentage Subdivided Bar Diagram;

(v) Deviation or Bilateral Bar Diagram;

(vi) Broken Bars.

Simple Bar Diagram:

It is used to compare two or more items related to a variable. In this case, the data are presented with the help of bars. These bars are usually arranged according to relative magnitude of bars. The length of a bar is determined by the value or the amount of the variable. A limitation of Simple Bar Diagram is that only one variable can be represented on it.

Multiple or Grouped Bar Diagram:

A multiple or grouped bar diagram is used when a number of items are to be compared in respect of two, three or more values. In this case, the numerical values of major categories are arranged in ascending or descending order so that the categories can be readily distinguished. Different shades or colours are used for each category.

Sub-divided or Component Bar Diagram:

A component bar diagram is one which is formed by dividing a single bar into several component parts. A single bar represents the aggregate value whereas the component parts represent the component values of the aggregate value. It shows the relationship among the different parts and also between the different parts and the main bar.

Percentage Sub-divided Bar Diagram:

It consists of one or more than one bars where each bar totals 100%. Its construction is similar to the sub-divided bar diagram with the only difference that where as in the sub-divided bar diagram segments are used in absolute quantities, in the percentage bar diagram the quantities are transformed into percentages.

PIE DIAGRAM OR ANGULAR DIAGRAM:

A pie diagram is a circular graph which represents the total value with its components. The area of a circle represents the total value and the different sectors of the circle represent the different parts. The circle is divided into sectors by radii and the areas of the sectors are proportional to the angles at the centre. It is generally used for comparing the relation between various components of a value and between components and the total value. In pie diagram, the data are expressed as percentages. Each component is expressed as percentage of the total value. A pie diagram is also known as angular diagram.

The name pie diagram is given to a circle diagram because in determining the circumference of a circle we have to take into consideration a quantity known as 'pie' (written as π).

Method of Construction: The surface area of a circle is known to cover 2 π radians or 360 degrees. The data to be represented through a circle diagram may therefore be presented through 360 degrees, parts or sections of a circle. The total frequencies or value is equated to 360° and then the angles corresponding to component parts are calculated (or the component parts are expressed as percentages of the total and then multiplied by 360/100 or 3.6). After determining these angles the required sectors in the circle are drawn. Different shades or colours of designs or different types of cross- hatchings are used to distinguish the various sectors of the circle.

Example: 120 students of a college were asked to opt for different work experiences. The details of these options are as under.

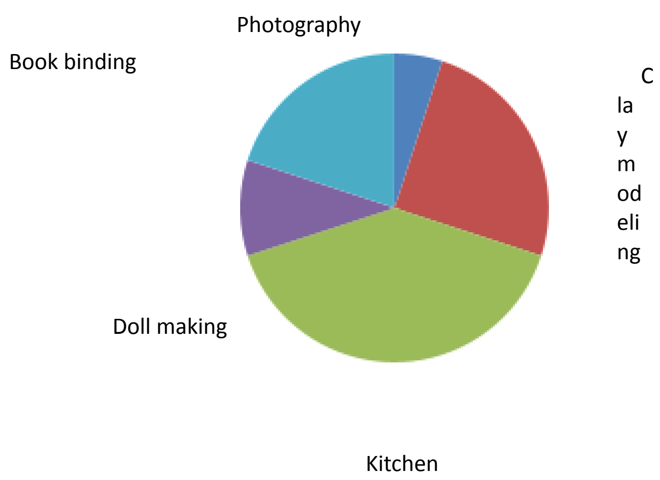
Areas of work experience	**No. of students**
Photography	6
Clay modeling	30
Kitchen gardening	48
Doll making	12
Book binding	24

Represent the above data through a pie diagram.

Solution:

The numerical data may be converted into the angle of the circles as given below:

Areas of work experience	**No. of students**	**Angle of the circle**
Photography	6	(6/120) x 360 = 18°
Clay modeling	30	(30/120) x 360 = 90°
Kitchen gardening	48	(48/120) x 360 = 144°
Doll making	12	(12/120) x 360 = 36°
Book binding	24	(24/120) x 360 = 72°
Total	120	360°



gardening

Areas of work experience opted by the students

FREQUENCY DISTRIBUTION

TALLY BARS AND FREQUENCY:

In order to make the data easily understandable, we tabulate the data in the form of tables or charts.
A table has three columns

(i) Variable

(ii) Tally marks

(iii) Frequency

(i) **Variable**: Any character which can vary from one individual to another is called a variable or a variate. For example, age, income, height, intelligence, colour etc. are variates. Some variates are measurable and others are not directly measurable. The examples of measurable variates are age, height, temperature, etc., where as colour and intelligence are the examples of those variates which cannot be measured numerically. Variables or observations with numbers as possible values are called quantitative variables, whereas those with names of places, quality, things etc., as possible values are called qualitative variables or attributes.

Variables are of two types i) Continuous; ii) Discontinuous or Discrete. Quantities which can take all numerical values within a certain interval are called continuous variables; But those variables which can take only a finite of values are called discrete variables; For example, number of students in a particular class, number of sections in a school etc.

(ii) **Tally**: It is a method of keeping count in blocks of five.

For example; 1 = 1; || = 2; ||| = 3; |||| = 4; ~~||||~~ = 5; ~~||||~~| = 6 and so on.

Tally Bars: These are the straight bars used in the Tally.

Each item falling in the class interval, a stroke (vertical Bar) is marked against it. This stroke (Vertical Bar) is called the Tally Bar. Usually, after every four strokes (Tally Bar), in a class, the fifth item is marked by a horizontal or slanted line across the Tally Bars (Strokes). For example the frequency 5, 6, 7 is represented by ~~||||~~ , ~~||||~~| , ~~||||~~|| respectively.

The above method of presentation of data is known as 'Frequency Distribution'. Marks are called variates. The number of students who have secured a particular number of marks is called Frequency of that variate.

In the first column of the table, we write all marks from lowest to highest. We now look at the first mark or value in the given raw data and put a bar (vertical line) in the second column opposite to it. We then, see the second mark or value in the given raw data and put a bar opposite to it in the second column. This process is repeated till all the observations in the given raw data are exhausted. The bars drawn in the second column are known as tally marks and to facilitate we record tally marks in bunches of five, the fifth tally marks is drawn diagonally across the first four.

For example, ~~||||~~||| = 8. We finally count the number of tally marks corresponding to each observation and write in the third column headed by frequency or number of students.

(iii) **Frequency**: The number of times an observation occurs in the given data is called the frequency of the observation.

Frequency Distribution: A frequency distribution is the arrangement of the given data in the form of a table showing frequency with which each variable occurs. In other words, Frequency distribution of a variable is the ordered set $\{x, f\}$, where f is the frequency. It shows all scores in a set of data together with the frequency of each score.

Types of frequency distributions:

Frequency distributions are of two types:

(i) Discrete Frequency Distribution

(ii) Grouped (or Continuous) Frequency Distribution

Discrete Frequency Distribution: The construction of discrete frequency distribution from the given raw data is done by the method of tally marks as explained earlier.

Construction of Discrete Frequency Distribution Table:

The frequency distribution table has three columns headed by

1. Variables (or classes)
2. Tally Mark or Bars
3. Frequency

The table is constructed by the following steps:

Step 1: Prepare three columns, viz., one for the variable (or classes), another for tally marks and the third for the frequency corresponding the variable (or class).

Step 2: Arrange the given data (or values) from the lowest to the highest in the first column under the heading variable (or classes)

Step 3: Take the first observation in the raw data and put a bar (or vertical line|) in the second column under Tally Marks opposite to it. Then take a second observation and put a tally marks opposite to it, continue this process till all the observations of the given raw data are exhausted. For the sake of convenience, record the tally marks in bunches of five, the fifth bar is placed diagonally crossing the other four (5 is represented by ~~||||~~) leave some space between each block of bars.

Step 4: Count the tally marks of column 2 and place this number opposite to the value of the variable in the third column headed by Frequency.

Step 5: Give a suitable title to the frequency distribution table so that it exactly conveys the information contained in the table.

SOME STATISTICAL TERMS:

Raw Data or Data: A raw data is a statistical data in original form before any statistical technique is applied to redefine process or summarize it.

Variable or Variate: Any character which can vary from one individual to another is called variable or variate. For example, age, income, height, intelligence, colour, etc., are variates. Some variates are measurable and others are not directly measurable. The examples of measurable variates are age, height, temperature, etc., whereas colour and intelligence are the examples of those variates which cannot be measured numerically. Variables or observations with numbers as possible values are called quantitative variables, whereas those with names of places, attributes, and things etc., as possible values are called qualitative variables.

Variables are of two types:

(i) Continuous;

(ii) Discontinuous or Discrete

Continuous Variable: A continuous variable is capable of assuming any value within a certain range or interval. The height, weight, age and temperature of any person can be expressed not only in integral part but also in fractions of any part. For example, the weight of a boy may 44.0 kg or 44.6 kg or 44.65 kg, similarly, his height may be 56 inches or 56.4 inches and age may be 10 years or 10.5 years. Thus, the height, weight, age or temperature etc. are continuous variables.

Discrete Variable: A discrete variable can assume only integral values and is capable of exact measurement. In other words, those variables which can take only a finite set of values are called discrete variables. For example: the number of students in a particular class, or the number of sections in a school, etc. are the examples of discrete variables. Discrete variables are also known discontinuous variables.

Continuous Series: When the continuous variables are arranged in the form of a series, it is called continuous series or exclusive series.

Discrete or Discontinuous Series: When the discrete variables are arranged in the form of a series, it is called a discrete or discontinuous series.

Array: An array is an arrangement of data in order of magnitude either in descending or ascending order.

Descending Order: When data is arranged from the highest value to the lowest value, the array so formed is in descending order.

Ascending order: When the data is arranged from the lowest value to the highest value, the array so formed is in ascending order.

Illustration: If the given data is 17, 7, 11, 5, 13, 9 then

Array in Ascending order: 5, 7, 9, 11, 13, 17.

Array in Descending order: 17,13,11,9,7, 5.

Range: It is the difference between the largest and the smallest number in the given data The

range of the data given in illustration is 17 – 5 = 12.

Class, Class-Interval and Class limits. If the observations of a series are divided into groups and the groups are bounded by limits, then each group is called a class. The end values of a class are called class limits. The smaller value of the two limits is called the lower limit and the higher value of the same is called the upper limit of the class. These two class limits are sometimes called the stated class limits.

Class Interval: The difference between the lower limit (L) and the upper limit (U) of the class is known as class interval (I).

Thus: I = U – L.

In other words, the range of a class is called its Class Interval.

Illustration: The given data is

Marks obtained	**Tally Marks**	**No. of Students**
1-10	卌 I	6
11-20	III	3
21-30	卌 II	7
31-40	II	2
Total		18

In the above data the classes are: 1-10, 11-20, 21-30, 31-40.

Class Interval: The range of the marks from 1 to 40 is grouped into four classes or groups viz: 1 – 10, 11, 20, 21-30, 31-40. Each group is known as class interval. The interval between one class and its adjacent class being 9. [as 10-1 = 9, 20 – 11 = 9, 30 – 21 = 9, etc.]

Class Limits: In the first class 1 – 10, its lower limit is 1 and upper limit is 10. Similarly, 31 is the lower limit and 40 is the upper limit of the class interval 31-40.

Actual Class Limit or Class Boundaries: In the illustration, there is a gap of 1 mark between the limits of any two adjacent classes. This gap may be filled up by extending the two limits of each class by half of the value of the gap. Thus

Lower class boundary = lower class limit – ½ of the

gap Upper class boundary = Upper class limit + ½ of

the gap The class boundary of the class 11 to 20 are

Lower class boundary = 11– ½ of 1 = 11 – 0.5 = 10.5

Upper class boundary = 20+ ½ of 1 = 20+ 0.5 = 20.5

In other words, the class boundaries are the limits up to which the two limits, (actual) of each class may be extended to fill up the gap that exists between the classes. The class boundaries of each class, so obtained are called the Actual class limits or True class limits.

True lower class limit = Lower class limit - ½ of the gap

True upper class limit = Upper class limit + ½ of the gap

Note: In the case of exclusive series True class limits are the same as class limits

Illustration:

Class Interval	Class Boundries
11-20	10.5-20.5
21-30	20.5-30.5
31-40	30.5-40.5

Class-mark or Mid-point or Mid-value: The central value of the class interval is called the mid-point or mid-value or class mark. It is the arithmetic mean of the lower class and upper class limit of the same class.

(or)

$$\text{Mid-value of Class} = \frac{\text{Lower class limit+Upper class limit}}{2}$$

$$\text{Class mark} = \frac{\text{True Upper class limit + True Lower class limit}}{2}$$

The class mark of the class 11-20 is $\frac{11+20}{2} = 15.5$

Class Magnitude: It is the difference between the upper class boundary and the lower class boundary of the class. In the illustration the class magnitude of the class, 20.5 – 30.5 is (30.5 – 20.5) = 10.

Inclusive and Exclusive Series: In the above illustration, all the marks we considered were integers. Hence, it was possible for us to choose classes 11 to 20, 21 to 30 etc. there is a gap of 1 between the upper limit of a class and the lower limit of its next consecutive class, which has not created any difficulty. But there can be situations where the raw data is not in integers. For example, in the information regarding maximum temperature of the city or time required to solve a statistical problem is recorded in the data, it may contain fractions as well. In such cases, the consecutive classes have to be necessarily continuous. We have the following:

Inclusive Series: When the class-intervals are so fixed that the upper limit of the class is included in that class, it is known as inclusive method of classification, e.g., 0-5, 6-10, 11-15, 16-20.

In the inclusive series, the upper limit and lower limit are included in that class interval. For example, in illustration, the marks 11 and 20 are included in the class 11-20. It is a discontinuous series or inclusive series. In order to make it a continuous one, some adjustment with the class limits is necessary. The class limits are extended to class boundaries by the adjusting adjustment factor, which is equal to half of the difference between the upper limit of the one class and lower limit of the next class. The series so obtained is continuous and is known as exclusive series.

Exclusive or Continuous Series: In this series the upper limit of the class is the lower limit of the other class, the common point of the two classes is included in the higher class. For example, 10-15, 15-20, 20-25, …, represent a continuous series or the exclusive series. In this series, 15 is included in the class 15-20 and 20 is included in 20-30. Here the class intervals overlap and the upper limit of each class is treated as less than that limit and lower limit of each class actually represents exact value. Thus

When the class-intervals are so fixed that the upper limit of one class is the lower limit of the next class, it is known as Exclusive method of classification.

RELATIVE FREQUENCY AND PERCENTAGE FREQUENCY OF A CLASS INTERVAL:

Relative Frequency: Frequency of each class can also be expressed as a fraction of percentage terms. These are known as relative frequencies. In other words, a relative frequency is the class frequency expressed as a ratio of the total frequency, i.e.,

$$\text{Relative frequency} = \frac{\text{Class frequency}}{\text{Total frequency}}$$

Percentage Frequency: Percentage frequency of a class interval may be defined as the ratio of the class frequency to the total frequency expressed as a percentage.

$$\text{Percentage frequency} = \frac{\text{Class frequency}}{\text{Total frequency}} \times 100$$

GRAPHICAL REPRESENTATION OF FREQUENCY

The graphs of frequency distribution are designed to present the characteristic features of a frequency data. They facilitate comparative study of two or more frequency distributions regarding their shape and pattern.

The most commonly used graphs are:

1. Histogram
2. Frequency Polygon
3. Frequency Curve
4. Cumulative Frequency Curve or Ogive.

HISTOGRAM (when C.I. are equal)

Let us consider a frequency distribution having a number of class intervals with their respective frequencies. The horizontal axis is marked to represent the C.I. and on these markings rectangles are drawn by taking the C.I. as breadth and corresponding frequencies as heights. Thus a series of rectangles are obtained whose total area represents the total of the class frequencies. The figure thus obtained is known as histogram.

It may be noted here that C.I. must be in continuous form. Even if this is not given, then the discrete C.I. must be transferred to class boundaries and hence to draw the histogram.

Example : Draw a histogram of the following frequency distribution showing the number of boys in the register of a school.

Age (in years)	No. of boys (in '000)
2–5	15
5–8	20
8–11	30
11–14	40
14–17	25
17–20	10

C.I. given are in class boundaries.

Histogram

Histogram (when C.I. are uneqal) : If the C.I. are unequal the frequencies must be adjusted before constructing the histogram. Adjustments are to be made in respect of lowest C.I. For instance if one C.I. is twice as wide as the lowest C.I., then we are to divided the height of the rectangle by two and if again it is three times more, then we are to divide the height of the rectangle by three and so on.

Aliter (with the help of frequency density) :

If the width of C.I. are euqal the heights of rectangles will be proportional to the corresponding class frequencies. But if the widths of C.I. are unequal (i.e. some are equal and others are unequal), then the heights of rectangles will be proportional to the corresponding frequency densities (and not with the class frequencies)

Frequency density = $\frac{\text{Class frequency}}{\text{Width of C.I.}}$

MEASURES OF CENTRAL TENDENCY AND DISPERSION

In the previous chapters, data collection and presentation of data were discussed. Even after the data have been classified and tabulated one often finds too much details for many uses that may be made of the information available. We, therefore, frequently need further analysis of the tabulated data. One of the powerful tools of analysis is to calculate a single average value that represents the entire mass of data. The word average is very commonly used in day-to-day conversation. For example, we often talk of average work, average income, average age of employees, etc. an 'Average' thus is a single value which is considered as the most representative or typical value for a given set of data. Such a value is neither the smallest nor the largest value, but is a number whose value is somewhere in the middle of the group. For this reason an average is frequently referred to as a measure of central tendency or central value. Measures of central tendency show the tendency of some central tendency show the tendency of some central value around which data tends to cluster.

Objectives of Averaging:

There are two main objectives of the study of averages:

(i) **To get one single value that describes the characteristic of the entire data**. Measures of central value, by condensing the mass of data in one single value, enable us to get an idea of the entire data. Thus one value can represent thousands, lakhs and even millions of values. For example, it is impossible to remember the individual incomes of millions of earning people of India and even if one could do it there is hardly any use. But if the average income is obtained, we get one single value that represents the entire population. Such a figure would throw light on the standard of living of an average Indian.

(ii) **To facilitate comparison**. Measures of central value, by reducing the mass of data in one single figure, enable comparisons to be made Comparison can be made either at a point of time or over a period of time. For example, the figure of average sales for December may be compared with the sales figures of previous months or with the sales figure of another competitive firm.

Characteristics of a Good Average:

Since an average is a single value representing a group of values, it is desirable that such a value satisfies the following properties:

(i) **It should be easy to understand**. Since statistical methods are designed to simplify complexity, it is desirable that an average be such that can be readily understood; otherwise, its use is bound to be very limited.

(ii) **It should be simple to compute**. Not only an average should be easy to understand but also it should be simple to compute so that it can be used widely. However, though case of computation is desirable, it should not be sought at the expense of other advantages, i.e., if in the interest of greater accuracy, use of a more difficult average is desirable one should prefer that.

(iii) **It should be based on all the observations**. The average should depend upon each and every observation so that if any of the observation is dropped average itself is altered.

(iv) **It should be rigidly defined**. An average should be properly defined so that it has one and only one interpretation. It should preferably be defined by an algebraic formula so that if different people compute the average from the same figures they all get the same answer (barring arithmetical mistakes).

(v) **It should be capable of further algebraic treatment**. We should prefer to have an average that could be used for further statistical computations. For example, if we are given separately the figures of average income and number of employees of two or more factories we should be able to compute the combined average.

(vi) **It should have sampling stability**. We should prefer to get a value which has what the statisticians call 'sampling stability'. This means that if we pick 10 different groups of college students, and compute the average of each group, we should expect to get approximately the same values. It does not mean, however, that there can be no difference in the value of different samples. There may be some difference but those averages in which this difference, technically called sampling fluctuation, is less are considered better than those in which this difference is more.

(vii) **It should not be unduly affected by the presence of extreme values**. Although each and every observations should influence it unduly. If one or two very small or very large observations unduly affect the average, i.e., either increase its value or reduce its value, the average cannot be really typical of the entire set of data. In other words, extremes may distort the average and reduce its usefulness.

The following are the important measures of central tendency which are generally used in business:

A. Arithmetic mean.

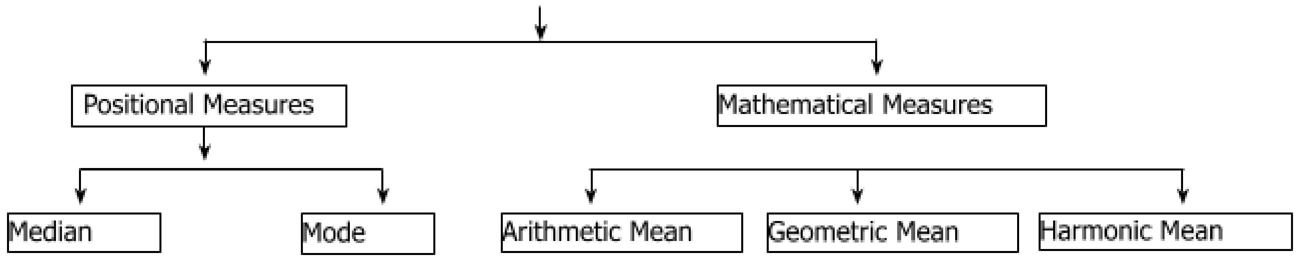
B. Median

C. Mode

D. Geometric mean, and

E. Harmonic mean

Measures of Central Tendency



I. ARITHMETIC MEAN: A.M is denoted by $\bar{X}$. It is a mathematical measurement. It is calculated by different methods of the following.

(a) **Individual Series: -**

(i) Direct Method: - $\bar{X} = \frac{\sum X}{N}$

(ii) Short Cut Method (or) Indirect Method: $\bar{X} = A + \frac{\Sigma dx}{N}$

(iii) Step Deviation method: $\bar{X} = A + \frac{\Sigma dx_1}{N} \times i$

Where 'N' is No. of terms in the given series.

Σx is Sum of terms, A is the assumed mean.

dx is the deviation of items from assumed mean i.e. $dx = x - A$

i is the common factor, $dx^1 = \frac{dx}{i}$

(b) **Discrete Series:**

(i) Direct Method: - $\bar{X} = \frac{\Sigma fx}{N \text{ or } \Sigma f}$

(ii) Short Cut Method (or) Indirect Method: $\bar{X} = A + \frac{\Sigma fdx}{N \text{ or } \Sigma f}$

(iii) Step Deviation method: $\bar{X} = A + \frac{\Sigma fdx_1}{N \text{ or } \Sigma f} \times i$

Where f is frequency, $N = \Sigma f$ = Total frequency

A is assumed mean, $dx = x - A$, $dx_1 = \frac{dx}{i}$

(c) **Continuous Series:**

(i) Direct Method: - $\bar{X} = \frac{\Sigma fm}{\Sigma f}$

(ii) Short Cut Method (or) Indirect Method: $\bar{X} = A + \frac{\Sigma fdx}{\Sigma f = N}$

(iii) Step Deviation method: $\bar{X} = A + \frac{\Sigma fdx'}{N} \times i$

Where A is assumed mean, m is the mid value of the class interval

f is the frequency, $N = \Sigma f$ = Total frequency, i is the common factor

$dx^1 = \frac{dx}{i}$, $dx = m - A$

(d) To Calculate Combined $\bar{X}$:

$$\bar{X}_{123\ldots n} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2 + \cdots + N_n\bar{X}_n}{N_1 + N_2 + \cdots + N_n}$$

$\bar{X}_{123\ldots\ldots..n}$ = Combined mean of the groups = $\bar{X}$

$\bar{X}_1$ = A.M. of first group, $\bar{X}_2$ = A.M. of second group $\bar{X}_n$ = A.M. of n^{th} group,

N_1 = No. of terms in the first group N_2 = No. of terms in the second group

N_n = No. of terms n^{th} group

Note: In case $N_1 = N_2 = N_3 = \cdots N_n$ then

$$\bar{X} = \frac{\bar{X}_1 + \bar{X}_2 + \cdots + \bar{X}_n}{n}$$

(e) TO DETERMINE CORRECT $\bar{X}$, WHEN SOME TERMS ARE INCLUDED WRONGLY:

Incorrect total of n terms = $n\bar{X}$

Correct Total = Incorrect total + Correct terms – In correct terms

$$\therefore \text{Correct } \bar{X} = \frac{\text{Correct Total}}{\text{No. of Terms}}$$

(f) WEIGHTED ARITHMETIC MEAN: $\bar{X}_w = \frac{\Sigma wx}{\Sigma w}$

Where x is variable, W is assigned weight, $\bar{X}_w$ is weighted A.M.

Examples:

1. Find Mean for the following figures.

30	41	47	54	23	34	37	51	53	47

Solution:

Adding all the terms and using the formula

$\bar{X} = \frac{\Sigma X}{N}$, Here N = 10. And Σ X = 417 n = number of observations = 10

$\therefore \frac{417}{10}$ = **41.7**

2. Calculate A.M. from the following data:

Marks obtained:	4	8	12	16	20
No. of students	6	12	18	15	9

Solution:

Marks X:	**No. of students**	
	F	fX
4	6	24
8	12	96
12	18	216
16	15	240

20	9	180
	N = 60	$\sum fx = 756$

As $\bar{X} = \frac{\sum fx}{N}$, $\bar{X} = \frac{756}{60}$ = **12.6**

3. Use direct method to find $\bar{X}$

Income	10-20	20-30	30-40	40-50	50-60	60-70
No. of Persons	4	7	16	20	15	8

Solution:

Income X	Mid Value m	No. of Persons f	fm
10-20	15	4	60
20-30	25	7	175
30-40	35	16	560
40-50	45	20	900
50-60	55	15	825
60-70	65	8	520

As $\bar{X} = \frac{\sum fx}{N} = \bar{X} = \frac{3040}{70}$ = **43.43**

4. $\bar{X}$ of 20 terms was found to be 35. But afterwards it was detected that two terms 42 and 34 were misread as 46 and 39 respectively. Find correct X.

Solution:

$\bar{X}$ of 20 terms = 35

(incorrect) Total of 20 terms = 35 x 20 = 700

Correct Total = 700 + 42 + 34 – 46 – 39 = 691

∴ Correct $\bar{X} = \frac{691}{20}$ = **34.55**

5. The mean of wages in factory A of 100 workers is ` 720 per week. The mean wages of 30 female workers in the factory was `650 per week. Find out average wage of male workers in the factory.

Solution:

N = 100, N_1 = 30, $\bar{X}_1$ = 650, $\bar{X}_2$ = ?

$N_1 + N_2 = N$ 30 + N_2 = 100 $N_2 = 70$

$\bar{X}_{12} = 720$

$\bar{X}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$

720 =

3
0

×

6
5
0

+

7
0
x
2
1
0
0

$$72000 = 19500 + 70\,\bar{X}_2$$

$$70\,\bar{X}_2 = 52500$$

$$\bar{X}_2 = \frac{52500}{70} = \mathbf{750}$$

Properties of Arithmetic Mean: The important properties of arithmetic mean are given below:

(i) The sum of the deviations of the terms from the Actual mean is always zero.

(ii) The sum of the squared deviations of the items from arithmetic mean is minimum i.e. less than the sum of the squared deviations of the items from any other value.

(iii) If we have arithmetic mean and the number of items of two or more than two groups, we can calculate the combined average of groups.

(iv) If the terms of a series are increased, decreased, multiplied or divided by some constant, the mean also increases, decreases, multiplied or are divided by the same constant.

(v) The standard error of the arithmetic mean is less than that of any other measure of central tendency.

II. **GEOMETRIC MEAN (g)**: The geometric mean is obtained by multiplying the values of the items together and then taking it to its root corresponding to the number of items. It is denoted by 'g'.

i.e., $g = \sqrt[n]{x_1 \times x_2 \times x_3 \ldots x_n}$

(a) Individual Series: $g = \text{Anti log}\left(\frac{\Sigma \log n}{n}\right)$

Where g is G.M, x is items, 'n' is No. of terms

(b) Discrete Series: $g = \text{Anti log}\left(\frac{\Sigma f \log x}{N}\right)$, Where N = total frequency

(c) Continuous Series: $g = \text{Anti log}\left(\frac{\Sigma f \log m}{N}\right)$, Where m is mid value of the C.I. $(N = \Sigma f)$

(d) Weighted Geometric mean: $g = \text{Anti log}\left(\frac{\Sigma W \log x}{\Sigma W}\right)$, Where W is weights

(e) Combined Geometric Mean: $g = \text{Anti log}\left(\frac{n_1 \log g_1 + n_2 \log g_2}{n_1 + n_2}\right)$

- The Geometric mean is relative value and is dependent on all items
- The geometric mean is never larger than the arithmetic mean. It is rare that it may be equal to the arithmetic mean.
- The Geometric mean of the products of corresponding items in two series is equal to product of their geometric mean.

III. **HARMONIC MEAN**: Harmonic mean of a given series is the reciprocal of the arithmetic average of the reciprocal of the values of its various items.

(a) **Individual Series**: H.M. = $\dfrac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$

where n is No. of items and x_i is $x_1, x_2, x_3 \ldots\ldots x_n$

(b) **Discrete Series**: H.M. = $\dfrac{N}{\sum \frac{f}{x}}$

Where N = ∑ f = Total frequency, f is frequency

(c) **Continuous Series**: $H.M = \frac{N}{\sum \frac{f}{m}}$, Where m is mid value of the C.I, $N = \sum f$

(d) **Weighted H.M**: $H.M_w = \frac{\sum W}{\sum \frac{W}{x}}$, Where W is weights

(e) **Let x, y are two numbers then**: $A.M = \frac{x+y}{2}$, $H.M = \frac{2xy}{x+y}$, $G.M = \sqrt{xy}$

(f) **Relationship among the Average**: In any distribution where the original items differ in size, then either the values of A.M > G.M > H.M (or)

H.M < G.M < A.M in case all items are identical then A.M. = G.M = H.M

(g) **Combined Harmonic Mean**: $\left[\frac{N_1 + N_2}{\frac{N_1}{H.M_1} + \frac{N_2}{H.M_2}}\right]$

IV. MEDIAN AND OTHER POSITIONAL MEASURES: Median is denoted by M. It is a positional measurement. Median is dividing a series in two equal parts. i.e., the middle most items are called median.

(a) Individual Series: The terms are arranged in ascending (or) descending order.

(1) When number of terms is odd then

Median (M) = size of $\left(\frac{N+1}{2}\right)$ th item

Where M is Median, N is No. of terms in the given series.

(2) When number of terms is even, then Median (M) =

$\left[\left(\frac{N}{2}\right) + \left(\frac{N}{2} + 1\right)\right]$

mean of $\left(\frac{N}{2}\right)$ and $\left(\frac{N}{2}+1\right)$ th terms = $\frac{\left[\left(\frac{N}{2}\right) + \left(\frac{N}{2}+1\right)\right]}{2}$ the terms.

(b) **Discrete Series**:

Median (M) = Size of $\left(\frac{N+1}{2}\right)$ th item

Where N = $\sum f$ = Total frequency

(c) **Continuous Series**:

Median (M) = $L_1 + \frac{N_1 - c.f}{f} \times c$

Where L_1 is lower limit of median class interval

c.f is the value in column just above $N_1 = \frac{N}{2}$

f is the frequency of median class, c is the class interval of median class

(d) **Quartiles, Quintiles, Octiles, Deciles and Percentiles**:

(1) Individual Series: First arrange the items in Ascending order

First Quartile (or) Lower Quartile Q_1 = size of $\left(\frac{N+1}{4}\right)^{th}$ term

Third Quartile (or) Upper Quartile Q_3 = size of $3\left(\frac{N+1}{4}\right)^{th}$ term

n^{th} quintile q_n $= \frac{n(N+1)}{5}^{th}$ term

n^{th} Octile o_n $= \frac{n(N+1)}{8}^{th}$ term

n^{th} Decile D_n $= \frac{n(N+1)}{10}^{th}$ term

n^{th} Percentile p_n $= \frac{n(N+1)}{100}^{th}$ term

Where N is total No. of items in the given series In case of quintile n = 1, 2, 3, 4

Octile n = 1, 2, 3, 4, 5, 6, 7

Decile n = 1, 2, 3, 4, 5, 6, 7, 8, 9

Percentile n=1, 2, 3,… 99

(ii) **Discrete Series**: First find out cumulative frequency column.

First Quartile (or) Lower Quartile Q_1 = size of $\left(\frac{N+1}{4}\right)^{th}$ item

Third Quartile (or) Upper Quartile Q_3 = size of $3\left(\frac{N+1}{4}\right)^{th}$ item

n^{th} quintile q_n = Size of $n(N+1)^{th}$ item

n^{th} Octile O_n $= \text{Size of } \frac{n(N \pm 1)}{8}^{th}$ item

n^{th} Decile D_n $= \text{Size of } \frac{n(N \pm 1)}{10}^{th}$ item

n^{th} Percentile P_n $= \text{Size of } \frac{n(N \pm 1)}{100}^{th}$ item

Where $N = \sum f$ = Total frequency

(iii) **Continuous Series**: i.e. $= L_1 + \frac{N_1 - c.f}{f} \times c$

Where $N_1 = \frac{N}{2}$ for Median

$= \frac{N}{4}$ for first Quartile (Q_1)

$= 3\frac{N}{4}$ for third quartile (Q_3)

$= \frac{nN}{5}$

$= \frac{nN}{8}$

)

f

o

r

n

t

h

q

u

i

n

t

i

l

e

(

q

n

)

for n^{th} Octile (O_n)

$= \frac{nN}{10}$ for n^{th} Deciles (D_n)

$= \frac{nN}{100}$ for n^{th} Percentile (P_n)

6. Find Median from following data:

17	19	21	13	16	18	24	22	20

Solution:

Arranging the terms in ascending order

13	16	17	18	19	20	21	22	24

Total number of terms = 9 or n = 9

Now $\frac{n+1}{2} = \frac{9+1}{2}$ 5

Median = 5^{th} term = 19

7. Compute Median for following data:

X:	10	20	30	40	50	60	70
F:	4	7	21	34	25	12	3

Solution:

X	**F**	**C_f**
10	4	=4
20	7	(4+7)=11

30	21	(11+21)=32
40	34	(32+34)=66
50	25	(66+25)=91
60	12	(91+12)=103
70	3	(103+3)=106
	N =106	

Now M = Size of $\left[\frac{N \pm 1}{2}\right]^{th}$ term

$= \text{size of } \left[\frac{106+1}{2}\right]^{th} \text{term}$

$= \text{Size of } 53.5^{th} \text{ term}$

Thus Median = **40**

8. Calculate Median from following data. Case of unequal class-internals.

Class Intervals	4-8	8-20	20-28	28-40	40-60	60-72
Frequency	7	12	42	56	39	22

Solution:

X	**f**	**C_f**
4-8	7	7
8-20	12	19
20-28	42	61
28-40	56	117
40-60	39	156
60-72	22	178
	N = 178	

$N_1 = \frac{178}{2} = 89$, Cf = 61, f = 56, L = 28, i = 12,

$$M = L + \frac{N_1 - c.f}{f} \times i = 28 + \frac{89 - 61}{56} \times 12 = \mathbf{34}$$

V. MODE: It is denoted by 'Z'. Mode may be defined as the value that occurs most frequently in a statistical distribution.

(i) **Individual Series**: The terms are arranged in any order, Ascending or Descending. If each term of the series is occurring once, then thee is no mode, otherwise the value that occurs maximum times are known as Mode.

(ii) **Discrete Series**: Here the mode is known by Inspection Method only. Here that variable is the mode, where the frequency is highest. For such a distribution we have to prepare (a) Grouping Table (b) Analysis Table.

(iii) **Continuous Series**:

$$\text{Mode (Z)} = L_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times C \text{ (or) } Z = L_1 + \frac{D_1}{D_1 + D_2} \times C$$

1 + D 2

Where L_1 is the lower limit of modal class interval

f_1 is the frequency corresponding to modal class interval

f_0 is the frequency preceding Modal class interval f_2 is the frequency succeeding Modal class interval C is the length of Modal class interval

$D_1 = f_1 - f_0$, $D_2 = f_1 - f_2$

Note: Class intervals must be exclusive, equal, in ascending order, not cumulative. If Modal value lies in any other interval than with highest frequency, then

$$\text{Mode } (Z) = L_1 + \frac{f_2}{f_0 + f_2} \times C$$

9. Find Mode from the following data.

12	14	16	18	26	16	20	16	11	12	16	16	20	24

Solution:

Arrange above data in ascending order

11	12	12	14	15	16	16	16	16	18	20	20	24	26

Here we get 16 four times, 12 and 20 two times each and other terms once only. Thus Z = 16.

10. Find Mode from the following data

X:	5	10	15	20	25	30	35	40	45
F:	1	3	4	9	11	12	3	2	1

Solution:

(Note: - Since we can't make use of inspection method as the frequencies are not most concentrated about highest frequency 12. Thus we will have to proceed for the tables.

Grouping Table:

X	f(I)	II	III	IV	V	VI
5	1					
10	3	4				
15	4				16	
20	9	13	20			24
25	11			32		
30	12	23	15		26	
35	3					17
40	2	5	4	7		
45	2					

Analysis Table:

Column X	**I**	**II**	**III**	**IV**	**V**	**VI**	**Total**
5							-
10							-
15						X	1
20			X	X		X	3
25		X	X	X	X	X	5
30	X	X		X	X		4
35					X		1
40							-
45							-

Here 25 has occurred maximum times (5), thus Modal Value is 25

11. Calculate Mode for the following data

C.I	0-10	10-20	20-30	30-40	40-50	50-60	60-70
F:	4	13	21	44	33	22	7

Solution:

Grouping Table

C.I	**f(I)**	**II**	**III**	**IV**	**V**	**VI**
0-10	4					
10-20	13		34	38		
20-30	21 f_0	65			78	
30-40	44 f_1		77			98
40-50	33 f_2			99		
50-60	22	55			62	
60-70	7		29			

Analysis Table

C.I	I	II	III	IV	V	Total
0-10						-
10-20					X	1
20-30		X			X	2
30-40	X	X	X	X	X	5
40-50			X	X		2
50-60				X		1
60-70						-

Thus Modal Interval is 30-40

L = 30

$D_1 = 44 - 21 = 23$; $D_2 = 44 - 33 = 11$

$i = 10$

As $Z = L + \dfrac{D_1}{D_1 + D_2} \times i$ or, $Z = L + \dfrac{Z_1 + Z_0}{2Z_1 - Z_0 - Z_2} \times i$, $(Z_1 = 44, Z_0 = 21, Z_2 = 33)$

$$Z = 30 + \frac{23}{23 + 11} \times 10 = 30 + \frac{230}{34} = 30 + 6.76 = \mathbf{36.76}$$

(iv) Mean, Median and Mode – Their Relation: $Z = 3M - 2\bar{X}$

Where Z is Mode, M is Median, $\bar{X}$ is Mean

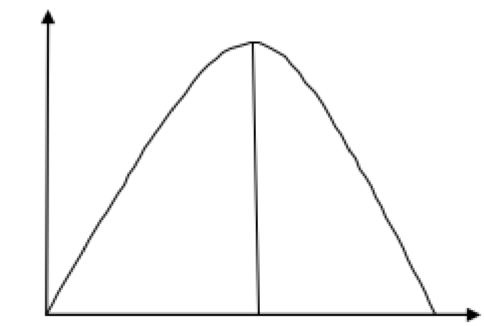
This formula was expressed by Karl Pearson. It is called Bi-mode.

(v) Incase of Symmetrical Series, the mean, median and mode coincide. i.e., $Z = M = \bar{X}$

(vi) Incase of Positive Skewed: If the tail is towards right, then it is called positive skewness. It means $Z < M < \bar{X}$

(vii) Negative Skewed: When the tail of a distribution is towards left, then the skew ness is negative i.e. $Z > M > \bar{X}$

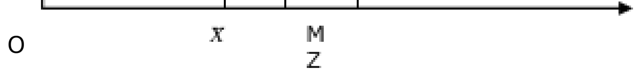
(viii) It should be noted that in both '+'ve and '-'ve skewed distributions that median lies in between the mode and the mean.



Symmetrical

O

Positive Skewed



Negative Skewed

Uses of various Averages: The use or application of a particular average depends upon the purpose of the investigation. Some of the cases of different averages are as follows:

(a) **Arithmetic Mean**: Arithmetic mean is considered an ideal average. It is frequently used in all the aspects of life. It possesses many mathematical properties and due to this it is of immense utility in further statistical analysis. In economic analysis arithmetic mean is used extensively to calculate average production, average wage, average cost, per capita income, exports, imports, consumption, prices etc. When different items of a series have different relative importance, then weighted arithmetic mean is used.

(b) **Geometric Mean**: Use of Geometric mean is important in a series having items of wide dispersion. It is used in the construction of Index Number. The averages of proportions, percentages and compound rates are computed by geometric mean. The growth of population is measured in it as population increases in geometric progression.

(c) **Harmonic Mean**: Harmonic mean is applied in the problems where small items must get more relative importance than the large ones. It is useful in cases where time, speed, values given in quantities, rate and prices are involved. But in practice, it has little applicability.

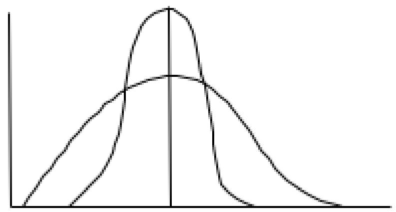
(d) **Median and Partition values**: Median and partition values are positional measures of central tendency. These are mainly used in the qualitative cases like honesty, intelligence, ability etc. In the distributions which are positively skewed, median is a more suitable average. These are also suitable for the problems of distribution of income, wealth, investment etc.

(e) **Mode**: Mode is also positional average. Its applicability to daily problems is increasing. Mode is used to calculate the 'modal size of a collar', modal size of shoes,' or 'modal size of ready-made garments' etc. It is also used in the sciences of Biology, Meteorology, Business and Industry.

The various measures of central tendency discussed in the previous chapter give us one single value that represents the entire data. But the average alone cannot adequately describe a set of observations, unless all the observations are alike. It is necessary to describe the variability or dispersion of the observations. Also in two or more distributions the average value may be the same but still there can be wide disparities in the formation of the distributions. Measures of variation help us in studying the important characteristic of a distribution, i.e., the extent to which the observations vary from one another from some average value. The significance of the measure of variation can best be appreciated from the following example:

	Factory A wages (`)	**Factory B Wages (`)**	**Factory C Wages (`)**
	2300	2310	2380
	2300	2300	2210
	2300	2304	2220
	2300	2306	2200
	2300	2280	2490
Total:	11,500	11,500	11,500
$\bar{X}$	2,300	2,300	2,300

The above data pertains to five workers each in three different factories. Since the average wage is the same in all factories, one is likely to conclude that the factories are alike in their wage structure, but a close examination shall revel that the wage distribution in the three factories differs widely from one another. In factory A, each and every worker is perfectly represented by the arithmetic mean, i.e., average wage or, in other words, none of the workers of factory A deviates from the arithmetic mean and hence there is no variation. In factory B, only one worker is perfectly represented by the arithmetic mean, the other workers vary from the mean but the variation is very small as compared to the workers of factory C. In factory C, the mean does not represent the workers as the individual wage figures differ widely from the mean. Thus we find there is no variation in the wages of workers in factory A, there is very little variation in factory B but the wages of workers of factory C differ most widely. For the student of social sciences, the mean wage is not so important as to know how these wages are distributed. Are there a large number receiving the mean wage or are there a few with enormous wages and millions with wages far below the mean? The following three diagrams represent frequency distribution with some of the characteristics we wish to emphasize:

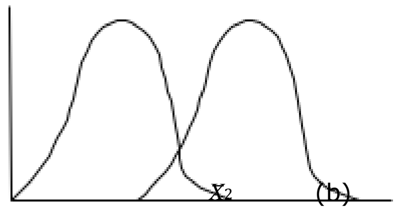


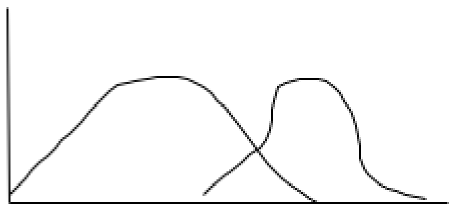
O O

$\bar{X}_1$

$\bar{X}$

(a





O $\bar{X}_1$ $\bar{X}_2$ (c)

The two curves in diagram (a) represent two distributions with the same mean $\bar{X}$, but with different variations. The two curves in (b) represent two distributions with the same variations but with unequal means, $\bar{X}_1$ and $\bar{X}_2$, Finally, (c) represent two distributions with unequal means and unequal variations.

The measures of central tendency are therefore, insufficient. They must be supported and supplemented with other measures. In this chapter, we shall be especially concerned with the measures of variation (or spread, or dispersion). Measure of variation is designed to state the extent to which the individual measures differ on an average from the mean. In measuring variation we shall be interested in the amount of the variation or its degree but not in the direction. For example, a measure of 6 centimeters below the mean has just as much variation as a measure of 6 centimeters above the mean.

Significance of Measuring Variation

Measures of variation are needed for four basic purposes:

(i) To determine the reliability of an average;

(ii) To serve as a basic for the control of the variability

(iii) To compare two or more series with regard to their variability; and

(iv) To facilitate the use of other statistical measures

A brief explanation of these points is given below:

(i) Measures of variation point out as to how far an average is representative of the entire data. When variation is small, the average is a typical value in the sense tht it closely represents the individual value and it sis reliable in the sense that it closely represents the individual value and it is reliable in the sense that it is good estimate of the average in the corresponding universe. On the other hand, when variation is large, the average is not so typical, and unless the sample is very large, the average may be quite unreliable.

(ii) Another purpose of measuring variation is to determine nature and cause of variation in order to control the variation itself. It matters of health, variation in body temperature, pulse beat and blood pressure are the basic guides to diagnosis. Prescribed treatment is designed to control their variation. In industrial production, efficient operation requires control of quality variation, the causes of which are sought through inspection and quality control programmes. Thus measurement of variation is basic to the control of cause of variation. In engineering problems, measures of variation are often specially important, in social sciences, a special problem requiring the measurement of variability is the measurement of "inequality" of the distribution of income and wealth, etc.

(iii) Measures of variation enable comparison to be made of two or more series with regard to their variability. The study of variation may also be looked upon as a means of determining uniformity or consistency. A high degree of variation would mean little uniformity or consistency whereas a low degree of variation would mean greater uniformity or consistency.

(iv) Many powerful analytical tools in statistics such as correlation analysis, the testing of hypothesis, the analysis of fluctuations, techniques of production control cost control, etc, are based on measures of variation of one kind or another.

Properties of a Good Measure of Variation: A good measure of variation should possess, as far as possible, the following properties:

(i) It should be simple to understand

(ii) It should be easy to compute

(iii) It should be rigidly defined.

(iv) It should be based on each and every observation of the distribution.

(v) It should be amendable to further algebraic treatment

(vi) It should have sampling stability

(vii) It should not be unduly affected by extreme observations.

DISPERSION: "Dispersion is a measure of variation of the items.

Methods of Measuring DISPERSION:

Types of Measures of Dispersion:

(A) **Absolute and Relative Measures**: Absolute measures of Dispersion are expressed in same units in which original data is presented but these measures cannot be used to compare the variations between the two series.

Relative measures are not expressed in units but it is a pure number.

It is the ratio of absolute dispersion to an appropriate average such as co-efficient of Standard Deviation or Co-efficient of Mean Deviation.

(B) **Methods of Measuring Dispersion**: Following methods are used to calculate dispersion.

(1) **Algebraic methods**:

(i) Methods of limits: (a) The Range (b) The Inter-quartile Range (c) The Percentile Range

(ii) Methods of moments: a) the first moment of dispersion or mean deviation (b) The second moment of dispersion or standard deviation.

(2) **Graphic Method**: Lorenz curve.

RANGE: It is the simplest of the values of Dispersion. It is merely the difference between the largest and smallest term. Symbolically; R = L - S (or) Range = Largest term – Smallest term and Coefficient of Range = $\frac{L - S}{L + S}$. It is also known as Ratio of Range or Co-efficient of scatteredness.

If the averages of the two distributions are close to each other, comparisons of the ranges show that the distribution with the smaller range has less dispersion. The average of that distribution is more typical of the group.

Methods of limits:

(i) **Range**: The difference between the largest and smallest term. It is denoted by 'R'.

Individual Series, Discrete Series and continuous series

Range = Largest term – smallest term

i.e., R = L-S and Coefficient of Range $\frac{L - S}{L + S}$

Note: Here Range is absolute measure and Co-efficient of Range is Relative measure.

Examples:

1. Find Range and coefficient of Range for following data

3	7	21	24	37	40	45

Solution:

Here $L = 45$ and $S = 3$

As Range = L-S

$\therefore$ Range = 45-3 = 42

And Coefficient of Range = $\frac{L - S}{L + S} = \frac{45 - 3}{45 + 3} = \frac{42}{48} = 0.875$

(Note Here range is absolute measure and Co-efficient of Range is Relative measure).

2. Find Range and Coefficient of Range for following data

X	5	10	15	20	25	30	35	40
f	4	7	21	47	53	24	12	6

Solution:

Going through the variables S = 5; and L = 40

∴ Range = 40-5 = 35 (R = L-S)

And Coefficient of Range $= \frac{40 - 5}{40 + 5}$ $\quad C.R = \frac{L - S}{L + S}$

$= \frac{35}{45} = 0.778$ (Approx.)

(Note Here range is absolute measure and Co-efficient of Range is Relative measure).

3. Calculate Range and Coefficient of Range for following data

X	20-30	30-40	40-50	50-60	60-70	70-80
F	4	9	16	21	13	6

Solution:

Here L = 80 and S = 20

∴ Range = L-S = 80 – 20 = 60

And Coefficient of Range = $\frac{L - S}{L + S} = \frac{80 - 20}{80 + 20} = \frac{60}{100} = 0.6$

(ii) **Inter Quartile Range** = $Q_3 - Q_1$

Semi Inter Quartile Range = $\frac{Q_3 - Q_1}{}$

(or) **Quartile Deviation**

Where Q_1 is lower Quartile, Q_3 is upper Quartile

Coefficient of Quartile Deviation $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

(iii) **Percentile and Deciles Range**: **Percentile Range** $= P_{90} - P_{10}$

Deciles Range $= D_9 - D_1$

Where P_{10} is 10^{th} percentile and P_{90} is 90^{th} percentile.

D_1 is first Deciles and D_9 is 9^{th} Deciles.

4. From the following data compute inter quartile range, quartile Deviation and Coefficient of Quartile Deviation.

24	7	11	9	17	3	20	14	4	22	27

Solution:

Arranging the series in Ascending Order

3	4	7	9	11	14	17	20	22	24	27

N = 11

Q_1 = Size of $\left[\frac{11+1}{4}\right]^{th}$ term

= Size of 3^{rd} term = 7

Q_3 = Size of $\frac{3(11+1)}{4}^{th}$ term

= Size of 9^{th} term

= 22

$\therefore$ Inter Quartile Range = $Q_3 - Q_1 = 22 - 7 = 15$

Semi – Inter Quartile Range or Quartile Deviation

(Q.D) = $\frac{Q_3 - Q_1}{2} = \frac{15}{2} = 7.5$

Coefficient of Q.D = $\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{22 - 7}{22 + 7} = \frac{15}{19} = 0.52$ (App.)

5. Compute Inter quartile Range, Coefficient of Quartile Deviation and Percentile Range for following data.

X	4	8	12	16	20	24	28	32
f	4	9	17	40	53	37	24	16

Solution:

X	f	C.f
4	4	4
8	9	13
12	17	30

16	40	70
20	53	123
24	37	160
28	24	184
32	16	200
	N =200	

$$Q_1 = \text{Size of } \left[\frac{200 \pm 1}{4}\right]^{th} \text{term}$$

$= \text{Size of } 50.25\text{th term}$

$\therefore Q_1 = 16$

$$Q_3 = \text{Size of } \frac{3(200 \pm 1)}{4}^{th} \text{term}$$

$= \text{Size of } 150.75^{th} \text{ term}$

$\therefore Q_3 = 24$

Inter Quartile Range = $Q_3 - Q_1 = 24 - 16 = 8$

Coefficient of Q.D = $\dfrac{Q_3 - Q_1}{Q_3 + Q_1} = \dfrac{24 - 16}{24 + 16} = \dfrac{8}{40} = 0.2$

To Compute Percentile Range : -

P_{90} = Size of $\dfrac{90\,(200 \pm 1)}{100}^{th}$ term

= Size of 180.9^{th} term = 28

P_{10} = Size of $\dfrac{10(200 \pm 1)}{100}^{th}$ term

= Size of 20.1^{th} term = 12

And Percentile Range = $P_{90} - P_{10} = 28 - 12 = 16$

6. From the following data find quartile deviation and its coefficient.

X	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50
f	8	10	12	15	10	7	8	5

Solution:

X	**f**	**C.f**
10-15	8	8
15-20	10	18
20-25	12	30
25-30	15	45
30-35	10	55
35-40	7	62
40-45	8	70
45-50	5	75

Now N_1 for Q_1 = $\dfrac{N}{4}^{th}$ item

= $\dfrac{75}{4}^{th}$ item = 18.75^{th} item

18.75^{th} item lies in (20-25)

Where L = 20, f = 12

N_1 = 18.75 Cf = 18; i=5

Interpolating for Q_1

$$Q_1 = L_1 + \frac{N_1 - Cf}{f} \times i$$

$$= 20 + \frac{18.75 - 18}{12} \times 5 = 20 + \frac{3.75}{12} = 23.125$$

Similarly N_1 for $Q_3 = \frac{3N}{4}$ th item 12

$N_1 = \frac{3 \times 74}{4}$ th item = 56.25th Item

56.25th Item lies in (35-40)

Where L = 35, I = 5, C.f = 55, f = 7, N_1 = 56.25

Putting value in the interpolation formula Where,

L = Lower limit of Median Class

Cf = Preceeding Cumulative frequency of median class f =

Corresponding frequency of median class.

i = Width of median class.

$$Q_3 = l + \frac{N_1 - C.f}{f} \times i$$

$$= 35 + \frac{56.25 - 55}{7} \times 5$$

$$= 35 + \frac{6.25}{7} = 35.893$$

Quartile Deviation (Inter Quartile Range)

= Q_3-Q_1

Q.D = 35.893 – 23.125

= 12.768

$$\text{Coefficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{35.893 - 23.125}{35.893 + 23.125} = \frac{12.768}{59.018} = 0.216$$

Methods of moments:

(i) **Mean Deviation (or) Average Deviation**: It is denoted by M.D.

$$\text{Mean Deviation (M.D)} = \frac{\sum f|D|}{N}$$

Where f is frequency, N is total frequency |D|=|X − A|

Coefficient of Mean Deviation:

(a) If deviations are taken from arithmetic mean $(\bar{X})$, Coefficient of M.D = $\frac{M.D}{\bar{X}}$; where $\bar{X}$ is mean

(b) If deviations are taken from median, (M), Coefficient of M.D = $\frac{M.D}{M}$; where M is median

(c) If deviations are taken from mode, (Z), Coefficient of M.D = $\frac{M.D}{}$;

Z

where Z is mode.

Individual Series:

Mean Deviation (M.D) = $\frac{\Sigma |D|}{n}$, Where $|D| = |x - \bar{X}|$, $\bar{X}$ = mean

= $|x - M|$, M = Median

= $|x - Z|$, Z = Mode

n = No. of terms in the given series. Coefficient of Mean Deviation = $\frac{M.D}{\bar{X} \text{ or } M \text{ or } Z}$

Discrete Series:

Mean Deviation (M.D) = $\frac{\sum f|D|}{N}$, Where N = ∑ f = Total frequency

$|D| = |x - \bar{X}| = |x - M| = |x - Z|$

Continuous Series: Mean Deviation (M.D.) = $\frac{\sum f|D|}{N}$

Where $|D| = |m - \bar{X}| = |m - M| = |m - Z|$, m is mid value of the class interval.

7. Compute M.D. and Coefficient of M.D. form mean and median for following series.

3	7	12	14	15	18	22

Solution:

(1) $\bar{X} = \frac{\sum x}{N}$

∑ X = 3 + 7 + 12 + 14 + 15 + 18 + 22 = 91

n = 7

$\bar{X} = \frac{91}{7} = 13$

X	dy = $\|X - \bar{X}\|$
3	10
7	6
12	1
14	1
15	2
18	5
22	9
	∑ dy = 34

Now ∑ dy = 34
N = 7

And M.D. = $\frac{\sum dy}{n} = \frac{34}{7} = 4.86$

∴ M.D. FROM Mean = 4.86

And Coefficient of Mean Deviation from Mean = $\frac{\text{M.D. from Mean}}{\text{Mean}}$

$= \frac{4.86}{13} = 0.374$

(2) n = 7; ∴ Median is 4^{th} term = 14.

X	dy = |X − M|
3	11
7	7
12	2
14	0
15	1
18	4
22	8
	∑ dy = 33

Now ∑ dy = 33, n = 7

And M.D. $= \frac{\sum dy}{} = \frac{33}{} = 4.71$

about median

And Coefficient of Mean Deviation from Median $= \frac{\text{M.D. from Median}}{\text{Median}}$

$= \frac{4.71}{14} = 0.336$

8. Compute M.D. from $\bar{X}$ and M for given series

X	5	10	15	20	25	30
f	3	4	8	12	7	2

Solution:

X	**f**	**fx**	**C_f**	**dy = |X – $\bar{X}$|**	**fdy**	**dy = |X – M|**	**fdy**
5	3	15	3	13.06	39.18	15	45
10	4	40	7	8.06	32.24	10	40
15	8	120	15	3.06	24.48	5	40
20	12	240	27	1.94	23.28	0	0
25	7	175	34	6.94	48.58	5	35
30	2	60	36	11.94	23.88	10	20
	N = 36	∑ f x = 650			191.64		180

$\bar{X} = \frac{\sum fx}{N} = \frac{650}{36}$ = 18.06 (Approx.)

Median will in $\frac{N+1}{2}$ th term = 18.5th term . Hence Median = 20

M.D about median $= \frac{180}{36}$

= 5.323

M.D. from $\bar{X} = \frac{\sum fdy}{N} = \frac{191.64}{36}$

9. Compute Coefficient of M.D. from $\bar{X}$, M and Z, for following series where N = 100

Class Intervals	0-10	10-20	20-30	30-40	40-50
Frequency	6	28	51	11	4

Solution:

A = 25; i = 10; N = 100

C.I	X	f	dx	fdx	dy = $\lvert X - \bar{X} \rvert$	fdy	fX
0-10	5		-2	-12	17.9	107.4	
10-20	15		-1	-28	7.9	221.2	
20-30	25		0	0	2.1	107.1	
30-40	35	6 } 34	1	11	12.1	133.1	30 } 450
40-50	45		2	8	22.1	88.4	
		28 51 11 66 4					420 1275 385 1840 180
		N = 100		-21		657.2	

As $\bar{X} = A + \frac{\sum fdx}{N} \times i$

$\bar{X} \;\; 25 \;\; - \frac{-21}{} \;\; 10 = 22.9$