Chapter 1

Introduction to Statistics

1.1. STATISTICS

1.1.1. Meaning

The origin of the word **'statistics'** can be traced back to either the Latin word *Status* or the Italian word ***Statista*** or German word ***Statistik***. All the three words mean, 'an organised political state'. 'Statistics' was regarded as the **'science of statecraft'** in early years because it was mainly used by the administration to keep track of births, death, crop yield, taxes, etc. It was referred to as the 'science of king' or 'king' as it was used as a tool for government record collection. In present time, statistics is regarded as an important tool for decision-making in uncertain times.

For example, suppose a sample data set, which records the heights of university students. This data is used differently by different users such as the manufacturer of graduation gowns and the architect. The gown manufacturer would be interested to know the average height of the students to decide the height of the gown whereas the architect would want to know the maximum height of the students to decide the height of doorways as he would base his measurements on it.

Any analysis of figures used for forecasting can be considered as statistics. This can also be done by diagrammatic representation of facts. Processing, analysis and application of data or quantitative facts can be referred to as statistics. There are three forms of statistics:

As a Product	Data.
As a Process	Statistical methods.
As an Application	Methods & theories that are used to study numerical data for inferential/decision making purposes.

1.1.2. Definition

Two separate concepts of statistics have emerged from various definitions by statisticians:

1) Descriptive Statistics or Statistical Data, and
2) Statistical Methods.

According to the first concept (Descriptive Statistics or Statistical data), statistics is expressed as pertaining to numerical data. This concept takes statistics in the plural sense.

According to the second concept (Statistical methods) statistics is expressed as a science. It takes 'Statistics' in the singular sense.

1.1.2.1. Statistics as Statistical Data

Statistical data refers to the collection of quantitative data such as price of a commodity, production, export, import, births, deaths, etc.

According to Bowley, "Statistics are numerical statement of facts in any department of enquiry placed in relation to each other".

This definition gives importance to quantitative aspects and helps in comparative study of figures.

According to Webster, "Classified facts respecting the conditions of the people in a state especially those facts which can be stated in numbers or in any other tabular or classified arrangements".

The above definition denotes that the facts which are expressed quantitatively can be termed as statistics and qualitative facts cannot be treated statistically.

According to Yule and Kendall, "By statistics we mean quantitative data affected to a marked extent by multiplicity of cause".

According to Boddington, "Statistics is a science of estimates and probabilities".

Though statistics is a general term used often but many people understand it on their own perceptions. Since statistics is a numerical facts however, it is considered as a system of methods used to take decisions at the time of uncertainty arises. This uncertainty arises due to the incompleteness of available information. The statistical science is a method used to judge the collective, natural or social phenomenon form the outcomes acquired from the collection or analysis the estimates.

Boddington has defined statistics as the science of "estimates and probabilities". The above definition provides the statements only to specific methods that derive the conclusions in this science. In many of the cases statistics are, estimates and probabilities without any doubt. However, it should be noted that only these things cannot confined the scope of science.

For example, consider a class of 100 students. We calculated the averages marks of the students by using the arithmetic formula and find out the average marks of the students are 45. It means all the students got the marks 45 nearly. However this is not true practically. It is because there are variations between the marks of students. Hence, if the conclusion is derived from the advanced statistical techniques then statistics will not be a science of estimates and probabilities.

1.1.2.2. Statistics as Statistical Methods

Statistics as a 'science of statistical methods' is based on the idea that, "Statistics is what statistics does or statistics is what statisticians do". This definition considers statistics as a subject or method.

According to A.L. Bowley, "Statistics may be called the science of counting".

In this definition only one aspect is covered, i.e., counting. But, in many cases data collection also depends on estimation, which helps in further classification, tabulation, etc.

According to A.L.Bowley, "Statistics may rightly be called the science of averages".

Statistics uses averages to summarise the collected data. However, it also uses, diagrams, graphs, correlation coefficient, etc. for analysis, which has not been included in the above definition of statistics and hence make it incomplete.

According to Wallis and Roberts, "Statistics is a body of methods for making decisions in the face of uncertainty".

According to Croxton and Cowden, "Statistics may be defined as the collection, presentation, analysis and interpretation of numerical data".

1.1.3. Types of Statistics

Statistics can be classified into two main categories:

1) **Descriptive Statistics:** This includes the methods of data collection, presentation and characterisation of a set of data. All these help in describing the various features of the collected sample data. Descriptive statistics includes graphical representation and quantitative measures. **For example,** graphical representation includes bar charts, line graphs and pie charts whereas measures of central tendency, dispersion, skewness and kurtosis denote quantitative measures.
2) **Inferential Statistics:** Inferential statistics includes the methods that help in characterising a population or help in decision-making which is based on the sample results of the population.

Sample and population are two relative terms. The larger unit about which analysis is to be done is referred to as 'population' or 'universe' and the fraction or a portion of that population/universe is called 'sample'.

1.1.4. Characteristics of Statistics

Basic characteristics of statistics are as follows:

1) Statistics are averages of numerical data. Single or isolated figures cannot be considered as statistics as they cannot help in analysis or drawing of conclusions. Single death or birth does not comprise statistics but a number of such figures help in determining the average death or birth rate.
2) Generally, statistics contains quantitative information which means it is always expressed numerically. Qualitative facts can never be measured numerically and therefore it cannot give accurate interpretations. Hence, this can never be considered as statistics.
3) Statistics are affected by various causes or factors. One cause or factor cannot be said to be responsible for the results after analysis of a given set of data.
4) Statistical data is collected in a systematic manner for a predetermined purpose. This means that the reason and plan for data collection is predetermined.
5) Statistical data may be used for comparative studies. **For example,** for comparing the efficiency of two departments in an office, the performance appraisal can be used.

1.1.5. Nature of Statistics

The question of whether Statistics is an art or science is often a subject of debate. In order to conclude whether it is a science or an art, it is necessary to know the true meaning of science and art.

1.1.5.1. Is Statistics as a Science

The systematic study of cause and effect relationship and the effort of making a generalisation in terms of scientific principles of laws is called **science**. It describes everything objectively without terming anything vague as good or bad. Like scientific study, statistical methods also answer various questions such as 'how an investigation should be conducted', 'what are the ways to find valid and reliable conclusions' and 'how far the conclusions are dependable'.

The science of statistics is very different from traditional sciences such as Physics, Chemistry, etc. Sometimes, it is referred to as the science of scientific methods. This is the reason why statisticians consider it as a scientific method instead of pure science.

According to Croxton and Cowden, "Statistics is not a science, it is a scientific method".

According to Tippet, "As science, the statistical method is a part of the general scientific method and is based on the same fundamental ideas and processes".

1.1.5.2. Is Statistics as an Art

Art is an applied knowledge. It refers to the expertise of handling facts to attain a given objective. It is concerned with ways and means of presenting and handling data, making logical inferences and drawing relevant conclusions. In this respect, statistics is also an art as it states how to use statistical rules and principles in order to study the problems and get their solutions. Various types of difficult problems can be solved with the use of statistical methods. The various rules of statistics help in attaining the desired results. Statistical methods are applied for obtaining the conclusion of the various aspects which are related to art such as 'which statistical methods are used to solve special type of problems' or 'how the two facts are compared or correlated'. Hence statistics is called an art.

1.1.5.3. Statistics is both Science and Art

As statistics has both the elements of art and science hence it can be said that statistics is both science as well as an art. It is used to draw conclusions from a set of facts and figures. This inference drawing or conclusive study is important for economic and social progress.

According to Tippet, "It (statistics) is both a science and an art. It is a science in that its methods are basically systematic and have general application and an art in that their successful application depends to a considerable degree on the skill and special experience of the statistician and his knowledge of the field of application, e.g., Economics".

1.1.6. Scope/Applications of Statistics

Statistics has affected almost all areas of life as it covers simple households to big businesses, even government also. Some of the areas where statistics has been used are as follows:

1) **Statistics and the Government:** Statistics has been used extensively since the beginning of organised society. Statistics has been used by the administrative heads and rulers of the states in the form of collecting data on different aspects for the purpose of formulating sound military and fiscal policies. This data includes figures of population, tax collection, military strength, etc. In the present times, the government is the biggest collector of data as well as the biggest user of statistics. A huge amount of data is collected and interpreted by various departments of the government for formulating efficient policies and decision-making.

2) **Statistics and Mathematics:** Statistics can be considered as a branch of science which is conceived on the foundation of mathematics. A person should have some knowledge of mathematics for understanding the basic fundamentals of statistics.

   According to Connor, "Statistics is a branch of Applied Mathematics which specializes in data".

   According to W. I. King, "Statistics may properly be considered as a branch of mathematics in as much as it attempts to formulate definite rules of procedure applicable in handling groups of data of many different varieties".

3) **Statistics and Economics:** Statistics is used as an important tool in economics study and research. Economics is mainly concerned with production and distribution of wealth and also savings and investments. Statistical tools are used in the following economic interest areas:
   i) Statistical methods are used for measuring and forecasting Gross National Product (GNP).
   ii) Statistical studies of business cycles give a clear picture about the economic stability.
   iii) Economic policy-making depends on the statistical analysis of population growth, unemployment figures, rural or urban population shifts, etc.
   iv) Optimum utilisation of resources is possible with the use of econometric models which uses statistical methods.
   v) For the study of finance, banking, consumer savings and credit availability, financial statistics is necessary.

4) **Statistics and Physical Science:** The use of statistical methods is continuously increasing in the field of physical sciences such as Biology, Physics, Chemistry, Astronomy, Medicine, etc. Statistical data are collected from different results of different experiments.

5) **Statistics and Natural Sciences:** Statistics is also very important in the study of natural sciences such as astronomy, biology, medicine, meteorology, zoology, botany, etc. **For example,** for diagnosing the exact disease of a patient, the doctor must believe on real data such as the body temperature, pulse rate, blood pressure, etc.

6) **Statistics and Research:** In the current scenario statistics is an essential part of research study. Improvement in knowledge has been possible because experiments are carried out with the help of statistical methods. **For example,** experiments about crop yields and their correlation with

different types of fertilisers and different types of soil are designed and studied with the help of statistical techniques. In the current time, statistical methods are used in all types of research work including medicine and public health.

7) **Statistics in Astronomy:** Astronomers were one of the first groups of people who used statistics in the study of movement of heavenly bodies and eclipses and other such astronomical issues. Astronomers earlier relied heavily on estimation but later statistics helped to turn these estimations into accurate ideas.

8) **Statistics in Education:** Statistics is used extensively in the field of education because research has become a common feature in all branches of activities. In education, statistics plays a vital role in formulation of new policies for new courses to be started, consideration of infrastructural requirements for new courses, etc. Apart from this, there are many people involved in research work who test past knowledge and develop new knowledge with the help of statistics.

9) **Statistics in Accounting and Auditing:** In accounting, exactness is an essential component but for decision-making purposes, approximation is taken into account. The current asset value is calculated on the basis of its current values and the corrected values are determined with the help of current purchasing power of money or the current value of it, while taking depreciation into consideration. This is done through the use of price indices which are based on the collection of statistics. For determining the trend of future profits it is required to use the study of correlation analysis between the profits and dividends. In auditing, sampling is generally used as it is not possible to examine voluminous transactions due to lack of human resources. An auditor will first co-relate the past error percentage with the current error rate after conducting a pilot audit. After this, the auditor decides on the sample size of books to be audited.

10) **Statistics in Planning:** Statistics is important for efficient planning in all modern economies, especially in the developing countries. This helps in successful planning by taking into consideration the correct analysis of complex statistical data. The plans, adopted for the economic development of a country, are also made on the basis of statistics available about the different economic problems being faced.

11) **Statistics in Business Management:** Statistics is used in some very typical areas of business operations. These are described as follows:

i) **Entrepreneur:** Statistics plays a very important role in the situation when an entrepreneur wants to start a new business or acquire any business. For opening a new business it is required to track the past and current market trends to make sure about the success of new business. An analysis of the needs and wants of the consumers, number of competitors and their marketing strategies, availability of resources, etc., help the entrepreneur to set up the new business. Unreliability of data, incorrect interpretations and analysis has sometimes led to failure of new enterprise.

ii) **Production:** The statistical techniques are used to forecast the demand of any item accurately and this demand is the basis of the production of that item. The feedback of market surveys which are analysed by statistical methods are helpful in taking decisions about what to produce and how much to produce.

iii) **Marketing:** In order to formulate an effective marketing strategy it is necessary to study the several variables with the help of statistical tools. These variables include population profile, shifts in population, disposable income, competition, social and professional status of target market, advertising, etc. The inter-relationships among all the variables are closely analysed.

iv) **Personnel:** Proper analysis of statistical data relating to human resource of a company helps the personnel departments in formulating personnel policies and manpower planning. Such data may include wage rates, employment trends, cost of living indices, work related accident rates and so on.

12) **Statistics and Commerce:** Commerce is now very much dependent on statistics for success of businesses. Any businessman cannot afford to over-stock or under-stock the goods. In the initial stage it is required to do a market survey using statistical techniques to understand the changing tastes of the consumers because a number of multinational companies have also entered with new products and services. Thus, the statistical techniques help in providing the present condition and forecast the future.

13) **Other Areas:** Statistics is used by many businesses such as insurance companies, stock brokerage houses, banks, public utility companies and so on. It also helps politicians to study their winning chances from a constituency through the use of sampling techniques in random selection of voter samples and studying their preferences on issues and policies.

1.1.7. Role/ Functions of Statistics

The functions of statistics can be divided into five main categories namely:

1) **Condensation:** Usually the word 'to condense' means to reduce or to lessen. In order to understand the characteristics of a huge amount of data with the help of some observation, 'condensation' is used. **For example,** if the marks obtained by a particular class of students are mentioned then it serves no purpose. But if the average marks obtained by a group of students are mentioned, it serves a better purpose. Range of marks obtained by the students is another measurement of data. Therefore, complexity is reduced with the use of statistical techniques and it facilitates to understanding the huge amount of data.

2) **Comparison:** The data is condensed with the use of two statistical techniques which is known as 'classification' and 'tabulation'. These methods help in data comparison, collected by various sources. For the comparative studies various techniques are used such as grand totals, measures of central tendency, measures of dispersion, graphs and diagrams, etc. **For example,** if rice production in two different districts of Uttar Pradesh is known, then it is easier for a comparative study between them. Comparison helps in understanding the data better.

3) **Forecasting:** Forecasting refers to the prediction or estimation of the future. The data which is related to the rainfall over the last 10 years in a particular district of Uttar Pradesh helps in predicting rainfall in that region in future. Similarly, businesses also do certain predictions related to production, sales, profits, etc. Time series and regression analysis play a vital role in forecasting activity.

4) **Estimation:** Inferences, drawn about a population from the analysis of sample which is selected from that population, is the central objective of statistics. Statistical inference can be classified as:
   i) Estimation Theory,
   ii) Tests of Hypothesis,
   iii) Non-parametric Tests, and
   iv) Sequential Analysis.

   In estimation theory, sample observations are used to estimate the value of unknown population parameter. **For example,** to estimate the average height of students in a school the sample of heights of 100 students is selected. On the basis of this sample the average height of students is estimated.

5) **Tests of Hypothesis:** A statistical hypothesis is a statement about some unknown population parameters of a given distribution. The information provided by the sample population is the basis of the statistical hypothesis involved in characterising a population. Following are some examples of statement of hypothesis which are tested by appropriate statistical techniques:
   i) Whether a certain drug is effective against typhoid in a certain locality, and
   ii) Whether the production of a certain item has increased due to the use of new machinery.

1.1.8. Importance/Benefits of Statistics

Following points highlight the importance of statistics:

1) Statistics provides an understanding and correct explanation of a phenomenon of nature.
2) Statistics aids in suitable planning and execution of statistical investigation in any field of study.
3) Statistics helps in assembling of appropriate numerical data.
4) Statistics helps in simplifying complex data in a fitting tabular, diagrammatic or graphic form for easier comprehension.
5) Statistics is used for drawing valid inference and calculation of consistency about the population parameters from the sample data.

1.1.9. Limitations of Statistics

Major limitations of statistics are as follows:

1) **Statistics does not Deal with Individual Measurements:** The study of individual measurements lies outside the range of statistics as it deals with aggregates of facts. **For example,** marks obtained by a single student in a class have no statistical significance but the average marks received by the students of the same class are relevant to statistics.

2) **Statistic Deals only with Quantitative Characteristics:** Statistics deals with the quantitative statements of the facts. Various characteristics cannot be expressed quantitatively and hence statistics is unable to analyse these characteristics. Qualitative characteristics such as honesty, efficiency, intelligence, blindness and deafness cannot be studied statistically.

3) **Statistical Results are True only on an Average:** Conclusions or inferences which are drawn after statistical study may not be true universally but only under certain conditions.

4) **Statistics is only One of the Methods of Studying a Problem:** In all conditions it is not necessary that statistical tools give the best solution. In a statistical problem it is important to take a country's culture, religion and philosophy into consideration.

5) **Statistics can be Misused:** Statistics may be misused when statistical conclusions are based on incomplete information. This may lead to incorrect conclusions.
6) **Statistics does not Reveal the Entire Story:** Statistics only simplifies and analyses complex data but does not give the real picture of the background about the data.
7) **Statistical Relations do not necessarily bring out the 'Cause and Effect' Relationship between Phenomena:** Statistics tells the relationships between variables but does not define which one is the cause and which one is the effect. It only reveals the association between the two variables.
8) **Statistics is Collected with a given Purpose and cannot be Indiscriminately Applied to any Situation:** The relevance of a statistical result in one situation does not guarantee its utility in another. Usage of secondary data leads to incorrect drawing of conclusions. It is therefore necessary to thoroughly scrutinise statistical data before inference.

1.1.10. Applications of Inferential Statistics in Managerial Decision-Making

1) **Marketing & Sales**
   i) Product selection and competence strategies.
   ii) Utilisation of resources including territory control.
   iii) Advertising decisions for cost & time effectiveness.
   iv) Forecasting & trend analysis.
   v) Pricing & market research.
2) **Production Management**
   i) Product mix & product positioning
   ii) Facility & production planning
   iii) Material handling & faculty planning
   iv) Maintenance policies
   v) Activity planning & resource allocation
   vi) Quality control decisions
3) **Material Management**
   i) Buying policy: sourcing & procurement
   ii) Material planning & lead time
4) **Finance, Investments & Budgeting**
   i) Profit planning
   ii) Cost flow analysis
   iii) Investment decisions
   iv) Dividend policy decisions
   v) Risk analysis
   vi) Portfolio analysis
5) **Personnel Management**
   i) Optimum organisation level
   ii) Job evaluation & assignment analysis
   iii) Social analysis
   iv) Salary/wage analysis
   v) Recruitment & training
6) **Research & Development**
   i) Area of thrust: analysis a & planning
   ii) Project selection criteria
   iii) Alternatives analysis
   iv) Trade-off analysis
7) **Defense**
   i) Optimisation of weapon system
   ii) Force deployment
   iii) Transportation cost analysis
   iv) Assignment suitabilities

1.2. EXERCISE

1.2.1. Very Short Answer Type Questions

1) Define Statistics.
2) Give the importance of statistics.
3) What are the types of statistics?
4) Give the characteristics of statistics.

1.2.2. Short Answer Type Questions

1) Discuss the limitations of statistics
2) Discuss the scope of Statistics.
3) What is the application of statistics?
4) What do you mean by descriptive statistics?
5) Give the importance of statistics in Business and Management. **(2018)**
6) Classify the importance of statistics. **(2022)**

1.2.3. Long Answer Type Questions

1) Discuss the meaning and scope of statistics bringing-out its importance in business.
2) Define statistics. How does it help a manager?
3) Discuss briefly the application of statistics. Pointing out their limitations, if any.
4) Comment on the following statements:
   i) "Statistics is the science of averages".
   ii) "Statistics is the science of counting".
   iii) "Statistics is the science of estimates and probabilities".
5) Discuss the meaning and scope of statistics bringing-out its importance particularly in the field of trade and commerce.
6) What role does statistics play in the management of a business enterprise? Examine its scope and limitations.

Chapter 2

Measures of Central Tendency

2.1. MEASURES OF CENTRAL TENDENCY

2.1.1. Meaning and Definition

Measure of Central tendency is also known as **'measure of central value'** or **'measure of location'** or **'average of first order'.** It is a statistical measure and calculates the location or position of central point to explain the central tendency of the whole quantity of data.

According to Simpson and Kafka, "A measure of location is a typical value around which other figures congregate."

According to R.F. Fesher, "The inherent inability of the human mind to grasp in its entirely a large body of numerical data, compels us to seek relatively few constants that will describe the data."

According to Bowley, "Statistical constants enable us to comprehend in a single effort the significance of the whole."

According to M.R. Speigal, "Average is a value which is typical or representative of a set of data."

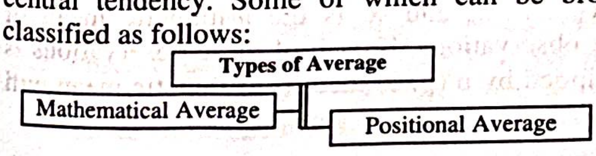
According to Clark and Sekkade, "Average is an attempt to find one single figure to describe whole of figure."

2.1.2. Properties of Good Measures of Central Tendency

1) It should be rigidly defined.
2) Its definition should be in the form of a mathematical formula.
3) It should be easy to understand and calculate.
4) It should remain unaffected by the extreme values.
5) It should be capable of further algebraic treatment.
6) It should be based on all items in the series.
7) It should be capable of being used in further statistical computation or processing.
8) It should possess sampling stability.

2.1.3. Types of Average

In statistics, there are various types of measures of central tendency. Some of which can be broadly classified as follows:



2.1.3.1. Mathematical Average

Mathematical averages are those averages which are calculated by taking into account the values of all the items in a series. When values of all the items in a series are considered in the calculation of average then it is called mathematical averages. Following are the different mathematical average:

1) **Mean/Arithmetic Mean:** Mean is also known as Arithmetic Mean (A.M.). To calculate the mean, summate all the observations and divide it by the total number of observations. Mean is denoted by $\overline{X}$.

   So, $\overline{X} = \dfrac{\text{Sum of all the observations}}{\text{Number of observations}} = \dfrac{\sum X}{N}$

   Where X and N are the observation and number of observation respectively.

2) **Geometric Mean:** The geometric mean of 'N' numbers is defined as the "N^{th} root of the product of 'N' numbers." If all the values of a series are multiplied and then N^{th} root of the product is extracted, then the geometric mean is obtained. Symbolically:

   $$G.M. = \sqrt[N]{(X_1)\times(X_2)\times(X_3).....\times(X_n)}$$

3) **Harmonic Mean:** The harmonic mean is the reciprocals of averaged numbers. It is defined as the reciprocal of the arithmetic mean of the reciprocal of the individual observations. Under certain conditions, harmonic mean is a better measure of central tendency e.g., computation of average speed, average price, etc.

   Mathematically,

   $$H.M. = \frac{N}{\frac{1}{X_1}+\frac{1}{X_2}+\frac{1}{X_3}+.....+\frac{1}{X_N}}$$

   Where, N = Number of observations;
   X_N = Value of observation

2.1.3.2. Positional Average

Positional average is the average which depends on the position of the items, rather than the values of the items. Following are the different positional average:

1) **Median:** When N observations are grouped and arranged in the sorting order (ascending or descending order) according to their values, then the central value of the observation is known as median. It is denoted by M or Me.

2) **Mode:** Mode is the value of the variable for which the frequency is maximum. It is denoted by Z or Mo.
3) **Quartiles:** The measure of central tendency which divides a group of data into four subgroups or parts, then it is called quartile. The three quartiles are denoted as Q_1, Q_2, and Q_3.
4) **Deciles:** If the values of the observation in a data set are arranged in ascending or descending order and divided it into ten equal parts by using nine points on the scale of observations, then it is called **decile**.
5) **Percentiles:** If the values of the observation in a data set are arranged in ascending or descending order and divided it into hundred equal parts by using ninety nine points on the scale of observations, then it is called **percentile.** It is represented by $P_1, P_2, \ldots\ldots P_{99}$

2.2. MEAN/ARITHMETIC MEAN

2.2.1. Introduction

Mean is also known as arithmetic mean (A.M.). To calculate the mean, summate all the observations and divide it by the total number of observations.

Mean is denoted by $\overline{X}$.

$$\text{So, } \overline{X} = \frac{\text{Sum of all observations}}{\text{Number of observations}} = \frac{\sum X}{N}$$

According to W. I. King, "The arithmetic average may be defined as the sum or aggregate of a series of items divided by their number".

Example 1: The arithmetic mean of the four numbers, 32, 40, 44 and 36 is given by,

$$\overline{X} = \frac{32+40+44+36}{4} = \frac{152}{4} = 38$$

Arithmetic Mean is considered as the best measure of central tendency because of the following reasons:

1) It is rigidly defined. Median might not be rigid when there are even number of observations and mode cannot be considered as rigid either when there are more than one value with the highest frequency or frequency density.
2) Only the mean directly uses all the observations; it changes when any one of the observations is changed.
3) The mean is generally the one that is the least affected by sampling fluctuations.

Example 2: Mr. Shyam spends ₹1,000 for apples costing ₹100 per kg and another ₹1,000 for apples costing ₹80 per kg. What is the average price per kg.?

Solution: Total Weight of Apples purchased

$$= \frac{1000}{100} + \frac{1000}{80} = 10 + 12.5 = 22.5 \text{ Kg}$$

Total Amount spends to purchase the Apples = 1000 + 1000 = ₹2000

Now average price per Kg $= \frac{2000}{22.5} =$ ₹88.89

Example 3: Average age of 50 students was 16 years. 5 students having their average age as 17 years left the group and 3 other students having their age 15, 16, and 19 years joined the group. Find the average age of present group.

Solution: Total age of 50 Students $= 16 \times 50 = 800$
Total age of 5 students left $= 17 \times 5 = 85$
Total age of 3 other students joined $= 15 + 16 + 19 = 50$
Total number of students after leaving 5 students and joining 3 students $= 50 - 5 + 3 = 48$
After leaving 5 students and joining 3 students the Total age of students $= 800 - 85 + 50 = 765$
Now average age of present group

$$= \frac{765}{48} = 15.94 \text{ or } 16 \text{ (Approx)}$$

Example 4: The average marks of 100 students were 45. Later it was found that the marks of one student were misread as 14 instead of 64. Find the correct mean marks.

Solution: The mean marks = 45
Total number of Students = 100
$\therefore$ Sum of marks $= 100 \times 45 = 4500$
Correct Total Marks $= 4500 - 14 + 64 = 4550$

Thus, correct mean marks $= \frac{4550}{100} = 45.5$

2.2.2. Properties of Arithmetic Mean

Property 1: Consider the n observations such as $x_1, x_2, x_3, \ldots\ldots x_n$ and $\overline{x}$ is the arithmetic mean of these observations; then

$$(x_1 - \overline{X}) + (x_2 - \overline{x}) + (x_3 - \overline{x}) + \text{------} + (x_n - \overline{X}) = 0$$

Property 2: Consider the n observations such as $x_1, x_2, x_3, \ldots\ldots x_n$ and $\overline{x}$ is the arithmetic mean of these observations. If there is the increase by p to each observation then its arithmetic mean will be $(\overline{x} + p)$.

Property 3: Consider the n observations such as $x_1, x_2, x_3, \ldots\ldots x_n$ and $\overline{x}$ is the arithmetic mean of these observations. If there is the decrement by p to each observation then its arithmetic mean will be $(\overline{x} - p)$.

Property 4: Consider the n observations such as $x_1, x_2, x_3, \ldots\ldots x_n$ and $\overline{x}$ is the arithmetic mean of these observations. If each of these observations is multiplied by p ($p \neq 0$) then its arithmetic mean will be $p\overline{x}$.

Property 5: Consider the n observations such as x_1, x_2, x_3, ...x_n and $\overline{x}$ is the arithmetic mean of these observations. If each of these observations is divided by p ($p \neq 0$) then its arithmetic mean will be $(\overline{x}/p)$.

2.2.3. Merits of Mean

1) Simple to calculate and understand.
2) Some value is always determined, i.e., it is never indefinite.
3) Can be used in other algebraic calculations.
4) No need of sorting or arrangement (ascending or descending order).
5) It is stable and not affected by the variation of sampling.

2.2.4. Demerits of Mean

1) Mean is greatly affected by the extreme values. **For example,** mean of 3, 7 and 200 is 70. Here, no value is present near this 70; hence this average is of no use.
2) Sometimes mean may provide confusing impressions. **For example,** in a hospital, per day average number of patients admitted is 5.7. Here given information is useful but doesn't provide the actual item because some values are of no use when expressed in fraction or decimal.
3) The 'mean' cannot be predicted by just inspecting the sample item.
4) If a single value is missing, mean cannot be calculated.
5) In case of open-end classes, mean cannot be calculated.
6) Graphical representation of mean is not possible.

2.2.5. Implications of Arithmetic Mean

1) The sum of the deviations of the individual items from the arithmetic mean is always zero. This means $\sum(X-\overline{X})=0$, where X is the value of an item and $\overline{X}$ is the arithmetic mean. 'Since the sum of the derivations in the positive direction is equal to the sum of the deviations in the negative direction, the arithmetic mean is regarded as a measure of location.'
2) The sum of the squared deviations of the individual items from the arithmetic mean is always minimum. In other words, the sum of the squared deviations taken from any value other than the arithmetic mean will be higher.
3) Since the arithmetic mean is based on all the observations in a series, a change in the value of any observation will lead to a change in the value of the arithmetic mean.
4) In the case of a highly skewed distribution, the arithmetic mean may get inaccurate on account of a few items with extreme values. In such a case, it may not represent the characteristic of the distribution.

2.2.6. Uses/Application of Arithmetic Mean

Following are the uses of arithmetic mean:

1) Arithmetic mean is used to measure the standard deviation.
2) Arithmetic mean is used in the construction of index numbers.
3) It is also used in the hypothesis testing.

2.2.7. Types of Arithmetic Mean

There are two types of arithmetic mean:

Types of Arithmetic Mean

Simple (or unweighted) Arithmetic Mean | Weighted Arithmetic Mean

2.2.8. Simple Arithmetic Mean

Simple arithmetic mean is also known as **un-weighted arithmetic mean.** In this mean is calculated by considering each and every value as equally important.

Method of Calculation of Simple Arithmetic Mean

The methods for calculating arithmetic mean vary according to various types of series such as:

1) Calculation of Arithmetic Mean - Individual Series
2) Calculation of Arithmetic Mean - Discrete Series (Ungrouped Data)
3) Calculation of Arithmetic Mean - Continuous Series (Grouped Data)

2.2.8.1. Calculation of Arithmetic Mean - Individual Series

The methods used are as follows:

1) **Direct Method:** The direct method of calculating arithmetic mean in the case of individual series involves the following steps:
   i) Add the various given values of the variable X and find the total value which is denoted by ΣX. After that we find the total No. of observation.
   ii) Divide this total value (ΣX) by the number of observations (N) i.e.,

$$\overline{X}=\frac{X_1+X_2+X_3+......+X_n}{N}=\frac{\Sigma X}{N}$$

Where,

$\overline{X}$ = Mean, ΣX = total of the observation, and

N = Number of all terms

Example 5: The weekly wage of 5 workers is as given below:

₹1,350, ₹1,400, ₹1,450, ₹1,370 and ₹1,480

Find arithmetic mean.

Solution: Computation of Arithmetic Mean

Serial Number	Weekly Wages (in Rupees)
1	1,350
2	1,400
3	1,450
4	1,370
5	1,480
N = 5	ΣX = 7,050

$$\overline{X}=\frac{\sum X}{N}=\frac{7{,}050}{5}=₹1{,}410$$

2) **Short-Cut Method:** With the use of short cut method also, arithmetic mean can be calculated. The calculation of arithmetic mean by short cut method includes the following steps:
   i) Any value may be taken as an assumed mean of the data. This assumed value is also known as working mean or arbitrary average (A = Assumed mean).
   ii) Subtract assumed mean from each value of the observation (d = X – A).
   iii) All the deviations are added and it is denoted by (Σd).
   iv) Apply the formula:

$$\overline{X}=A+\frac{\sum d}{N}$$

Where,

$\overline{X}$ = Arithmetic mean ; A = Assumed mean

$\sum d$ = Sum of the deviation ; N = Number of terms

Example 6: Calculate mean by shortcut method from the following data:

Years	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Price of Rice (in ₹)	40	50	55	80	58	60	75	35	45	52

Solution: Let the assumed mean be A = 50

Years	Price of Rice (X)	d = (X – 50)
2001	40	–10
2002	50	0
2003	55	5
2004	80	30
2005	58	8
2006	60	10
2007	75	25
2008	35	–15
2009	45	–5
2010	52	2
N = 10		Σd = 50

$$\overline{X}=A+\frac{\sum d}{N}=50+\frac{50}{10}=50+5=₹55.$$

2.2.8.2. Calculation of Arithmetic Mean - Discrete Series (Ungrouped Data)

The methods used are as follows:

1) **Direct Method:** In the discrete series, the sum of items is determined by multiplying each value with the respective frequency. The values got after multiplication are totalled up. To find the arithmetic mean, total value is divided by the total number of items. Following are the steps which are involved in the calculation of mean:
   i) Variable Denoted by X and Frequency denoted by f.
   ii) Multiply each item by its frequency, denoted by (fX).
   iii) Add all the fX, denoted by (ΣfX).
   iv) ΣfX is divided by the total number of items.

The formula is $\overline{X}=\frac{\sum fX}{\Sigma f}$ Or $\overline{X}=\frac{\Sigma fX}{N}$

Where, $\overline{X}$ = Arithmetic mean;

ΣfX = The sum of products;

Σf = Total number of items.

Example 7: Find average wages of 10 workers

Daily Wage (in ₹)	4	6	10	11	14	Total
No. of Workers	2	1	4	2	1	10

Solution:

Daily Wage (X)	No. of Workers (f)	fX
4	2	8
6	1	6
10	4	40
11	2	22
14	1	14
Total	Σf = 10	ΣfX = 90

∴ Arithmetic mean (Average Wage)

$$=\frac{\sum fX}{f}=\frac{90}{10}=₹9.00$$

2) **Short-Cut Method:** For calculating the arithmetic mean by short-cut method in a discrete series the following steps are taken:
   i) Any value may be taken as an assumed mean of the data.
   ii) Subtract assumed mean from each value of the observation (d_x = X – A).
   iii) Each value of deviation is multiplied by its respective frequency, denoted by fd_x.
   iv) Following is the formula which is used for the calculation of arithmetic mean:

$$\overline{X}=A+\frac{\sum fd_x}{\sum f}\text{ or } A+\frac{\sum fd_x}{\sum N}$$

Where, $\overline{X}$ = Arithmetic Mean

A = Assumed Mean

Σfd_x = Total of deviations multiplied with the respective frequencies

Σf = Total of frequencies (N)

Example 8: From the following frequency distributions find out the mean weight of the 100 persons:

Weight (in kg.)	64	65	66	67	68	69	70	71	72	73
No. of Persons	15	13	18	5	20	11	7	6	3	2

Solution: Let assumed mean (A) = 68;

Short-cut Method

Weight in Kg. (X)	No. of Persons (f)	Deviation (d_x) = (X – A)	fd_x
64	15	– 4	– 60
65	13	– 3	– 39
66	18	– 2	– 36
67	5	– 1	– 5
68 (A)	20	0	0
69	11	1	11
70	7	2	14
71	6	3	18
72	3	4	12
73	2	5	10
	Σf = 100		Σfd_x = – 75

$$\overline{X}=A+\frac{\Sigma fd_x}{\Sigma f}=68+\frac{(-75)}{100}=68-0.75=67.25\text{ Kg.}$$

2.2.8.3. Calculation of Arithmetic Mean - Continuous Series (Grouped Data)

The methods used are as follows:

1) **Direct Method:** For calculating the arithmetic mean in a continuous series the following steps are taken:

   i) First mid value of each class-interval is determined by adding the lower and upper limit of each class and dividing the total by two.

   For example, in a class interval say 0 – 10, the mid value is $\frac{0+10}{2}$=5.

   ii) Multiply these mid values of each class with the respective frequency of each class. In other words X will be multiplied by f.

   iii) Add up all the products and obtain ΣfX.

   iv) ΣfX is divided by the sum of the frequencies i.e., Σf.

   Apply the formula: $\overline{X}=\frac{\Sigma fX}{\Sigma f}\left(\frac{0+10}{2}=10/2=5\right)$

Example 9: From the following find out the mean profits:

Profits per Share	100-200	200-300	300-400	400-500	500-600	600-700	700-800
Number of Shares	12	20	18	30	32	26	22

Solution: Calculation of Mean

Profits	Mid-Point (X)	No. of Shares (f)	fX
100-200	150	12	1800
200-300	250	20	5000
300-400	350	18	6300
400-500	450	30	13500
500-600	550	32	17600
600-700	650	26	16900
700-800	750	22	16500
		Σf = 160	ΣfX = 77600

$$\overline{X}=\frac{\Sigma fX}{\Sigma f}=\frac{77600}{160}=485$$

The average profit is ₹485.

2) **Step-Deviation Method:** For calculating the arithmetic mean by step-deviation method in case of grouped series the following steps are taken:

   i) Find the mid-point of each group or class, denoted by X.

   ii) Any value may be taken as an assumed mean of the data (A = Assumed mean).

   iii) Subtract assumed mean from the mid-value of each class, denoted by d_x = (X – A).

   iv) Deviations (d_x) are divided by their common factor (i) and step deviation is determined for each class-interval.

   $$d'=\frac{d_x}{i}.$$

   $$\text{Thus, } d'=\frac{X-A}{i}$$

   v) The step deviation of each class is multiplied by respective frequency of that class to find the total value Σfd′.

   vi) Find the total frequency, N =Σf.

   vii) Its all sum of all products (Σfd') divided by frequency(Σf).

   viii) Use the formula:

   $$\overline{X}=A+\frac{\Sigma fd'}{\Sigma f}\times i$$

   Where,

   $\overline{X}$ = Arithmetic mean,

   A = Assumed mean:

   Σfd′ = Total of product of step-deviations and frequencies;

   Σf = Total number of frequencies;

   i = Common factor in x or in d_x

Example 10: Calculate the mean from the following table:

Weights of Children (in kg.)	0–10	10–20	20–30	30 – 40	40–50	50 – 60	60 – 70	70 – 80	80 – 90
Number of Children	4	12	8	21	32	28	10	3	2

Solution: Let assumed mean (A) = 55, let i = 10 (Common Factor)

C.I (Weight in kg.)	Mid Value (X)	Frequency (f)	d_x = X – A	d′ = d_x/10	fd′
0 – 10	5	4	– 50	– 5	– 20
10 – 20	15	12	– 40	– 4	– 48
20 – 30	25	8	– 30	– 3	– 24
30 – 40	35	21	– 20	– 2	– 42
40 – 50	45	32	– 10	– 1	– 32
50 – 60	55	28	0	0	0
60 – 70	65	10	10	1	10
70 – 80	75	3	20	2	6
80 – 90	85	2	30	3	6
Total		Σf =120			Σ fd′= – 144

$$A.M. = A + \frac{\sum fd'}{\sum f} \times i = 55 + \frac{(-144)}{120} \times 10$$

$$= 55 - 12 = 43 \text{ kg}$$

2.2.9. Weighted Arithmetic Mean

In the economic studies, weighted arithmetic mean plays a very important role. In the calculation of weighted arithmetic mean, items are assigned weights according to their relative importance in the group. The relative importance of various items in a distribution may be different; some items may be relatively more important than others.

For example, to calculate mean-expenditure of a family, it would be wrong to give equal importance to different items of family expenditure. The family may give different importance to the various items depending upon how much monthly expenditure is incurred on those items. They may give more importance to food and less importance to entertainment or clothes. Hence, the arithmetic mean should be calculated according to their relative importance.

Methods of Calculation of Weighted Arithmetic Mean

1) **Direct Method:** The following steps are taken in the calculation of weighted arithmetic mean by direct method:
   i) Multiply size of the variable (X) and their respective weights and add up the products.
   ii) Divide the total by sum (ΣWX) of the weights (ΣW).
   iii) Following formula is used for the calculation of weighted arithmetic mean:

$$\overline{X}_w = \frac{\sum WX}{\sum W}$$

Where,

$\overline{X}_w$ = Weighted arithmetic mean

W = Weights,

X = Variable

ΣW = Sum of Weights

Example 11: An examination was held to decide the award of a scholarship. The weights given to the various subjects were different. Only three applicants for the scholarship obtained over 50% marks in each subject. The marks were as follows:

Subject	Weight	Marks of A	Marks of B	Marks of C
Statistics	4	65	58	66
Accountancy	3	64	65	72
Economics	2	58	57	58
Mercantile Law	1	72	79	54

Of the candidates the one getting the highest marks is to be awarded the scholarship, who should get it?

Solution:

Subject	Weight W	Marks of A X_1	Marks of A WX_1	Marks of B X_2	Marks of B WX_2	Marks of C X_3	Marks of C WX_3
Statistics	4	65	260	58	232	66	264
Accountancy	3	64	192	65	195	72	216
Economics	2	58	116	57	114	58	116
Mercantile Law	1	72	72	79	79	54	54
Total	ΣW = 10		ΣWX_1 = 640		ΣWX_2 = 620		ΣWX_3 = 650

Weighted Mean of A,

$$\overline{X}_{W_A} = \frac{\sum WX_1}{\sum W} = \frac{640}{10} = 64$$

Weighted Mean of B,

$$\overline{X}_{W_B} = \frac{\sum WX_2}{\sum W} = \frac{620}{10} = 62$$

Weighted Mean of C,

$$\overline{X}_{W_C} = \frac{\sum WX_3}{\sum W} = \frac{650}{10} = 65$$

The Weighted Mean of 'C' is the highest; hence he is entitled for scholarship.

2) **Short-Cut Method:** The following steps are taken in the calculation of weighted arithmetic mean by short-cut method:
   i) Any value may be taken as an assumed weighted mean of the data (A_w = Assumed mean).
   ii) Subtract assumed mean from each value of the observation ($d_x = X - A_w$)
   iii) Each value of deviation is multiplied with its respective weight, denoted by Wd_x and then added up, $\sum(Wd_x)$.
   iv) Following is the formula which is used in the calculation of weighted arithmetic mean:

$$\overline{X}_w = A_w + \frac{\sum Wd_x}{\sum W}$$

Where, $\overline{X}_w$ = Weighted arithmetic mean;

A_w = Assumed weighted mean

ΣWd_x = Total of deviations multiplied with the respective weights;

ΣW = Total of weights

Example 12: Determine the weighted mean from the below data using the direct and short-cut method:

Item	81	76	74	58	70	73
Weights	2	3	6	7	3	7

Solution: Table 2.1: Calculation of Weighted Mean

Direct Method			Short-Cut Method			
X	W	WX	X	W	Deviations d = X − A_w A_w = 74	Wd
81	2	162	81	2	+7	+14
76	3	228	76	3	+2	+6
A_w =74	6	444	74	6	0	0

58	7	406	58	7	−16	−112
70	3	210	70	3	−4	−12
73	7	511	73	7	−1	−7
	$\Sigma W = 28$	$\Sigma WX = 1961$		$\Sigma W = 28$		$\Sigma Wd = -111$

Direct Method: $\bar{X}_w = \frac{\sum WX}{\sum X} = \frac{1961}{28} = 70.04$

Short-Cut Method:

$$\bar{X}_w = A_w + \frac{\sum Wd}{\sum w} = 74 + \frac{-111}{28} = 70.04$$

2.2.10. Combined Mean/Grouped Mean

If the means of two or more than two sub-groups or different samples from the same universe along with their numbers are given, then a combined mean can be calculated with the help of following formula:

$$\text{Combined Mean}(\bar{X}) = \frac{N_1\bar{X}_1 + N_2\bar{X}_2 + N_3\bar{X}_3 .... + N_n\ \bar{X}_n}{N_1 + N_2 + N_3 .... + N_n}$$

$N_1, N_2, N_3, \ldots$ are the frequencies of different groups and $\bar{X}_1, \bar{X}_2, \bar{X}_3 ....$ are the arithmetic means of these groups.

Example 13: An analysis of the average speed of the buses in two transport companies A and B, gives the following results:

	Transport Company A	Transport Company B
Number of Buses	595	645
Average Speed (Km/h)	52.5	47.5

Find the combined average speed.

Solution: Combined Average Speed in the two transport companies would be, $\bar{X}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$

Where,

N_1 and N_2 = the number of items in the two series

$\bar{X}_1$ and $\bar{X}_2$ = the mean of the two series respectively

$\bar{X}_{12}$ = the combined mean of the two series.

Given, $N_1 = 595$ and $N_2 = 645$;

$\bar{X}_1 = 52.5$ and $\bar{X}_2 = 47.5$

Substituting the values,

$$\text{Combined Average Speed} = \frac{(595 \times 52.5) + (645 \times 47.5)}{595 + 645}$$

$$= \frac{31.237.5 + 30637.5}{1240} = \frac{61875}{1240} = ₹49.9\text{km/h}$$

Example 14: The mean rainfall of a place from Monday to Saturday in a week was 4.5cm. On account of heavy rains on Sunday the mean rainfall for the whole week shot up to 6cm. Determine the rainfall for Sunday.

Solution: A week can be treated as composed of two groups: First group consisting of 6 days excluding Sunday for which $N_1 = 6$ and $\bar{X}_1 = 4.5$; Second group consisting of only Sunday for which $N_2 = 1$. Also, mean of this group will be equal to the observation itself. Let this be X. We have to determine the value of X.

We are also given N = 7 and $\bar{X}_1 = 6$ (for the whole week).

$$\therefore\ 6 = \frac{6 \times 4.5 + 1 \times X}{7} \quad \text{Or} \quad 27 + X = 42$$

$\Rightarrow X = 42 - 27 = 15$ cms.

Thus, the rainfall on Sunday was 15 cms.

2.3. GEOMETRIC MEAN (G.M.)

2.3.1. Introduction

For a number 'N' geometric mean is defined as the N^{th} root of the product of N number. To calculate the geometric mean one needs to multiplying all the values of a series and taking the N^{th} root. Symbolically G.M. can be written as,

$$\text{G.M.} = \sqrt[N]{(X_1) \times (X_2) \times (X_3) ..... \times (X_n)}$$

Where, $X_1, X_2, X_3 \ldots X_n$ are the different values in the series and N are the Number of items. To perform the calculation logarithms and antilog of the mean of the logarithms values are taken. So it can be also written as,

$$\text{G.M.} = \text{Antilog of} \frac{(\text{Log}X_1 + \text{Log}X_2 + \text{Log}X_3 ...... + \text{Log}X_N)}{N} \text{ or } \frac{\Sigma \text{Log}X}{N}$$

2.3.2. Merits of Geometric Mean

1) It is rigidly defined and depends on the entire observations.
2) It is less affected by the extreme observation as compared to mean.
3) Also when sampling is fluctuated there is no effect on GM
4) For further algebraic treatment it is appropriate.

2.3.3. Demerits of Geometric Mean

1) Geometric mean is neither easy to understand nor to calculate.
2) If anyone observation is zero, geometric mean becomes zero and if any one of the observations is negative, geometric mean becomes imaginary regardless of the magnitude of the other items.
3) It cannot be computed if any value is missing.

2.3.4. Applications of Geometric Mean

Following are the applications of Geometric mean:

1) To find the rate of population growth Geometric Mean is effective.
2) Also to find out the rate of interest Geometric Mean is effective.
3) During the index number construction, Geometric Mean is used.

2.3.5. Method of Calculation of Geometric Mean

Following are the different methods of calculation of Geometric mean:

1) Calculation of Geometric mean – Individual Series
2) Calculation of Geometric mean – Discrete Series (Ungrouped Data)
3) Calculation of Geometric mean – Continuous Series (Grouped Data)

2.3.5.1. Calculation of Geometric Mean – Individual Series

Following steps are used to calculate the GM in case of individual series:

1) Take the logarithm of every value or the size of item using, log table.
2) Now perform the addition operation on the all the values of Log X (denote as ΣLog X).
3) Then divided the ΣLog X by the number of items. $\left(\text{denote as, } \frac{\text{LogX}}{N}\right)$.

At the end, take the antilog of the quotient (from step 3).

Example 15: Calculate geometric mean of the following:

40	60	44	74	83

Solution:

X	log of X
40	1.60206
60	1.778151
44	1.643453
74	1.869232
83	1.919078
	ΣLog X = 8.811974

$$G.M. = \sqrt[5]{40\times60\times44\times74\times83} \quad \text{Or}$$

$$G.M. = \text{Antilog } \frac{\Sigma \text{LogX}}{N} = \text{Antilog} \frac{8.811974}{5} = \text{Antilog} 1.7624 = 57.8629$$

2.3.5.2. Calculation of Geometric Mean – Discrete Series (Ungrouped Data)

Following steps are used to calculate the GM in case of discrete series:

1) Take the logarithm of every value (Using, log table = Log X).
2) Now multiply the frequency 'f' by the every size.
3) Now perform the addition operation and we get Σ (f log X).
4) Then divided the Σ f log X by the frequency 'Σf'. $\left(\text{denote as, } \frac{\sum f \text{ logX}}{\sum f}\right)$.
5) At the end, take the antilog of result (from step 4).

$$G.M. = \text{Antilog} \frac{(\log X_1 f_1 + \log X_2 f_2 + \log X_3 f_3 ...... + \text{Log} X_N f_N)}{N}$$

$$= \text{Antilog} \frac{\Sigma f \log X}{\Sigma f}$$

Example 16: Weight of the 31 people in sample survey is given in the following table. Calculate the geometric mean from the following data:

Weight	120	140	150	135	125	158	139	145	167
No. of student	3	4	6	6	3	5	2	1	1

Solution:

Size of item (X)	Frequency f	log X	f log X
120	3	2.0792	6.2376
140	4	2.1461	8.5844
150	6	2.1761	13.0566
135	6	2.1303	12.7818
125	3	2.0969	6.2907
158	5	2.1986	10.993
139	2	2.1430	4.286
145	1	2.1613	2.1613
167	1	2.2227	2.2227
	Σf = 31		Σf log X= 66.6141

$$G.M. = \text{Antilog } \frac{\Sigma f \log X}{\Sigma f} = \text{Antilog} \frac{66.6141}{31}$$

= Antilog of 2.1488 ⇒ G.M. weight = 140.865 lbs.

2.3.5.3. Calculation of Geometric Mean – Continuous Series (Grouped Data)

In case of continuous series following steps are involved in the computing process of Geometric mean:

1) Calculate the middle value of every class, denoted by 'm'.
2) Now find the logarithm value of every m, denoted by log m.
3) Next multiply the frequency by the log m, represent as f log m.
4) Sum up the all-Σf log m.
5) Divide the result of step 4 by Σf. it can be represent as -Σf = $\frac{\Sigma f \log m}{\Sigma f}$
6) Finally take the antilog of the result generated after step 5. It is written as,

$$G.M. = \text{Antilog} = \frac{\Sigma f \log m}{\Sigma f}$$

Example 17: Find out the geometric mean from the following data:

Yield of wheat	No. of farms
7 – 10	5
10 – 13	9
13 – 16	19
16 – 19	23
19 – 22	7
22 – 25	4
25 – 28	1

Solution:

Yield of wheat (mounds)	No. of farms (f)	mid value (m)	log m (x)	f(m)	f log m
7-10	5	8.5	0.929419	5	4.647095
10-13	9	11.5	1.060698	9	9.546282
13-16	19	14.5	1.161368	19	22.06599
16-19	23	17.5	1.243038	23	28.58987
19-22	7	20.5	1.311754	7	9.182278
22-25	4	23.5	1.371068	4	5.484272
25-28	1	26.5	1.4246	1	1.4246
				$\Sigma f = 68$	$\Sigma f \log m = 80.93901$

$$G.M. = \text{Antilog}\ \frac{\Sigma f \log m}{\Sigma f} = \text{Antilog}\ \frac{80.93901}{68}$$

$= \text{Anti log } 1.19028 = 15.49.$

2.4. HARMONIC MEAN (H.M.)

2.4.1. Introduction

When number of terms i.e., 'N' is divided by the Average of reciprocals of number i.e., $\frac{1}{X_1}+\frac{1}{X_2}+\frac{1}{X_3}+.....+\frac{1}{X_N}$ is known as harmonic mean.

It is better in the calculation of average speed, price, etc.

So, Harmonic mean is represented as,

$$H.M. = \frac{N}{\frac{1}{X_1}+\frac{1}{X_2}+\frac{1}{X_3}+.....+\frac{1}{X_N}}$$

Where, N = Number of observations and X_N = Value of observation

2.4.2. Merits of Harmonic Mean

1) It is rigidly defined.
2) It is created using all observations and also appropriate for further mathematical treatment.
3) Sampling fluctuations does not much affect the Harmonic Mean.

2.4.3. Demerits of Harmonic Mean

1) Harmonic mean is not easy to understand and to calculate.
2) Its value can't be obtained if any one of the observation is zero.
3) It is not a representative figure of the sample.

2.4.4. Applications of Harmonic Mean

1) Harmonic mean is generally used to determine the patterns in the Fibonacci series.
2) Whenever average multiples are to be evaluated then it is also used in finance.
3) It is also used in calculating some quatities like speed since speed is expressed in the form of ratio of two measuring units such as km/hr.
4) It is used to determine the average of rates because it assigns equal weight to all data points in a sample.

2.4.5. Method of Calculation of Harmonic Mean

Following are the different methods of calculation of harmonic mean:

1) Calculation of Harmonic mean – Individual Series
2) Calculation of Harmonic mean – Discrete Series (Ungrouped Data)
3) Calculation of Harmonic mean – Continuous Series (Grouped Data)

2.4.5.1. Calculation of Harmonic Mean – Individual Series

Following steps are involved in the calculation process of harmonic mean for individual series:

1) Calculate the reciprocal of every size i.e., $\frac{1}{X}$.

   (To make the calculation easy refer to log table)
2) Next calculate the summation of all reciprocals $\left(\Sigma\frac{1}{X}\right)$. Now apply the formula to calculate the HM i.e.,

$$H.M. = \frac{N}{\left(\frac{1}{X_1}+\frac{1}{X_2}+\frac{1}{X_3}+.....+\frac{1}{X_N}\right)}$$

And it can be written as,

$$H.M. = \frac{N}{\Sigma\frac{1}{X}}$$

Where, X_1, X_2, X_3 .... X_N, refer to the various values in the observations.

Example 18: The monthly incomes of 10 families in rupees in a certain village are given below:

Family	1	2	3	4	5	6	7	8	9	10
Income	75	60	15	65	400	7	35	150	35	26

Calculate the Harmonic mean.

Solution: From the following data:

Calculation of Harmonic Mean

Family	Income (X)	Reciprocals (1/X)
1	75	0.01333333
2	60	0.01666667
3	15	0.06666667
4	65	0.01538462
5	400	0.0025
6	7	0.14285714
7	35	0.02857143
8	150	0.00666667
9	35	0.02857143
10	26	0.03846154
		$\Sigma 1/X = 0.359679$

$$H.M. = \frac{N}{\left(\frac{1}{X_1}+\frac{1}{X_2}+\frac{1}{X_3}+.....+\frac{1}{X_N}\right)} = \frac{N}{\Sigma\frac{1}{X}} = \frac{10}{0.359679} = ₹27.80$$

2.4.5.2. Calculation of H.M. – Discrete Series (Ungrouped Data)

Following steps are involved in the calculation process of harmonic mean for discrete series:

1) Calculate the reciprocal of every size i.e., $\frac{1}{X}$.

2) Multiply the frequency by every reciprocal i.e., $(f\frac{1}{X})$.
3) Next calculate the summation of all products $\Sigma\left\{f\left(\frac{1}{X}\right)\right\}$.
4) Now apply the formula to calculate the HM i.e.,

$$H.M.=\frac{\Sigma f}{\Sigma f\left(\frac{1}{x}\right)}$$

Example 19: Calculate H.M. from the following data.

Size	6	7	8	9	10	11
Frequency	5	7	5	2	6	4

Solution:

Size of Item X	Frequency (f)	Reciprocal (1/X)	Product of Reciprocal f (1/X)
6	5	0.16666667	0.83333333
7	7	0.14285714	1
8	5	0.125	0.625
9	2	0.11111111	0.22222222
10	6	0.1	0.6
11	4	0.09090909	0.36363636
	Σf= 29		Σf(1/X) = 3.644192

$$H.M.=\frac{\Sigma f}{\Sigma f\left(\frac{1}{x}\right)}=\frac{29}{3.644192}=7.96$$

2.4.5.3. Calculation of Harmonic Mean – Continuous Series (Grouped Data)

In case of continuous series the harmonic mean is calculated using following steps:

1) Take the mid value of the each class. The mid value is denoted by 'm'.
2) Next calculate the reciprocal of each mid value i.e., (1/m).
3) Multiply the frequency by every mid value reciprocal i.e., $(f\frac{1}{m})$.
4) Next calculate the summation of all products, $\Sigma\left\{f\left(\frac{1}{m}\right)\right\}$.
5) Now apply the formula to calculate the Harmonic mean:

$$H.M.=\frac{\Sigma f}{f_1\frac{1}{m_1}+f_2\frac{1}{m_2}+......f_n\frac{1}{m_n}}=\frac{N}{\Sigma f\frac{1}{m}}$$

Example 20: Calculate H.M. of the following data:

Marks	20-30	30-40	40-50	50-60	60-70	70-80	80-90
Frequency	12	15	6	8	18	5	4

Solution:

Marks	Mid value (m)	Frequency (f)	Reciprocal (1/m)	f × (reciprocal) {f (1/m)}
20-30	25	12	0.04	0.48
30-40	35	15	0.02857143	0.42857143
40-50	45	6	0.02222222	0.13333333
50-60	55	8	0.01818182	0.14545455
60-70	65	18	0.01538462	0.27692308
70-80	75	5	0.01333333	0.06666667
80-90	85	4	0.01176471	0.04705882
		Σf = 68		Σf (1/m) = 1.5780079

$$H.M.=\frac{\Sigma f}{\Sigma f\frac{1}{m}}=\frac{68}{1.5780079}=43.092$$

2.5. RELATIONSHIP BETWEEN A.M., G.M. AND H.M.

G.M. = $\sqrt{A.M.\times H.M.}$ for two positive numbers. To prove the above relations let take 'a' and 'b' two positive numbers.

So, $A.M.=\frac{a+b}{2}, G.M.=\sqrt{ab}, H.M.=\frac{2ab}{a+b}$

Now, $A.M.\times H.M.=\frac{a+b}{2}\times\frac{2ab}{a+b}=ab=(G.M.)^2$

$\therefore\quad G.M.=\sqrt{A.M.\times H.M.}$

Example 21: If the A.M. and G.M. are two numbers are 20 and 16 respectively then find the harmonic mean of two numbers and also find two numbers.

Solution: Given: A.M. = 20, G.M. = 16

The relation between A.M., G.M. and H.M is given as:

$$\sqrt{(A.M.)(H.M.)}=(G.M.)\Rightarrow\sqrt{(20)(H.M.)}=(16)$$

Square the both means, we get, $H.M.=\frac{256}{20}=12.8$

Let the two numbers be X_1 and X_2. We are given that

$$A.M.=\frac{X_1+X_2}{2}=20 \text{ and } G.M.=\sqrt{X_1.X_2}=16$$

$\Rightarrow X_1+X_2=40$ .....(1)

$\Rightarrow X_1.X_2=256$ .....(2)

We can write,

$(X_1-X_2)^2=(X_1+X_2)^2-4X_1.X_2$

$=(40)^2-4\times256$

$=1600-1{,}024=576$

$\Rightarrow X_1-X_2=24$ .....(3)

Adding equation (1) and (3), we get

$2X_1=64\quad\therefore X_1=32$

Also $X_2=8$

The different cases are as follows:

1) When all the values of the series differ in size, A.M. is greater than G.M. and G.M. is greater than harmonic mean, i.e.,

A.M. > G.M. > H.M.

Example 22: Using the values 4, 8 & 16, verify that A.M. > G.M. > H.M.

Solution: $A.M.=\frac{4+8+16}{3}=9.33$ approx.

$$G.M. = 3\sqrt{4 \times 8 \times 16} = (512)^{1/3} = 8$$

$$H.M. = \frac{3}{\frac{1}{4}+\frac{1}{8}+\frac{1}{12}} = \frac{3}{\frac{11}{24}} = \frac{3 \times 24}{11} = \frac{72}{11} = 6.54$$

Thus, A.M. > G.M. > H.M.

2) If all the values of the series are equal, then A.M. is equal to G.M. and G.M. is equal to H.M., i.e.,

A.M. = G.M. = H.M.

Example 23: Using the values 12 & 12, verify that A.M. = G.M. = H.M.

Solution: Mean = $\frac{12+12}{2} = 12$

Geometric Mean = $\sqrt{12 \times 12} = 12$

Harmonic Mean = $\frac{2}{1/12+1/12} = \frac{2}{2/12} = \frac{2 \times 12}{2} = 12$

Thus A.M. = G.M. = H.M.

2.6. MEDIAN

2.6.1. Introduction

When N observations are grouped and arranged in the sorting order (ascending or descending order) according to their values, then the central value of the observation is known as median. It is denoted by M or Me. So,

$$M = \left(\frac{N+1}{2}\right)^{th} \text{Observation}$$

According to Prof. L R. Conner, "The median is that value of the variable which divides the group into two equal parts, one part comprising of all values greater and the other all values less than the median".

According to Croxton and Cowden, "The median is that value which divides a series so that one half or more of the items are equal to or less than it and one half or more of the items are equal to or greater than it."

Circumstances of Using Median

Median is preferred when:

1) There are few extreme scores in the distribution.
2) Some scores have undetermined values.
3) There is an open ended distribution.
4) Data are measured in an ordinal scale.

2.6.2. Merits of Median

1) It is unaffected by the extreme values.
2) In case of individual observation and discrete series, it is easy to understand and calculate.
3) It can be used in other algebraic calculations and also useful during calculation of mean deviation.
4) After sorting the values of variable, it can be located by inspection.
5) It can also be measured by using 'graphical representation'.
6) Median of open-end classes can be measured.
7) Median is clearly definite in nature i.e., it is always clearly defined.

2.6.3. Demerits of Median

1) Sorting (ascending and descending order) is necessary for the calculation.
2) It cannot be used to measure the combined mean of two or more groups.
3) When sampling is fluctuated then median also gets fluctuated (affected more than mean).
4) It is not based on all the data samples. It is a positional average.

2.6.4. Uses/Applications of Median

1) Median is used to find middle most data.
2) It is used to determine a point from where 50% of data is more & 50% data is less.
3) It is used where extreme cases can be ignored. **For example,** to find the performance of a cricketer where his worst & best extreme performance can be ignored to give his consistent performance.

2.6.5. Method of Calculation of Median

Following are the different methods of calculation of median:

1) Calculation of Median – Individual Series
2) Calculation of Median – Discrete Series (Ungrouped Data)
3) Calculation of Median–Continuous Series (Grouped Data)

2.6.6. Calculation of Median – Individual Series

In case of individual series, the following steps are taken for calculating the median:

1) If the given observations are not arranged in ascending or descending order of magnitude then it has to be arranged for calculation.
2) Locate the middle value:
   i) There will be a single value in the middle which is selected as median when the number of observation N is **odd**. Mathematically

   $$\text{Median} = \text{Value of} \left(\frac{N+1}{2}\right)^{th} \text{items}$$

   ii) If N is **even**, there will be two middle values, $\left(\frac{N}{2}\right)^{th}$ and $\left(\frac{N}{2}+1\right)^{th}$. The arithmetic mean of these two middle values is taken as median. That is

   $$\text{Median} = \frac{\left(\frac{N}{2}\right)^{th} \text{value} + \left(\frac{N}{2}+1\right)^{th} \text{value}}{2}.$$

2.7. MODE

2.7.1. Introduction

Mode is the value of the variable for which the frequency is maximum and it is denoted by Z or Mo.

According to Zizek, "Mode is the value occurring most frequently in a series (or group) of items and around which other items are distributed most densely".

According to Croxton and Cowden, "The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded as the most typical of a series of values".

Circumstances of Using Mode

Mode is the preferred measure when data are measured in a nominal scale. Geometric mean is the preferred measure of central tendency when data are measured in a logarithmic scale.

2.7.2. Merits of Mode

1) Easy to understand and calculate.
2) Can be easily found out by using inspection method.
3) It is an actual value, which most frequently occurs in the series.
4) Not affected by extreme values.
5) It is simple and accurate and can be measured in an open end class-interval without determining the class limits.

2.7.3. Demerits of Mode

1) It is ill defined and in few cases it is not possible to find a definite value.
2) Not a good representative because it is not based on all observations.
3) In further algebraic calculation it is of no use.
4) It may be possible that there is no mode or more than one mode present in the sample.
5) Fluctuation in sampling effects is made to a higher degree than mean or median.
6) Before calculating the mode, data sorting (ascending or descending order) is necessary.

Here, $L_1 = 65$, $f_m = 20$, $f_1 = 14$, $f_2 = 16$ and $i = 5$;

$$\therefore \quad \text{Mode} = 65 + \frac{20-14}{2\times20-14-16}\times5 = 65 + \frac{30}{10} \text{ or}$$

Modal value = 65 + 3 = ₹68 thousand

Note: Generally, modes are used for nominal scores, medians for ordinal scores, and means for interval scores.

2.8. RELATION BETWEEN MEAN, MEDIAN AND MODE

The empirical relation between Mean, Median and Mode is,

Mode = 3 Median – 2 Mean.

Karl Pearson has expressed this relationship as follows:

Mode = Mean – 3[Mean- Median]

Mode = 3 Median – 2 Mean Or,

Median = Mode + 2/3[Mean - Mode]

Example 32: Calculate the Median when Mean and Mode of Distribution are 38.6 and 32.6 respectively.

Solution: Given, Mean = 38.6 and Mode = 32.6.

$$\text{Median} = \text{Mode} + \frac{2}{3}[\text{Mean} - \text{Mode}]$$

$$= 32.6 + \frac{2}{3}[38.6 - 32.6] = 32.6 + \frac{2}{3}[6]$$

$$= 32.6 + 4 \Rightarrow \text{Median} = 36.6$$

Example 33: For a given set of values of a variable, the mean and mode are 28.16 and 24 respectively, calculate the median.

Solution: Mean = 28.16, Mode = 24

$$\text{Median} = \text{Mode} + \frac{2}{3}[\text{Mean} - \text{Mode}]$$

$$= 24 + \frac{2}{3}[28.16 - 24] = 24 + 2.77 = 26.77$$

$\Rightarrow$ Median = 26.77

Example 34: In a moderately asymmetrical distribution estimate the following measures of central tendency then find:

1) **Median:** If mode and arithmetic mean are 32.1 and 35.4 respectively.
2) **Arithmetic Mean:** If mode and median are 22 and 21.4 respectively.

Solution:

1) $\text{Median} = \text{mode} + \frac{2}{3}[\text{mean} - \text{mode}]$

$$= 32.1 + \frac{2}{3}[35.4 - 32.1]$$

$$= 32.1 + \frac{2}{3}\times3.3 = 32.1 + 2.2 = 34.3$$

2) Mode = 3 median – 2 mean

$\Rightarrow$ given that : mode = 22, median = 21.4

$22 = 3 \times 21.4 - 2$ mean

$22 = 64.2 - 2$ mean $\Rightarrow 2$ mean $= 64.2 - 22$

2 mean = 42.2

$$\text{Mean} = \frac{42.2}{2} \Rightarrow \text{mean} = 21.1$$

2.9. COMPARISON OF MEAN, MEDIAN AND MODE

Table 2.1 shows the comparison among mean, median and mode:

Table 2.1: Comparison of Mean, Median and Mode

Mean	Median	Mode
1) It is highly affected by extreme values.	It is not affected by extreme values as in case of mean.	It is also not affected by extreme values as in case of mean.
2) It is based on all the observations.	It is not based on all observations.	It is also not based on all observations.
3) It is not necessary to arrange the observations in ascending or descending order.	It is necessary to arrange the data in ascending or descending order of magnitude.	It is also arrange the data in ascending or descending order of magnitude.
4) It is capable of further algebraic treatments.	It is also capable of further algebraic treatment.	It is not capable of further algebraic treatment.
5) It can hardly be located by inspection.	It can be located by inspection, after arranging the data in order of magnitude.	It can also be found out by inspection.

2.10. QUARTILES

2.10.1. Introduction

The measure of central tendency which divides a group of data into four subgroups or parts then it is called quartiles. Firstly data arranging into ascending or descending order for calculating quantities after that it's divided into four equal parts. The three quartiles are denoted as Q_1, Q_2, and Q_3. The first quartile, Q_1, divides a frequency distribution in such a way that one-fourth (25%) of the distribution has a value less than Q_1 and three-fourth (75%) have a value more than Q_1. The second quartile, Q_2 divides a frequency distribution in such a way that it has equal number of observations above and below it. Hence, it is equal to the median of the data. The third quartile, Q_3 divides a frequency distribution in such a way that three-fourth (75%) of the observations have a value less than Q_3 and one-fourth (25%) have a value more than Q_3.

2.10.2. Advantages of Quartiles

Salient advantages of quartiles are as follows:

1) It is simple to understand and easy to calculate.
2) It is less affected by the extreme values.
3) It can be calculated exactly in case of open-end classes.

2.10.3. Disadvantages of Quartiles

Quartiles have some disadvantages too, which are as follows:

1) It is not based on all observations. So it may not be a good representative. In fact, 50% of the items in any series are ignored. In other words, it does not cover the first 25% and the last 25% items of a series.
2) It is not capable of further algebraic treatment.
3) It is affected to a greater extent by fluctuations of sampling.

2.10.4. Calculation of Quartiles

Following are the different methods of calculation of quartiles:

1) Calculation of Quartiles - Individual Series
2) Calculation of Quartiles - Discrete Series (Ungrouped Data)
3) Calculation of Quartile - Grouped Series (Grouped Data)

2.10.4.1. Calculation of Quartiles - Individual Series

Let the series be in ascending or descending order. Let N be the number of observations. Then

First quartile $(Q_1) = \left(\frac{N+1}{4}\right)^{th}$ item ;

Second quartile $(Q_2) = \left[\frac{2(N+1)}{4}\right]^{th}$ item or $\left(\frac{N+1}{2}\right)^{th}$ item

Third quartile $(Q_3) = \left[\frac{3(N+1)}{4}\right]^{th}$ item

Example 35: Calculate quartiles from the following data:

Marks obtained: 06, 30, 37, 18, 14, 42, 34, 11, 09, 26, 22, 03, 28, 52, 48

Solution: Arranging the series in ascending order, we get

Serial No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Marks Obtained	03	06	09	11	14	18	22	26	28	30	34	37	42	48	52

Number of observations, N = 15

$$Q_1 = \left(\frac{N+1}{4}\right)^{th} \text{item} = \left(\frac{15+1}{4}\right)^{th} \text{item} = 4^{th} \text{ item} = 11$$

$$Q_2 = \left[\frac{2(N+1)}{4}\right]^{th} \text{item} = \left[\frac{2(15+1)}{4}\right]^{th} \text{item} = 8^{th} \text{ item} = 26$$

$$Q_3 = \left[\frac{3(N+1)}{4}\right]^{th} \text{item} = \left[\frac{3(15+1)}{4}\right]^{th} \text{item} = 12^{th} \text{ item} = 37$$

2.10.4.2. Calculation of Quartiles - Discrete Series

The computation of quartiles from discrete series involves the following steps:

1) Arrange the data in ascending or descending order of magnitude (if not arranged).

3.1. MEASURES OF DISPERSION / MEASURES OF VARIATION

3.1.1. Meaning and Definition

Distribution cannot be clearly depicted by measuring the averages or central tendency. Averages provide the observations of only the central part of the distribution. So, study of scatteredness of observations is very important and this study is known as **measure of dispersion.**

The word 'Dispersion' literally means 'fluctuation', 'scatter', 'variation', 'deviation', or 'spread'. So, measure of dispersion shows the variation of an individual item from its average. It is also known as **measure of variation** as it explains the extent of scatterness in an observation and it measures the mean deviation about some central value. Dispersion is based on values from first order (the average or mean). So, it is known as **second order**. Deviation is calculated by measuring the difference between mean and actual value.

According to A.L. Bowley, "Dispersion is the measure of the variation of the items."

According to Reigleman, "Dispersion is the extent to which the magnitudes or qualities of the items differ; that is the degree of diversity".

According to Connor, "Dispersion is a measure of the extent to which the individual vary."

According to Kafka, "The measurement of a scatteredness of the mass of figures in a series about an average is called measure of dispersion or measure of variation".

3.1.2. Properties of an Ideal Measure of Dispersion

1) In should be capable of treating it by Algebraic or Statistical techniques.
2) It should be easy to calculate i.e. by simple methods.
3) It should be easy to understand i.e. Even a layman must understand about its message or what it demonstrates.
4) It must not be affected by different samples or fluctuation of sampling. Every sample should give same type of information.
5) The quality and quantity of each term must affect it. As in median, last value may be 15 or 15000 in the series 3, 5, 7, 9, 15 (15000); does not effect at all.

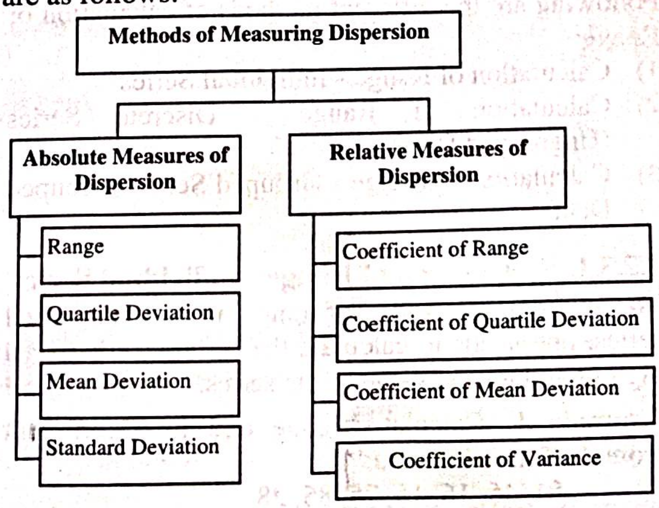
3.1.3. Significance of Measuring Variation

Measuring variation is significant for some of the following purposes:

1) Measuring variability determines the reliability of an average by pointing out as to how far an average is representation of the entire data.
2) Another purpose of measuring variability is to determine the nature and cause variation in order to control the variation itself.
3) Measures of variation enable comparisons of two or more distributions with regard to their variability.
4) Measuring variability is of great importance to advanced statistical analysis. **For example,** sampling or statistical inference is essentially a problem in measuring variability.

3.1.4. Methods of Measuring Dispersion

The two types of methods for measuring dispersion are as follows:



3.1.4.1. Absolute Measures of Dispersion

The measure of dispersion which is expressed in terms of the units of the observations (e.g., Rupees, Meter, Years, etc.) is called absolute measure of dispersion.

1) Range
2) Quartile Deviation
3) Mean Deviation
4) Standard Deviation

3.1.4.2. Relative Measures of Dispersion

The measure of dispersion which is independent of unit or may involve the point about which the deviations are taken, is known as relative measure of dispersion or coefficient of dispersion. There is only one relative measure corresponding to an absolute measure of dispersion.

1) Coefficient of Range
2) Coefficient of Quartile Deviation
3) Coefficient of Mean Deviation
4) Coefficient of Variance

3.2. RANGE

3.2.1. Introduction

Range is the simplest absolute measure of dispersion which shows the difference between the highest and the lowest value in a series.

Mathematically: R = L – S

Where, R = Range,
L = Maximum (largest) value,
S = Minimum (smallest) value.

3.2.2. Coefficient of Range

Relative measure of variation (coefficient of range) is used to compare the series. Coefficient of range is defined as:

Coefficient of Range (C.R.) =

$$\frac{\text{Largest Value} - \text{Smallest Value}}{\text{Largest Value} + \text{Smallest Value}} = \frac{L-S}{L+S}$$

3.2.3. Methods of Calculation of Range

Following are the different methods of calculation of Range:

1) Calculation of Range - Individual Series
2) Calculation of Range - Discrete Series (Ungrouped Data)
3) Calculation of Range - Grouped Series (Grouped Data)

3.2.3.1. Calculation of Range - Individual Series

To find out the value of range in the individual series, one needs to calculate the difference between highest and lowest value of the series.

Example 1: Calculate Range and its Coefficient from the following data:

54, 45, 16, 15, 75, 85, 28

Solution:

Here, Largest value (L) = 85, Smallest value (S) = 15

Range = L – S = 85 – 15 = 70

Coefficient of Range

$$= \frac{L-S}{L+S} = \frac{85-15}{85+15} = \frac{70}{100}$$

= 0.7 or 70%

3.2.3.2. Calculation of Range - Discrete Series (Ungrouped Data)

In case of individual series, it can also be calculated by taking the difference of highest and lowest value. In this case frequency of the series is not taken into consideration for the calculation.

Example 2: Find range and coefficient of range from the following data:

X	5	15	25	35	45	55
f	7	14	18	10	4	1

Solution: Here, Largest value (L) = 55,
Smallest value (S) = 5

∴ Range = L – S = 55 – 5 = 50

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{55-5}{55+5} = \frac{50}{60}$$

= 0.83 or 83.33%

3.2.3.3. Calculation of Range - Grouped Series (Grouped Data)

In case of grouped series it can be calculated by taking the difference of lower limit of the lowest class interval and upper limit of the highest class interval. It can also be measured by taking the difference of mid value of the lowest class interval and mid value of the highest class interval.

Example 3: Calculate Range and the Coefficient of range from the following data:

X	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90	90 – 100
f	6	4	15	24	11	3	10	16	20

Solution: Largest value (L) = 100;
Smallest value (S) = 10

∴ Range (L – S) = 100 – 10 = 90

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{(100-10)}{(100+10)} = \frac{90}{110}$$

= 0.82 or 82%

3.3. INTERQUARTILE RANGE

In a data set 'Interquartile range' is the spread of values between third quartile (Q_3) and first quartile (Q_1). In other words, it can be defined by the range of middle 50% of the data. Interquartile range is expressed as follows:

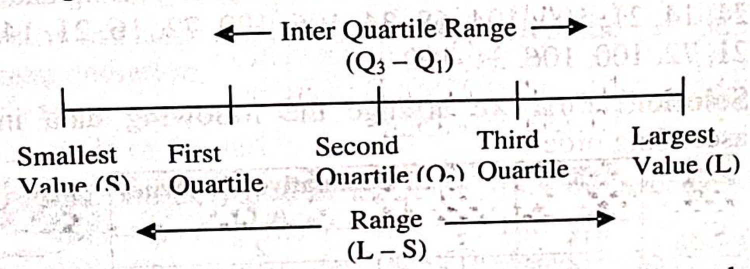
Interquartile Range = $Q_3 - Q_1$

3.4. QUARTILE DEVIATION

3.4.1. Introduction

Quartile deviation is another measure of variation which gives the solution to overcome the limitation of range. In a data set, it calculates the spread over the middle half of the values. This measure of variation minimises the effect of extreme values (also known as outliers).

In this method, the study of **Interquartile range** is necessary because a large amount of values in the data set lie in the central part of the frequency distribution. To calculate this value, all the data set is divided into four parts. Every part of the data set contains 25% of the observed value. In these values the highest one is known as quartile.



The half of the difference between third quartile and first quartile (Q_3 – Q_1) is known as **'semi-interquartile** or **quartile deviation'**.

Mathematically,

$$\text{Quartile Deviation, Q.D.} = \frac{Q_3 - Q_1}{2}$$

Where, $(Q_3 - Q_1)$ = Interquartile range

3.4.2. Coefficient of Quartile Deviation

If the difference of the third and first quartiles is divided by the sum of the third and first quartiles then it is known as the **'coefficient of quartile deviation'**. In case of open-ended distribution (when the frequency distribution has the extreme class limits and no specific class limits), it is a very useful measure. Mathematically, coefficient of quartile deviation is defined as follows:

Coefficient of Quartile Deviation $(\text{CQD}) = \frac{Q_3 - Q_1}{Q_3 + Q_1}$

Where,

CQD = Coefficient of Quartile Deviation

Q_3 = Third quartile

Q_1 = First quartile

3.4.3. Methods of Calculation of Quartile Deviation

Following are the different methods of calculation of quartile deviation:

1) Computation of Quartile Deviation - Individual Series

Q_1 = Size of $\left(\frac{N}{4}\right)^{th}$ item $=\left(\frac{56}{4}\right)^{th}$ item

= 14^{th} item which lies in class interval 30-40.

Here,
$L_1 = 30, L_2 = 40, c = 11, f = 15$ and $i = L_2 - L_1 = 40 - 30 = 10$.

So, $Q_1 = L_1 + \frac{\frac{N}{4} - c}{f} \times i = 30 + \frac{14-11}{15} \times 10 = 30 + \frac{30}{15} = 32$

Q_3 = Size of $\left(\frac{3N}{4}\right)^{th}$ item $=\left(\frac{3\times56}{4}\right)^{th}$ item

= 42^{th} item which lies in class interval 60-70.

Here, $L_1 = 60$, $L_2 = 70$, $c = 41$, $f = 9$ and
$i = L_2 - L_1 = 70 - 60 = 10$.

So, $Q_3 = L_1 + \frac{\frac{3N}{4} - c}{f} \times i = 60 + \frac{42-41}{9} \times 10$

$= 60 + 1.1 = 61.1$

Quartile Deviation $= \frac{Q_3 - Q_1}{2}$

$= \frac{61.1-32}{2} = \frac{29.1}{2} = 14.55$

3.5. MEAN DEVIATION/ MEAN ABSOLUTE DEVIATION (M.D.)

3.5.1. Introduction

The arithmetic average of the absolute deviation of a series is known as the mean deviation. This deviation is calculated from any one of the measures of averages such as mean, mode and median.

Mathematically,

$$\text{Mean Deviation (MD)} = \delta = \frac{\sum|d|}{N}$$

Where, $\sum|d|$ = Sum of all deviations (taken either from mean or median, ignoring ± signs.)

In a series, it can be defined as the arithmetic average of the deviation of various items from mean or median of that series. For calculating the mean deviation, median is preferred because the addition of the deviation from the mean is greater than that from the median. So, the value of mean deviation measured from the mean is greater than the value measured from the median. Mean deviation is also known as **First Moment of Dispersion.**

Note: It is generally measured either from median or mean.

3.5.2. Coefficient of Mean Deviation

$$\text{Coefficient of mean deviation from mean} = \frac{\text{Mean Deviation(M.D.)}}{\text{Arithemtic mean}}$$

$$\text{Coefficient of mean deviation from median} = \frac{\text{Mean Deviation(M.D.)}}{\text{Median}}$$

$$\text{Coefficient of mean deviation from mode} = \frac{\text{Mean Deviation(M.D.)}}{\text{Mode}}$$

3.5.3. Method of Calculation of Mean Deviation

Following are the different methods of calculation of mean deviation:

1) Computation of Mean Deviation - Individual Series
2) Computation of Mean Deviation - Discrete Series
3) Computation of Mean Deviation - Continuous Series

3.5.3.1. Computation of Mean Deviation - Individual Series

Following steps are involved in the calculation of the mean deviation:

1) First, measure the average mean, median or mode of the series.
2) Find the deviation of the items from the average, while ignoring the positive (+) and negative (–) signs. The resultant deviation is denoted by $|dx|$.
3) Next, measure the total sum of these deviations. It is represented by $\sum|dx|$.
4) In the last step, divide the sum obtained by the number of items.

Symbolically,

$$\text{Mean Deviation} = \frac{\sum|d_x|}{N}$$

Where, $|d_x|$ = deviation from mean (or median) ignoring ± signs.

N = Number of items

Example 8: The following are the monthly expenditure of six families. Calculate mean deviation from mean and mean deviation from median.

Expenditure (₹)	4,260	4,980	8,460	5,240	4,780	6,480

Solution: First we arrange the given data in ascending order:

Expenditure (₹)	4,260	4,780	4,980	5,240	6,480	8,460

Here, N = 6 which is even, so

$$\text{Median} = \frac{\left(\frac{N}{2}\right)^{th}\text{value} + \left(\frac{N}{2}+1\right)^{th}\text{value}}{2}$$

$$= \frac{\left(\frac{6}{2}\right)^{th}\text{value} + \left(\frac{6}{2}+1\right)^{th}\text{value}}{2}$$

$$= \frac{3^{rd}\text{value} + 4^{th}\text{value}}{2} = \frac{4980+5240}{2} = 5110$$

$$\overline{X} = \frac{4{,}260+4{,}780+4{,}980+5{,}240+6{,}480+8{,}460}{6}$$

If $\overline{X}$ is the mean of $X_1, X_2, \ldots, X_N$, then,

$$\sigma=\sqrt{\left[\frac{1}{N}\left\{(X_1-\overline{X})^2+\ldots\ldots\ldots+(X_N-\overline{X})^2\right\}\right]}=\sqrt{\sum_{i=1}^{N}\frac{(X_i-\overline{X})^2}{N}}$$

According to Yule and Kenndall, "The standard deviation is the square-root of the arithmetic mean of the square of all the deviations, deviations being measured from the arithmetic mean of the observation."

3.6.2. Properties of Standard Deviation

1) Standard deviation is always positive.
2) Standard deviation will not change if all the variables are summated with the same number.
3) If researcher multiplies the variable with the same number, then the output received will be multiplied by the square of that same number.
4) In case of multiple distribution (with same mean and known standard deviation), SD can be calculated as:
   i) If samples size are same
   $$\sigma=\sqrt{\frac{\sigma_1^2+\sigma_2^2+\ldots+\sigma_n^2}{N}}$$
   ii) If samples size are different
   $$\sigma=\sqrt{\frac{k_1.\sigma_1^2+k_2.\sigma_2^2+\ldots.+k_N.\sigma_N^2}{k_1+k_1+\ldots+k_N}}$$

3.6.3. Coefficient of Standard Deviation

The standard deviation is the absolute measure of dispersion. Its relative measure is called the standard coefficient of dispersion or coefficient of standard deviation. It is defined as:

$$\text{Coefficient of Standard deviation} = \frac{\sigma}{\overline{X}}$$

3.6.4. Difference between Mean Deviation and Standard Deviation

Table 3.1 shows the difference between mean and standard deviation:

Table 3.1: Difference between Mean Deviation and Standard Deviation

Mean Deviation	Standard Deviation
Deviations are calculated from mean, median or mode	These are calculated from the arithmetic median mean only.
The algebraic signs have to be ignored - only values of deviations are taken.	Since the deviations are squared, the plus and minus signs need not deviations be omitted.
It is based on simple average of the sum of absolute deviations	It is based on the square root of the average of the squared deviations.
It is simple to calculate when mean is a round number. The short-cut method is somewhat cumbersome.	This is somewhat complex because of squaring of the deviations but it is suitable in all cases - whether the mean is a round number or a fraction, since a shortcut method is also available.
It lacks mathematical properties since only absolute values are considered.	It is mathematically sound on account of the fact that algebraic signs are not ignored.

3.6.5. Methods of Calculation of Standard Deviation

Following are the methods of calculation of standard deviation:

1) Calculation of Standard Deviation – Individual Series
2) Calculation of Standard Deviation – Discrete series
3) Calculation of Standard Deviation – Continuous Series

3.6.5.1. Calculation of Standard Deviation – Individual Series

There are two methods of calculating standard deviation in an individual observation or series:

1) **Deviation Taken from Actual Mean:** This method is adopted when the mean is a whole number.

   The following are the steps:

   i) Find out the actual mean of the series.
   ii) Find out the deviation of each value from the mean (d_x or $x = X - \overline{X}$).
   iii) Square the deviations and take the total of squared deviations Σx^2.
   iv) Divide the total (Σx^2) by the number of observations. The square root of the quotient is standard deviation. Thus apply the following formula:

$$\sigma=\sqrt{\frac{\Sigma x^2}{N}} \text{ or } \sqrt{\frac{\Sigma(X-\overline{X})^2}{N}}$$

Example 11: Find the standard deviation from the given data:

X:	16,	20,	18,	19,	20,	20,	28,	17,	22,	20

Solution:

Calculation of Standard Deviation

X	$\overline{X}=20$ $x=X-\overline{X}$	x^2
16	–4	16
20	0	0
18	–2	4
19	–1	1
20	0	0
20	0	0
28	8	64
17	–3	9
22	2	4
20	0	0
$N = 10, \Sigma X = 200$	$\Sigma x = 0$	$\Sigma x^2 = 98$

$$\overline{X}=\frac{\Sigma X}{N}=\frac{200}{10}=20$$

$$\Rightarrow \quad \sigma=\sqrt{\frac{\Sigma x^2}{N}}=\sqrt{\frac{98}{10}}=\sqrt{9.8}=3.13$$

3.9. COMPARISON OF VARIOUS MEASURES OF DISPERSION

Following table illustrates the comparison of various measures of dispersion:

Parameters	Range	Mean Deviation	Standard deviation
1) **Concept**	It is the difference between the largest and the smallest value of the variable.	It is the arithmetic average of the deviation of all the values of a series.	It is the sequence root of th arithmetic mean of the squares all the deviations.
2) **Criteria**	It is highly useful in quality control measures.	Deviations of values may be taken from any measure of central tendency.	Deviations of values are taken onl form actual mean of the series.
3) **Objective**	It only calculates the difference between the largest and the smallest number.	In mean deviation only deviation are added.	In standard deviation only squar of deviations are added.
4) **Output**	Range provides the minimum and maximum value.	Absolute values of deviations are taken. All the deviation are taken to be positive and negative sign are ignored.	Negative sign are neutralized b squaring the deviation.
5) **Functionality**	It provides knowledge about dispersion in the series.	Simply average of the absolute values of deviations is taken.	Square root of the arithmetic averag of square of deviation is taken.
6) **Capability**	It is not capable for algebraic calculations.	It is not capable for algebraic calculations.	It is capable for algebrai calculations.
7) **Usability**	It is not useful for statistical anaiysis.	It is less useful for statistical analysis.	It is more useful for statistic analysis.
8) **Reliability**	Range has a lack of reliability.	It is more reliable than range.	It is very much reliable than ran and mean deviation.
9) **Stability**	Range is unstable.	Mean deviation is stable.	Standard deviation is stable.
10) **Cost Effectiveness**	It is time and cost effective.	It is not time effective.	It is not time effective.
11) **Data Distribution**	It is not useful for frequency distribution.	It is useful for frequency distribution.	It is very much useful f frequency distribution.

Chapter 4

Skewness and Kurtosis

4.1. MEASURES OF SHAPE

The principal measures of distribution shape used in statistics are skewness and kurtosis.

Measures of Shape

Skewness

Moments

Kurtosis

4.2. SKEWNESS

4.2.1. Meaning

Measure of skewness is a measure that gives the direction and extent of symmetry or asymmetry in a series and permits us to compare the two (or more) series. In other words, skewness tells the shape of a distribution. Thus, "Lack of asymmetry or symmetry of a given distribution of a random variable is called skewness". The idea about the shape of the curve which can be drawn with the help of the given data is provided by skewness.

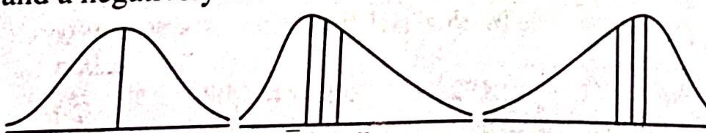
A distribution is said to be skewed if:

1) The values of mean, median and mode are not equal, i.e., Mean ≠ Median ≠ Mode.
2) Quartiles are not equi-distant from median, and
3) The curve which is drawn with the help of the given data is not symmetrical and stretched more to one side in comparison to other side.

According to Croxton and Cowden, "When a series is not symmetrical it is said to be asymmetrical or skewed."

According to Morris Hamburg, "Skewness refers to the asymmetry or lack of symmetry in the shape of a frequency distribution".

The concept of skewness will be clear with the help of following three figures (**figure 4.1**), showing a Normal distribution, a positively skewed distribution and a negatively skewed distribution:



$\overline{X} = M = Z$ **Normal Distribution**

$\overline{X} > M > Z$ **Positively Skewed Distribution**

$\overline{X} < M < Z$ **Negatively Skewed Distribution**

Figure 4.1

1) **Normal Distribution:** It is also called symmetrical distribution. Its curve called notmal curve or curve of error. It is clear from the given figure that the values of mean, median and mode are equal in a 'symmetric distribution'. Equal number of items is spread on both sides of the mid-point of the curve. If a distribution of a random variable is not symmetric then it is referred as skewed distribution and this type of distribution could either be positively skewed or negatively skewed.

   Here value of $\overline{X}, M, Z$ are same.

2) **Positively Skewed Distribution:** It is clear from the given figure that the value of mean is maximum and the value of mode is minimum in a positively skewed distribution. Median lies in between the mean and mode.

   Here, $\overline{X}$ is greater than M and M is greater than Z and it is positive skewness.

3) **Negatively Skewed Distribution:** The value of mode is maximum and the value of mean is minimum in case of negatively skewed distribution. Median lies in between the mean and mode. It should be noted that in moderately symmetrical distributions, the interval between the mean and the median is approximately one-third of the interval between the mean and the mode. It is the relationship which provides a means of measuring the degree of skewness.

   Here, Z is maximum value and M is less than Z and $\overline{X}$ is the smallest value.

4.2.2. Objectives of Skewness

1) Measures of skewness tell about the degree of concentration.
2) Moderately skewed distribution is the basis of the empirical relations of mean, median and mode.

   The measure of skewness tells about the extent to which the empirical relationship holds true.
3) Skewness helps to know about the normality of the distribution. Various statistical measures such as error of mean, sampling error are based on the assumption of a normal distribution.

4) Measures of skewness gives an idea about the nature of variation of the item from the central value.
5) Measures of skewness help in knowing about the dispersion on either side of mode, which differs in the arrangement of their frequencies.

4.2.3. Tests of Skewness

Following are the tests which are applied to find that a distribution is skewed or not:

1) If a distribution is skewed then the value of mean, median and mode would not coincide. The value of median generally lies between the mean and the mode. In a moderately asymmetrical distribution,

$$\text{Mean} - \text{Mode} = 3\ (\text{Mean} - \text{Median})$$
$$\Rightarrow 3\ \text{Median} = 3\ \text{Mean} + \text{Mode} - \text{Mean}$$
$$= 2\ \text{Mean} + \text{Mode}$$
$$= 2\ \text{Mean} + \text{Mode} + 2\ \text{Mode} - 2\ \text{Mode}$$
$$= 3\ \text{Mode} + 2\ (\text{Mean} - \text{Mode})$$
$$\Rightarrow \text{Median} = \text{Mode} + \frac{2}{3}(\text{Mean} - \text{Mode})$$

2) If a distribution is skewed then the two quartiles would not be equi-distant from the median. In other words, it can be said that $(Q_3 - M) - (M - Q_1) \neq 0$.
3) If a distribution is skewed then its graph would not give a symmetrical bell-shaped curve.
4) The sum of positive deviations and the sum of negative deviations from the median would not be equal.
5) At various points, frequencies are not equally distributed which are equi-distant from the mode. In an asymmetrical distribution (Mean – Median) = 3 (Mean – Mode).

4.2.4. Advantages of Skewness

1) If a given distribution is normal then the skewness would be zero and it is called symmetric distribution. But generally, data points are not perfectly symmetric.
2) With the help of skewness, we know that the deviation from the mean is whether positive or negative.
3) D'Agostino's K-squared test is a goodness-of-fit normality test based on sample skewness and sample kurtosis.

4.2.5. Methods of Computing Skewness and their Coefficient/ Measures of Skewness

Measure of skewness may be subjective or relative. In a frequency distribution, absolute measures tell us about the direction and extent of asymmetry. Relative measures, generally called coefficient of skewness, provides the facility to compare two or more frequency distributions. We can find the skewness by two methods:

Measures of Skewness
- Karl Pearson's Coefficient of Skewness
- Bowley's Coefficient of Skewness
- Kelly's Coefficient of Skewness

4.2.5.1. Karl Pearson's Coefficient of Skewness

1) **When Measure Depends upon Difference of Mean and Made.**

Absolute Measure, Sk = Mean – Mode

$$\text{Relative Measure, } J = \frac{\text{Mean} - \text{Mode}}{\text{Standard Deviation}} \text{ or } \frac{\overline{X} - Z}{\sigma}$$

Where, $\overline{X}$ = Arithmetic mean; Z = Mode; and σ = standard deviation.

2) If mode is not determined, then

Absolute Measure, Sk = 3(Mean – Median) or $3(\overline{X} - M)$

$(Sk) = 3(\overline{X} - M)$

Relative measure, coefficient

$$J = \frac{3(\text{Mean} - \text{Median})}{\text{Standard Deviation}} \text{ or } \frac{3(\overline{X} - M)}{\sigma}$$

Where, M = Median.

Note: Mathematically there are no limits for J, but practically

$J = \dfrac{\text{Mean} - \text{Mode}}{\text{Standard Deviation}}$ takes values between ± 1;

$J = \dfrac{3(\text{Mean} - \text{Median})}{\text{Standard Deviation}}$ takes values between ± 3

Example 1: From the following data find out the Karl Pearson's Coefficient of Skewness:

Measurement	20	21	22	23	24	25
Frequency	1	3	8	11	6	1

Solution: Let the assumed mean (A) = 22

Calculation of Karl Pearson's Coefficient of Skewness

Measurement (X)	Frequency (f)	d_x = (X – 22)	fd_x	d_x^2	fd_x^2
20	1	– 2	– 2	4	4
21	3	– 1	– 3	1	3
22	8	0	0	0	0
23	11	+ 1	+ 11	1	11
24	6	+ 2	+ 12	4	24
25	1	+ 3	+ 3	9	9
	N = 30		Σfd_x = + 21		$\sum f(d_x)^2 = 51$

$$\text{Arithmetic Mean, } (\overline{X}) = A + \frac{\sum fd_x}{N} = 22 + \frac{21}{30} = 22.7$$

Mode (by inspection) Z = 23

Standard Deviation,

$$\sigma = \sqrt{\frac{\sum f(d_x)^2}{N} - \left(\frac{\sum fd_x}{N}\right)^2} = \sqrt{\frac{51}{30} - \left(\frac{21}{30}\right)^2}$$

$$= \sqrt{1.7 - 0.49} = \sqrt{1.21} = 1.1$$

Absolute Measure, Sk = $\bar{X} - Z = 22.7 - 23 = -0.3$

Karl Pearson's Coefficient of Skewness,

$$J = \frac{\bar{X} - Z}{\sigma} = \frac{22.7 - 23}{1.1} = -\frac{0.3}{1.1} = -0.273$$

Example 2: From the information given below, calculate Karl Pearson's Coefficient of Skewness:

Measure	Place A	Place B
Mean	160	150
Median	142	145
Standard deviation	28	52
Third quartile	198	255
First quartile	68	78

Solution: Place A:

$$J = \frac{3(\bar{X} - M)}{\sigma} = \frac{3(160 - 142)}{28} = \frac{3 \times 18}{28} = 1.93$$

Place B:

$$J = \frac{3(\bar{X} - M)}{\sigma} = \frac{3(150 - 145)}{52} = \frac{15}{52} = 0.3$$

Example 3: Calculate Karl Pearson's Coefficient of Skewness, with use of following observations:

Score: 45, 41, 50, 48, 45, 43, 47, 40, 45, 49,

Solution: Let be assumed mean (A)= 45

X	$d_x = (X - A)$	d^2_x
40	–5	25
41	–4	16
43	–2	4
45	0	0
45	0	0
45	0	0
47	+2	4
48	+3	9
49	+4	16
50	+5	25
N = 10	$\Sigma d_x = +3$	$\Sigma d^2_x = 99$

Mean, $\bar{X} = A + \frac{\Sigma d_x}{N} = 45 + \frac{3}{10} = 45 + 0.3 = 45.3$

Mode = 45

Standard Deviation

$$= \frac{1}{N}\sqrt{\Sigma d_x^2 \cdot N - (\Sigma d_x)^2} = \frac{1}{10}\sqrt{99 \times 10 - (3)^2}$$

$$= \frac{1}{10}\sqrt{990 - 9} = \frac{1}{10}\sqrt{981} = \frac{1}{10} \times 31.32 = 3.132$$

Coefficient of Skewness $J = \frac{\bar{X} - Z}{\sigma} = \frac{45.3 - 45}{3.132}$

$$= \frac{0.3}{3.132} = +0.0958$$

Example 4: Determine Karl Pearson's coefficient of skewness for given data:

Size	:	10	11	12	13	14	15
Frequency	:	3	8	12	8	6	3

Solution: Let be Assumed Mean (A) = 12

Size (X)	Frequency (f)	Deviation from Assumed Mean $d_x = (X - 12)$	d^2_x	$f \times d_x$	$fd_x \times d_x = fd^2_x$
10	3	–2	4	–6	12
11	8	–1	1	–8	8
12	12	0	0	0	0
13	8	+1	1	+8	8
14	6	+2	4	+12	24
15	3	+3	9	+9	27
Total	N = 40	–	–	$\Sigma fd_x = +15$	$\Sigma fd^2_x = 79$

Arithmetic Mean,

$$(\bar{X}) = A + \frac{\Sigma fd_x}{N} = 12 + \frac{15}{40} = 12 + 0.375 = 12.375$$

Mode by observation = 12.

Standard Deviation

$$(\sigma) = \sqrt{\frac{\Sigma fd_x^2}{N} - \left(\frac{\Sigma fd_x}{N}\right)^2} = \sqrt{\frac{79}{40} - \left(\frac{15}{40}\right)^2}$$

$$= \sqrt{1.975 - (0.375)^2} = \sqrt{1.975 - 0.1406}$$

$$= \sqrt{1.8344} = 1.3544$$

Skewness $= \bar{X} - Z = 12.375 - 12 = 0.375$

Coefficient of Skewness

$$(J) = \frac{\bar{X} - Z}{\sigma} = \frac{12.375 - 12}{1.3544} = \frac{0.375}{1.3544} = 0.2769$$

Example 5: Calculate Coefficient of Skewness from the following table:

Marks	0–10	10–20	20–30	30–40	40–50	50–60	60–70
f :	5	10	25	40	26	11	3

Solution: Let be Assumed mean = 35

Marks	f	X	$d_x = (X - \bar{X})$	fd_x	fd^2_x
0–10	5	5	–30	–150	45
10–20	10	15	20	–200	4000
20–30	25	25	–10	–250	2500
30–40	40	35	0	0	0
40–50	26	45	+10	+260	2600
50–60	11	55	+20	+220	4400
60–70	3	65	+30	+90	2700
Total	N= 120	-	-	$\Sigma fd_x = -30$	$\Sigma fd^2_x = 20700$

Mean ($\bar{X}$)

Mean,

$$\bar{X} = A + \frac{\Sigma fd_x}{N} = 35 + \frac{-30}{120} = 35 - 0.25 = 34.75$$

Mode (Z)

Mode Group = 30 – 40

Made,

$$= \ell_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} = 30 + \frac{40 - 25}{80 - 25 - 26} \times 10$$

$$= 30 + \frac{150}{29} = 30 + 5.17 = 35.17$$

Standard Deviation (σ)

$$\sigma = \frac{1}{N}\sqrt{\Sigma f d_x^2 \times N - (\Sigma f d_x)^2} = \frac{1}{120}\sqrt{20700 \times 120 - (-30)^2}$$

$$= \frac{1}{120}\sqrt{2484000 - 900} = \frac{1}{120}\sqrt{2483100}$$

$$= \frac{1}{120} \times 1575.79 = 13.13$$

Coefficient of Skewness(J)

$$J = \frac{\bar{X} - Z}{\sigma} = \frac{34.75 - 35.17}{13.13} = \frac{-0.42}{13.13} = -0.032$$

Example 6: With the help of given table, calculate Karl Pearson's Coefficient of Skewness:

Height (in inches)	58	59	60	61	62	63
No. of persons	10	18	30	42	35	28

Solution: Let us consider Assumed mean (A)=60

Calculation of Karl Pearson's coefficient of Skewness

Height (X)	f	d = (X – 60)	fd	fd²
58	10	–2	–20	40
59	18	–1	–18	18
60	30	0	0	0
61	42	+1	+42	42
62	35	+2	+70	140
63	28	+3	+84	252
	N = 163		Σfd=158	Σfd² = 492

$$\bar{X} = A + \frac{\Sigma fd}{N} = 60 + \frac{158}{163} = 60 + 0.969 = 60.969$$

$$\sigma = \sqrt{\frac{fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} = \sqrt{\frac{492}{163} - \left(\frac{158}{163}\right)^2}$$

$$= \sqrt{3.0184 - 0.9395} = \sqrt{2.0789} = 1.4418$$

By inspection, mode is 61 (as its frequency is maximum)

Thus, $\bar{X} = 60.969, \sigma = 1.4418, Z = 61$

∴ Coefficient of Skewness

$$= \frac{\bar{X} - Z}{\sigma} = \frac{60.969 - 61}{1.4418} = \frac{-0.031}{1.4418} = -0.0215$$

4.2.5.2. Bowley's Coefficient of Skewness

An alternative measure of skewness has been given by **Professor A.L. Bowley**. It is based on the quartiles. If a distribution is symmetric then the first and third quartiles are equi-distant from the median which is shown in the following diagram. So if the distribution is asymmetric then, both quartiles are not equidistant from the median. If the distribution is symmetric then:

$Q_3 - \text{Median} = \text{Median} - Q_1$ or $Q_3 + Q_1 - 2\,\text{Median} = 0$

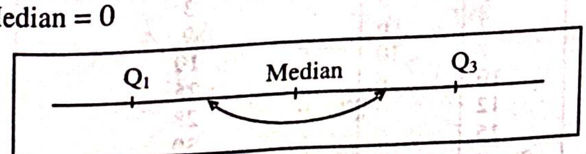


Figure 4.2

If the given distribution is positively skewed then the top 25% of the items will tend to be farther from median than the bottom 25%. This means that Q_3 will be farther from median than Q_1 and the reverse will be for negative skewness. Hence a possible measure is:

$$SK_B = \frac{(Q_3 - \text{Med.}) - (\text{Med.} - Q_1)}{(Q_3 - \text{Med.}) + (\text{Med.} - Q_1)} \text{ or } \frac{Q_3 + Q_1 - 2\,\text{Med.}}{Q_3 - Q_1}$$

SK_B = Bowley's coefficient of skewness.

Properties of Bowley's Coefficient of Skewness

1) If the distribution has open end classes or unequal class intervals then the Bowley's measure is very useful. In this condition Pearson's coefficient of skewness cannot be used.
2) Bowley's measure is based on the central 50% of the data while ignoring the remaining 50% of the data towards both the extremes. This is one of the major limitations of Bowley's measure.
3) Continuous frequency distribution is the basis of the Bowley's measure.
4) Limits for Bowley's coefficient of skewness are: $-1 \le SK_B \le 1$, i.e., Bowley's coefficient of skewness ranges from –1 to 1.
5) The values of the coefficient of skewness obtained by Bowley's formula and Pearson's formula cannot be compared. If the distribution is symmetrical then $SK_B = 0$, which means the absence of skewness.

Example 7: Find Bowley's Coefficient of Skewness, if difference between two quartiles = 7, Sum of two quartiles = 24 and Median = 11.

Solution: Coefficient of skewness can be determined by applying Bowley's method:

Coefficient of Skewness, $(SK_B) = \frac{Q_3 + Q_1 - 2\,\text{Med.}}{Q_3 - Q_1}$

$Q_3 - Q_1 = 7, Q_3 + Q_1 = 24$, Median = 11

Coefficient of Skewness, (SK_B)

$$= \frac{24 - 2 \times 11}{7} = \frac{24 - 22}{7} = 0.286$$

Example 8: Calculate Bowley's Coefficient of Skewness from the data given below:

Profits per Shop (₹)	100-200	200-300	300-400	400-500	500-600	600-700	700-800
Number of Shops	10	18	20	26	30	28	18

Solution: **Calculation of Coefficient of Skewness**

Profits per Shop (₹) (X)	Number of Shops (f)	Cumulative Frequency (c.f.)
100-200	10	10
200-300	18	28
300-400	20	48
400-500	26	74
500-600	30	104
600-700	28	132
700-800	18	150

4.2.6. Difference between Skewness & Dispersion

Table 4.1 shows the difference between Skewness and Dispersion:

Table 4.1: Difference between Skewness and Dispersion

Skewness	Dispersion
Skewness gives the idea about the direction of variation.	Dispersion gives the idea about the amount of the variation.
The tendency of the variation of data points into a certain direction can be understood with the use of skewness.	Dispersion is used to know the range of the data points and offset from the mean.
Skewness gives an idea about the shape of the series.	Dispersion gives an idea about the composition of the series.

4.3. MOMENTS

4.3.1. Concepts

According to Trederic Mills, "Moment is a familiar mechanical term for the measure of a force with reference to its tendency to produce rotation. The strength of this tendency depends, obviously, on the amount of the force and the distance from the origin of the point at which the force is exerted".

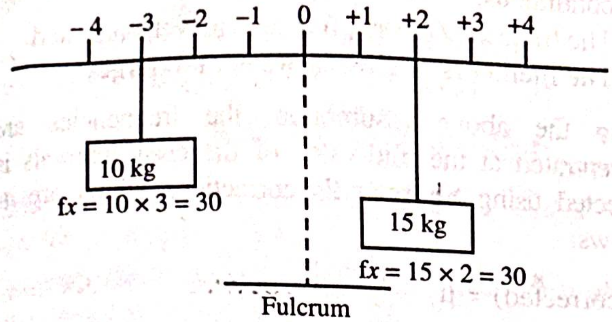
The definition of moments shows that it includes two factors such as the quantum of the force and the distance from which it is applied.

If f_1, f_2 and f_3 are the forces applied with the distances of x_1, x_2, x_3 respectively, then f_1x_1 is the moment of the first force, f_2x_2 is the moment of the second force and f_3x_3 is the moment of the third force.

If these moments are added then we get the total moments which is represented by $\sum fx$. If this total value is divided by the total of the force of $\sum f$, then we get a value $\frac{\sum fx}{\sum f}$, which is called a 'moment'.

If the total force on both sides of the origin (which is zero) is equal then the above scale will be balanced.

The following **diagram** below will illustrate the point:



The above scale will balance if the total force on either side of the origin (which is zero) is equal.

4.3.2. Uses of Moments

Different uses of moments are as follows:

1) It is used to define the central tendency of a distribution.
2) The dispersion or scatter of items of a series is defined by the moments.
3) It is used to identify the symmetry of the distribution whether the distribution is skewed or non-skewed.
4) Moments helps in determining the nature of the symmetrical distribution in terms of kurtosis.

4.3.3. Types of Moments

The moments are classified into three types on the basis of the method of computation. These are as follows:

1) Moments about the mean (central moment),
2) Moments about an arbitrary value (assumed moments), and
3) Moments about the origin (zero moments).

How moments help in analysing a frequency distribution is summarized below:

Moment	What it Measures
1) First moment about origin	Mean
2) Second moment about the mean	Variance
3) Third moment about the mean	Skewness
4) Fourth moment about the mean	Kurtosis

4.3.3.1. Moments about the Mean/Central Moments

Moments about the mean are generally represented by μ (read as mu)

Thus the various moments about the mean would be:

Individual Series	Frequency Distribution
First Moment about the Mean: $\mu_1 = \frac{\sum(X-\bar{X})}{N}$ or $\frac{\sum d}{N} = 0$	$\frac{\sum f(X-\bar{X})}{\sum f}$ or $\frac{\sum fd}{\sum f}$
Second Moment about the Mean: $\mu_2 = \frac{\sum(X-\bar{X})^2}{N}$ or $\frac{\sum d^2}{N} = \sigma^2$	$\frac{\sum f(X-\bar{X})^2}{\sum f}$ or $\frac{\sum fd^2}{\sum f} = \sigma^2$
Third Moment about the Mean: $\mu_3 = \frac{\sum(X-\bar{X})^3}{N}$ or $\frac{\sum d^3}{N}$	$\frac{\sum f(X-\bar{X})^3}{\sum f}$ or $\frac{\sum fd^3}{\sum f}$
Fourth Moment about the Mean: $\mu_4 = \frac{\sum(X-\bar{X})^4}{N}$ or $\frac{\sum d^4}{N}$	$\frac{\sum f(X-\bar{X})^4}{\sum f}$ or $\frac{\sum fd^4}{\sum f}$

Where, $N = \Sigma f$

4.3.3.2. Moments about Arbitrary Point

We know that the deviations of the values can be obtained from any arbitrary value A, corresponding moments about A can also be defined. The r^{th} moment about A is represented by μ'_r. It is defined by:

$\mu'_r = \frac{1}{N}\Sigma f_i (X_i - A)^r$, i.e., μ'_r is the mean of $(X_i - A)^r$ values.

If r = 0, we have $\mu'_0 = \frac{1}{N}\Sigma f_i (X_i - A)^0 = 1$

If r = 1, we have

$$\mu'_1 = \frac{1}{N}\Sigma f_i (X_i - A) = \frac{1}{N}\Sigma f_i X_i - A\frac{\Sigma f_i}{N} \text{ or } \mu'_1 = \bar{X} - A$$

Thus, the first moment about A is equivalent to the difference of $\bar{X}$ and A. In the same way, we have

$$\mu'_2 = \frac{1}{N}\Sigma f_i (X_i - A)^2 \text{ when } r = 2,$$

$$\mu'_3 = \frac{1}{N}\Sigma f_i (X_i - A)^3 \text{ when } r = 3, \text{ etc.}$$

The raw moments are those moments which are about any arbitrary value.

Conversion of Moments about Assumed Mean into Central Moments

We can obtain the four central moments from moments about an assumed mean by the given relations:

$$\mu_1 = \mu'_1 - \mu'_1 = 0$$
$$\mu_2 = \mu'_2 - (\mu'_1)^2$$
$$\mu_3 = \mu'_3 - 3\mu'_1\,\mu'_2 + 2(\mu'_1)^3$$
$$\mu_4 = \mu'_4 - 4\mu'_3\,\mu'_1 + 6\mu'_2(\mu'_1)^2 - 3(\mu'_1)^4$$

4.3.3.3. Moments about Zero or Origin

Moments about Zero are other types of assumed moments where deviations are taken from origin which means from zero. When the values of the variable are of small size then this type of moments are desirable. These moments are signified by given formulae:

Individual Series	Frequency Distribution
$m_1 = \frac{\Sigma x}{N}$	$m_1 = \frac{\Sigma fx}{N}$
$m_2 = \frac{\Sigma x^2}{N}$	$m_2 = \frac{\Sigma fx^2}{N}$
$m_3 = \frac{\Sigma x^3}{N}$	$m_3 = \frac{\Sigma fx^3}{N}$
$m_4 = \frac{\Sigma x^4}{N}$	$m_4 = \frac{\Sigma fx^4}{N}$

Relation between Moments about Zero and Moments about Mean

$$m_1 = A + \mu'_1 = \overline{X}$$

$$m_2 = \mu_2 + (m_1)^2$$

$$m_3 = \mu_3 + 3m_1v_2 - 2m_1^3$$

$$m_4 = \mu_4 + 4m_1m_3 - 6m_1^2v_2 + 3m_1^4$$

Relation between Central Moments and Moments about Origin

We have $m_1 = \overline{X} - 0 = \overline{X}$ ...(1)

The expressions for moments of various orders can be obtained by replacing μ_i by m_i. As a result, the expression for second, third and fourth central moments in terms of moments about origin are written as follows:

$$\mu_2 = m_2 - m_1^2$$

$$\mu_3 = m_3 - 3m_2m_1 + 2m_1^3$$

$$\mu_4 = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4$$

We can similarly write the expressions for second, third and fourth moments about origin in terms of central moments as:

$$m_2 = \mu_2 + m_1^2$$

$$m_3 = \mu_3 + 3\mu_2m_1 + m_1^3$$

$$m_4 = \mu_4 + 4\mu_3m_1 + 6\mu_2m_1^2 + m_1^4$$

4.3.3.4. Coefficients Based on Moments

On the basis of relative proportions of different moments, Alpha (α) Beta (β) and Gamma (γ) coefficients are calculated with the help of the following formulae:

Alpha Coefficients	Beta Coefficients	Gamma Coefficients
$\alpha_1 = \frac{\mu_1}{\sigma} = 0$	$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \alpha_3^2$	$\gamma_1 = \sqrt{\beta_1} = \alpha$
$\alpha_2 = \frac{\mu_2}{\sigma^2} = 1$	$\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}} = \alpha_3$	$\gamma_2 = \beta_2 - 3 = \frac{\mu_4 - 3\mu_2^2}{\mu_2^2}$
$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{\mu_2^{3/2}}$	$\beta_2 = \frac{\mu_4}{\mu_2^2} = \alpha_4$	
$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{\mu_4}{\mu_2^2}$		

These coefficients are used to measure skewness and Kurtosis.

4.3.4. Calculation of Moments/ Sheppard's Corrections for Moments

Higher moments of grouped frequency distributions are computed by assuming that the frequencies are concentrated at the mid-values of class-intervals. However, it produces some errors at the time of calculating the moments. **W.E. Sheppard** proved that, in case:

1) The frequency curve of the distribution is continuous,
2) The frequency tapers off to zero at both ends, and
3) The members of classes are not too large.

Using the above assumption, the frequencies are concentrated at the mid-value of the class intervals is corrected using **Shappard's** corrections which are as follows:

$$\mu_2\text{ (corrected)} = \mu_2 - \frac{h^2}{12}$$

$$\mu_4\text{ (corrected)} = \mu_4 - h^2 - \mu_2 + \frac{7}{240}h_4$$

Here h denotes the width of the class interval; μ_1 and μ_3 require no correction.

Example 17: The first four moments of a distribution about the origin are given as 1, 4, 10, and 46 respectively. Find the different characteristics of the distribution on the basis of the information given. Comment the nature of the distribution.

Solution: In the usual notations, we have:

$A = 0,\ \mu'_1 = 1,\ \mu'_2 = 4,\ \mu'_3 = 10$ and $\mu'_4 = 46$

$\overline{x}$ = First moment about origin = $\mu'_1 = 1$

The variance is calculated as

$$(\sigma^2) = \mu_2 = \mu'_2 - (\mu'_1)^2 = 4 - 1 = 3$$

The standard deviation is given as

$$\text{S.D. } (\sigma) = \sqrt{\sigma^2} = \sqrt{3} = 1.732$$

$$\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2(\mu'_1)^3 = 10 - 3(4)(1) + 2(1)^3 = 0.$$

Karl Pearson's coefficient of skewness is calculated as:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \text{ or } \gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\sigma^3}$$

By putting the values of μ'_1, μ_2 and σ in above formula, we get:

$$\gamma_1 = 0\ (\beta_1 = 0)$$

From the above expression, we observe that the given distribution is symmetrical, and thus for the given distribution:

Mean = Median = Mode

Example 18: Determine the first four moments about mean from the following data:

X: 5 10 15 20 25

Solution:

Computation of First Four Moments about Mean

X	d = X – 15	d^2	d^3	d^4
5	–10	100	–1000	10000
10	–5	25	–125	625
15	0	0	0	0
20	5	25	125	625
25	10	100	1000	10000
ΣX = 75	**Σd = 0**	**$\Sigma d^2 = 250$**	**$\Sigma d^3 = 0$**	**$\Sigma d^4 = 21250$**

The form of any distribution is given by kurtosis. If any distribution is plotted on the graph then it represents either a normal curve or a curve more flat than the normal curve, or a curve more peaked than the normal curve. Kurtosis is a measure which tells how much 'peaked' or 'flat' the data is in relation to a normal distribution. In other words, kurtosis depicts a distinct peak near the mean, shows a rapid decline and ends up with a heavy tail, ending up with thin tail.

According to Simpson and Kafa, "The degree of Kurtosis of a distribution is measured relative to the peakedness of a normal curve".

4.4.2. Difference between Skewness and Kurtosis

Table 1 shows the difference between Skewness and Kurtosis:

Table 1: Difference between Skewness and Kurtosis

Skewness	Kurtosis
A measure of the asymmetry of a distribution.	A measure of the extent to which observations cluster around a central point.
A distribution with a significant positive skewness has a long right tail.	Positive kurtosis indicates that the observations cluster more and have longer tails.
A distribution with a significant negative skewness has a long left tail.	Negative kurtosis indicates the observations cluster less and have shorter tails.
A symmetrical distribution has a skewness of zero.	A Gaussian distribution has a kurtosis of 0.

4.4.3. Types of Curves

The terms **'Mesokurtic', 'Leptokurtic'** and **'Platykurtic'** were introduced by **Karl Pearson** in 1905. A peaked curve is termed as 'Leptokurtic' and a flat topped curve is termed as 'Platykurtic'.

1) **Leptokurtic:** A frequency curve which is more peaked than the normal curve is known as **Leptokurtic**. In comparison to the normal distribution, the Leptokurtic distribution has higher peak around the mean which leads to thick tails on both sides. These peaks result from the data which is highly concentrated around the mean because the variations are lower within the observations.
2) **Platykurtic:** A frequency curve which has flatter peak than the normal curve is known as **Platykurtic**. In comparison to the normal distribution, the Platykurtic distribution have flatter peak around the mean which leads to thin tails on both sides. These peaks result from the data which is less concentrated around the mean because the variations are large within the observations.
3) **Mesokurtic:** A frequency curve with a normal curve is known as **Mesokurtic**. The kurtosis coefficient of a normal distribution is 3. If the kurtosis coefficient is high then the distribution will be more peaked around the mean. When the kurtosis coefficient is greater than 3 then the distribution is called leptokurtic and if it is less than 3 then it is called platykurtic.

In the **figure 4.3, curve No. 1** is normal or mesokurtic, **curve No. 2** is more peaked than the normal curve, hence is leptokurtic and **curve No. 3** is more flat than the normal curve, and is platykurtic.

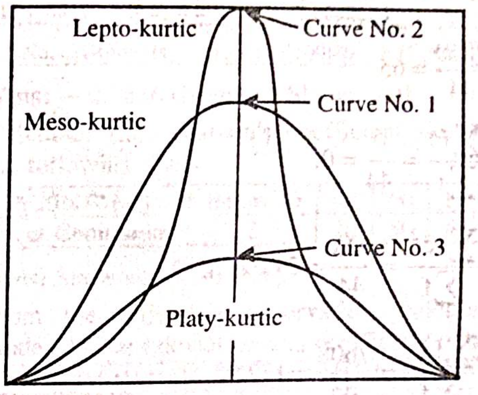


Figure 4.3: Kurtosis

4.4.4. Measure of Kurtosis

Kurtosis is measured by the 'Moment ratio', which is based on the second and fourth moments. Karl Pearson has given the following formulae for measure of kurtosis:

Measure of Kurtosis or β_2 (Beta-two) or α_4 (alpha-four)

$$= \frac{\mu_4}{\mu_2^2} = \frac{\text{Fourth moment}}{\text{Second moment}^2} \quad \text{or} \quad \frac{\mu}{\sigma^4}$$

In a normal distribution β_2 will be equal to 3. If it is greater than 3, the curve is more peaked than normal, if less than 3, the curve is flatter at the top than normal.

Thus:

If $\beta_2 = 3$, the curve is normal or Mesokurtic,

If $\beta_2 > 3$, the curve is Leptokurtic or more peaked,

If $\beta_2 < 3$, the curve is Platykurtic or flat topped.

The measure of kurtosis is also represented by γ_2 (Greek letter gamma two) Therefore if:

γ_2 or $\beta_2 - 3 = 0$, the curve is Mesokurtic.

γ_2 is positive, the curve is Leptokurtic,

γ_2 is negative, the curve is Platykurtic.

Example 25: Calculate first four moments from the following data and find out β_1 and β_2

Weight (in Kg.)	60	61	62	63	65	70	75	80
No. of Workers	3	4	7	9	12	6	1	2