CHAPTER – 1 – WHAT IS ECONOMICS

Meaning of Economics

Economics revolves mainly around what **Alfred Marshall** called "the study of man in the ordinary business of life"

Economics is a social science of human behaviour which aims at allocation of scarce resources in such a way that consumers can maximise their satisfaction, producers can maximise their profits and society can maximise social welfare.

In simple words, Economics is about making choice in the presence of scarcity .

According to **Robbins,**

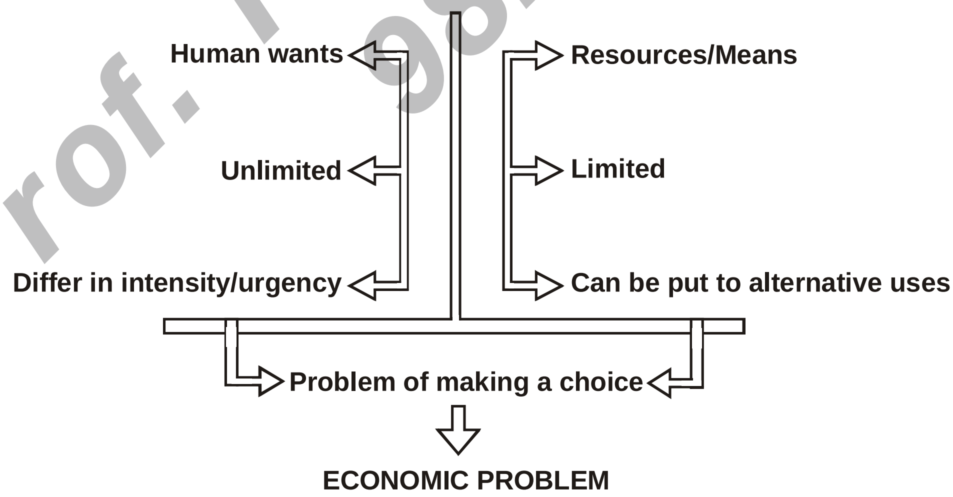
"Economics is the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses."

SCARCITY AND ECONOMIC PROBLEM

Meaning of Scarcity

In daily life, scarcity means acute shortage of a certain commodity but in economics, it means limitation of supply of a commodity in relation to its demand.

Meaning of Economic Problem



For example,

1. You have a ten rupee note (means/resources).
2. You want to buy a copy, a pen, a chocolate and a pencil (wants).
3. Your wants are unlimited in comparison to your means or your means are limited in relation to demand for them.
4. As your wants differ in urgency, also your ten rupee note can buy

anything you want to, you face a problem, "a problem of making a choice" as to how should you spend 10 rupee note in such a way that you can buy maximum. "The problem of making a choice is called economic problem."

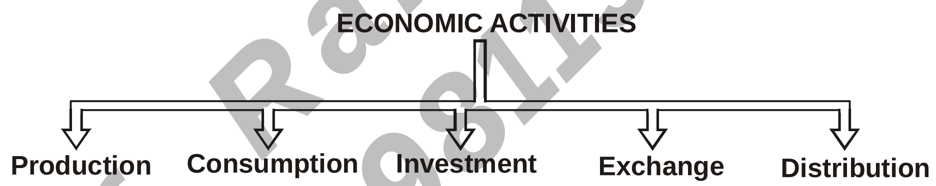
According to **Prof. Erich Roll,**
"Economic problem is essentially a problem arising from the necessity of choice ; choice of the manner in which limited resources with alternative uses are disposed off. It is a problem of husbandry of resources."

Thus, it is basically a problem of scarcity of resources on one hand and also that resources can be put to alternative uses on the other hand.

Economic Activities

An economic activity refers to the activity which is relate to the use of scarce resources to satisfy maximum of unlimited human wants. These are undertaken for a monetary gain? These are what economists mean by ordinary business of life.

Economic activities are classified as under :



Production : Production is that economic activity which is related to increasing the utility of goods and services. e.g., converting cloth into shirt is an activity of production. Here utility is created through conversion of cloth into shirt with the help of labour, sewing machine, etc.

Consumption : Consumption is that economic activity which is related to the use of goods and services for the direct satisfaction of individual and collective wants. There would have been no production if there was no consumption.
For example, drinking water or milk, services of teacher or doctor etc. are consumption activities.

Investment : Investment is that economic activity which is related to production of capital goods for further production of goods and services.
For example, the production of car to be used as taxi.

Exchange : Exchange is that economic activity which is related to sale and purchase of commodities. This buying and selling is mostly done in terms of price.
Therefore, it is also called "Product Pricing".
For example, buying a notebook by paying Rs. 10 for it.

Distribution: Distribution is that economic activity which relates to determination of price of factors of production (land, labour, capital and enterprise). This is known as the 'Factor Pricing'.
Distribution is the study to know how national income is distributed through salaries, wages, profits and interest.

Difference between economic activities and non-economic activities
Economic activities are different from non-economic activities as economic activities have economic aspect and are undertaken for monetary gain . *For example*, Production, Investment, etc.
Whereas **non-economic activities** do not have economic asp and are not concerned with monetary gain. *For example*, religious activities, charitable activities, etc.

DEFINITION OF ECONOMICS

1. Wealth Definition

The main feature or subject matter of this definition is wealth.
Adam Smith the father of modern economics, in his book 'An Enquiry into the Nature and Causes of Wealth of Nations' defined that : *"Economics is an enquiry into the factors that determine the wealth of a country and its growth".*

According to **J.B. Say,**
"Economics is the science which deals with wealth".
According to J.S. Mill,
"Economics is the practical science of the production and distribution of wealth".

Criticism : This definition is not a complete definition. It gives importance to wealth rather than social welfare and scarcity of resources.

The wealth definition of economics was discarded towards the end of the 19th century.

2. Welfare Definition

This definition gives importance to human welfare in economic activities undertaken.

Alfred Marshall in his book 'Principles of Economics', published in 1890, has defined

Economics in these terms :

"Economics is a study of mankind in the ordinary business of life, it examines that part of individual and social action which is most intimately connected with attainment and use of material requisites of well-being."

Criticism : This definition ignores the measurement of Welfare in monetary terms. It also ignores scarcity of resources and economic growth.

3. Scarcity Definition

This definition considers economics as science which studies human behaviour as relationship between unlimited wants and scarce means.

Lord Robbins in his book 'An Essay on Nature and Significance of Economics Science', published in 1932, has defined Economics in these words :

"Economics is a science that studies human behaviour as a relationship between ends and scarce means which have alternative uses".

According to **Robbins,**

" Economics is a science of choice. It deals with how the resources of society should be allocated to the satisfaction of different wants".

Criticism : This definition does not relate to welfare and growth aspects.

4. Growth Definition

"This definition considers economic growth and social welfare. Paul A. Samuelson defines :

"Economics is the study of how man and society choose with or without the use of money to employ scarce productive resources that could have alternative uses, to produce various commodities overtime and distribute them for consumption now, or in the future among various persons and groups in society."

Accordingly, economics is concerned with the efficient allocation and use of scarce means as a result of which economic growth is enhanced and social welfare is increased.

Q. Who is the father of economics ?

Ans. Adam Smith is regarded as the father of economics.

Q. What is the basis of difference between economic activities and non-economic activities?

Ans. Motive of earning of wealth is the basis of difference between economic and non-economic activities.

Q. What is the nature of economics according to Adam Smith?

Ans. According to Adam Smith, economics is a science of wealth.

Q. Explain the difference between economic goods and free goods with the help of examples.

Ans. Goods which are in such plentiful supply that they do not command any market price are called free goods, for example, air, sunshine, rain, etc. Goods which are scarce and command price, are called economic goods; for example, wheat, cloth, car, furniture, tea, etc.

Q. What is the difference between economic and non-economic activities? Explain it with the help of examples.

Ans. Activities which have earning of wealth as their prime objective are economic activities; for example, activities of farmers, weavers, labourers, shopkeepers, teachers, etc.

Activities which do not have earning of wealth as the*, main objective are non-economic activities; for example, playing football for health or recreation, singing for self-entertainment; nursing of baby by the mother, etc.

Q. Can the same activity be economic or non-economic? Explain with example.

Ans. The same activity can be economic or non-economic, depending on the motive for the activity. If a player plays football in order to get remuneration, it is economic activity and if the same player plays it for the sake of his health, it becomes non-economic activity.

Q. What are the two branches of economics ?

Ans. The two main branches of economics are : Microeconomics and Macroeconomics.

Q. What is Microeconomics ?

Ans. Microeconomics is the study of the behaviour of individual economic units such as price determination of a commodity, behaviour of a consumer, a producer or a firm.

Q. What is Macroeconomics ?

Ans. Macroeconomics is the study of the aggregates and averages relating to the whole economy, e.g., total consumption, total employment, national income, general price level, etc.

Q. Is economics a positive science or normative science or both?

Ans. As positive science answers what is? What was? Where we study human decisions as facts which can be verified with actual data. A normative science which refers to what ought to be ? What ought to have happen ?

The normative statements, in fact, are the opinions of different persons relating to Tightness or wrongness of particular thing or policy. Thus, economics is both positive and normative science.

Q. Is economics a science or an art ?

Ans. Economics is a social science that studies economic problems and policies in a scientific manner. Economics is an art as it gives practical guidance in solution to various economic problems. Economics, thus, is as science as well as an art.

CHAPTER – 2 – WHAT IS STATISTICS

The systmatic treatment of quantitative expression is known as 'statistics'. Not all quantitative expressions are statistics we will see that certain conditions must be fulfilled for a quantitative statement to be called statistics.

Statistics can be defined in two ways :

a. In plural sense

b. In singular sense

STATISTICS DEFINED IN PLURAL SENSE (as statistical data)

According to Horace Secrist, "By statistics we mean aggregates of facts affected to a marked extent by multiplicity of cause numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other".

1. **Statistics are aggregates of facts :** A single observation is not statistics, it is a group of observation. Eg., A single age of 30 years is not statistics but a series relating to the ages of a group of persons is statistics.
2. **Statistics are affected to a marked extent by multiplicity of causes :** Statistics are generally not isolated facts they are dependant on, or influenced by an number of phenomena.
3. **Statistics are numerically expressed :** Qualitative statements are not statistics unless they are supported by numbers.
4. **Statistics are enumerated or estimated according to reasonable standard of accuracy :** Enumeration means a precise and accurate numerical statement. But sometimes, where the area of statistical enquiry is large, accurate enumeration may not be possible. In such cases, experts make estimations on the basis of whatever data is available. The degree of accuracy of estimates depends on the nature of enquiry.
5. **Statistics are collected in a systematic manner :** Statistics collected without any order and system are unreliable and inaccurate. They must be collected in a systematic manner.

6. **Statistics are collected for a pre-determined purpose :** Unless statistics are collected for a specific purpose they would be more or less useless.
7. **Statistics are placed in relation to each other :** Statistical data are often required for comparisons. Therefore, they should be comparable periodwise, regionwise, commoditywise etc.

   When the above characteristics are not present a numerical data cannot be called statistics. Thus, "all statistics are numberical statements of facts but all numberical statement of fact are not statistics.

Statistics defiend in singular sense (as a statistical method)

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

1. **Collection of data :** Data should be gathered with maximum care by the investigator himself or obtained from reliable published or unpublished sources.
2. **Organisation of data :** Figures that are collected by an investigator need to be organised by editing, classifying and tabulating.
3. **Presentation of data :** Data collected and organised are presented in some systematic manner to make statistical analysis easier. The organised data can be presented with the help of tables, graphs, diagrams etc.
4. **Analysis of data :** The next stage is the analysis of the presented data. There are large number of methods used for analysing the data such as averages, dispersion, correlation etc.
5. **Interpretation of data :** Interpretation of data implies the drawing of conclusions on the basis of the data analysed in the earlier stage.

FUNCTIONS OF STATISTICS

1. **Statistics simplifies complex data :** With the help of statistical methods a mass of data can be presented in such a

manner that they become easy to understand.

2. **Statistics presents the facts in a definite form :** This definiteness is achieved by stating conclusions in a numerical or quantitative form.
3. **Statistic provides a technique of comparison**. Comparison is an important function of statistics i.e., Period wise, country wise etc..
4. **Statistics studies relationship :** Correlation analysis is used to discover functional relationship between different phenomena, for example, relationship between supply and demand, relationship between sugarcane prices and sugar, relationship between advertisement and sale.
5. **Statistics helps in formulating policies :** Many policies such as that of import, export, wages, production, etc., are formed on the basis of statistics.
6. **Statistics helps in forecasting :** Statistics also helps to predict the future behaviour of phenomena such as market situation for the future is predicted on the basis of available statistics of past and present.

IMPORTANCE OF STATISTICS

1. **Statistics in Economics**

   A number of economic problems can easily be understood by the use of Statistics.

   It helps in formulation of economic policies, e.g., basic economic activities like production, consumption etc. use Statistics.

The importance of Statistics in various parts of economics has been discussed as follows :

(a) ***Statistics in consumption.*** To obtain the knowledge of how different groups of people spend their income form Statistics relating to consumption. The data of consumption are useful and helpful in planning their budget and improve their standard of living.

(b) ***Statistics in production.*** The Statistics of production are very useful and helpful for adjustment of demand and supply and

determining quantity of production of the commodity.

(c) ***Statistics in distribution.*** Statistical methods are used in solving the problem of distribution of national income among various factors of production i.e., land, labour, capital and entrepreneur.

2. Statistics in Economic Planning

Economic planning is done to achieve certain targets for growth of the economy using scarce resources of the nation. Statistics helps in evaluating various stages of economic planning through statistical methods.

According to **Tippett,**

"Planning is the order of the day, and without Statistics planning is inconceivable."

Statistics helps in comparing the growth rate. It helps to formulate plans to achieve predetermined objectives. It measures the success and failure of plans and accordingly guides to apply corrective measures.

3. Statistics in Business

Statistical tools play very important role in major business activities.

The producer depends upon market research to estimate market demand and the market research is based on Statistics. The trader depends heavily on methods of statistical analysis to study market.

Statistical tools are very important for the detailed analysis of money transactions in the business.

4. Statistics in Administration

Formulation of a policy involves Statistics. The state gathers the facts relating to population, literacy, employment, poverty, per capita income etc., with the help of statistical methods and principles.

It helps the state to achieve targets with the help of optimum utilisation of scarce resources.

LIMITATIONS OF STATISTICS

1. **It does not study the qualitative aspect of a problem :** The most important condition of statical study is that the subject of investigation and inquiry should be capable of being uantiatively measured. Qualitiative phenomena e.g., honesty, intelligence, poverty, etc., cannot be studied in statistics unless these attributes are expressed in terms of numberals.

2. **It does not study individuals :** Statistics is the study of mass data and deals with aggregates of facts which are ultimately reduced to a single value for analysis. Individual values of the observation has no specific importance. For example, the income of a family is, say Rs. 1000, does not convey statistical meaning while the average income of 100 families say Rs. 400, is a statistical statement.

3. **Statistical laws are true only on an average :** Laws of statistics are not unversally applicable like the laws of chemistry, physics and mathematics. They are true on an average because the results are affected by a large number of causes. The ultimate results obtained by statistical analysis are true under certain circumstances only.

4. **Statistics can be misused :** Statistics is liable to be misused. The results obtained can be manipulated according to one's own interest and such manipulated results can mislead the community.

5. **Statistics simply is one of the methods of studying a phenomenon :** Statistical calculations are simple expression which should be supplemented by other methods for a complete compreshension of the results. Thus statistics is only a means and not the end.

6. **Statistical results lack mathematical accuracy :** The result drawn from statistical analysis are normally in approximates.

As the statistical analysis is based observations of mass data, number of inaccuracies may be present and it is difficult to rectify them. Therefore, these results are estimates rather than exact statements. Statistical studies are a failure in the fields where one humberd percent accuracy is desired.

Origin and Growth of Statistics

The term 'STATISTICS' has been derived from the Latin word 'STATUS', which means political state. Germans have spelled it as 'STATISTIK'. The term 'Statistics' was first used by German scientist **Gottfried Achenwall** in 1749. He is known as **the Father of Statistics**.

1. 'Oil prices are rising globally'. How will Statistics help us to know how much oil should be imported ?

Ans. In the present time of rising global oil prices, it might be necessary to decide how much oil India should import in 2013. The decision to import would depend on the expected domestic production of oil and the likely demand for oil in 2013. Without the use of Statistics, it cannot be determined what the expected domestic production of oil and the likely demand for oil would be. Thus, the decision to import oil cannot be made unless we know the actual requirement of oil. This vital information that helps to make the decision to import oil can only be obtained statistically.

2. 'Statistics can prove anything'. Explain.

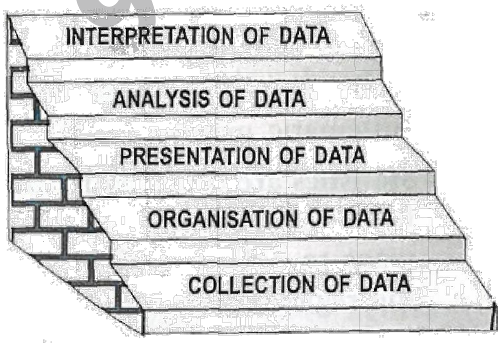
Ans. Although Statistics has become an integral part of business and economic analysis, layman has a distrust of economics. This is so because it is possible to misuse Statistics by manipulating data. It is similar to the state as a lawyer does interpret law to prove his own point. That is why some people say that with Statistics anything can be proved. Some political parties or governments misuse statistical tools for their selfish motives. Misuse of Statistics generally takes place at the time of selecting samples and while making comparisons and interpreting analysis of data. Though Statistics can be misused, it is for the economist to find out the truth with the help of his expertise in Statistics.

3. Statistics prove that there is oil shortage in India. What values do you suggest to people to meet the growing demand for oil ?

Ans.
- There should be optimum use of oil, as it is scarce.
- We should make people aware of shortage of oil and encourage them to use public transport system, car pools etc.

STAGES OF STATISTICAL STUDY

According to the figure, interpretation of data the last stage in order to draw some conclusion. One has to go through the four stages to arrive at the final stage; they are collection, organisation, presentation and analysis. First stage collection of data refers to gathering some statistical facts by different methods. The second stage is to organise the data so that collected information is easily intelligible. This is the arrangement of data in a systematic order after editing. Third stage of statistical study is presentation of data.



After collection and organisation the data are to be reproduced by various methods of presentation, namely, tables, graphs, diagrams, etc. so that different characteristics of data can easily be understood on the basis of their quality and uniformity. Fourth stage of statistical study is the analysis of data. Calculation of a value by different methods and tools for various purposes is made to arrive at the last stage of study, viz., interpretation of data.

In brief, statistics is a method of taking decisions on the basis of numerical data properly collected, organised, presented, analysed and interpreted.

Role of Computer in Statistics

Today, Computer technology has revolutionised our modern living and simplified a lot of our manual efforts in day-to-day life like reservation of railway and airline tickets, payment of electricity, water and mobile bills by using the Internet. Computer is a boon to the modern civilisation. The development of computer softwares, especially programrned for statistical studies has been very Helpful in collection, organisation, presentation, analysis and interpretation of data.

Q. Why does an educated man depend on statistics ?

Ans. Modern educated man wants to put his views in a focussed and precise manner and provide full information on the subject he is discussing or writing about. Hence, he uses statistical statements.

Q. Give examples of a statistical statement. Mention two of its features.

Ans. A statement like, "on 18th July 08, because of heavy rain, out of 52 students of -. Class XI B, only 8 boys out of the 32 enrolled were present and the rest of the class including the girls were absent", is a statistical statement.

Q. Explain three features of statistical data. Three distinct features of statistical data are :

Ans. (a) Aggregate of facts. Statistical data are the aggregate figures of a fact over a period of time like per capita income of India (not the income of a single man) since 1950-51.

(b) Numerically expressed. Statistical data are quantitatively expressed not qualitatively like high or low.

(c) Collected or estimated with some degree of accuracy. A theoretically sound method should be used to collect statistical data.

Q. How is statistics useful to a policy maker ?

Ans. The policy makers use statistical tools for policy appraisal. If a new policy is implemented or modifications are done in an existing policy, its effect on the beneficiaries is judged by statistical methods.

Q. What role does statistics play in economic modelling ?

Ans. Economic modelling is a technique to explain economic activities like consumption, production, etc. The association between demand and its determinants or the amount produced and the inputs is studied by economic modelling. Statistics help in estimating those relationships quantitatively.

Q. How does statistics help the government in reducing inequality in income distribution in a country ?

Ans. Statistical tools measure the extent of inequality present in an economy. Based on those estimates the government adopts different steps to reduce the inequality in the country.

— *Notes* —

CHAPTER – 3 – COLLECTION OF PRIMARY AND SECONDARY DATA

Q. What is Statistical Inquiry ?

Inquiry means a search for truth, knowledge or information.

Thus, statistical enquiry means statistical investigation; one who conducts this type of inquiry is called an investigator. The investigators needs the help of certain persons to collect information, they are known as enumeratores, and repondents are those from whom the statistical information is collected.

SOURCES OF DATA

There are different sources of collection of data. Following are the various sources of collection of data.

SOURCES OF DATA

Internal Sources External Sources

Primary Data Secondary Data

Internal Sources

Q. Explain Internal sources of data with E.g..

Ans. 1. Different organisations and Government department generate the data as their regular function which is the internal information.

2. Internal data may be available in the organisation about sales, production, interest, profit, dividends etc.

3. Such data are complied and used for future planning.

Q. External sources of data with E.g..

Ans.
- Information collected from outside (other) organisations and institutions is called external data.
- External data can be obtained from primary source or secondary source.

PRIMARY AND SECONDARY DATA

Primary data is original and first hand information and secondary data is collected through other sources. Primary data is first hand information for a particular statistical enquiry while the same data is second hand information for an another enquiry.

METHODS OF COLLECTING PRIMARY DATA

1. Direct personal interview

According to this method, data are collected by the investigator personally from persons who are the subject to enquiry. He interviews personally every one who is in a position of supply information he requires. We can use this method of collection of data when area of enquiry is limited or when a maximum degree of accuracy is needed. The investigator must be skilled, tactful, accurate, pleasing and should not be biased.

Merits

1. Original data are collected by this method.
2. There is uniformity in collection of data.
3. The required information can be properly obtained.
4. There is flexibility in the enquiry as the investigator is personally present.
5. Information can be obtained easily from the informants due to a personal interview.
6. Since the enquiry is intensive and in person the result obtained are normally reliable and accurate.
7. Informants relations to questions can be properly studied.
8. Investigators can use the language of communication according to the educational standard of the information.

Limitations

1. This method can be used if the field of enquiry is small. It cannot be used when field of enquiry is wide.
2. It is costly method and consume more time.
3. Personal bias can give wrong results.
4. Investigators need to be trained and supervised for the job, otherwise results obtained may not be reliable.
5. This method is lengthy and complex.

2. Indirect personal interview

Direct personal enquiry cannot be used in the case of the refusal and reluctance of the persons who are to be interviewed. Then an indirect personal enquiry can be conducted to get necessary information from an indirect source. Persons who have the knowledge of relevant material or event are interviewed and asked questions for collecting the data.

In indirect personal enquiry information is gathered by interviewing friends, neighbors, employers, relatives etc.

In this method a precaution is necessary in selecting the informants. An informant should be a person :

a. Who is not biased or prejudiced
b. Must know the facts of problem
c. Must be capable of answering correctly for giving true informantion,
d. Is not motivated to give colour to the fact.

To get success of collecting of data by this method, it is necessary that the evidence of one person alone is not to be relied upon; the opinions of various persons should be obtained to find out the real and true picture of the situation.

Merits

1. This method covers a wide area of investigation. whenever the informant in direct investigation is relucant to give information, or cannot be contacted, this method is a good alternative.
2. As the information is obtained from the third party, it is more or less free from biased or prejudiced approach of the investigator and the informant.
3. It saves labour, time and money.
4. As the information covers a wide range, different aspects of problems can properly be studied.

Limitations

1. As the information is obtained from the third-party and not by the person directly concerned, there exists as possibility of not getting true information.
2. Various evidences obtained may be somethimes manipulated according to the interest of the person answering the question or supplying the information.

Difference between Direct Personal Investigation and Indirect Oral Investigation

1.**Coverage**	This method is suitable for smaller areas.	This method is suitable for Wider areas.
2. **Originality**	The data collected are original in nature.	This method lacks originality as witnesses provide information.
3. **Cost**	It is more costlier in terms of time, money and efforts.	It is less costlier in terms of time, money and efforts.
4. **Reliability**	This method is more reliable as information is collected directly from the informants.	This method is less reliable as information is collected from the witnesses.

3. Information from correspondents

In this method, local agents or correspondent are appointed in the different parts of the investigation area. These agents regularly supply the information to the central office or investigator. They collect the information according to their own judgements and own methods. Radio and newspaper agencies generally obtain information by this method. It is also adopted by government departments. It is suitable when the information is to be obtained from a wide area and where a high degree of accuracy is not required.

Merits

1. This method is comparatively cheap.
2. It give results easily and promptly.
3. It can cover a wide area under investigation.

Limitations

1. In this method original data is not obtained.
2. It gives approximate and rough reults.
3. As the correspondent use his own judgement, his personal bias may affect the accuracy of the information sent.
4. Different attitude of different correspondents and agents may increase errors.

4. Mailed questionnaries

A list of questions relating to inquiry, which is called schedule or questionnaire, is prepared. This list of questions provides a space for each answer. Schedules are sent to informants by post, with a request to answer and return it within a specific time. Such schedules gererally have preparid postage stamp affixed to them. If necessary, an assureance is given that the answer will be kept confidential. The success of this method depeds on the cooperation that the informant is prepared to extend and the manner in which the questionnaire is drafter.

Merits

1. Large field can be studied by this method.
2. This is not an expensive method. It is cheap as mailing cost is less than the cost of personal visits.
3. We can obtain original data by this method.
4. It is free from the bias of the investigator as the information is given by the informants themselves.

Limitations

1. It is difficult to presume the cooperation on the part of the informatnts.

They may fail to send back the schedules or may misinterpret or may not understand some questions.

2. Schedules sent back by the informants may be incomplete or inaccurate and it may be difficult to verufy the accuracy.
3. There may be delays in getting replies to the questionnaires.
4. This method can be used only when the informants are educated or literate, so that they return the questionnaries duly read, understood, answered.
5. There is a possibility of getting wrong results due to partial responses, and those who do answer may not include certain type of persons from whom the specific information is required.

Suitability

a. When it is compulsory by law to fill the questionnarie, e.g., government agencies compel bank and companies etc., to supply information regularly to the government in a prescribed from.
b. This method can be successful when the informants are educated.

Following are some suggestions for making this method more effective and successful.

i. Questions should be simple and easy so that the informants may not find it a burder to answer them.
ii. Informants should not be required to spend for posting the questionnairs back therefore, prepaid postage stamp should be affixed.
iii. This method should be used in a large sample or wide universe.
iv. This method is preferred in such enquiries where it is compulsory by law to fill the schedule. Thus, there is little risk of non-response.
v. The language of the schedule should be polite and should not hurt the sentimetnts of the informants.

5. Questionnaire to be filled by enumerators

Mailed questionnaire method poses a number of difficlties in collection of data. Generally, these filled questionnaries received are incomplete, inadequate and unrepresentative.

The second alternative approach is to send trained investigators or investigator to informants with standardised questionnaire which are to be filled in by the nvestigators. The investifgator helps the informants in recording their answers. The investigators should be honest, tactful and painstaking. This is the most common method used by research organisations. They train with different persons tactfully, to get proper answers to the questions put to them. The statistical information collected under this method is highly reliable.

Merits

1. It can cover a wide area.
2. The results are not affected by personal bias.
3. True and reliable answer to dificult questions can be obtained through establishment of personal contact between the enumerator and informant.
4. As the infromation is collected by trained and experienced enumerators, it is reasonably accurate and reliable.
5. This method can be adopted in those cases also where the informants are illiterate.
6. Personal presence of investigator assured complete response and respondents can be persuaded to give the answers to the questionnaire.

Limitations

1. It is an expensive method as compared to other methods of primary collection of data, as the enumerators are required to be paid.
2. This method is time consuming since the enumerator is required to visit people spread out over a wide area.
3. This method need the supervision of investigators and enumerators.
4. Enmuerators need to be trained. Without good interview and proper training, most of the collected information is vague and may lead to wrong conclusions.
5. It needs a crmy of investigation to cover the wide area of universe and therefore it can be used by bigger organisations.

DRAFTING THE QUESTIONNAIRE

Following are the basic principles of drafting questionnaire :

1. Covering Letters

The peron conducting the survey must introduce himself and make the aims of the objective of the enquiry clear to the informant. A personal letter can be enclosed indicating the purposes and aims of enquiry. The informant should be taken into confidence. A self - addressed and stamped envelope should be enclosed for the convenience of the informant to return the questionnaire.

2. Number of Questions

Minimum number of questions based on the objectives and scope of enquiry should be asked.
Therefore, normally fifteen to twenty - five questions should be asked. Lenghty questions should preferably be divided into parts. Irrelevant questions should be eliminated.

3. Personal questions should be avoided

The informant may not desire to answer such questions which may disclose his confidential, private or personal information. Questions affecting the sentiments of informants should not be asked.

4. The questions should be simple and clear

The language of the questions should be easy to understand.

5. The questions should be arranged logically

It helps in classification and tabulation of data. It is not logical to ask a man his income before asking him whether he is employed or not. There should be a proper sequence of the questions.

6. Instructions and Informations

Definite instructions for filling in the questionnaire should be given.

7. The questions may be divided and sub

Divided under different heads and subheads and should be properly numbered for the convience of the informant and the investgator.

8. Multiple Choice Questions

Questions should be capable of objective answers. For this the informant should be able to give the answers simply by using a tick - mark in the blank space.

9. Simple Alternative Question (Yes/No)

As far as possible the questions should be framed in such a way that they are answered in 'Yes' or 'No' or 'Right' or 'Wrong'.

10. Open Question

It makes the informant free to give any reply he chooses. Such questions should be minimum in number in the questionnaire.

11. The questions should be directly related to the point under enquiry for which the data is being collected.

12. Avoidance of leading questions

As far as possible leading questions should be avoided. Why do you like 'Broke Bond Tea' ? Instead of such simle question, two questions can be framed for enquiry, namely :

a. Which brand of tea do you take ?

b. Why do you prefer it ?

13. Attractive layout

The questionnaire should be made to look as attractive as possible. Keeping in view the possible answer the questions of schedule, sufficient space should be provided.

14. Avoidance of questions of calculations

As far as possible no question should be asked which require mathematical calculations like percentage, ratio etc. it gives strain to the informant and he may avoid sending the questionnaire back.

15. Cross Check

Some questions should provide the mean of checking inaccuracies in the answers.

For example, question on age and date of birth is a cross check. It helps to decide whether the informant is answering the questions correctly and consistently.

16. Questions on familiar topics

Questions which require strain should be avoided. Too much reliance on memories of distant past may elict wrong answers. Informants should be able to answer from their own memory and knowledge.

17. Pre - testing of questionnaire

Before taking the enquiry on a large scale the questionnaire drafted should be pre - tested with a small number of a group of persons.

COLLECTION OF SECONDARY DATA

Secondary data are those which are collected by some other agency and are used for further studies. It saves cost and time which is involved in collection of primary data. Secondary data may be either (a) published or (b) unpublished.

Published Sources

1. Government Publications

Different ministries and departments of Central and State Governments publish regularly current information along with statistical data on a number of subjects.

This information is quite reliable for related studies. The example of such publications are : Annual Survey of Industries, LAbour Gazette, Agriculture Statistics of India, Indian Trade Journal, etc.

2. Pulications of International organisations

We can obtain valuable international statistics from official publication of different international organisations, like, (U.N.O.), (I.L.O), (I.M.F)

3. Semi Official Publication

Official publications. Local bodies such as Muncipal Corporations, District Boards etc. publish periodical reports which give factual information about health, sanitation, births, deaths etc.

4. Reports of Committees and Commissions

Various Committees and Commissions are appointed by the Central and State Governments for some special study and recommendations. The report of such committes and commissions contain valuable data.

5. Private Publications

a. Journals and Newspapers : Journals like Eastern Economists, Journals of Industry and Trade, Monthly Statistics of Trade ; and newspapers, like Financial Express, Economic Times, collect and regularly publish the data on different fields of economics, commerce and trade.

b. Research Institutions : There are a number of institutions doing research on allied subjects.

c. Proffesional Trade Bodies : Institute of Chartered Accountants, Sugar Mills Association, Bombay Mill Owners Association, Stock Exchanges, Bank and Cooperative Socieities, Trade Unions, etc. publish statistical data.

d. Annual report of joint stock companies are also useful for obtaining statistical information. These are published by comapnies every year.

e. Articles, market review and reports also provide valuable data for research study.

Unpublished Data

Research institutions, trade associations, universities, labour bureaus, Research workers and scholars do collect data but they normally do not publish it. A part from the above sources we can get the information from records and files of Government and Private Offices.

Limitations of Secondary Data

1. They may not have been collected by proper procedure.
2. They may not be suitable for a required purpose. The information which was collected on a particular base may not be suitable and relevant to an enquiry.
3. They may have been influenced by the biased Investigation or personal prejudices.
4. They may be out of date and not suitable to the present period.
5. They may not satisfy a reasonable standard of accuracy.
6. They may not cover the full lertour of Investgation.

Precautions in the Use of Secondary Data

The investigator should consider the following points before using the secondary data :

1. Are the data reliable ?
2. Are the data suitable for the purpose of investigation ?
3. Are the data adequate ?
4. Are the data collected from proper method ?
5. From which source were the data collected ?
6. Who has collected the data ?

7. Are the data biased ?

Thus, the secondary data should not be used at its face value. It is risky to use such statistics collected by others unless they have been properly scrutinised and found reliable, suitable and adequate.

DIFFERENCES

	Quantitative Data	Qualitative Data
1.	They are expressed in numerical terms. Explain population of India has increased from 84.7 crores in 1991 to 102.7 crores in 2001.	They describe attributes to a single person or group of persons even though they cannot be expressed in numerical terms. They are relative measures they can only be ranked.
2.	They are exact.	They may not be exact.
3.	Example : Sita is taller than geeta.	Example : Sita's height is 5.6´ where as geeta is just 5.0´.

Basis	Primary Data	Secondary Data
1. Defination	Primary data are those data which are collected for the first time.	Secondary data to those data which have already be collected by some other persons.
2. Originality	Primary data are original in character. They are first hand information.	Secondary data are not original.
3. Nature of data	Primary data are in the form of raw material to which statistical methods are appied.	Secondary data are in the form of finished product as they have already been statistically applied.
4. Object	Primary data have been collected for a definite purpose.	Secondary data are collected from published or unpublished sources.
5. Time and Money	A lot of time and money is required for collection of primary data.	Since secondary data are collected from other sources. So both time and money are saved.
6. Suitability	Primary data are strictly in accordan as with the investigation.	Secondary data are made according to the object of the investigation.

	Basis	Internal Data	External Data
1.	Meaning	Refers to those data which have been obtained from the internal recordes of an organisation.	External data refer to those which have been obtained from other sources.
2.	Example	Reliance Industry Ltd. publishes its annual report which would constitute an internal source of data for the company.	When Reliance Industries Ltd. obtaines data from sources other than its own records, this would be external source of data.
3.	Cost	These are less expensive	These are more expensive.
4.	Reliability	These data are more reliable.	These data are comparatively less reliable.

	Exclusive Series	Inclusive Series
1.	In it, the upper limit of one class is the lower limit of the next class interval.	In case of inclusive services, there exists difference between the upper limit of one class and lower limit of next class interval.
2.	The upper limit of the class is not included.	The upper limit of the class is included.
3.	There is no need to convert it into exclusive series.	It generally becomes necessary to convert it into exclusive series.
4.	It is useful in all cases whether the value is in complete number or in fractions.	It is useful only when the value is a complete number.

Pilot Survey or Pre-Test : from practical point of view it is found useful to conduct a pre-test a guiding survey known as Pilot Survey, on a small scale before starting the main survey. The information supplied by the pilot survey helps in :

(i) Estimating the cost of the main sample survey and also the time needed for the availability of the results.

(ii) Improving the organisation of the Held Work by removing the defects or faults observed in the pilot survey.

(iii) Formulating effective methods of asking questions and also in the improvement of the questionnaire.

(iv) Training of field staff.

(v) Disclosing certain problems or troubles that may otherwise be of a serious nature in a large scale main survey.

Q. What are the important sources of secondary data ?

Ans. There are some agencies both at the national and the state level to collect, process and tabulate statistical data. The notable sources are :

(a) The census of India provides most complete and demographic record of population. The census is being regularly conducted every ten years since 1881. The census officials collect information on various aspects of population such as sex ratio, litera etc. Census data is interpreted and analysed to understand many economic and social issues in India.

(b) The NSSO was established by the Government of India to conduct nationwide surveys on socio-economic issues. The data collected by NSSO are released through reports arid its quarterly journal 'Sarvekshana'. NSSO provides periodic estimates of literacy, school enrolment, unemployment etc.

NSSO conducts three types of surveys :

- Socio-economic surveys
- Annual survey of industries
- Agricultural surveys

NSSO conducts the following functions :

- Carries out socioeconomic surveys
- Collects data on price level in rural and(urban sector
- Follows up surveys of economic census
- Designs research activities.

Q. How primary data is collected ?

Ans. The most popular and common method is questionnaire/ interview schedule to collect the primary data. The questionnaire is managed by the enumerator, researchers, or trained investigators.

Q. What is basic difference between primary and secondary data.

Ans. Primary data is original and first-hand information collected originally while secondary data is collected through other sources like published reports, website, Government or company departments, etc.

Q. Write any five principles of drafting questionnaire.

Ans. Principles: {a) covering letter, (b) question should be simple and clear, (c) multiple choice questions, {d) minimum number of questions, (e) avoidance of questions of calculations.

Q. Distinguish between population and sample.

Ans. Population or Universe means the inclusion of all the items in the field of statistical equiry and sample means selection of few items as representatives of all the items.

Q. Name two methods of obtaining the simple random sample.

Ans. (a) Lottery method,

(b) Tables of random numbers like Tippett's random numbers, Fisher and Yates tables.

Q. What is stratified random sampling?

Ans. In this method the universe is divided into strata or homogeneous groups and equal sample is drawn from each stratum or layer at random.

Q. What is systematic or quast-random sampling?

Ans. This is used when a complete list of. the population is available in order, e.g., alphabetical order. The method consists of selecting every nth item from the list, viz., 15th, 25th or 35th and so on.

Q. What is the difference in the meaning of mistake and error in statistics? What are the sources of errors?

Ans. Mistake means a wrong calculation or use of inappropriate method in the collection or analysis of data. Error means the difference between the true value and the estimated value.

Source of errors: Errors in origin, errors of inadequacy, errors of calculation and errors of interpretation.

Q. Give names of types of errors.

Ans. a. Absolute and relative errors,

b. Biased and unbiased errors,

c. Sampling and non-sampling errors.

Q. Which are the important sources of secondary data in India?

Ans. a. Census of India,

b. National Sample Survey Organisation (NSSO).

LAKME LIMITED, MUMBAI
Questionnaire

Note : Please tick (✓) mark in the squares which apply to you.

I. General. (This information will be kept confidential).

Name .... Age ....

Address .... Sex : Male ☐

.... Female ☐

....

Marital Status :

Married ☐ Unmarried ☐ Any other ☐

Monthly income		No. of members in the family	
Rs. 0-10,000	☐	1-3	☐
Rs. 10,000-20,000	☐	4-6	☐
Rs. 20,000-30,000	☐	Above 6	☐
Rs. Above 30,000	☐		

II. Which of our cosmetics do you like most?

Lipstick ☐ Cream ☐ Powder ☐

Nail Polish ☐

III. Do you like our cosmetics because :

(i) They are reasonably priced ? Yes ☐ No ☐

(ii) They are easily available? Yes ☐ No ☐

(iii) They are liked and used by your friends ? Yes ☐ No ☐

(iv) You have a fancy for them ? Yes ☐ No ☐

IV. Do you prefer Lakme mainly due to :

(i) Perfume Yes ☐ No ☐

(ii) Material Yes ☐ No ☐

(iii) Colour Yes ☐ No ☐

(iv) Packing Yes ☐ No ☐

V. Do you feel that our cosmetics add to your grace and beauty ?

Yes ☐ No ☐

VI. Do you find our product costly ? Yes ☐ No ☐

VII. Do your friends often ask you about the cosmetics you have used in parties, functions, marriages and get-togethers ? Yes ☐ No ☐

VIII. Do you visit our showrooms in various parts of the city ?

Yes ☐ No ☐

IX. Do you think our company should manufacture a new cosmetics ?

Yes ☐ No ☐

X. Which product do you think needs improvement :

Lipstick	☐	Cream	☐
Powder	☐	Nail Polish	☐

CHAPTER – 4 – CENSUS AND SAMPLING

1. ***Population***

Population or universe in statistics mean the inclusion of all the items in the field of statistical enquiry

Sampling

It means selection of few items as representative of all the item. Apart of whole population is called sample and the process is turned as sampling.

2. ***Census Method***

A survey which includes every element of the population is known as Census Method. It is also known as the 'Method of Complete Enumeration' or '100% Enumeration'.

The essential feature of this method is that it covers every individual unit in the entire population. For example, if certain agencies are interested in studying the total population in India, they, have to obtain information from the household in rural and urban India.

Example : CENSUS OF INDIA, which is carried out every ten years includes a house-to-house enquiry covering all the households in India. Demosgraphic data on birth and death rates, literacy, published by the Registrar General of India. The last Census of India was held in February, 2011.

Suitability

Census method is suitable when :

(i) size of population is small.
(ii) extensive study of diverse items is required.
(iii) high degree of accuracy is needed.
(iv) selection of sample items from universe is not possible.
(v) reliable data is required.

Merits

a. **Accurate and reliable data**

Since under this method the information relating to each unit of the universe is collected.

b. **Less element of bias**

Since, information collected about each unit the possibility personal element or bias is minimum.

c. **Comprehensive information**

This method facilitates the collection wide and comprehensive information.

d. **Appropriations**
For certain types of surveys this method is more appropriate such as pop. census.

e. **Characteristics of Universe**
It contains all the characteristics of a universe.

Demerits

a. **Costly**
It is a costly method.

b. **Time consuming**
This method involves more time and labour.

c. **Unusable**
This method cannot be used in many situations. For e.g., it cannot be used where universe is finite or where unis may finish during the process of testing. e.g., perishable commodities.

d. **Unverifiable**
In case any problem arises it is difficult to verify the inf. because. A census survey can nto be easily reconducted.

e. **Huge Organisazation**
A very large organization is needed for successfully conducting a census survey.

f. **Unit for infinite universe**
This method cannot be used in cases of infinite universe. In such a situaton it is not feasible to contact all the units of a universe.

Sample Survey

In a sample survey information is collected about a part of the universe and on the basis result and conclusions are drawn about the whole universe or the part of the universe about which contains all the broad characteristics of the universe.

A sample survey is generally preferred where the size of population is large, where pop, is infinite, where different units of the universe are broadly similar and where very high degree of accuracy is not required.

Suitability

(i) size of the universe is very large.
(ii) when more accuracy is not required
(iii) when an extensive study is not necessary.
(iv) when different items of the population are broadly similar.
(v) when census method is not applicable.

Merits

a. **Economical** : It saves time, money and labour.

b. If sample can be taken out properly with due care and caution, it's conclusion, and inferences can also be reliable and accurate like that of census survey.

c. **Scientific** : This method is more appropritate for survey whose scope or area is very wide.

d. **Quick results** : Results can be obtained very quickly with the help of this method.

e. Even a small organization can use sampling method.

Demertis

a. Under this method data are not very accurate.

b. If a sample is not representative then all conclusion become wrong.

c. For the proper selection of sample, special knowledge and understanding is required.

d. This method is of no use for such surveys where information about the whole universe is required such as pop. census.

Q Which method is Better ?

Sampling method is genrally regarded better and more appropriate in comparison to census method because of the following reasons.

a. This method is more appropritate because it involves less time and labour.

b. This method is more appropritate for such data about which inf. is regularly available such as information about newly born babies.

c. In many cases method is not possible.

Keeping in view the above facts, generally the sample survey is more appropritate than census survey. These days, sample survey is popular so much that even the accuracy of census survey is tested with a sample survey. Keeping this in view the above facts, generally the sample survey is more appropritate than census survey. These days, sample survey is more popular so much so that even the accuracy of census survey is tested with a sample survey.

Difference between Census Method and Sample Method Census Method Sample Method ?

Basis	*Census Method*	*Sample Method*
Coverage	The whole universe is taken for collection of data in census method.	Only representative sample is taken for collection of data in sample method.
Suitability	This method is suitable when the area under investigation is relatively small.	This method IS suitable when the area under investigation is large.
Accuracy	Since all items are studied under census method highest degree of accuracy is possible	Since only representative samples are studied under sample method, it is less accurate However errors can be easily detected and removed.
Cost	As all items are studied under census method this method is very expensive and involves a lot of money and efforts.	As only few samples are studied under sample method this method is comparatively less expensive.
Time	Census method is very time consuming as all items are studied.	Sample method is less time-consuming as only samples are studied.
Nature of Items	Census method is suitable when items in the universe have diverse characteristics.	Sample method is suitable when items in the universe are homogeneous.
Verifica-tion	Verification of data obtained through census method is not generally possible as it involves huge expenses and repetition of the whole exercise.	Verification of data obtained through sample method is possible, so errors can be easily identified and corrected.
Organisa-tional Skills	Census method needs more organisational skills as all items are to be studied.	Sample method needs less organisational skills as few items are to be studied.

METHODS OF SAMPLING

Broadly speaking, various methods of sampling can be grouped under two main heads:

(a) Random Sampling, and

(b) Non-Random Sampling.

Let us discuss now the various sampling methods which are popularly used in practice.

METHODS OF SAMPLING

Random Sampling	Non-Random Sampling
(a) Simple or Unrestricted Random Sampling	(a) Judgement Sampling
(b) Restricted Random Sampling	(b) Quota Sampling
(i) Stratified Sampling	
(ii) Systematic Sampling or Quasi-Random Sampling	(c) Convenience Sampling
(iii) Cluster Sampling or Multi-stage Sampling	

RANDOM SAMPLING

Random Sampling is one where the individual units (samples) are selected at random. It is called as probability sampling.
Random sampling does not mean unsystematic selection of units. It means the chances of each item of the universe being included in the sample is equal.
Following are the methods of random sampling.

Simple or Unrestricted Random Sampling
This method is also known as simple random sampling. In this method the selection of item is not determined by the investigator but the process used to select the terms of the sample decides the chances of selection. Each item of the universe has an equal chance of being included in the sample It is free from discrimination and human judgement. It depends on the law of probability which decides the inclusion of items in a sample. There are two methods of obtaining the simple random sample They are :

(a) Lottery Method, and
(b) Table of Random

(a) Lottery Method : All the items of the universe are numbered and these numbers are written on identical pieces of paper (slip). They are mixed in a bowl and then there starts the selection by draw one by one by shaking the bowl before every draw. The numbers are picked out blind folded. All silps must be identical in size, shape and colour to avoid the biased selection.

A special kind of rotating drum is used for Ending random numbers. It is called the Electronic Random Numbers Indicator Equipment. On which numbers 0 to 9 are written. The drum is rotated by a mechanical device and each time one piece comes out. The process is repeated to get the fill number of digits.

(b) Table of random numbers : A table of random digits is simply a table of digits which have been generated by a random process. The following tables of random digits are available :

(a) Tippett's Random Sampling Numbers. There are 10400 numbers arranged 4 digits a time.

(b) Rand Corporation's a million random digits.

(c) Fisher and Yates Table having 15000 digits.

Tippett's table of random numbers is most popular which can be used in taking out sample. The first thirty sets of numbers out of 10400 are given below :

Suppose, we want to decide the sample of 15 students out of 2000 students in a college. We will first number all 2000 students from 1 to 2000. After numbering the students, now we will consult a page of Tigpett Table. We can get sample by taking any 15 successive number either horizontally or vertically.

Merits :

1. It is more scientific method of taking; out samples from a universe . Every item in the universe has equal chance of being selected.
2. It is more representative. When size of sample increases, it is more representative of the population as the Law of Inertia of large numbers and the Law of Statistical Regularity begin to operate.
3. This method is economical as it saves sum, money and labour in investigating a population.
4. The theory of probability is applicable, if the sample is random.
5. Sampling error can be measured.

Demerits

1. This requires complete lisp of population but up-to-date lists are not available in many enquiries.
2. If the size of the sample is .small, then it will not be a representative of a population.
3. When the distribution between items is very large, this method cannot be used.

Restricted Random Sampling

They are as follows :

(i) Sratified random sampling : In this method the universe is divided into strata or homogeneous groups and an equal sample is drawn from each stratum. For example, suppose we want to know how much pocket money an average university student gets every month will be taken equal sample from various strata, namely, BA. students, M.A. students and Ph.D. students, etc.

There are different types of stratified sampling :

(a) Proportional stratified sampling is one in which the items are taken, from each stratum in the proportion of the units or the stratum to the total population;

(b) Disproportionate stratified sampling is one in which units in equal numbers are taken from each stratum irrespective of its size.

(c) Stratified weighted sampling is one where units are taken in equal number from each stratum, but weights are given to different strata on the basis of their size.

Merits

I. The sample taken under this method is more representative of the universe as it has been taken from different groups of universe.

2. It ensures greater accuracy as each group (stratum) is so formed that it consists of uniform or homogeneous items.

3. It is easy to administer as universe is sub-divided.

4. For non-homogeneous population, it is more reliable.

Demerits

1. Stratified sampling is not possible unless some information concerning the population and its strata is available.

2. It proper stratification is not done the sample will have an effect of bias

(ii) Systematic sampling or quasi-random sampling : This is used when a complete list of the population is available. This is called a quasi-random method because a kind of randomness is achieved by preparing this list in some random order, for example, alphabetical order.

The method consists of selecting every nth item from the list, n stands for any number. Suppose We have a universe of 10,000 items and We Want a sample of 1000, then we take n = 10. The method of selecting the first item from the list is to decide at random from the first sampling interval, i.e., between one and ten.

Suppose we pick up the 5th item. Then the other items will be 15th, 25th, 35th, and so on until we have got our full sample.

Merits

1. It is systematic, very simple, convenient and checking can also be done quickly.
2. In this method time and work is reduced much.
3. The results are also found to be generally satisfactory.

Demerits

1. Systematic selection may or may not approach chance or random selection as random will not be a determining factor in the selection of a sample.
2. It is feasible only if the units are systematically managed.
3. The universe is arranged in wrong manner, the results will be misleading.

(iii) Cluster sampling or multi-stage sampling : In this method Sampling is carried out in a number of stages. This is done when we know that for getting reliable results we have to divide and sub-divide a universe according to its characteristics. Thus, if a survey is to be conducted in a country it will first be divided into zones or states or regions, then into smaller units cities, towns and villages and then into localities and households. At each stage sampling is done by a suitable method, say simple random sampling. This method of sampling is very helpful in many large scale survey where the preparation of the list of all units in the population is difficult, time consuming and expensive.

Non-Random Sampling

It is done on the basis of convenience and judgement of the investigator and not on the basis of probility. The following are imp. methods.

a. ***Deliberate Sampling or Purpositive or Judgement Samplery*** This method is also called judgement sampling. According to this method, for selecting a sample, no specific procedure is used rather the investigator according to his desire and needs select those units of the universe as sample which fully represents the Universe. Units which are to be included in the sample inclusively depends upon the discreation of the person who results the sample. For e.g., If he has to select a sample of 40 students out of 1000 students which 40 students will be selected absolutely depends upon the direction of peoples concern. In this method, there is every possibility that only those units may be included in the sample.

Merits

a. This method is simple

b. It involves less expenditure

c. This method is more appropriate for such surveys where all the items of the Universe are similar.

Demerits

a. This method contains an element bias.

b. Data collected using this method is not reliable.

b. ***Quota Sampling***

It is a kind of judgemetn sampling. According to this method.

a. Universe or population is divided into various groups on the basis of different characters. Such as income, age or religion.

b. Quota is fixed for each group such as, how many units are to be taken from income groups for sample.

c. According to the prescribed Quota for various groups the investigation chosen by the unit according to his discretion. Under this method the sampling of units is different from that of Random sampling.

c. ***Convienence Sampling***

This method is selected purely on the basis of conviencnce for e.g., for study the running away students from school the investigator may select a school or schools in the neighbourhood because it is convienient him to go to the school.

TYPE OF ERRORS

a. **Absolute and relative errors** : Absolute error is the difference between the actual true value and estimated approximate value while relative error is the ratio of absolute error to the approximated value.

Absolute error = Actual value – Estimated value

$$\text{Relative error} = \frac{\text{Actual value} - \text{Estimated value}}{\text{Estimated value}}$$

Symbolically,

$$e = \frac{U' - U}{U}$$

Here, U_e = Absolute error

e = Relative error

U´ = Actual value

U = Approximate value

Relative error is generally used in statistical calculations because absolute error gives wrong or misleading calculations.

b. **Biased and unbiased errors :** Biased errors arise due to some prejudice or bias in the mind of investigator or the informant or any measurement instrument. Suppose the enumerator used the deliberate sampling method in place of simple random sampling method; then it is called biased error. Biased errors arise due to faulty process of selection, faulty work during the collection of information and faulty method of analysis.

Unbiased errors are not the result of any prejudice or bias. They are those which arise acccidently just on account of chance in the normal course of investigation. Unbiased errors are generally compensating.

c. **Sampling and non-sampling errors :** The errors arising on account, of drawing inferences about the population on the basis of few observations (sampling) are called sampling errors. The errors mainly arising at the stages of ascertainment and processing of data, are called non sampling errors. They are common both in census enumeration and sample surveys.

Discuss the various qualities of Good Sample.

Essentials/Qualities of a Good Sample

For obtaining impartial and accurate results, a sample should have the following qualities :

1. **Representative.** A good sample is one which represents the characteristics of all the items in the universe. This is possible only when each and every item in population has a fair and equal chance of being selected as a sample.
2. **Homogeneity.** The items that are selected as samples should be homogeneous nature so that they can truly help in investigation. However, these samples should not be contradictory to each other.
3. **Independent.** Selection of one item of the universe in the sample should not depend selection of some other item in the sample, e.g.., as done in systematic sampling. Items in the population should be independent of each other.
4. **Sufficiency.** To get accurate and reliable conclusions, the number of items in the sample should be adequate enough to cover all characteristics of the universe.

Reliability of Sample Data

For ensuring reliability in sampling certain principles must be followed. In sampling method it is presumed that whatever conclusions are drawn from a sample are also true for the whole population. This presumption is based mainly on the following two laws :

a. The Law of Statistical Regularity
b. The Law of Inertia of Large Numbers.

a. **The Law of Statistical Regularity**

The law of statistical regularity is derived from the mathematical theory of probability. It says that a comparatively small group of items chosen at random from a very large group will, on the average, represent the characteristics of the large group.

In the process of sampling each unit of the universe has an equal chance of being selected. Therefore, the selected items can be said to be representative of the universe. Although the law is not as accurate as a scientific law is, it does insure a reasonable degree of accuracy.

b. **The Law of Inertia of Large Numbers**

This law is also called the law of stability of mass data. It is based on the law of statistical regularity. Basically, it states that if the numbers involved are very large, the change in a sample is likely to be very small in other words, the individual units of a universe very continually but the total universe changes slowly. That is, large aggregates are most stable than small ohes. Because of the slow change in the nature of total universe this law is called the law of inertia (laziness) of large numbers.

For example, sugar production of factory will vary significantly from year to year but the sugar production of a country as a whole will remain comparatively stable. Or a great change may take place in the male-female ratio of family may appreriably change over a short period, but the male-female ratio of a country as a whole will remain almost for the period.

Statistical Errors

There is a great difference in the meaning of mistake and error in statistics. Mistake means a wrong calculation or use of inappropriate method in the collection or analysis of data. Error means " the difference between the true value and the estimated value." In other words the difference between the true approximated (estimated value) and the actual value (true value) is called statistical error in a technical sense. For example, we make an estimation that in a particular meeting , 1000 persons are there. But we count persons, it may be wrongly counted as 1,030. There is a difference of 30 between the estimated value and counted value. This difference is called 'error' in statistics. But when we make wrong calculation, following wrong method, draw wrong conclusions, etc. They are known as 'mistake'. For example, there is a meeting, we sent a person to count the audience, he counts the number of persons as 600, but actually, there are 590 persons. This is called 'mistake' in counting.

Types of Errors

(a) Absolute and relative errors 3 Absolute error is the difference between the actual true value and estimated approximate value while relative error is the ratio of absolute error to the approximated value.

Absolute error = Actual value - Estimated value

Symbolically,

$$Ue - U' - U$$

$$\text{Relative error} = \frac{\text{Actual value} - \text{Estimated value}}{\text{Estimated value}}$$

Symbolically,

$$e = \frac{U' - U}{U}$$

Here, Ue = Absolute error
e = Relative error
U' = Actual value
U = Approximate value

Relative error is generally used in statistical calculations because absolute error gives wrong or misleading calculations.

(b) Biased and unbiased errors ; Biased errors arise due to some prejudice or bias in the mind of investigator or the informant or any measurement instrument.

Suppose the enumerator used the deliberate sampling method in place of simple random sampling method; then it is called biased error. Based errors arise due to faulty process of selection, faulty work during the collection of information and faulty method of analysis.

Unbiased errors are not the result of any prejudice or bias. They are those which arise accidently just on account of chance in the normal course of investigation. Unbiased, errors are generally compensating.

(c) Sampling and non-sampling errors : The errors arising on account of drawing inferences about the population on the basis of few observations (sampling) are called sampling errors. The errors mainly arising at the stages of ascertainment and processing of data, are called non-sampling errors. They are common both in census enumeration and sample surveys,

Methods of sampling

Random Sampling	*Non - Random Sampling*
a. Simple or Unrestricted Random Sampling	a. Judgment Sampling
b. Restricted Random Sampling	b. Quota Sampling
i. Stratified Sampling	
ii. Systematic Sampling or Quasi Random Sampling	c. Convenience Sampling
iii. Cluster Sampling or Multistage Sampling	

Q. Give names of types of errors.

Ans. (a) Absolute and relative errors,
(b) Biased and unbiased errors,
(c) Sampling and non-sampling errors.

Q. Which are the important sources of secondary data in India?

Ans. (a) Census of India,
(b) National Sample Survey Organisation (NSSO).
Methods of Sampling

— Notes —

CHAPTER – 5 – TABULAR PRESENTATION – TABLES

Table

A table is a systematic organisation of statistical data is columns and rows. Rows are hrizontal arrangements and columns are vertical arrangements.

Tabulation

The process of arranging the statical data in rows and columns is known as tabulation. It is scientific process of presentation of clasified data in a proper order to facilitate comparison.

Advantage or objectives of tabulation :

The tabulated data performs the following functions :

(i) The tabulated faCts can be more easily understood

(ii) The data become attractive and leaves a lasting impression as compared to data which are not tabulated.

(iii) The tabulated data facilitate quick comparison

(iv) The tabulated arrangement makes the summation of items possible

(v) Detection of errors and omissions in data becomes easier.

(vi) In the tabular form of data the repetition of explinatory phases and headings becomes unnecessary

(vii) They facilitate computation of different stastistical mainly average, despersion, correlations etc.,

PARTS OF TABLE

(1) **Table Number**

When there are more than one table, each table must have a number. So that they may be easily refferred to.

(2) **Title**

The title is placed above the table . it must be brief clear, self explanatory unmbigous and consist. A complete title explain.

a) Nature of data

b) Time period of data

c) The field to which data is related

d) Basis of classification of data

(3) **Captions (Column heading) and stubs (Row heading)**

Captions are the disignation (headings or sub heading) given to a vertical coulumns, while stubs are the designations given to horizontal rows. both should be brief, descriptive and clearly defined.

(4) **Body of the table**

It contains the numerical information, irrelavant matter must be avoided. Body should be made as comprehensive as possible keeping in view the purpose of presentation.

(5) **Head Note (prefatory Note)**

It is a statement normally given below the title. It explains the contains of the table or main parts of it.

(6) **Foot Notes**

It is placed at the bottom of the table. Footnotes are used to explain the things that may not be clear in the main body.
Footnotes can be identified by putting stars or signs.

(7) **Source**

The source appears below the footnotes if they are used. In case of tabulation of secondary data, a source note should be provided indicating place of publication, page, edition, name of the publisher or author, etc.

Example 1 : In XIth class, there are 120 students, out of which 70 are boys. Out of total 70 boys, 30 boys belong to Science stream and 25 are students of Art steam. There are 10 girls in Science stream and 35 in Commerce stream. Present these facts in the table.

Sol.

Stream	No. of Students		Total
	Boys	Girls	
Science	30	10	40
Commerce	15	35	50
Arts	25	5	30
Total	70	50	120

Example 2 : In 2007, out of a total of 2,000 applicants in a college, 1,200 were from Commerce background. The number of grils was 750, out of which 330 were from Science steam. In 2008, the total number of applicants was 3,500 of which 2,200 were boys. The number of students from Science Steam was 1,100 of which 610 were girls. Tabulate the given information.

Sol.

Stream	2007			2008			Total		
	Boys	Grils	Total	Boys	Girls	Total	Boys	Grils	Total
Science	470	330	800	490	610	1100	960	940	1900
Commerce	780	420	1200	1710	690	2400	2490	1110	3600
Total	**1250**	**750**	**2000**	**2200**	**1300**	**3500**	**3450**	**2050**	**5000**

Example 3 : 2002, out of total 700 employees of a factory, 475 employees were skilled. The number of woman employed was 450 of which 175 were unskilled. In 2007, the number of unskilled employees fell down to 190 of which 80 were women. Tabulate the given information and gives a suitable title.]

Sol.

Years	Number of Employees						Grand Total		
	Man			Women					
	Skilled	Unskilled	Total	Skill	Unskilled	Total	Skill	Unskilled	Total
2002	200	50	250	275	175	450	475	225	700
2007	120	110	230	240	80	320	360	190	550
Total	**320**	**160**	**480**	**515**	**255**	**770**	**835**	**415**	**1250**

Unsolved Practicals

Q1. In a sample study about coffee habit in two towns, the following information was received.

Town A : Femals were 40%; Total coffee drinkers were 45% and Males non-coffee drinkers were 20%

Town B : Males were 55%; Males non-coffee drinkers were 30% and females coffee drinkers were 15%. Present the data in tablular form.

Q2. Of the 1,125 students studying in a school during 2005-2006, 720 are Hindus, 628 are boys and,440 are science students. The number of Hindu boys is 392, that of boys studying science 205 and that of Hindu students studying science 262; finally, the number of science students among the Hindu boys was 148. Enter these frequencies in a table and complete the table by obtaining the frequencies of the remaining cells.

Q3. In a sample study about the smoking habits of people in two towns following data were derived :

	Town A	Town B
Males in Total Population	52%	54%
Smokers	26%	28%
Male Smokers	18%	20%

Tabulate the above data.

Q4. In a sample study about the coffee habits in two towns following data were observed : Tabulate the observations.

Town A	:	60% people were males
		40% were coffee drinkers, and
		26% were male coffee drinkers
Town B	:	55% people were males,
		30% were coffee drinkers, and
		20% were male coffee drinkers

Q5. Census of India 2001 reported that Indian population had risen to 102 crore of which only 49 crore were females against 53 crore males. 74 crore people resided in rural India and only 28 crore lived in towns or cities. While there were 62 crore non-worker Population against 40 crore workers in the entire country, urban population had an even higher share of non-workers (19 crore) against the workers (9 crore) as compare to the rural population where there were 31 crore workers out of 74 crore population. Represent the above information in a tabular form.

Q6. Present the following information in a suitable table :
In 1995 out of a total of 1750 workers of a factory 1200 workers were members of a trade union. The number of women employed was 200 of which 175 did not belong to a trade union. In 2000, the number of union workers increased to 1580 of which 1290 were men. On the other hand, the number of non-union workers fell down to 208 of which 180 were men. In 2005, :here were on the pay rolls of factory, 1850 workers of whom 1800 belong to a trade union. Of all the employees in 2005, 300 were women of whom only 8 did not belong to a trade union.

CHAPTER – 6 – DIAGRAMMATIC PRESENTATION

Diagrams & charts occupy an important place in statistical analysis. Data may be presented in a simple and attractive manner in the form of diagrams. Diagrams are helpful in communicating the entire meaning of the complex facts, which they represent. This facilitates analysis & comparisions.

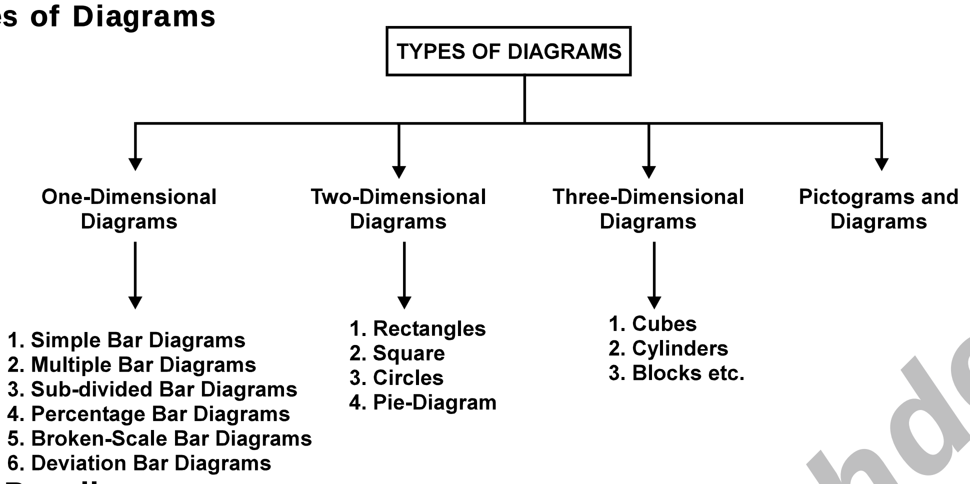
Advantanges of Diarammetric Presentation

1. **Diagrams are attractive and impressive :** Data presented in the form of diagrams are able to attract the attention of even a comon man.
2. **Easy to Remember :** Diagrams have a great memorising effect. The picture created in the mind by diagrams last much longer than those created by figures presented through tabular form.
3. **Diagrams are useful in making comparisions :** It becomes easier to compare two sets of data by presenting them through diagrams.
4. **Universal Applicability :** This technique can be used universally at any place and at any time.
5. **Diagrams are more information :** Diagrams not only depicts the characteristics of data, but also bring out other hidden facts and relations which are not possible from the classified and tabulated data.

Limitations of Diagramatic Presentation

1. **No Utility to Experts :** Diagrams give only a vague idea of the problem which may be useful for a common man but not for an expert, who wishes to have an exact idea of the problem.
2. **Limited information :** Diagrams provide limited and approximate information. For detailed and precise information, we have to refer the original statistical tables.
3. **Lack of further analysis :** Diagram cannot be further analysed.
4. **Minute Difference Presentation Impossible :** In case of large figure (observations) diagrams fail to reveal small differences in them. For example, if two large values like 8,560 and 8,500 are to be shown in a diagram, the difference will not be very apparent.

Types of Diagrams



1. Bar diagram

Bar diagram is the most commonly used diagrams for presenting data. They are called bar diagrams because the information is presented through means of one dimensional bars. The width of the bar does not matter.

Kinds of bar diagrams

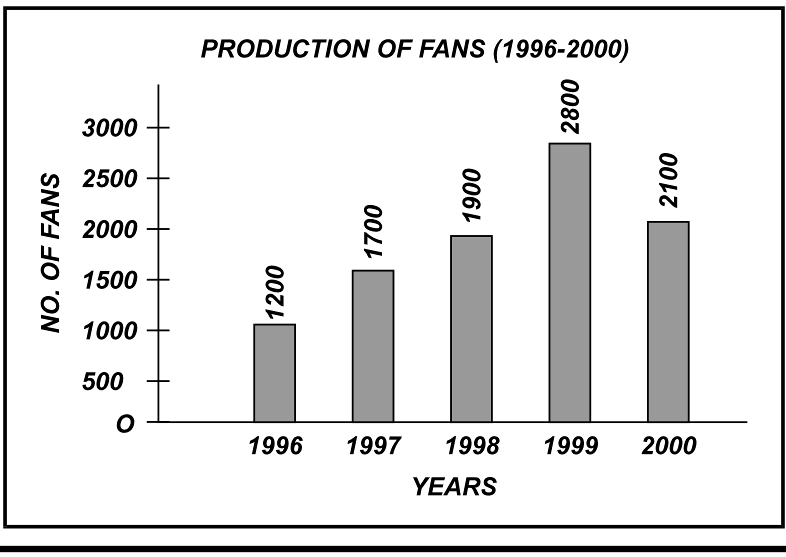
1. Simple bar diagrams

In these diagrams values of a variable such as quantity, population, demand etc is presented with the help of bars. The length of the reactangle represents the value of the variable. In this kind of diagram the width of the reactangle is meant for beauty and attraction only. These bar diagrams present only a single set of numerical data.

Illustration 1 : Draw a bar diagram to represnt the following figures relating to manufacturing of fans.

Solution :

Years :	1996	1997	1998	1999	2000
No. of Fans :	1200	1700	1900	2800	2100



2. Multiple bar diagrams

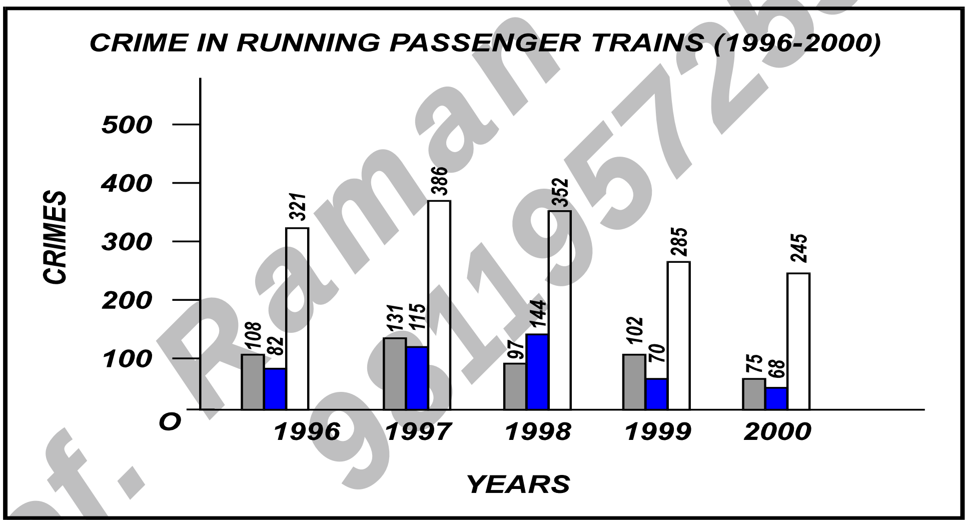
They are used while comparing data relating to two or more facts at various times or places. All the bars presenting various variable to a year are made together adjacently. In this the length of the bars depend upon the value of the variables they represent.

Illustration. Draw a suitable diagram of the following data :

Statement of Crime in Running Passenger Trains

Year	Loot	Murder	Robbery
1996	321	108	82
1997	386	131	115
1998	352	97	144
1999	285	102	70
2000	245	75	68

Solution



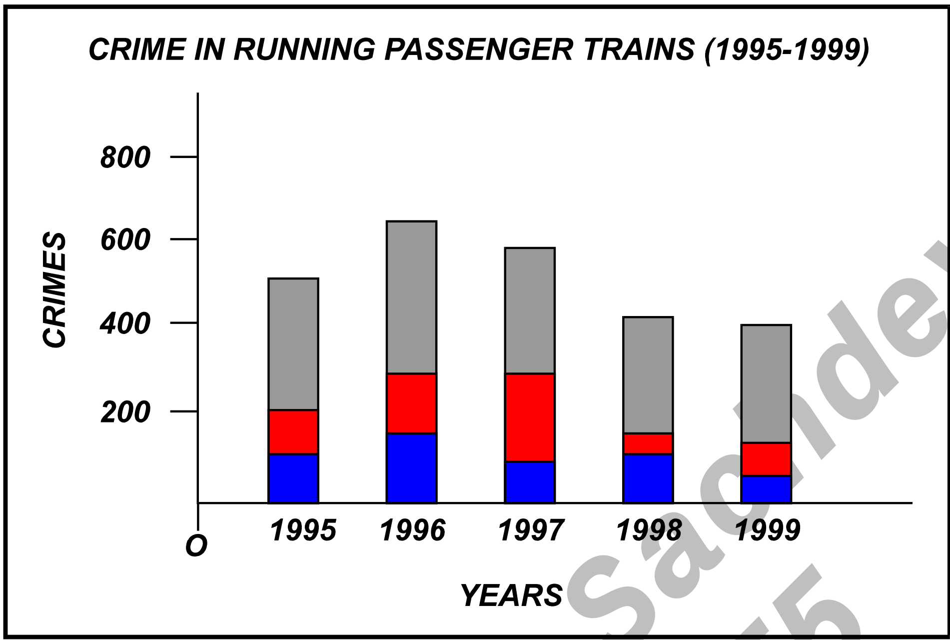
3. Subdivided bar diagrams

They are those which present simultaneously total value parts of sets of data to indicate different parts of bars these are shared with colours.

Illustration. Draw a suitable diagram to represent the following information.

Satement of Crimes in Running Passanger Trains

Year	Murder	Robbery	Loot	Total
1995	108	82	321	511
1996	131	115	386	623
1997	97	144	352	593
1998	102	70	285	457
1999	75	68	245	388



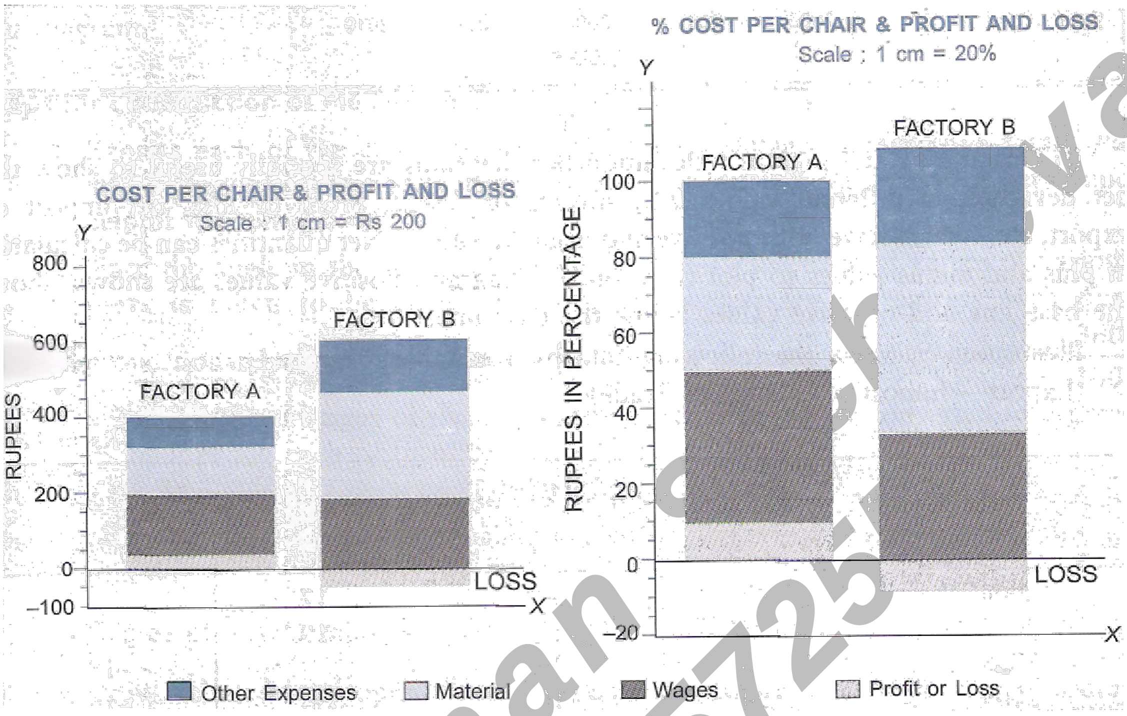
4. Percentage bar diagrams

These are those diagrams which show simultaneously different parts of the values of a set of data in terms of percentage. If we have to show the precentage share of different crops in the total production, then percentage of sub-divided bar diagram will- be used. Total value indicates by total length of the bar is assumed to be 100. Each part then is shown as a part of 100. These parts may be shaded with different colours in order to highlight the difference.

Illustration 6. Represent the following data by (i) sub-divided bar diagram, and (ii) by sub-divided bar diagram on percentage basis. Show also the profit and loss.

Proceeds per Chair	Factory A (Rs)	Factory B (Rs)
Wages	160	200
Material	120	300
Other Expenses	80	150
Total	360	650
Selling Price	400	600
Profit or Loss (±)	(+) 40	(–) 50

Solution. To get the percentage bar diagram all the values are required to be converted into percentages on the basis of selling price considering the selling price 100%. The percentages are calculated as under :



Proceeds per Chair	Factory A (%)	Factory B (%)
Wages	40	33.3
Material	30	50
Other Expenses	20	25
Total	90	108.3
Selling Price	100	100
Profit or Loss (±)	(+) 10	(–) 8.3

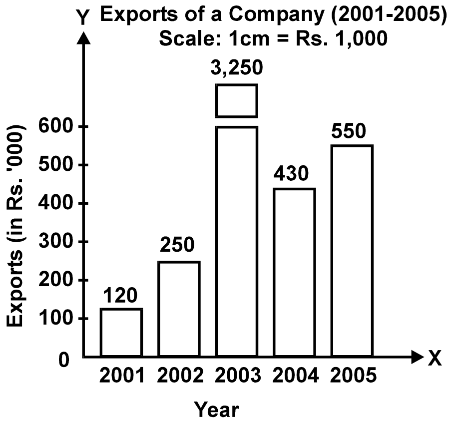
(i) **Broken-Scale Bar Diagram**

This diagram is used when value of some variable is very high or low as compared to others. In order to gain space for the smaller bars of the series, largest bars may be broken. The value of each bar is written on the top of the bar.

Ex. From the following data of exports of a company, prepare a suitable diagram.

Year	2001	2002	2003	2004	2005
Exports (Rs. in '000)	120	250	3,250	430	550

Sol. In the given example, the smallest variable is 120 and the largest one is 3,250. So there is much difference between the largest and other values. Thus, the diagrammatic representation is done by using a Broken-scale as shown in figure.



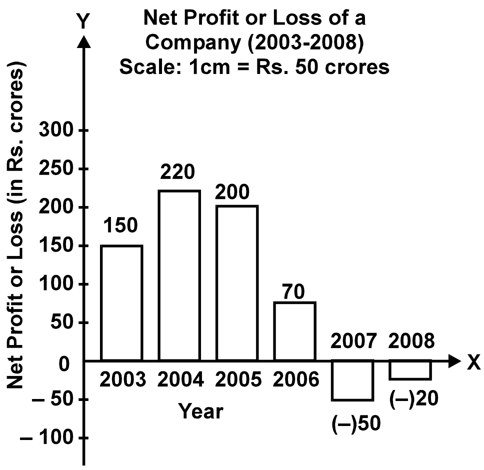
(ii) **Deviation Bar Diagram**

Deviation bars are used for representing net changes in data like net profit, net loss, net exports, net imports, etc. Only changes are represented not the original data. Positive values are shown above the X-axis and negative values below it.

Ex. Represent the following data relating to net profit and loss of company for a period of five years by deviation bar diagram.

Year	2003	2004	2005	2006	2007	2008
Net Profit / Loss (Rs. crores)	150	220	220	70	(–)50	(–)20

Sol. The positive deviations, i.e., net profit is presented by bars above the base line while negative deviations, i.e., net loss is represented by bars below the base line.



6. Pie Diagram

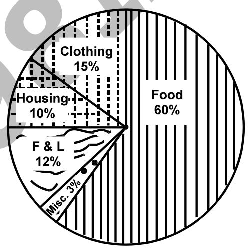
Similar to sub-divided bar diagram, a circle can also be partitioned into sections to show proportions of various components. Such a diagram is known as a pie diagram. The circle is divided into as many parts as there are components by drawing straight lines from the centre to the circumsference. In the words of Herbert and Raymond.

"A pie diagram is a chart of circular shape broken into sub-divisions. The size of the section indicates the proportion of each component part to the whole". Pie Diagram is also known as Angular Circle Diagram.

Ex.1 Draw a piec digram to represent the following data of expenditure of an average working class family.

Items of Expenditure	Food	Clothing	Housing	Fuel & Lightning	Misc.
% of Total Expenditure	60	15	10	12	3

Items of Expenditure	Percentage(%) Expenditure	Proportionate angles
Food	60%	$60 \times 3.6 = 216$
Clothing	15%	$20 \times 2.6 = 54$
Housing	10%	$16 \times 3.6 = 36$
Fuel and Lightning	12%	$18 \times 3.6 = 43.2$
Miscellaneous	3%	$6 \times 3.6 = 10.8$
	100%	360°



The circle is divided into 5 parts according to the degrees of angles at the centre.

III. Represent the following data by a pie diagram.

	Items of Expenditure	Family X	Family Y
1.	Food	400	640
2.	Clothing	250	480
3.	Rent	150	320
4.	Education	40	100
5.	Miscellaneous (Including Saving)	160	60
	Total	1000	1600

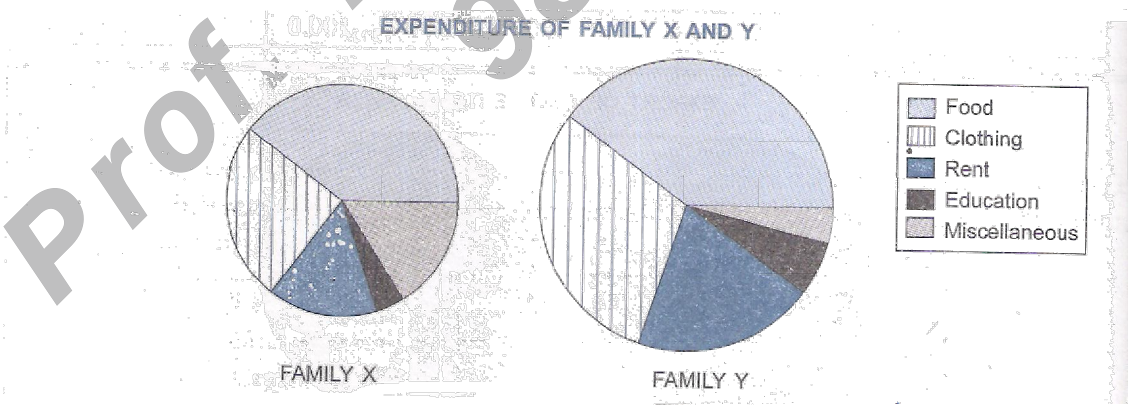
Solution. To construct pie diagram we should get the following calculations on the basis of 360 taken as equal to the total of the values.

Items of Expenditure	Family X		Family Y	
	Rs.	Degree	Rs.	Degree
1. Food	400	$\frac{400}{1000}\times 360=144$	640	$\frac{640}{1600}\times 360=144$
2. Clothing	250	$\frac{250}{1000}\times 360=90$	480	$\frac{480}{1600}\times 360=108$
3. Rent	150	$\frac{150}{1000}\times 360=54$	320	$\frac{320}{1600}\times 360=72$
4. Education	40	$\frac{40}{1000}\times 360=14.4$	100	$\frac{100}{1600}\times 360=22.50$
5. Miscellaneous (Including Saving)	160	$\frac{160}{1000}\times 360=57.6$	60	$\frac{60}{1600}\times 360=13.50$
Total	1000	360	1600	360
Square root	31.6		40	

Radii of circle are determined in proportion 3.2 : 4 (31.6 : 40).
Therefore the radii of circle according to availability of space are

Family X : Radius $\frac{3.2}{2}$ = 1.6 cm

Family Y: Radius $\frac{4}{2}$ = 2 cm



Unsolved Practicals :

Q1. **Average wages of some firms are given below. Represent this by Simple Bar Diagram.**

Firms	A	B	C	D	E
Average Wages (Rs.)	2,725	3,250	3,750	4,680	5,320

Q2. Present the following data by multiple bar diagram.

Year	B.Com(H)	Eco (H)	B.Com(P)
2002	420	200	140
2003	320	240	300
2004	380	360	480

Q3. Represent the following data with the help of sub-divided bar diagram :

Production (in '000 metre)

Year	Wheat	Rice	Cotton
2002	35	22	10
2003	15	25	16
2004	40	12	20

Q4. Represent thd data relating to the cost of production in a factory by means of percentage diagram:

Element of cost (Rs. in '000)	Items A	B	C
Raw Materials	75	55	60
Wages	45	30	25
Factory Overheads	15	10	30
Office Overheads	5	15	5

Q5. From the following data of imports of a firm, prepare a suitable diagram

Year	2003	2004	2005	2006	2007
Imports (Rs. in lakhs)	15	22	34	170	25

Q6. Represent the following data relating to net exports of a company by deviation bar diagram.

Year	2000	2001	2002	2003	2004	2005
Net Exports (Rs. crores)	250	160	(–)80	50	(–)110	70

Q7. Draw two pie diagrams to represent the marks obtained by Isha and Utkarsh (out of 50) in an examination.

Subject	Isha	Utkarsh
Maths	40	37
Economics	35	45
Accounts	45	40
Commerce	46	34
English	34	44

Q8. Represent by means of percentage bar diagram from the following data :

Item of Expenditure	Family A (Rs.)	Family B (Rs.)
Food	150	350
Clothing	38	120
Rent	56	130
Education	24	68
Miscellaneous	70	95

Q9. **Draw a bar diagram from the following data :**

Year	2007	2008	2009	2010
Export (Rs. in Crores)	73	80	85	80
Import (Rs. in Crores)	70	72	74	85

Q10. **Draw a pie diagram to represent the following data for the year 2005.**

Items	Food	Clothing	Rent	Medical	Others	Total
Expenditure (Rs.)	644	200	420	80	96	1440

Q11. **Where the diagrammatic presentations can be seen.**

Ans. Diagrammatic presentations normally seen in financial reports, in newspapers, magazines, and journal and in exhibitions.

Q12. **Distinguish between simple bar diagram and sub-divided bar diagram.**

Ans. Simple bar diagram is a presentation of one type of variable for visual comparison of one variable over different years, months, week, etc. It can present variables like students, prices population, sales, export, etc. by drawing either horizontal or . vertical base. All bars can be beautified by single colour or shading to make them more attractive. Sub-divided bar diagrams are also called component bar diagram. First of all a bar representing total of values is drawn, then it is divided into various parts in proportion to the values given in the data. Different colours, shades of colours, crossing, dotting or designs can be used to sub-divisions of a bar.

Q13. **Is pie diagram a bar diagram ?**

Ans. No. Pie diagram is a circular diagram which draws percentage breakdowns by portioning a circle into various parts.

Q14. **Draw a bar diagram from the following: data :**

Year	2007	2008	2009	2010
Export (Rs. in crores)	73	80	85	80
Import (Rs. in crores)	70	72	74	85

Q15. Represent by means of percentage bar diagram from the following data :

Item of Expenditure	Family A (Rs.)	Family B (Rs.)
Food	150	350
Clothing	38	120
Rent	56	130
Education	24	68
Miscellaneous	70	95

Q16. Draw a pie diagram to represent the following data for the year 2005.

Items	Food	Clothing	Rent	Medical	Others	Total
Expenditure (Rs.)	644	200	420	80	96	1440

Q17. Draw a pie diagram to represent the following data :

Items	2009 (Average expenditure per week in Rs.)	2010 (Average expenditure per week in Rs.)
Food	644	945
Clothing	200	322
Rent	420	525
Medical	80	210
Other item	96	518
Total	1440	2520

Q18. Represent the following information on the cost composition of a product in bar chart :

Particulars	2008	2009
Cost per table:		
(a) Wages	21	9
(b) Other costs	14	6
(c) Polishing	7	3
Total Cost	42	18
Proceeds per table	40	20
Profit (+) loss (−)	−2	+2

— Notes —

CHAPTER – 7 – GRAPHIC PRESENTATION

Graphic presentation is an attractive and simple method of presenting data According to this method, graphs are made on graph paper.

Advantages of Graphic Presentation

1. **Attractive and impressive :** Graphs are always attractive and impressive than table of figures. A fact, that an ordinary man cannot understand easily, can be understood by graphs.
2. **Simple and understandable presentation of data :** Graphs help to present complex data in a simple and understandable way. It saves time and energy of both, the statistician as well as the observer.
3. **Useful in comparison :** Graphs provide easy comparison of two or more phenomena.
4. **No need for mathematical knowledge :** There is no much technicality involved in understanding the graphs as they are very simple and easily comprehensible. Even a layman, who has no mathematical knowledge, can easily graph it.
5. **Helpful in Predictions :** Through graphs, tendencies that could occur in near future can be predicted in a better way.
6. **Universal utility :** In modern era, graphs can be used in all spheres such as trade, economic government departments, advertisements, etc.

Kinds of Graphs: There are two kinds of graphs:-

(i) Time series Graphs

(ii) Frequency Graphs

(i) Time series Graphs

A statistical series resulting to the time of its occurence (such as days, weeks, months, years etc.) is known as a time.

When time series are graphically represented, such graphs are known as time series graphs.

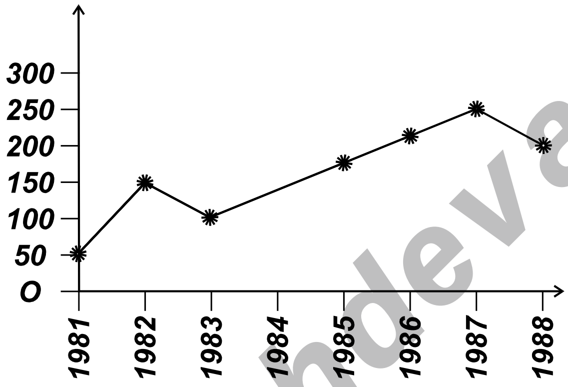
Time series Graphs are of 3 kinds:

(a) One variable Graph

In one variable time series graph, time (days, weeks, months, years) is represented on X-axis and values (such as production figures, amount of profits, sales etc.) are represented on y-axis. There are as many points as there are number of years. By joining all such points, a curve is formed which is known as the time series graph or a ***historigrams***.

Eg.

Years	Students
1981	50
1982	150
1983	100
1984	150
1985	200
1986	250
1987	200

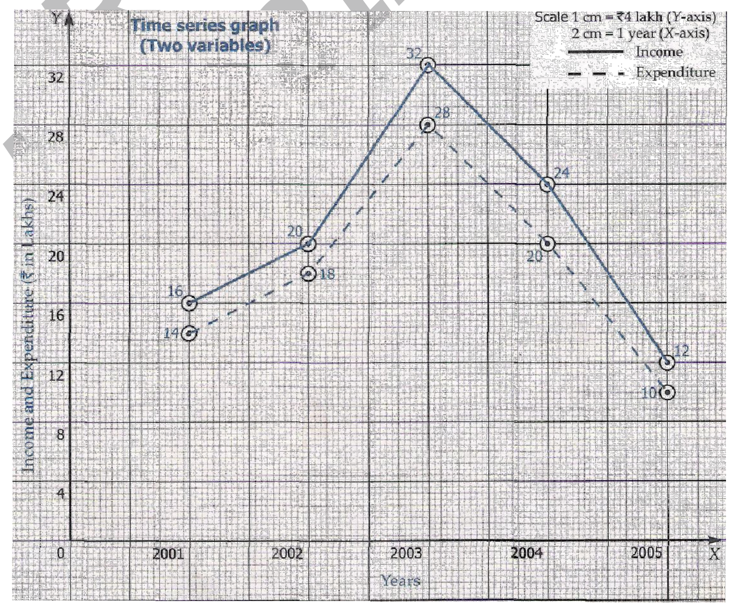


(b) Two or more than two variable graphs

Two or more than two variables can be shown on a graph paper provided that all the variables are in the same unit of measurement.

III3. Construct a time series graph to present the following data relating to expenditure and income of person in a town :

Years	2001	2002	2003	2004	2005
Expenditure (Rs.Lakhs)	14	18	28	20	10
Income (Rs.Lakhs)	16	20	32	24	12



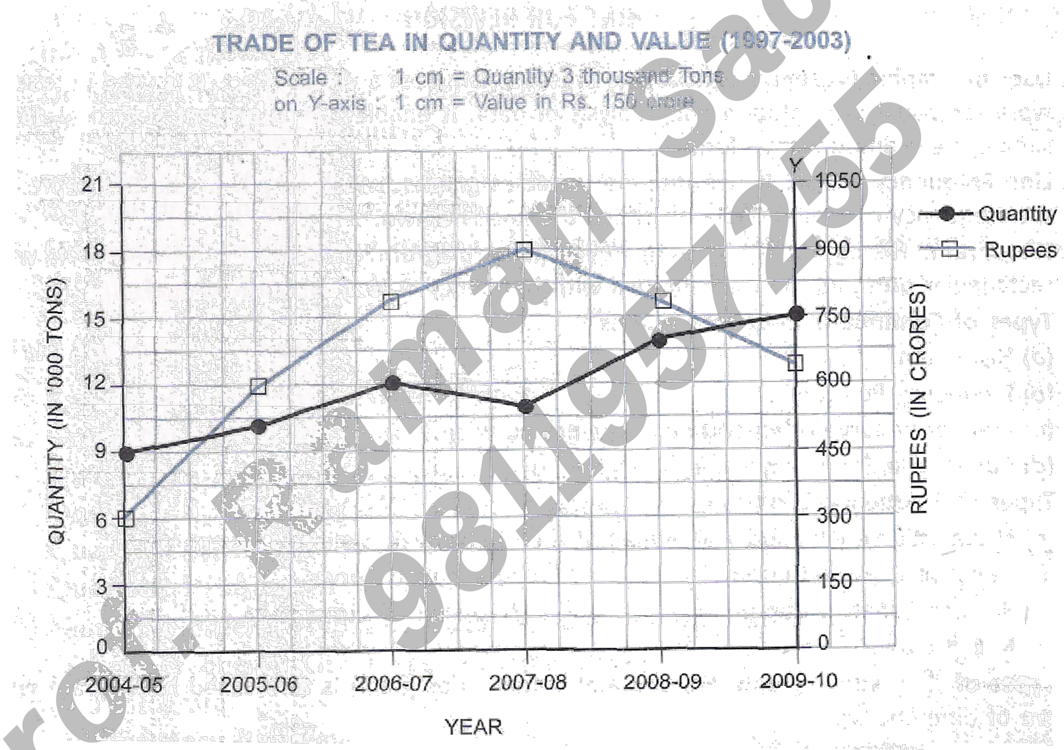
(c) Graphs of different units

When two values are given in two different units, we will have two different scales. We can take one scale on left and the other on the right side of the graph paper. As far as possible, average values of the given two different groups are kept in the middle of y-axis for better presentation and comparision.

Eg.

Year	Qty(in'000 tons)	Value (Rs. crores)
1983	9	300
1984	10	596
1985	12	782
1986	11	900
1987	14	762
1988	15	640

Avg. of Qty.: 12 approx   Avg. of Value:695 approx.



2. FREQUENCY GRAPHS

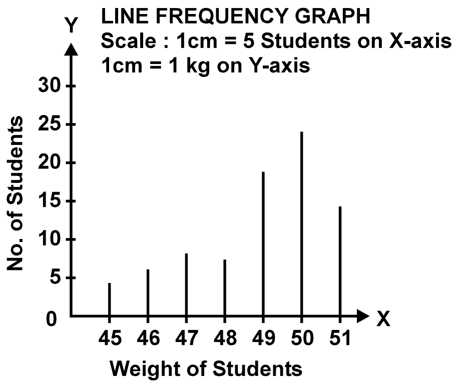
Frequency graphs represent frequencies of items in the frequency distribution. We should take class intervals mid-values and measurements on X-axis and frequencies on Y-axis. Following are the various kinds of frequency distribution graphs.

(a) Line Frequency Graphy : In this diagram, discreate series are graphically represented. Frequencies are represented through straight lines. Values are represented on X-axis & frequencies on Y-axis. Perpendiculars representing every frequency is drawn on the related value represented on Y-axis. The height of perpendicular is equal to the value of frequency. This method cannot be used for a continuous is series.

Ex. Prepare Line Frequency graph from the following data :

Weight (in kg)	45	46	47	48	49	50	51
No. of Students	4	6	9	8	18	26	15

Sol. The given data is presented in figure. Weight of students is taken on the X-axis and number of students (frequencies) on the Y-axis.



(b) Histogram : It is used to represent a continuous series. In a histogram frequency of classes are represented through adjacent rectangles. While making histogram, classes are represented on X-axis and frequencies on Y-axis. If class intervals are
equal, then the width of the rectangle will also be the same otherwise they would differ. The heights of the rectangles depend upon the values of frequencies.

Q. What is the difference between histogram and line frequency diagram ?

Ans. a. In the , line frequency diagram only the length of the line is significant where as in histogram, both width and length of the rectangle are significant.

b. Thus, we can say that line frequency graphs are one dimensional while histograms are two dimensional diagram.

Technique of Constructing Histogram

A. Histogram of Equal Class Intervals

B. Histogram when mid-points are given

C. Histogram of Unequal Class Intervals

D. Histogram when class intervals are given by Inclusive Method

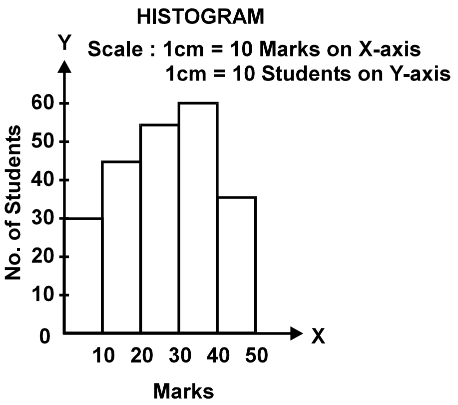
A. Histrogram of Equal Class Intervals

Ex. The frequency distribution of marks obtianed by 225 students of a college is given below :

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of Students	30	45	55	60	35

Draw a histogram for the distribution.

Sol. For representing the above data by histogram (figure), marks are plotted on the X-axis and number of studetns on the Y-axis.



B. Histogram when mid-points are given

If the mid-points of various classes are gien in place of class-intervals, then these must first be converted into classes.

Ex. Construct a histrogram from the following distribution of total wages obtained by 50 workers in a company.

Daily Wages(Mid – points)	150	160	170	180	190	200
No. of Workers	5	10	20	12	2	1

Sol. In the given example, we are given the mid-points The difference between two mid-ponts is 10. To find out lower and upper limits of various classes, class-interval of 10 is divided by 2, i.e. $\frac{10}{2} = 5.$ Now, 5 will be added and subtracted from each mid-point, to get the following class-intervals

Daily wages	145 – 155	155 – 165	165 – 175	175 – 185	185 – 195	195 – 205
No. of workers	5	10	20	12	2	1

Taking daily wages on X-axis and number of workers on the Y-axis, we get the folloiwng histogram.

C. Histogram of Unequal Class Interval

If the class-intervals are unequal, then frequencies are first adjusted before constructing the histogram. The frequency distribution is adjusted in accordance to equal class width. It is done with the help of adjustment factor, taking width of lowest class-interval as the standard one.

$$\text{Adjustmnet factor for any Class} = \frac{\text{Width of the Class}}{\text{Width of the Lowest Class}}$$

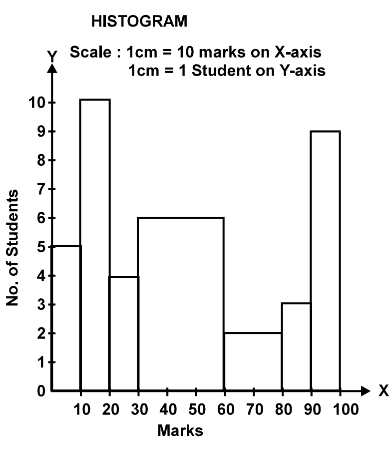
Ex. Prepare Histogram for the following data :

Marks	0 – 10	10 – 20	20 – 30	30 – 60	60 – 80	80 – 90	90 – 110
No. of Students	5	10	4	18	4	3	9

Sol. In the given example, class intervals are unequal. Therefore, for preparing histogram, the frequencies are to be adjusted. The minimum class difference is 10. If it is treated as standard class difference, then the determination of frequency for preparing histogram in the given table :

Marks	No. of Students	Adjusted Frequency for Histogram
0 – 10	5	5
10 – 20	10	10
20 – 30	4	4
30 – 60	18	$18 \div 3 = 6$
60 – 80	4	$4 \div 2 = 2$
80 – 90	3	3
90 – 100	9	9

The class-interval of the class of 30-60 is 30, which is thrice in comparison to the lowest class-interval. So, its frequency will be divided by 3. Similarly, the class-interval of 60-80 is 2 times more than the minimum class-interval. Therefore, frequency of this class will be divided by 2. The histogram for adjusted frequency is shown as under :



D. Histogram when class intervals are given by Inclusive Method

In the class-intervals are equal, but the series is inclusive one, then the inclusive series should be first converted into exclusive series, so that the class limits may have continuity. Refer the following example.

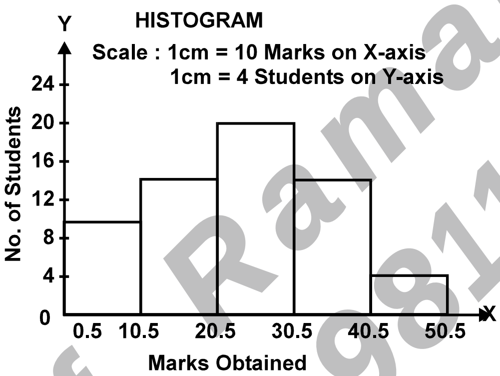
Ex. Draw a histogram from the following series :

Marks Obtained	1 – 10	11 – 20	21 – 30	31 – 40	41 – 50
No. of Students	10	14	20	12	4

Sol. First of all, we have to convert this series into exclusive series. The difference between lower limit of second class and upper limit of first class (11 – 10 = 1). This difference will be divided by 2, i.e. 1 ÷ 2 = 0.5. Now 0.5 will be subtracted from the lower limit and added to the upper limit of each class to get the following table :

Marks Obtained	0.5 – 10.5	10.5 – 20.5	20.5 – 30.5	30.5 – 40.5	40.5 – 50.5
No. of Students	10	14	20	12	4

Marks obtained are taken on the X-axis and number of studetns on the Y-axis.

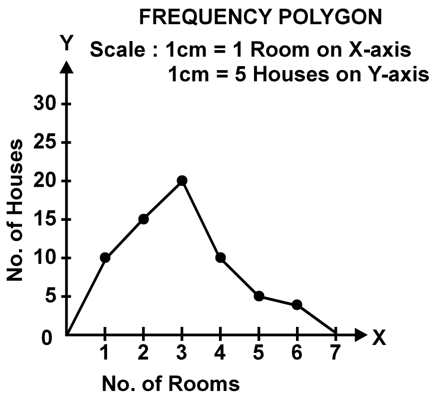


(c) FREQUENCY POLYGON

Frequency Polygon in Discrete Series

Ex. The following data shows number of rooms in 64 houses of a colony. Construct a frequency polygon

No. of Rooms	1	2	3	4	5	6
No. of Houses	10	15	20	10	5	4



Sol. This is a case of discrete frequency distribution. So, we will plot the number of rooms on X-axis and number of houses on the Y-axis. The point obtained are joined by straight line giving a frequency polygon.

Frequency Polygon in continuous Series

There are two ways of making frequency polygon.

(i) With the help of Histogram

(ii) Without of Histogram

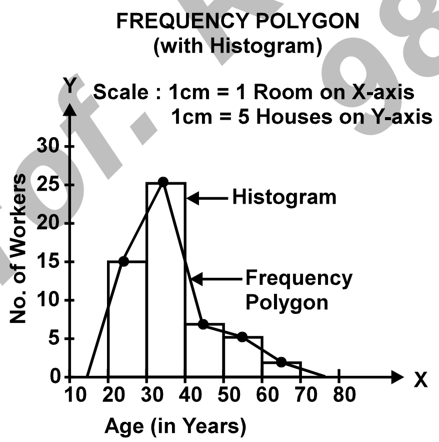
A frequency polygon is formed by joining mid-points of the tops of rectangle of a histogram and further joining these points on either side with the mid-points of the adjacent classes(class before the first class & class after the last class). The adjacent classes before first class and after last class have zero frequency on the base line. Under discrete series frequency polygon is made by joining the tops of the lines. Frequency polygon is constructed both for continous and discute series.

(i) With histogram: Firstly histogram is made on the basis of continuous series, after that mid-points of the top of the rectangles are joined and after that these points are joined with the mid-points, of the adjacent classes before the last and after the first class with zero frequencies.

Ex. The following is the age distribution of workers of a factory. On the basis of this information, construct a histogram and convrt into a frequency polygon.

Age (in Years)	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
No. of Workers	15	25	7	5	2

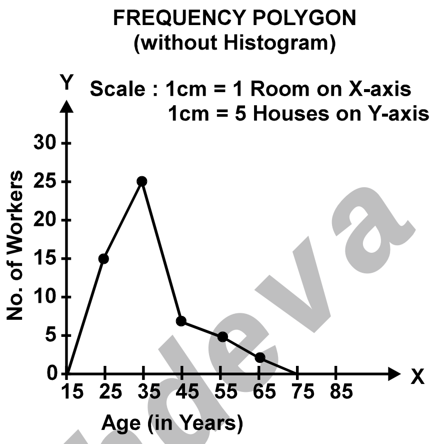
Sol. The frequency distribution is continuous with equal class-intervals. We will first prepare a histogram and then frequency polygon. Here, we plot the age on the X-axis and number of workers (frequency) on the Y-axis.



(ii) Without Histogram : In continuous series mid values are plotted on the bases of their frequencies, By joining these mid value points based on frequencies, polygon is formed. The first and the last mid value points are joined with the mid-points of adjacent classes on either sides with zero frequencies. According to this method while making frequency polygon histogram is not made.

Ex. Construct a frequency polygon, without making the histogram, from the data given in above example.

Sol. The mid-point of various class-intervals (age) are taken along the X-axis and frequency (number of workers) corresponding to each mid-points is shown in the Y-axis. The points, thus obtained are joined by straight lines to give a frequency polygon.



(d) Frequency Curve

It is the smoothened forms of frequency polygon. with a free hand, the angularities of a frequency polygon are eliminated in such a way so, that the area under curve remains the same. Both the sides of the curve are joined with X-axis on either sides.

Method of construction: Firstly, frequency polygon is made secondly, with a free hand a line couering frequency polygon is drawn in such a way so that its anguiarities are smooth ened.

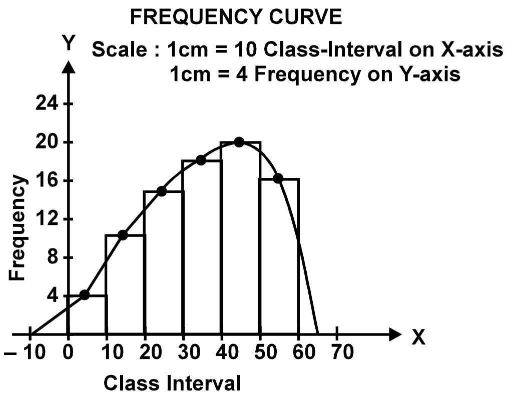
Ex. Draw a frequency curve from the following distribution

Class interval	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	4	10	15	18	20	16

Sol. The frequency distribution is continuous with equal class-intervals. We will first prepare a histogram and then frequency curve the clas-intervals are shown along the X-axis and frequency on the Y-axis.

(e) Cumulative Frequency Curve (Ogive):

Comulative frequency curve is also known as "Ogive". Cumulative frequency curve is the graphical representation of cumulative of frequency distribution. With the help of this curve frequencies of values of higher or lower than a given values can be ascertained.



Less than Ogive

In less than method, the frequencies of all preceding class-intervals are added to the frequency of a class. Then these less than cumulated frequencies are plotted against the upper class boundaries of the respective classes. The points so obtained are joined by a smooth free hand curve to give 'less than' ogive. Obviously, 'less than' ogive is an increasing curve, sloping upwards from left to right.

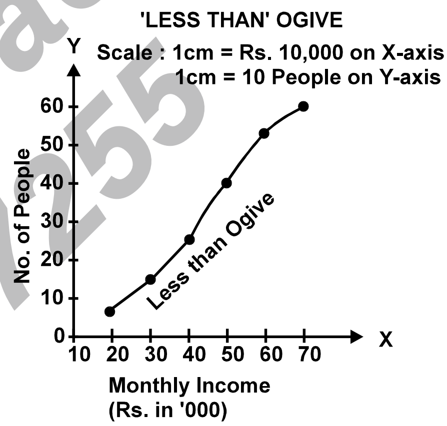
Ex1. From the following distribution of monthly income of 60 people in a company, draw a 'less than' ogive curve.

Monthly Income (Rs. in '000)	No. of People
10-20	6
20-30	9
30-40	10
40-50	15
50-60	12
60-70	8

Sol. For depicting the 'Less than' ogive curve, above frequency distribution has to be converted into 'less than' cumulative frequency in the following manner :

Commulative Frequency Distribution

Monthly Income (Rs in '000)	No. of People (c.f.)
Less than 20	6
Less than 30	15
Less than 40	25
Less than 50	40
Less than 60	52
Less than 70	60



For constructing 'less than' ogive, upper limit of the class-interval is shown along the X-axis and cumulative frequency of the respective class on the Y-axs. The cumulative frequency will be plotted against the upper limit of the class-interval. These points are joined with the help of free hand to get 'Less than' Ogive. See figure.

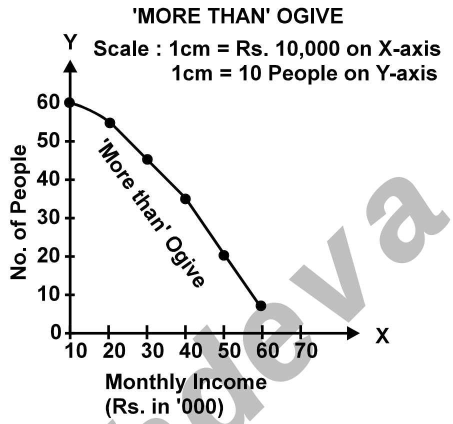
More than Ogive

In more than method, the frequencies of all succeeding class-intervals are added to the frequency of a class. In other words, we start with the lower limits of the classes and go on subtracting the frequencies of each class. The more than cumulative frequencies are plotted against the lower limit of corresponding class-intervals. The points so obtained are joined by a smooth free hand curve to give 'more than' ogive.

Ex.2 Construct a 'more than' ogive from the data given in **Ex.1**.

Sol. To construct the 'More than' ogive, the frequency distribution will be converted as shown in the following table

Monthly Income (Rs. in '000)	No. of People (c.f.)
More than 10	60
More than 20	54
More than 30	45
More than 40	35
More than 50	20
More than 60	8



IN the 'more than' ogive, X-axis represents the lower limit of the class-interval and Y-axis represents the cumulative frequency of the respective class in the descending order. The more than cumulative frequencies are plotted against the lower limit of the class-interval. These plotted points are joined with the help of smooth curves to get 'more than' ogive.

Both 'Less than' and 'More than' Ogives

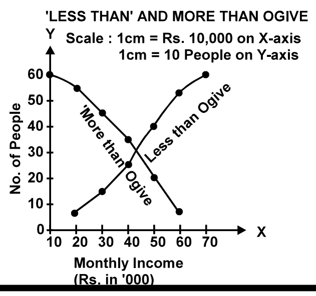
It is possible to draw both the 'less than' and 'more than' ogives on the same graph. The intersection of these two curves will give the value of 'Median'.

Ex.3 Draw 'Less than' and 'More than' ogive curves of the data given in **Ex.1**, on the same graph.

Sol. For depicting the 'Less than' and 'More than' ogive curves, frequency distribution has to be converted into 'less than' and 'more than' cumulative frequency in the following manner :

Cumulative Frequency Distribution

Monthly Income (Rs. in '000)	No. of People (c.f)	Monthly Income (Rs. in '000)	No. of People (c.f)
Less than 20	6	More than 10	60
Less than 30	15	More than 20	54
Less than 40	25	More than 30	45
Less than 50	40	More than 40	35
Less than 60	52	More than 50	20
Less than 70	60	More than 60	8



False Base Line

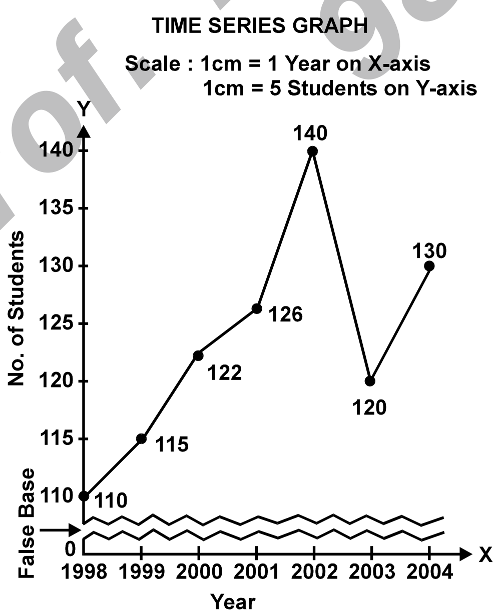
As discussed before, the scale of the Y-axis should begin from zero. However, if this rule is strictly followed, the curve would be very much pulled up away from the point of origin, if the gap between zero and smallest value of the variable is large. For example, if the variable starts from 2,000 and there are very small fluctuations in the remaining values of the variable, then a lot of space would be required to show the variable. In order to solve this difficulty, false base is used. Normally, in such cases, vertical line is broken in two parts and some blank space is left between them. This blank space is shown by zig-zag or linked line.

Ex. The following table shows the number of students, who have taken admission in a Central School in different years. Plot the data with a suitable graph

Year	1998	1999	2000	2001	2002	2003	2004
No. of Students	110	115	112	126	140	120	130

Sol. The datea of number of students starts with a lowest value of 110 students and touches a peak value of 140 students. In this case, we start with the lowest value of 110 after making a false base (as indicated in figure). This will make our presentation more meaningful. If we do not do this and start from a zero base instead, the presentation would require a lot of space and would become unwidely.

Taking years on the X-axis and number of students on the Y-axis, we get the following graph :



GRAPHS OF TIME SERIES

Time series can be shown on the graph paper. The information arranged over a period of time (e.g., years, months, weeks, days etc.) is termed as a time series.

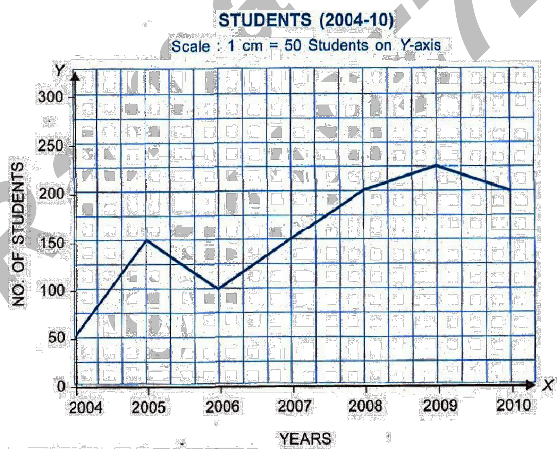
Presentation of this type of information by line or curve on the graph paper is of great use of economics statistics, These graphs are known as line graphs or histograms, or arithmetic line graph.

A. GENERAL RULES TO CONSTRUCT A LINE GRAPH

1. As the time (year, month, week) is never in negative (i.e., in minus figures), there is no need fusing Quadrant II and III.
2. Year, month or week according to the problem, is taken on X-axis. Give titles to X-axis and Y-axis.
3. Start Y-axis with zero and decide the scales for both the axes.
4. The pair values will give different dots on the graph paper. For example, values corresponding to time factor are :

Years	2004	2005	2006	2007	2008	2009	2010
Students	50	150	100	150	200	225	200

These dost obtained of pair values are joined by straight line which is called line graph or historigram.



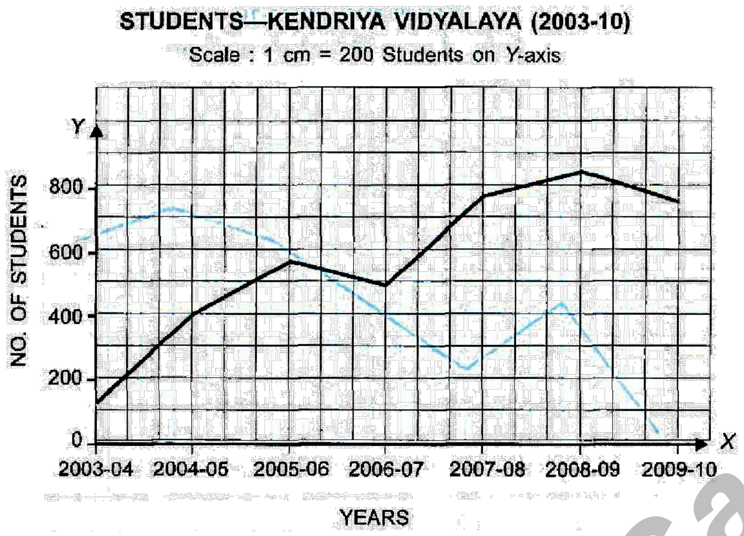
B. ONE VARIABLE GRAPH

Q. Present the following data on the graph paper :

Kendriya Vidyalaya

Year	Students
2003 – 04	120
2004 – 05	400
2005 – 06	567
2006 – 07	490
2007 – 08	760
2008 – 09	834
2009 – 10	750

Ans.



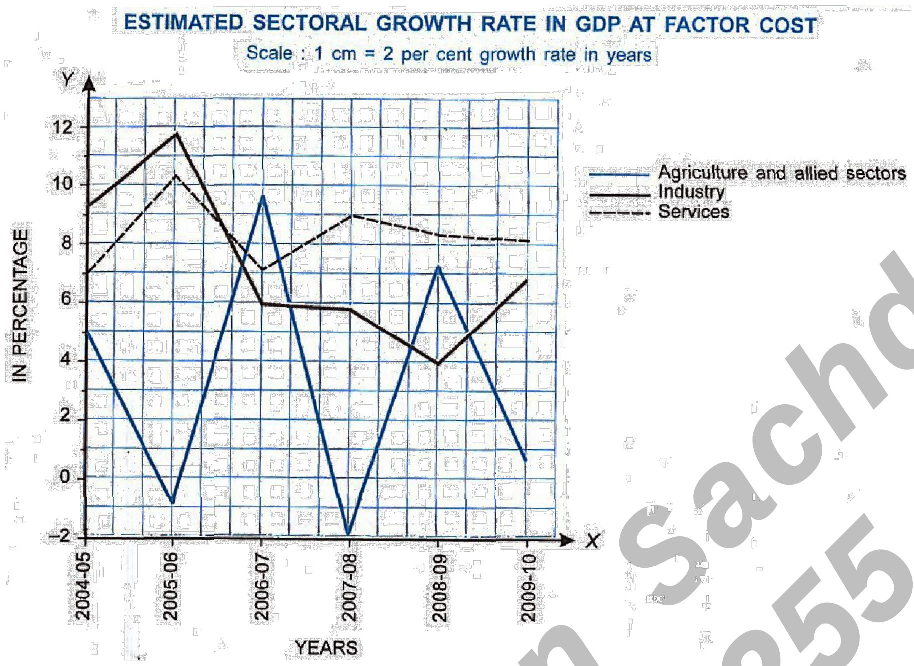
C. TWO OR MORE THAN TWO VARIABLE GRAPHS

Q. The following table shows the estimated sectoral real growth rates (percentage changes over the previous year) in GDP at factor cost.

Year (1)	Agriculture and allied sectors (2)	Industry (3)	Services (4)
2004 – 05	5.0	9.2	7.0
2005 – 06	–0.9	11.8	10.3
2006 – 07	9.6	6.0	7.1
2007 – 08	–1.9	5.9	9.0
2008 – 09	7.2	4.0	8.3
2009 – 10	0.8	6.9	8.2

Represent the data as multiple time series graph.

Ans.



D. GRAPHS OF DIFFERENT UNITS

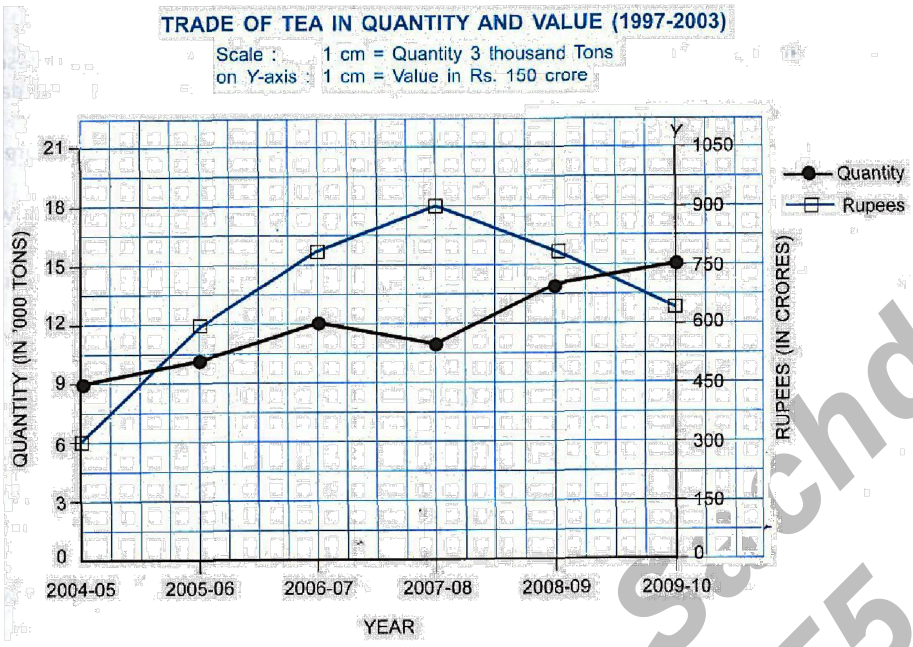
When two values are given into two different units, we will have two different scales. We can take one scale on left and the other on the right side of the graph paper.

Q. Present the following figures of Trade in Tea on graph.

Year	Quantity (in 000 tons)	Value (Rs. in crores)
2004 – 05	9	300
2005 – 06	10	596
2006 – 07	12	782
2007 – 08	11	900
2008 – 09	14	762
2009 – 10	15	640

Ans. Average of Quantity : 12 Approximately

Average of Value : 695 Approximately



Unsolved Praticals :

Q1. **Draw a line frequency graph of the following data :**

Marks	10	20	30	40	50	60	70
Frequency	3	7	9	11	12	14	15

Q2. **Represent the following data by a histogram :**

Marks in Statistics	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100
No. of Students	12	28	60	48	30

Q3. **The following data relates to the marks in Economic of 70 students. Depict it through Histogram.**

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of Students	7	10	24	18	11

Q4. **Represent the following frequency distribution by a histogram :**

Mid – Values	2.5	7.5	12.5	17.5	22.5
Frequency	5	10	30	15	6

Q5. **The frequency distribution of marks obtained by 60 students of a class in a college is given below :**

Marks	30 – 34	35 – 39	40 – 44	45 – 49	50 – 54	55 – 59	60 – 64
No. of Students	3	5	12	18	14	6	2

Q6. **Draw a frequency polygon for the following data :**

Items	4	5	6	7	8	9	10
Frequency	4	6	10	25	22	18	12

Q7. The frequency distribution of marks obtained by students in a class test is given below :

Marks (mid – points)	45	55	65	75	85
No. of Students	5	9	12	8	2

Q8. Depict the following frequency distribution with the help of frequency polygon

Mid – values	5	15	25	35	45	55	65	75
Frequency	4	10	16	22	25	12	7	2

Q9. Make a frequency curve of the following data :

Class – Interval	20 – 40	40 – 60	60 – 80	80 – 100	100 – 120	120 – 140	140 – 160
Frequency	3	7	11	15	13	6	7

Q10. The table given below shows the amount of sales of 100 companies :

Sales(Rs. in crores)	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
No. of Companies	7	12	15	30	22	14

Q11. Represent the following data relating to annual profits of a company with the help of suitable graph:

Year	2001	2002	2003	2004	2005	2006	2007
Profit (in Rs.crores)	25	37	45	35	50	54	60

Q12. Prepare a graph to represent the following data of imports and export of a commodity from 2001 to 2007 :

Year	2001	2002	2003	2004	2005	2006	2007
Imports (Rs. crores)	15	22	35	45	52	55	60
Exports (Rs. crores)	20	28	42	60	65	70	75

Q13. Make histogram and frequency polygon from the following distribution :

Class – Interval	0 – 20	20 – 30	30 – 40	40 – 60	60 – 100
Frequency	10	4	6	14	16

Q14. In a certain colony, 40 household were selected. The data on monthly income (Rs. '000) is given below:

20	12	35	55	40	14	35	8	18	11
60	11	35	50	45	20	16	7	15	70
18	65	62	16	21	19	24	11	19	35
13	25	43	15	30	40	25	55	22	35

(i) Make a frequency distribution taking a class interval of 10 marks. (Take first class interval as 0-10)

(ii) Draw a histrogram and a frequency polygon from the frequency distribution.

Q15. **Draw a 'Less than' and 'More than' ogive from the following distribution.**

Profits (Rs. in lakhs)	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
No. of Companies	4	7	10	20	17	2

Q16. **Name the different types of frequency distribution graphs :**

a. Line Frequency Graph (For discrete Series)

b. Histogram

c. Frequency Polygon

d. Frequency Curve or Smoothed Frequency Curve

e. Cumulative Frequency Curve or 'Ogive'

These are for continuous series.

Q17. **Define histogram ?**

Ans. Histogram is joining rectangular diagram of a continuous series in which each rectangular represents the class interval with frequency.

Q18. **How is Frequency Polygon drawn ?**

Ans. Frequency Polygon can be drawn by two ways :

(a) With histogram,

(b) Without histogram.

(a) After drawing suitable histogram, we get the mid-points of upper horizontal side of each rectangle. Joining these mid-points of adjacent rectangles of the histogram by straight line and ends of frequency polygon be extended to be base line at the mid-point of classes at both ends, the frequency polygon is obtained.

(b) Scale of X-axis can either be decided on the basis of class interval or mid-points of classes. Joining the points plotted for the mid-points corresponding to their frequencies by straight lines, we get frequency polygon, without histogram.

Q19. **How is Cumulative Frequency Curve (Ogive) is drawn ?**

Ans. Ogive can be obtained by calculated either less than cumulative frequencies or more than cumulative frequency. By plotting the various points on graph according to less than or more than value on graph paper and join them to get a curve, called Ogive.

Q20. What is histogram? Present the data given in the table below in the form of a Histogram:

Mid-points :	115	125	135	145	155	165	175	185	195
Frequency :	625	48	72	116	60	38	22	3	

Q21. In a certain colony a sample of 40 households was selected. The data on daily income for this sample are given as follows :

200	120	350	550	400	140	350	85
180	110	110	600	350	500	450	200
170	90	170	800	190	700	630	170
210	185	250	120	180	350	110	250
430	140	200	400	200	400	210	300

a. Construct a Histogram and a frequency polygon.

b. Show that the area under the polygon is equal to the area under the histogram.

(Hint. Get a frequency distribution table to obtain a continuous series).

Q22. Present,the data given in the table below in Histogram :

Marks	25 – 29	30 – 34	35 – 39	40 – 44	45 – 49	50 – 54	55 – 59
Frequency	4	5	23	31	10	8	5

Q23. A survey showed that the average daily expenditures (in rupees) of 30 households in a city were :

11, 12, 14, 16, 16, 17, 18, 18, 20, 20, 20, 21, 21, 22, 22, 23, 23, 24, 25, 25, 26, 27, 28, 28, 31, 32, 32, 33, 36, 36.

a. Prepare a frequency distribution using class intervals : 10-14, 15-19, 20-24, 25-29, 30-34 and 35-39.

b. What percent of the households spend more than Rs 29 each day ?

c. Draw a frequency histogram for the above data.

Q24. Draw a Histogram and Frequency Polygon of the following information.

Wages in Rupees	75 – 80	80 – 85	85 – 90	90 – 95	95 – 100	100 – 105	105 – 110	110 – 115
No. of workers	9	12	15	11	20	20	11	2

Q25. Draw ogive – (a) less than, and (b) more than of the following data :

Weekly wages of No. of workers	100 – 105	105 – 110	110 – 115	115 – 120	120 – 125	125 – 130	130 – 135
Workers (Rs.)	200	210	230	320	350	520	410

Q26. Prepare a less than ogive from the following data :

Class	0 – 6	6 – 12	12 – 18	18 – 24	24 – 30	30 – 36
Frequency	4	8	15	20	12	6

Q27. From the following frequency distribution prepare the 'less than' 'ogive'.

Capital (Rs. in laks)	0 – 10	10 – 20	20 – 30	39 – 40	40 – 50	50 – 60	60 – 70	70 – 80
No. of companies	2	3	7	11	15	7	2	3

Q28. Arrange the following information on a time-series graph :

Year	200 – 04	2004 – 05	2005 – 06	2006 – 07	2007 – 08	2008 – 09	2009 – 10
NDP (Rs. in 000 crores)	35	36	37	40	41	44	44

Q29. Present graphically the following sales of Delhi branch of USHA FANS.

Year	2002	2003	2004	2005	2006	2007	2008	2009	2010
Sales (Rs. in 000)	13	15	12	19	25	31	29	27	35

Q30. Prepare a graph showing total cost and total production of a scooter manufacturing company.

Year	2006	2007	2008	2009	2010
Production (in units)	8500	9990	11700	13300	15600
Total cost (Rs. in lakh)	24	29	34	45	49

CHAPTER – 8 – ORGANISATION OF DATA

CLASSIFICATION

The collected data (either by primary or secondary method) are always in an unorganised form in schedules or questionnaries. So it is necessary to make them available for comparision, analysis and appreciation by proper and suitable grouping and arrangement in condensed form. The process of grouping into different classes or sub-classes according to charactertistics is called ***classification***. The classified information arranged in a logical and systematic order in a particular sequence is called ***seriation*** or ***statistical series***. The classified information presented in precise and systematic tables is called ***tabulation***.

In other words, classification is for division of data, ***seriation*** is for arrangement of data in a systematic order and ***tabulation*** is for presentation of data in a table.

OBJECTIVES OF CLASSIFICATION

1. **To present the facts in a simple form.**
   Classification process eliminates unnecessary details and makes the mass ofcomplex data, simple, brief, logical and understandable.

2. **To bring out clearly points of similarity and dissimilarity.**
   So that they can be easily grasped. Facts having similar characteristics are placed in a class, such as educated, in educated employed, unemployed etc.

3. **To facilitate comparision.**
   This is not possible in an un-organised and unclassified data.

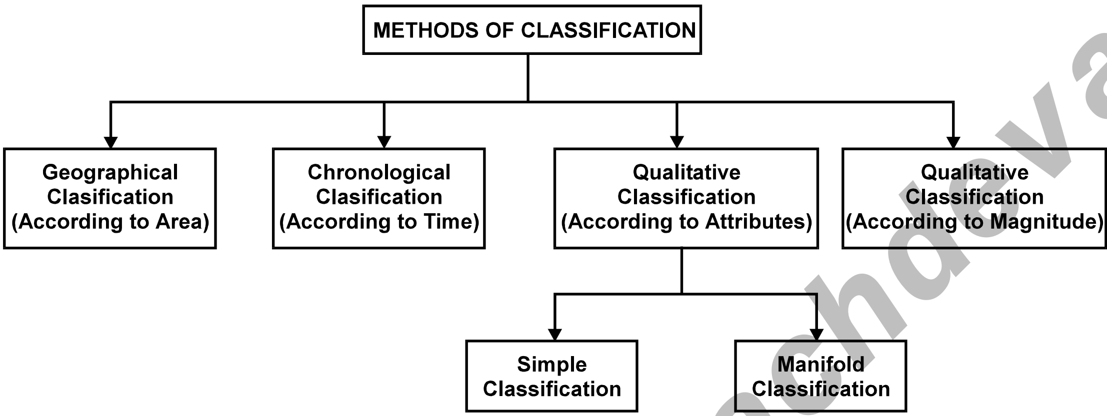
4. **To bring out cause and effect relationship.**
   For example, data of small-pox patients can help in finding out whether small pox cases occured more on vaccinated or unvaccinated population.

5. **To present a mental picture.**
   Summarisied data can easily be understood and remembered.

6. **To prepare the basis for tabulation.**
   Classification prepared the basis for tabulation and statistical analysis of the data. Unclassified data cannot be presented in tables.

METHODS OF CLASSIFICATION

Statistical data can be classified according to their characteristics. It can be grouped on the following basis :



1. Chronological Classification

When data are classified on the basis of time, it is called chronological Classification. Eg- time series showing population, production etc. It is classified normally in ascending order of time but in certain cases one can decide even by decending order time.

Population of Delhi (1951-2001)

Year	1951	1961	1971	1981	1991	2001
Population (in '000)	1,744	2,659	4,066	6,220	9,421	13,851

2. Geographical Classification

When data are classified on the basis or location of region, it is called geographical classification. Eg. Production of food grains state wise, population etc.

Population of 5 States as per Census 2001

City	Andhra Pradesh	Tamilnadu	Rajasthan	Karnataka	Gujrat
Population (in '000)	76,210	62,406	56,507	52,851	50,671

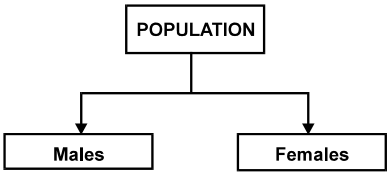
3. Qualitative Classification

Data may be classified on the basis of their qualitative differences or attributes which cannot be measured. Such as difference in religion, intellegence, beauty etc.

It may be of two types.-

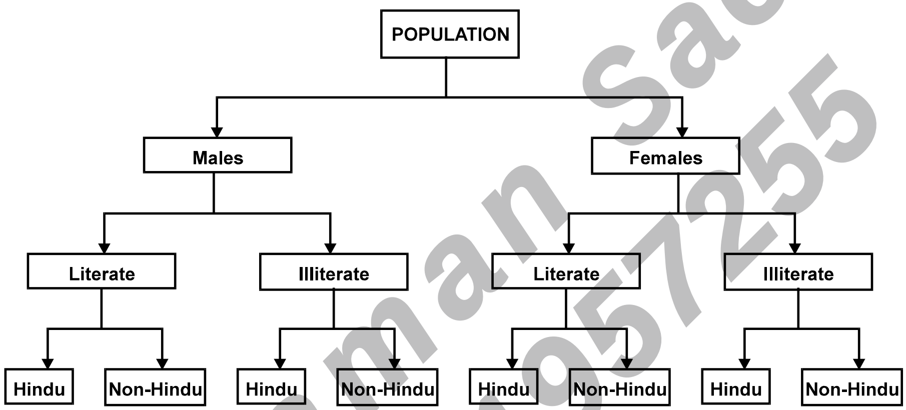
a. A simple Classification

A simple classification is one where data are classified on the basis of presence or absence of an attribute such as married & unmarried. This is also known as two fold or dischotomous classification.



b. *Manifold Classification*

Where data are classified on the basis of more than one attribute it is called Manifold Classification. Eg : Students of a college may be first classified as girls & boys and further they may be classified as good as bad in studies. Attributes should be very clearly defined to avoid confusion in classification.



4. Quantitative Classification

The collected data is grouped with reference to characteristic which can be measured and numerically described such as height, weight, sales, import, age, income, etc. This type of classification is made in the formation of statistical series, eg,

Income Group (Rs.)	100 – 199	200 – 299	300 – 399	400 – 499	500 – 599
No. of workers	15	27	58	72	40

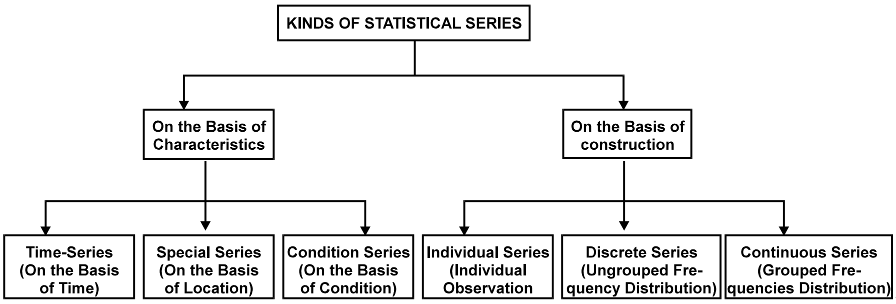
Thus, there are 15 workers in the income group of Rs. 100 to 199; 27 workers in income group of Rs. 200-299 and so on.

STATISTICAL SERIES

DEFINITION

Series or seriation is a logical and systematic arrangement of items into a particualr order of sequence in different classified categories while classification is concerned with the division of the data into various groups. Thus, statistical series are prepared to present the collected and classified data in a properly arranged way.

Statistical series can be classified in the following way:



SERIES ON THE BASIS OF GENERAL CHARACTER

1. Time Series

A series of values of some variable according to successive points in time is called time series. Data are presented with reference to some time unit, viz, year, month, week, or day.

Example:

Sugar Production of a Factory

Year	***Production (in '000 tons)***
1994	78
1995	75
1996	94
1997	86
1998	89
1999	92
2000	95

Sale in Supar Bazar (1st week of Jan.2000)

Day	***Sale*** (Rs.)
Mon.	1,892
Tues.	2,757
Wednes.	3,090
Thurs.	2,650
Fri.	2,592
Satur.	3,822

2. Spatial series

A series of values of some variable according to geographical division of the universe under study is called a spatial series or geographical series. Data are presented with reference to some geographical division, viz; country, sate, city, town, village or colony.

Example:

Per Capita Income

Country	**Per Capita Income (Rs.)**
U.S.A	5,100
France	3,900
Japan	2,800
Canada	2,100
India	500

Number of Schools

City	**No. of Schools**
Delhi	792
Mumbai	649
Chennai	573
Calcutta	532
Bangalore	459

3. **Condition Series**

A series of values of some variable made according to a condition is called condition series. Data are presented with reference to some condition, viz, height, age, weight, income etc.

Example:

Income of 100 Families

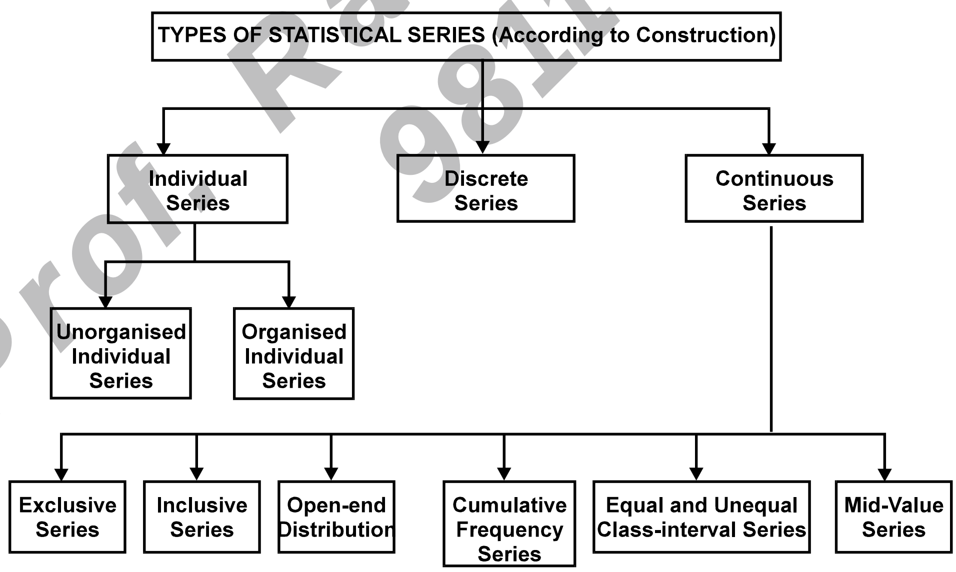
Income(Rs.)	***No. of Families***
500-999	35
1000-1499	25
1500-1999	15
2000-2499	20
2500-2999	5

Series on the basis of construnction

The data should be arranged in order for further study of presentation, analysis and interpretation. This arrangement can be done in three ways:

a. Series of Individual Observation
b. Discrete Series
c. Continuous Series

b and c are with reference to frequency distribution.



Series of Individual Observation

Mass data in its original form is colled **raw data** or **unorganised data** which can be arranged in any of the following ways:

i. Serial order of alphabetical order
ii. Ascending order

iii. Descending order

The mass data when put in ascending or descending order of magnitude is called an **array**. A series of individual observation is a series where items are listed singly afte collection. They are not listed in groups.

FREQUENCY DISTRIBUTION

Two types of distributions are grouped under frequency distribution which are:

a. Discrete Frequency Distribution

b. Continuous Frequency Distribution.

Set I. Children in Families

Children	***No. of families(f)***
0	25
1	45
2	37
3	15
4	8
Total	130

Set II. Height of Students

Height in Inches	***No. of Students(f)***
56-58	12
58-60	16
60-62	15
62-64	4
64-66	10
Total	57

Series is a systematic arrangement of items into a particualr order or sequence in different classified categories, such as Set I. for which children in families and set II for height of students.

Frequency The number of times given value in a **observation** appears is the frequency For example, in the above sets there are 25 families having no child and 8 families having four children and 10 students in the group of 64" to 66" and 16 students in group of 58" to 60" etc.

Class Frequency. The number of values in each of the quantitative **classes** is called the class frequency , e.g. out of the five classes of set II students in a group of 58" to 60" are 16. So the class frequency of the class 58" to 60" is 16. There is no instance of a class in set I.

Class It is a decided group of magnitudes, e.g. 100-200, 10-19, 4-8, 7-13 etc.

Upper and lower limits of the classes.The lowest and the highest magnitudes, which form the boundaries of a class, are known as the upper and lower limits, respectively., for example, for a class of 62-64, 62 is lower limit and 64 is upper limit. Left hand side magnitudes are the lower limits and right hand side magnitudes are the upper limits of their respective classes.

Class Interval The magnitude spread between the lower and upper class limits is called class interval. It is the span or width of a class which can be obtained by finding the difference between the upper and lower limits of the class.

Mid-Point The mid-value which lies half way between the lower and the upper class limits is known as mid-point.

$$\text{Mid Point} = \frac{\text{Upper Class Limit} + \text{Lower Class Limit}}{2} = \frac{\ell_2 - \ell_1}{2}$$

Variable A quantity which varies from one individual to another is known as a **variable** or **variate**. Quantitative characteristic such as income, height, weight, number of units sold etc. are variables. A variable may be either discrete or continuous.

Discrete and Continuous Variables

Discrete and Discontinuous Variables are those which are exact or finite and are not normally fractions. For example, children in a family can be either 2 or 3, but cannot be 2.2, 2.8, or 2.7. It is a discrete variable which is not expressed in a fraction.The occurence of the observation will be integers, i.e. 1, 2, 3 4, 5, 6 ....and so on.

Continuous Variables are those that one in units of measurement which can be broken down into infinite gradations. They fall in any numerical value with in a certain range. All the fractional values are continuous variables.

Discrete Series. Any series represented by discrete variable is called a discrete series.

Continuous Series.Any series described by continuous variables is called continuous series.

It is to be noted that a discrete variable series can be presented in a continuous type of series also, but continuous variables cannot be presented in a discrete series. When ever the range of values in a discrete series is too wide, one can have the choice of a continuous frequency distribution.

CONSTRUCTION OF DISCRETE FREQUENCY DISTRIBUTION

Illustration 1. Prepare frequency table of ages of 25 students of XI class in your school.

15, 16, 16, 17, 18, 18, 17, 15, 15, 16, 16, 17, 15, 16, 16, 15, 16, 16, 15, 17, 17, 18, 19, 16, 15

Solution.

Frequency Distribution of ages of 25 students.

Age	***Tally bars***	***Total***
15	1111 11	7
16	1111 1111	9
17	1111	5
18	111	3
19	1	1
	Total	25

Illustration 2. Prepare a frequency distribution table of 25 families occupying no. of rooms from the following information.

No. of rooms in the Houses Kept by 25 Families

1	2	4	3	4
2	5	3	2	2
4	1	2	3	5
1	3	5	1	3
3	1	3	1	1

Construction of Continuous Frequency Distribution

Observations are divided in to groups having class intervals. There are two methods of classifying the data according to class intervals.

a. Inclusive Method and

b. Exclusive Method

a. Inclusive Method

Under this method upper class limits of classes are included in respective classes.

b. Exclusive Method

Under this method upper limits are excluded. The upper limits of class interval is the lower limit of the next class. Some times lower limits are excluded from their respective classes.

The Classes of type 10-20, 20-30, 30-40, etc. are known as exclusive classes. In this method, upper limit of one class-interval becomes the lower limit of the next class.

For example, if the lower limit of a class is 10 and its upper limit is 20, then this class, written as 10-10 and its upper limit is 20, then this class, written as 10-20, includes all the observations which are greater than or equal to 10 but less than 20. The observations with magnitude 20 will be included in the next class.

(b) By Exclusive method :

(i) Lower limit excluded :

Marks : 5-10, 10-15, 15-20, 20-25, 25-30

Lower limits 5, 10, 15, 20, 25, 30 of their respective groups are excluded.

(ii) Upper limit excluded :

Marks : 5-10, 10-15, 15-20, 20-25, 25-30

Upper limits 10, 15, 20, 25, 30 of their respective groups are excluded.

However, if the class intervals are given as 5-10, 10-15, 15-20, 20-25, etc., it is always presumed that upper limits are excluded in absence of any specific instructions.

(c) By mentioning lower limits (followed by a dash) :

Marks : 5-, 10- 15-, 20-, 25-, These are to be read as 5-10, 10-15,15-20,20-25 and 25-30.

(d) By mentioning upper limits (preceded by a dash) :

Marks : -10, -15, -20, -25, -30. These are to be read as 5-10, 10-15, 15-20, 20-25 25-30.

(e) By mid-points of class interval :

Marks : 7.5 12.5 17.5 22.5 27.5

These mid-points are required to be converted into class intervals. Say for first midpoint (12.5-7.5) and divide the difference by 2, i.e., (5/2). The quotient is added and subtracted to first mid-point we get, (7.5-2.5 = 5) and (7.5 + 2.5 = 10). We get thus the class interval 5-10. In the same way intervals of all the mid-points can be obtained, i.e., 10-15, 15-20, 20-25, 25-30.

(f) 'Open-end' class intervals

In certain frequency distributions 'open-end' class intervals are given as we find in the example given below :

Marks Below	10	10-15	15-20	20-25	25-30	30-35	35 and above	Total
Frequency (f)	7	10	13	18	8	5	3	64

In such cases, values are put on the basis of construction of series. In the above series '5' in place of 'below' and '40' in place of 'above' may be put. Thus makingthe classes as :

Marks	0-10	10-15	15-20	20-25	25-30	30-35	35-40

Principles of Grouping

There is no hard and fast rule for grouping the data, but following general principles may be kept in mind for satisfactory and meaningful classification of data:

a. It is advisable to have total number of classes between 5 and 15. The preference for the total number of classes depends on the numbers and figures to be grouped, the magnitude of the figure and possibility of simplified calculations of further statistical studies.

b. Odd figures, for example, 3, 7, 9, 11, 27, 33 etc. should be avoided for class intervals. The choice for the class intervals should be either 5 or a multiple of 5. It simplifies our further statistical calculations.

c. Lower limit of the class as far as possible, should be 0 or a multiple of 5.

d. For maintaining continuity and correct classes exclusive method of preparing classes is adopted.

e. The class interval should be equal for all classes.

f. As far as possible open-end classes should be avoided. For example,

Marks
Below 5
5-10
10-15
15-20
Above 20

The first and the last classes are open-end classes; the first is open at the lower-end and last at the upper-end. For statistical calculations the open-ends should be closed. Maintaining the regularity of the class intervals we can close these groups as 0-5 and 20-25.

g. For frequency distribution, we prepare a table having three columns first for variables, second for 'Tally bars" and the third for the total representing corresponding frequency to each class.

Simple Series and Cumulative Series : We have seen in the above illustrations the patterns of simple series of discrete type and continuous type (using inclusive and exclusive methods of class intervals). In simple series the frequency is shown against each value or class, in cumulative series the frequencies are progressively totalled. See the following illustration :

SIMPLE SERIES

Discrete Type Marks	Discrete Type No. of students	Continuous Type Marks	Continuous No. of students
10	4	0 – 10	4
20	8	10 – 20	8
30	15	20 – 30	15
40	20	30 – 40	20
50	13	40 – 50	13

CUMULATIVE SERIES

Less Than Marks	Less Than No. of students (c.f.)	More Than Marks	More Than No. of students (c.f.)
Less than 10	4	More than 0	60
Less than 20	12 (4 + 8)	More than 10	56 (60 – 4)
Less than 30	27 (4 + 8 + 15)	More than 20	48 (60 – 12)
Less than 40	47 (4 + 8 + 15 + 20)	More than 30	33 (60 – 27)
Less than 50	60 (4 + 8 + 15 + 20 + 13)	More than 40	13 (60 – 47)

Q. Following is the data of weekly pocket money (in Rs) of teenagers

102	125	76	111	71	121	154	81	129	72
91	118	175	62	144	91	107	64	128	82
83	74	135	149	119	125	100	84	120	115

(i) Prepare a frequency table for the given data taking of each class-interval as 20 and mid-value of the first class as 70. Also calculate mid-values of each class.

(ii) Find the number of teenagers, with weekly pocket money :

(a) More than Rs. 140

(b) Less than Rs. 100

(iii) How much percent of teenagers have weekly pocket money between Rs 80 to Rs. 120.

(iv) What is the maximum and minimum amount of weekly pocket money ?

Sol. (i) Given : Width of a class is 20 and mid-point value of first class is 70.

Mid-value of each class is $\frac{20}{2} = 10$ units away from the class limits.

Hence, lower limit of the first class = 70 – 10 = 60 and upper limit is 70 + 10 = 80. Similarly, the other classes are 80-100, 100-120 etc.

Frequency Distribution showing Weekly Pocket Money of 30 Teenagers

Weekly Pocket Money (Rs.)	Mid – Value	Tally Marks	Frequency
60 – 80	$\frac{60+80}{2} = 70$	|||||	6
80 – 100	$\frac{80+100}{2} = 90$	|||||	6
100 – 120	$\frac{100+120}{2} = 110$	||||||	7
120 – 140	$\frac{120+140}{2} = 130$	||||||	7
140 – 160	$\frac{140+160}{2} = 150$	|||	3
160 – 180	$\frac{160+180}{2} = 170$	|	1
		Total Teenagers	30

(ii) (a) Number of teenagers with weekly pocket money more than Rs. 140 = 3 + 1 = 4

(b) Number of teenagers with weekly pocket money less than Rs. 100 = 6 + 6 = 12

(iii) Percent of teenagers with weekly pocket money between Rs. 80 to Rs. 120.

Number of teenagers with weekly pocket money between Rs. 80 to Rs. 120 = 6 + 7 = 13

Total number of teenagers = 30

Percent of teenagers with weekly pocket money (80-120 = $\frac{13}{30} \times 100 = 43.33\%$

(iv) Maximum amount of weekly pocket money = Rs. 175

Minimum amount of weekly pocket money = Rs. 62.

Q. The following data relates to monthly household expenditure (in Rs.) on food of 50 households

1904	1559	3473	1735	2760	2041	1612	1753	1855	4439
5090	1085	1823	2346	1523	1211	1360	1110	2152	1183
1218	1315	1105	2628	2712	4248	1812	1264	1183	1171
1007	1180	1953	1137	2048	2025	1583	1324	2621	3676
1397	1832	1962	2177	2575	1293	1365	1146	3222	1396

On the basis of given data, answe rthe following questions :

(i) Obtain the range of monthly household expenditure on food.

(ii) Divide the range into appropriate number of class-intervals and obtain the frequency distribution of expenditure.

(iii) Find the number of households whose monthly expenditure on food is

(a) Less than Rs. 2000

(b) More than Rs. 3000

(c) Between Rs. 1500 and Rs. 2500

Sol. (i) Range = Largest value – Smallest value

= 5,090 – 1,007 = Rs. 4,083

(ii) The number of observations are 50 with smallest value as 1,007 and the largest is 5,090. Taking the class-interval as 500, the number of classes can be determined in the following manner :

$$\text{Width of Class – Interval} = \frac{\text{Largest Observation} - \text{Smallest Observation}}{\text{Number of Classes Desired}}$$

$$500 = \frac{5,090 - 1,007}{\text{Number of Classes Desired}}$$

$$\text{Number of Classes Desired} = \frac{4,083}{500} = 8.166 = 9$$

With the help of tally marks, we get the following frequency distribution by

Exclusive Method :

Monthly Expenditure (in Rs.)	Tally Marks	Frequency
1000 – 1500	|||| |||| |||| ||||	20
1500 – 2000	|||||||| |||	13
2000 – 2500	|||||	6
2500 – 3000	||||	5
3000 – 3500	||	2
3500 – 4000	|	1
4000 – 4500	||	2
4500 – 5000	–	0
5000 – 5500	|	1
	Total Households	50

(iii) (a) Number of households with monthly expenditure on food less than Rs. 2,000.

= 20 + 13 = 33

(b) Number of households with monlty expenditure more than Rs. 3,000.

= 2 + 1 + 2 + 0 + 1 = 6

(c) Number of households with monthly expenditure between Rs. 1,500 and Rs. 2,500.

= 13 + 6 = 19

Inclusive Series

The classes of the type 10-19, 20-29, 30-39 etc, are known as inclusive classes. Under inclusive series, all observations with magnitude greater than or equal to the lower limit and less than or equal to the upper limit of a class are included in it. Thus, uder this series, overlapping of intervals is avoided. In this series, the value of upper limit of a class never equals the value of lower limit of the next class.

Refers the following examples

Classes	Frequency
10 – 19	6
20 – 29	5
30 – 39	9
40 – 49	10
Total	30

In the given example, a student getting 29 marks is included in 20-29 class-interval and similarly a students getting 39 marks is included in 30-39 class intervals.

Q. Marks obtained by 50 boys of a class are as follows :

33	52	10	25	53	55	11	42	47	34	47
21	39	25	32	19	10	16	41	46	36	15
30	27	59	30	14	6	17	40	49	52	54
40	31	45	24	30	34	52	36	12	16	19
40	42	3	6	44	26	45				

(i) Construct a frequency table with class intervals 0-9, 10-19, 20-29 and so on.

(ii) Find the number of student securing marks

(a) More than Rs. 30

(b) Less thna Rs. 19

(iii) How much percent of students have secured marks between 20 and 49.

Sol. (i) Frequency Distribution showing Marks 50 Students (Inclusive Method)

Marks (Classes)	Tally Marks	Frequency
0 – 0	|||	3
10 – 19	|||||||||||	11
20 – 29	|||||	5
30 – 39	|||||||||||	11
40 – 49	||||||||||||	12
50 – 59	||||||||	8
	Total Students	50

(ii) (a) Number of students with marks more than 30 = 11 + 12 + 8 = 31

(b) Number of students with marks less than 19 = 11 + 3 = 14

(iii) Percent of students scoring marks between 20 and 49.

Number of students scoring marks between 20 and 49 = 5 + 11 + 12 = 28

Total number of students = 50

Percent of students scoring marks between 20 and 49 = $\frac{28}{50} \times 100$ = 56%

Q. The following data shows the number of computers sold by a company during a six-week period.

22	65	65	67	55	50	65	77	73	30	62
54	58	65	79	60	63	45	51	68	79	83
33	41	49	28	55	61	65	75	55	75	39
87	45	50	66	65	59	25	35	53		

Group the data into a frequency distribution by inclusive method.

Sol. Since the number of observations are 42, it seems reasonable to choose 7 classes. Again, since the smallest value is 22 and the largest is 87, therefore width of the class-interval is given by :

$$\text{Width of Class – Interval} = \frac{\text{Largest Observation} - \text{Smallest Observation}}{\text{Number of Classes Desired}} =$$

$$\frac{87-22}{7} = 9.28 \simeq 9$$

Now, counting the number of values in each with the help of tally marks, we get the following frequency distribution :

Frequency Distribution

Class – Interval (Number of Computers)	Tally Marks	Frequency (Number of Daya)
20 – 29	|||	3
30 – 39	||||	4
40 – 49	||||	5
50 – 59	||||||||	9
60 – 69	|||| |||| |||	13
70 – 79	|||| |	6
80 – 89	||	2
	Total Students	42

Conversion of Inclusive Series into Exclusive Series

Sometimes, it becomes necessary to have exclusive classes to apply some statistical tools. Then, if the given classes are inclusive type, they are to be transformed into exclusive classes. Inclusive series may be converted into exclusive series by applying the 'Correction Factor' to the lower and upper limits.

$$\text{Correction Factor (d)} \quad \frac{\text{Upper limit of a Class} - \text{Lower limit of the next higher Class}}{2}$$

Now, correction factor will be subtracted from the lower limits of all the classes and added to the upper limits of all the classes.

The upper and lower class limits of the new 'exclusive type' classes as called class boundaries.

Q. Convert the following inclusive series into exclusive series.

Marks	No. of Students (f)
10 – 19	4
20 – 29	2
30 – 39	12
40 – 49	10
50 – 59	9
60 – 69	3
Total	40

Sol. In the given example, Correction Factor (d) = $\frac{20-19}{2} = \frac{30-29}{2} = \frac{40-39}{2} = 0.5$.

So, we will subtract 0.5 from the lower limits of all the classes and add 0.5 to the upper limits. The adjusted classes would then be as shown in the following table :

Exclusive Series

Marks	No. of Students (f)
9.5 – 19.5	4
19.5 – 29.5	2
29.5 – 39.5	12
39.5 – 49.5	10
49.5 – 59.5	9
59.5 – 69.5	3
Total	40

Q. Following marks where obtained by 25 students of a class in Mathematics paper carrying 50 marks

19	13	12	25	32	12	31	19	21	23	27	41	29
30	45	39	33	40	17	11	20	26	14	41	15	

(i) Construct a frequency distribution when class intervals are inclusive, taking the lowest class as 10-19. Also construct class boundaries.

(ii) Construct a frequency distribution when class intervals are exclusive,taking the lowest class as 10-20.

Sol. (i) **Frequency Distributition by Inclusive Method**

Marks	Tally Marks	Frequency	Class Boundaries
10 – 19	| | | | | | | | |	9	9.5 – 19.5
20 – 29	| | | | | | |	7	19.5 – 29.5
30 – 39	| | | |	5	29.5 – 39.5
40 – 49	| | | |	4	39.5 – 49.5
	Total Students	25	

(ii) **Frequency Distribution by Exclusive Method**

Marks	Tally Marks	Frequency
10 – 20	| | | | | | | |	9
20 – 30	| | | | | | |	7
30 – 40	| | | |	5
40 – 50	| | | |	4
	Total Students	25

Difference between Exclusive Method and Inclusive Method

S.No.	Exclusive Method	Inclusive Method
1.	The upper limit of a class-interval is counted in the next immediate class.	Both the limits of a class-intervals is counted in the same class.
2.	The upper limit of a class-interval and the lower limit of next class are the same.	The upper limit of a class-interval and lower limit of next class are different. The difference is generally of one.
3.	There is no need of converting it to inclusive method prior to calculation.	For simplicity in calculation, it is necessary to change it into exclusive method.

Open-End Distribution

In a frequency distribution, if the lower limit of the first class and the upper limit of last class are not given, it is known as open-end distribution. In this series, in place of lower limit of first class, word 'below' or 'less than' is written and in the last class, in place of upper limit, word 'over' or 'more than' is written.

Example of Open-End Series

Marks	Below 20	20 – 30	30 – 40	40 – 50	Above 50
No. of Students	7	6	12	15	18

Conservation of Open-end Series into a Continuous Series

Marks	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. of Students	7	6	12	15	18

Q. From the following data, Calculate the lower limit of the first class and upper limit of the last class.

Class Interval	Frequency
Less than 10	12
10 – 20	15
20 – 30	4
30 – 40	16
40 – 50	3
Total	50

Sol. In the given example, the class-intervals of 2nd, 3rd and 4th classes are uniform, i.e. 10. We can therefore, assume that class-intervals of open-end classes are also equal to 10. It means, the lower limit of the first class-interval is 10 - 10 = 0 and the upper limit of the last class is 40 + 10 = 50.

So, Lower limit of first class = 0 and Upper limit of last class = 50.

Illustrations 3. Form a frequency distribution from the following data by inclusive method taking 4 as the magnitude of class intervals.

31	23	19	29	22	20	16	10	13	34
38	33	28	21	15	18	36	24	18	15
12	30	27	23	20	17	14	32	26	25
18	29	24	19	16	11	22	15	17	10

Bivariate Frequency Distribution

So far, our study was confined to frequency distribution of a single variable. Such frequency distributions are also known as univariate frequency distribution. This have to two characteristidcs. When the data are classified on the basis of two variables such as height and weight, marks in statistics and economics etc., the distribution is known as Bivariate frequency distribution or Two-way frequency distribution.

Q. The following data represent the marks in English (X) and Maths (Y) of 25 students. Prepare a two-way frequency distribution with class-intervals of 15 - 25, 25 - 35, 35 - 45 and so on.

Marks in English (X)	35	28	30	18	40	45	27	36	52	38	40	38	35
Marks in Maths (Y)	56	42	40	30	60	62	18	52	74	60	60	56	52

Marks in English (X)	35	70	17	42	46	35	35	27	32	41	46	58
Marks in Maths (Y)	48	80	18	58	65	48	50	38	46	68	74	88

Sol. Steps for construction of bivariate frequency distribution :

1. Write one of the variables on the left-hand side of the table and the other at the top.
2. The first student gets 35 in English and 56 in Maths. A tally mark has to be put in the cell where the column showing 35-45 marks in English Inter-sects the row showing 55-65 marks in Maths.
3. Repeat the procedure for all the 25 students.
4. Total the tallies at the bottom and to the right side.
5. Totals at the right at the extreme column are for Maths and those at the bottom row are for English.

Marks in English(X) → Marks in Maths (Y)↓	15 – 25	25 – 35	35 – 45	45 – 55	55 – 65	65 – 67	Total
15 – 25	I (1)	I (1)	–	–	–	–	2
25 – 35	I (1)	–	–	–	–	–	1
35 – 45	–	III (3)	–	–	–	–	3
45 – 55	–	I (1)	IIII (5)	–	–	–	6
55 – 65	–	–	IIIII (6)	I (1)	–	–	7
65 – 75	–	–	I (1)	III (3)	–	–	4
75 – 85	–	–	–	–	–	I (1)	1
85 – 95	–	–	–	–	I (1)	–	1
Total	2	5	12	4	1	1	25

Note : Figure within brackets denote the frequency corresponding to each cell.

Unsolved Practicals

Q1. **Following are the figures of marks obtained by 39 students. You are required to arrange them in ascending and in descending order,**

15	18	16	14	10	6	5	3	8	7	22
18	14	19	17	8	6	4	10	3	12	16
15	13	11	10	18	22	14	19	11	18	22
14	25	17	8	9	10					

Q2. **Heights (in inches) of 35 students of a class in given below. Classify the following data in a discrete frequency sereis**

58	60	72	68	58	55	58	72	55	68	66
60	60	55	72	66	58	72	68	60	55	68
55	58	62	66	72	58	60	68	55	58	65
72	68									

Q3. **The following are the marks of the 30 students in Statistics. Prepare a frequency distribution taking the class-intervals as 10-20, 20-30 and so on.**

12	33	23	25	18	35	37	49	54	51	37
15	33	42	45	47	55	69	65	63	46	29
18	37	46	59	29	35	45	27			

Q4. **Prepare a frequency table taking class intervals 20-24, 25-29, 30-34 and so on, from the following data:**

21	20	55	39	48	46	36	54	42	30	29
42	32	40	34	31	35	37	52	44	39	45
37	33	51	53	52	46	43	47	41	26	52
48	25	34	37	33	36	27	54	36	41	33
23	39	28	44	45	39					

Q5. **From the following data, calculate the lower limit of the first class and upper limit of the last class.**

Daily Wages	Less than 120	120 – 140	140 – 160	160 – 180	Above 180
No. of Workers	35	12	10	40	13

Q6. **Convert the following 'more than' cumulative frequency distribution inot a 'less than' cumulative frequency distribution.**

Class-Interval (More than)	10	20	30	40	50	60	70	80
Frequency	124	119	107	84	55	31	12	2

Q7. **Prepare a frequency distribution from the following data :**

Mid – points	25	35	45	55	65	75
Frequency	6	10	9	12	6	5

Q8. **You are given below a mid-value series, covert into a continuous series**

Mid – Value	15	25	35	45	55
Frequency	8	12	15	10	5

Q9. **From the following data of the ages of different persons, prepare less than and more than cumulative frequency distribution.**

Age (in years)	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
No. of Persons	5	12	10	6	4	11	2

Q10. **The ages of 20 husbands and wives are given below. Form a two-way fre-**

quency table showing the relationship between the ages of husbands and wives with the class-intervals 20-25, 25-30, etc.

S.No.	Age of husband	Age of wife	S.No.	Age of husband	Age of wife
1	28	23	11	27	24
2	37	30	12	39	34
3	42	40	13	23	20
4	25	26	14	33	31
5	29	25	15	36	29
6	47	31	16	32	35
7	37	35	17	22	23
8	35	25	18	29	27
9	23	21	19	38	34
10	41	38	20	48	47

Q11. **What do you mean by classification of data? Name the methods of classification.**

Ans. Classification is grouping of data according to their identity, similarity or resemblances. We can classify the data (a) Chronological, [b) Geographical, (c) Qualitative, and (d) Quantitative.

Q12. **Define an individual series.**

Ans. In individual series, the observations are all put into one column or are kept as they are collected.

For example, figures of population of India in different census years starting from 1901 to 2001, Marks of 5 students in row represent individual series.

Q13. **What do you mean by a discrete series? What does the frequency indicate in these series?**

Ans. In discrete series, data is presented in two columnsthe first one showing the exact value of different observations and the second column showing the frequency. The frequency indicates the number of times that observation occurs in data.

Q14. **Define a continuous series. What does the frequency represent in this case ?**

Ans. In continuous series the observations are presented in different class intervals rather than by single number and the frequency represents the number of observations falling in that class interval.

Q15. **Prepare a statistical table from the following data taking the class width as 7 by inclusive method :**

28 17 15 22 29 21 23 27 18 12

7	2	9	4	6	1	8	3	10	5
20	16	12	8	4	33	27	21	15	9
3	36	27	18	9	2	4	6	32	31
29	18	14	13	15	11	9	7	1	5
37	32	28	26	24					

Q16. **Prepare a discrete series from the following data :**

62	50	57	58	51	53	62	64	60	61
60	51	64	55	55	52	60	65	58	60
59	52	63	56	56	58	64	63	62	60
58	54	62	54	54	60	65	60	62	59
65	56	63	52	53	62	53	61	61	59

Q17. **Change the following into continuous series and convert the series into 'less than' and 'more than' cumulative series :**

Marks (mid values)	5	15	25	35	45	55
No. of students	8	12	15	9	4	2

Q18. **The marks obtained by 20 studends in Statistics and Economics are given below. Prepare a bivariate frequency distribution.**

Marks in Statistics	Marks in Economics	Marks in Statistics	Marks in Economics
10	20	13	24
11	21	12	23
10	22	11	22
11	21	12	23
11	23	10	22
14	23	14	22
12	22	14	24
12	21	12	20
13	24	13	24
10	25	10	23

Q19. **Prepare 'less than' and 'more than' cumulative frequency distributions of the following data :**

Wages (Rs.)	140 – 150	150 – 160	160 – 170	170 – 180	180 – 190	190 – 200
No. of workers	5	10	20	9	6	2

Q20. **Find out the frequency distribution and 'more than' cumulative frequency table :**

Price (Rs.) below	10	20	30	40	50	60
Quantity (Kg.)	17	22	29	37	50	60

CHAPTER – 9 – MEASURES OF CENTRAL TENDENCY – ARITHMETIC MEAN

Meaning

A meaning of central tenency or average or measure of location is a single value, which is used to represent an entire set of data. For example, we often talk of an average marks in a class, average age, average speed of a vehicle et. The most important objective of statistical analysis is to get one single value is called average or central value.

Objective and Functions of Average

1. **To present huge mass of data in a summarised form**
2. **To make comparison easier**
3. **To help in decision making**
4. **To know about universe from a sample**
5. **To trace precise relationship**
6. **Base for computing measures**

ARITHMETIC MEAN

Meaning

It is ordinarily known as 'Average' by the common man. It is defined as the sum of the values of all observations divided by the number of observations and is usually denoted by $\overline{X}$.

Kinds of Arithmetic Mean

Arithmetic mean can be computed in two ways;

(i) Simple Arithmetic Mean;

(ii) Weighted Arithmetic meanl.

Q. Calculation of Arithmetic Mean (Direct Method), Short cut method and step deviation method.

Students	Marks (X)
A	85
B	60
C	50
D	75
E	55
F	40
G	55
H	70
I	45
J	65
$N = 10$	$\Sigma X - 600$

Ans : 60

Ex.1 Following is the marks of 8 students. Find out arithmetic mean by (i) Direct Method; (ii) Short-cut Method; (iii) Step Deviation Method.

30, 45, 60, 40, 15, 65, 85, 20

DISCRETE SERIES

In case of discrete series (ungrouped frequency distribution), the values of variable shows the repetitions, i.e., frequencies are given corresponding to different values of variables.

Ex.2 From the following data of the marks obtained by 60 students of a class, calculate the average marks by the direct method, Shortcut method and step deviation method.

Marks	20	30	40	50	60	70
No. of Students	8	12	20	10	6	4

Ans : 41

Ex.3 The following data gives the daily earnings (in rupees) of workers in a factory Calculate the average income per workers by : (i) Direct Method (ii) Short-cut Method; (iii) Step Deviation Method.

Daily earning (in Rs.)	2	3	4	5	6	7	8	9
No. of workers	10	16	11	8	6	4	3	2

Ex.4 The following table gives the marks in English secured by 30 studetns a class in their weekly test :

Marks	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25
No. of students	2	8	6	10	4

Calculate the average marks of students by the direct Methods.

Ex 5 The following table shows the marks obtained by 90 students in a certain examination. Calculate the average marks by : (i) Direct Method; (ii) Short-cut Method; (iii) Step Deviation Method.

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
No. of students	3	8	12	16	19	16	11	5

CALCUALTION OF MISSING VALUE

Individual Series

Ex.6 Mean of a series of 5 items is 25, and values of four items are 10, 25, 35, 40. Find out the values of 5th term. Ans : 15

Discrete and Continuouis Series

Ex.7 Find the missing item, if mean of the following series is 50 :

Marks	30	35	40	?	50	75
No. of students	1	2	3	4	6	4

Sol. Let the missing items be x :

Calculating of Missing item

Marks (X)	No. of students (f)	fX
30	1	30
35	2	70
40	3	120
x	4	x
50	6	300
75	4	300
	$\Sigma f = 20$	$\Sigma fx = 820 + 4x$

$$\bar{X} = \frac{\Sigma fx}{\Sigma f}$$

$$50 = \frac{820 + 4x}{20}$$

or $50 \times 20 = 820 + 4x$

$\therefore$ x = 45 **Ans.** Missing marks = 45

Ex.8 Find out the missing frequency, if mean = 29.

X	5	15	25	35	45	55
f	5	7	?	18	5	3

Ans. f = 12

Ex.9 In the following frequency , the frequency of the class interval 50-60 is not known. Find it out, if the arithmetic mean of the distribution is 52.

Class Interval	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
No. of students	5	3	4	7	?	6	13

Sol. be f

Calculation of Missing frequency

Class Interval	No. of students (f)	Mid value (m)	fm
10 – 20	5	15	75
20 – 30	3	25	75
30 – 40	4	35	140
40 – 50	7	45	315
50 – 60	f	55	55f
60 – 70	6	65	390
70 – 80	13	75	975
	$\Sigma f = 38 + f$		$\Sigma fm = 1,970 + 55f$

$$\bar{X} = \frac{\Sigma fm}{\Sigma f}$$

$$52 = \frac{1,970 + 55f}{38 + f}$$

or 1,976 + 52f =1,970 + 55f

3f = 6

$\therefore$ f = 2

Ans. Missing frequency = 2

Arithmetic Mean in Special Cases

Mid-Values are given

Ex.10 Calculate arithmetic mean with the help of following data :

Mid – Value	5	15	25	35	45
Frequency	4	8	10	6	2

Sol.

Mid – value m	Frequency f	d = m – A (A=25)	$d = \frac{m-A}{C}$ C=10	fd′
5	4	–20	–2	–8
15	8	–10	–1	–8
25(A)	10	0	0	0
35	6	+10	+1	+6
45	2	+20	+2	+4
	Σf = 30			Σfd = –6

$$\bar{X} = A + \frac{\Sigma fd'}{\Sigma f} \times C$$

$$= 25 + \frac{-6}{30} \times 10 = 23$$

Ans. Arithmetic mean = 23.

Cumulative Series ('Less than' or 'More than')

When the data given in the form oof less than or more than for all items in thy series, it is called cumulative frequency distribution. To calculate the arithmetic mean, such cumulative frequency distribution has to be converted into a simple frequency distribution.

Ex.11 The following table gives the profit earned by varioius companies in the year 2007-08. Calculate the average profit earned.

Profit (Rs. in crore)	No. of Companies
Less than 20	5
Less than 30	22
Less than 40	48
Less than 50	60
Less than 60	83
Less than 70	100

Sol. Since were given the cumulative frequencies, we first find the simple frequency

Profit (Rs. in crore) X	No. of companies f	Mid – value m	d = m – A (A=45)	$d' = \frac{m-A}{C}$ C=10	fd′
10 – 20	5	15	–30	–3	–15
20 – 30	17	25	–20	–2	–34
30 – 40	26	35	–10	–1	–26
40 – 50	12	45 (A)	0	0	0
50 – 60	23	55	+10	+1	+23
60 – 70	17	65	+20	+2	+34
	Σf = 100				Σfd′ = –18

$$\bar{X} = A + \frac{\Sigma fd'}{\Sigma f} \times C$$

$$= 45 + \frac{-18}{100} \times 10 = \text{Rs. } 43.2 \text{ crores}$$

Ans. Rs. 43.2 crore.

Ex.12 From the following distribution of marks obtained by 50 students in statistics, calculate average marks.

Marks	No. of students
More than 0	50
More than 10	45
More than 20	38
More than 30	26
More than 40	10
More than 50	4

Sol. Since we are given the cumulative frequencies, we find the simple frequencies as shown in the following table :

Marks X	No. of students f	Mid – value m	d = m – A A=25	$d' = \frac{m-A}{C}$ C=10	fd
0 – 10	5	5	–20	–2	–16
10 – 20	7	15	–10	–1	–7
20 – 30	12	25(A)	0	0	0
30 – 40	16	35	+10	+1	+16
40 – 50	6	45	+20	+2	+12
50 – 60	4	55	+30	+3	+12
	Σf = 50				Σfd = +23

$$\bar{X} = A + \frac{\Sigma fd'}{\Sigma f} \times C$$

$$= 25 + \frac{+23}{50} \times 10 = 29.60 \text{ marks}$$

Ans. Average marks = 29.60 marks.

Inclusive Class-Intervals

When the data is given in inclusive series, then it is not necessary to adjust the classes for calculating arithmetic mean as the mid-value remains the same whether the adjustment made or not.

Ex.13 Find mean of the following data

Class – Interval	50 – 59	40 – 49	30 – 39	20 – 29	10 – 19	0 – 9
Frequency	1	3	8	10	15	3

Sol. In the given example, it is neither necessary to converty the data into exclusive class-interval series nor to arrange the data in ascending order :

Computation of Mean

Class – inverval X	Frequency f	Mid – value m	d = m – A (A=24.5)	$d' = \frac{m-A}{C}$ (C=10)	fd′
50 – 59	1	54.5	+30	+3	3
40 – 49	3	44.5	+20	+2	6
30 – 39	8	34.5	+10	+1	8
20 – 29	10	24.5 (A)	0	0	0
10 – 19	15	14.5	–10	–1	–15
0 – 9	3	4.5	–20	–2	–6
	Σf = 40				Σfd = –4

$$\bar{X} = A + \frac{\Sigma fd'}{\Sigma f} \times C$$

$$= 24.5 \times \frac{-4}{40} \times 10 = 23.50$$

Ans. Mean = 23.50

Open-end Class Intervals

Open-end class-intervals are those which do not have the lower limit of the first class-interval and the upper limit of the last class-interval. For example, 'less than 10', or 'more than 100' are open end class-interval.

Ex.14 Calculate mean of the following series

Marks	Below 20	20 – 50	50 – 90	90 – 140	Above 140
No. of students	10	20	40	15	15

Sol.

Marks X	No. of students f	Mid – value m	fm
0 – 20	10	10	100
20 – 50	20	35	700
50 – 90	40	70	2,800
90 – 140	15	115	1,725
140 – 200	15	170	2,550
	$\Sigma f = 100$		$\Sigma fm = 7,875$

$$\bar{X} = \frac{\Sigma fm}{\Sigma f} = \frac{7,875}{100} = 78.75$$

Ans. Average marks = 78.75

Unequal Class-interval

Sometimes the class-interval fo the distribution is unequal. In such cases, mean can be determined in the usual manner after calculating the md-values of each interval.

Ex.15 Calculate arithmetic mean from the following data :

Marks	0 – 10	10 – 20	20 – 40	40 – 70	70 – 100
No. of Students	8	12	30	6	4

Sol. In the given example, the class-intervals are unequal. Mean will be calculated directly after calculating the mid-point.

Marks	Mid – Point (m)	No. of students (f)	fm
0 – 10	5	8	40
10 – 20	15	12	180
20 – 40	30	30	900
40 – 70	55	6	330
70 – 100	85	4	340
		$\Sigma f = 60$	$\Sigma fm = 1,790$

$$\bar{X} = \frac{\Sigma fm}{\Sigma f} = \frac{1,790}{60} = 29.83$$

Ans. Average marks = 29.83

Ex.16 Find out combined mean from the following data :

	Seriex X_1	Series X_2
Mean	12	20
No. of items	80	60

Sol. Combined Mean $(\bar{X}_{1,\,2}) = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$

Given $\bar{X}_1 = 12,\ \bar{X}_2 = 20,\ N_1 = 80,\ N_2 = 60$

$$(\bar{X}_{1,2}) = \frac{(80 \times 12) + (60 \times 20)}{80 + 60}$$

$$= \frac{960 + 1,200}{140} = \frac{2,160}{140}$$

$$= 15.43$$

Ans. Combined Mean = 15.43

Ex.17 The average rainfall of a city from Monday to Saturday is 0.3 cms. Due to heavy rainfall on Sunday, the average for the whole week rose to 0.5 cms. How much was the rainfall on Sunday ?

Sol. Consider the rainfall from Monday to Saturday (6 days) as first group and rainfall on Sunday (1 day) as second group.

Then, $N_1 = 6,\ N_2 = 1,\ \bar{X}_1 = 0.3,\ \bar{X}_2 = 0.5$

$$(\bar{X}_{1,2}) = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$

$$0.5 = \frac{(6 \times 0.3) + (1 \times \bar{X}_2)}{6 + 1}$$

$$3.5 = 1.8 + \bar{X}_2$$

$$X_2 = 1.7$$

Ans. Rainfall on Sunday = 1.7 cms

Ex.18 The average marks of 50 students in class is 4. The pass result of 40 student who took up a class test is given below. Calculate mean marks of 10 students who failed.

Marks	4	5	6	7	8	9
No. of students	8	10	9	6	4	3

Sol. $\bar{X}_{1,2} = 4$; Mean of Pass Students $(\bar{X}_1) = ?$ Mean of fail students $(\bar{X}_2) = ?$

$N_1 = 40, N_2 = 10$

Calculation of Mean Marks of 40 Students $(\bar{X}_1)$

Marks (X)	No. of students (f)	fX
4	8	32
5	10	50
6	9	54
7	6	42
8	4	32
9	3	27
	$\Sigma f = 40$	$\Sigma fX = 237$

$$\bar{X}_1 = \frac{\Sigma fX}{\Sigma f} = \frac{237}{40} = 5.925 \text{ marks}$$

$$(\bar{X}_{1,2}) = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$

$$4 = \frac{(40 \times 5.925) + (10 \times X_2)}{40 + 10}$$

$$200 = 237 + 10\bar{X}_2$$

$$\bar{X}_2 = 3.7$$

Ans. Average marks of 10 students who failed = 3.7 marks.

Ex.19 The mean wage of 100 workers is Rs. 284. The mean wage of 70 workers is Rs. 290. Find the mean wage of remaining 30 workers.

Ans : Rs. 270.

Ex.20 The mean age of a combined group of men and women is 30 years. If the mean age of group of men is 32 and that of the group of women is 27, find out the percentage of men and women in the group.

Sol. Let x be the percentage of men in the combined group. Therefore, percentage of women = 100 – x.

We are given that,

$\bar{X}_1$(Men) = 32 years; $\bar{X}_2$(Women) = 27 years; $\bar{X}_{1,2}$(combined) = 30 years

$$(\bar{X}_{1,2}) = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$

$$30 = \frac{32x + 27(100 - x)}{x + (100 - x)}$$

32x – 2x = 3,000 – 2,700

5x = 300

x = 60

It means men are 60% and women = 100 – 60 = 40%

Ans. Men = 60%, Women = 40%

Corrected Mean

Ex.21 The average weight of a group of 25 boys was calculated to be 52 kg. It was later discovered that one weight as 45 kg instead of 54 kg. Calculate the correct average weight. **Ans.** 52.36 kg

Ex.22 The mean salary paid to 1000 employees of a factory was found to be Rs. 180.4. Later on it was discovered that the wages of two employees were wrongly taken as 297 and 165 instead of 197 and 185. Find the correct mean.

Ans. Rs.180.32

Ex.23 The average marks in statistics of 10 students of a class were 68. A new student took admission with 72 marks whereas two existing students left the college. If the marks of these students were 40 and 39, find the average marks.

Ans :74.78 marks.

Ex.24 The arithmetic mean of a given set of data on expenditure was found to be Rs. 100. Later it was found, due to an error of omission, that all the observations are Rs 5 less than the true values. Calculate the mean expenditure for the corrected data:

(i) in terms of rupees, and (ii) in terms of paise.

Ans. Since $\bar{X} = \frac{\sum X}{N}$

$\therefore \quad \sum X = N\bar{X}$

Here, $\bar{X}$ = 100, N = 5 (Assume any number of observations)

SX = 100 × 5 = 500

Calculated SX, i.e., 500 is wrong as all the observations are Rs 5 less than the true values. Let us correct SX by adding 25 incorrect SX (5 observations × Rs 5 less in true value = 25).

Incorrect SX = 500

Add : Corrected balance of 5 observations (5 × 5) = 25

Correct SX = 525

a. Hence, corrected mean expenditure in terms of rupees

$$\frac{\text{Corrected}\sum X}{N} = \frac{525}{5} = \text{Rs.}105$$

b. Corrected mean expenditure in terms of paise

Rs. 105 × 100 = 10500

= Paise 10500

Ex.25 The average age of class having 35 students is 14 years. When the age of the class teacher is added to the sums of the ages of students, the average rises by 0.5 year. What must be the age of the teacher?

Sol. $\bar{X} = \frac{\Sigma X}{N}$

or $\Sigma X = \bar{X}N$

Total age of 35 students = 35 × 14 = 490

Total age of students and the teacher together = 36 × 14.5 = 522

Age of teacher = 522 – 490 = 32 years

Ans. Teacher's age = 32 years.

Ex.26 What will be the new mean, if it is known that for a group of 10 students, scoring an average of 60 marks, the best paper was wrongly marked 80 instead of 75 ?

Sol. $\bar{X} = \frac{\Sigma X}{N}$

or $\Sigma X = \bar{X}N$

Given, $\bar{X} = 60, N = 10$

So, $\Sigma X = 60 \times 10 = 600$

Corrected ΣX = 600 – 80 + 75

= 595

Corrected Mean $(\bar{X}) = \frac{\Sigma X}{N} = \frac{595}{10}$ = 59.5 marks

Ans. 59.5 marks.

Ex.33 Calculate the weighted mean of the following data :

Items	10	15	20	25	30	35
Weight	6	9	4	10	5	2

Sol.

Items (X)	Weight (W)	WX
10	6	60
15	9	135
20	4	80
25	10	250
30	5	150
35	2	70
	$\Sigma W = 36$	$\Sigma WX = 745$

$$\bar{X}_w = \frac{\Sigma WX}{\Sigma W} = \frac{745}{36} = 20.69$$

Ans. Weighted Mean = 20.69

Ex.27 Calculate weighted mean by weighting each price by the quantity consumed

Food items	Quantity consumed (in kg)	Price in Rupees (per kg)
Wheat	300	10
Rice	400	20
Sugar	200	15
Potato	500	7

Sol.

Food items	Quantity consumed (in kg) (W)	Price in Rupees (per kg) (X)	WX
Wheat	300	10	3000
Rice	400	20	8000
Sugar	200	15	3000
Potato	500	7	3500
	$\Sigma W = 1400$		$\Sigma WX = 17,500$

$$\bar{X}_w = \frac{\Sigma WX}{\Sigma W} = \frac{17500}{1400} = 12.5$$

Ans. Weighted mean 12.5

Ex.28 Calculate the value of weighted mean from the given details of a college.

Course	Students appeared	Students passed
B.Com(H)	200	180
B.Com(P)	400	320
B.A.	700	490
M.Com	300	150

Sol.

Course	Appeared (W)	Passed	Percentage Pass (X) $\frac{\text{Passed}}{\text{Appeared}}\times 100$	WX
B.Com(H)	200	180	90	18,000
B.Com(P)	400	320	80	32,000
B.A.	700	490	70	49,000
M.Com	300	150	50	15,000
	$\Sigma W = 1,600$			$\Sigma WX = 1,14,000$

$$\bar{X} = \frac{\Sigma WX}{\Sigma W} = \frac{1,14,000}{1,600} = 71.25\%$$

Ans. Weighted mean = 71.25%

Ex29. An examination was held to decide the award of a scholarship. The weight of various subjects were different . The marks obtained by 3 candidates (out of 100 in each subject are given below :

Subject	Weights	Marks of Student A	Marks of Student B	Marks of Student C
Mathematics	4	60	57	62
Business Studies	3	62	61	67
Economics	2	55	53	60
English	1	67	77	49

Calculate the weighted Arithmetic Mean to award the scholarship.

Sol.

Student	Weights W	Student A Marks (X_A)	WX_A	Student B Marks(X_B)	WX_B	Student C Marks (X_C)	WX_C
Mathematics	4	60	240	57	228	62	248
B. Studies	3	62	186	61	183	67	201
Economics	2	55	110	53	106	60	120
English	1	67	67	77	77	49	49
	10	244	603	248	594	238	618

Simple Arthemetic Mean — Weighted Mean

Student A $\bar{X}_A = \frac{\Sigma X_A}{N} = \frac{244}{4} = 61$ $\quad$ $\bar{X}_{WA} = \frac{\Sigma WX_A}{\Sigma W} = \frac{603}{10} = 60.3$

Student B $\bar{X}_B = \frac{\Sigma X_B}{N} = \frac{248}{4} = 62$ $\quad$ $\bar{X}_{WB} = \frac{\Sigma WX_B}{\Sigma W} = \frac{594}{10} = 59.4$

Student C $\bar{X}_C = \frac{\Sigma X_C}{N} = \frac{238}{4} = 59.5$ $\quad$ $\bar{X}_{WC} = \frac{\Sigma WX_C}{\Sigma W} = \frac{618}{10} = 61.8$

Ans. From the above calculations, C should get the scholarship as his weighted mean is the highest

Note : According to simple arithmetic mean, B should get the scholarship. But all the subjects of examination are not of equal importance. Therefore, weighted mean is to be considered for award of scholarship.

FORMULAE AT A GLANCE

	Individual Series	Discrete Series	Contin. Series
1. Arithemetic Mean			
Direct Method	$\bar{X} = \frac{\Sigma X}{N}$	$\bar{X} = \frac{\Sigma fX}{\Sigma f}$	$\bar{X} = \frac{\Sigma fm}{\Sigma f}$
Short-cut Method	$\bar{X} = A + \frac{\Sigma d}{N}$	$\bar{X} = A + \frac{\Sigma fd}{\Sigma f}$	$\bar{X} = A + \frac{\Sigma fd}{\Sigma f}$
Step Deviat. Method	$\bar{X} = A + \frac{\Sigma d'}{N} \times C$	$\bar{X} = A + \frac{\Sigma fd'}{\Sigma f} \times C$	$\bar{X} = A + \frac{\Sigma fd'}{\Sigma f} \times C$

2. Combined Mean $(\bar{X}_{1,2}) = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$

3. Weighted Mean $\bar{X}_w = \frac{\Sigma WX}{\Sigma W}$

UNSOLVED PRACTICALS

Q1. **The following figures are the heights in cms of 7 children chosen at random. Calculate the simple arithmetic mean of the heights by (i) Direct Method (ii) Short-cut Method and (iii) Step Deviation Method.**

[Mean Height = 66 cm]

64, 59, 67, 69, 65, 70, 68

Q2. **The distribution of age at marriage of 50 males is given below :**

Age in years	20	21	22	23	24	25
No. of males	1	2	4	5	15	23

Calculate arithmetic mean using direct method.

[Mean at Marriage age = 24 years]

Q3. **Calculate the arithmetic average from the following data :**

Daily wages (in Rs.)	2 – 4	4 – 6	6 – 8	8 – 10	10 – 12	12 – 14	14 – 16	16 – 18
No. of workers	11	14	20	32	25	7	5	2

[Avg. wages Rs. 8.67]

Q4. **Find mean for the following by data by using :** **[Mean = 412]**

(i) Direct Method (ii) Short-cut Method

(iii) Step Deviation Method

X	100 – 200	200 – 300	300 – 400	400 – 500	500 – 600
F	10	18	12	20	40

Q5. **Find the mean from the following data :** **[Mean 34.40 Marks]**

Marks	No. of students
Less than 10	5
Less than 20	20
Less than 30	45
Less than 40	70
Less than 50	80
Less than 60	88
Less than 70	98
Less than 80	100

Q6. **The following table shows the age of workers in factory. Find out the average age of workers.**

Age(in years)	20 – 29	30 – 39	40 – 49	50 – 59	60 – 69
Workers	10	8	6	4	2

[Avg. Age is 23.86 years]

Q7. **Find the mean from the following data :** **[Mean = 51.75 Marks]**

Marks	No. of students
Above 0	80
Above 10	77
Above 20	72
Above 30	65
Above 40	55
Above 50	43
Above 60	28
Above 70	16
Above 80	10
Above 90	8
Above 100	0

Q8. **Calculate the mean from the following data :**

Mid – Value	10	20	30	40	50
Frequency	8	12	15	9	6

[Mean = 28.6]

Q9. **Calculate mean from the following data :**

X	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
Frequency	5	9	20	12	4

Q10. **The mean height of 25 male workers in a factory is 61 inches and the mean height of 35 female workers in the same factory is 58 inches. Find the combined mean height of 60 workers in the factory.**

Q11. **The mean age of a combined group of men and women is 30.5 years. If the mean age of the group of men is 35 and that of the group of women is 25, find out the percentage of men and women in the group.**

Q12. **The mean of 100 observations was found to be 40. Later on, it was discovered that two items were wrongly taken as 30 and 27 instead of 3 and 72. Find correct mean.**

Q13. **The average height of a group of 40 students was calculated as 155cm. It was later discovered thta the height of one student was read as 157 instead of 137 cm. Calculate the correct average height.**

Q14. **From the followuing data, calculate the weighted mean.**

Marks	62	77	65	62	57
Weights	2	1	2	3	4

[Weight Mean = 62.08 Marks]

Q15. **An examination was held to decide the most eligible student for scholarship. The wights of various subjects along with the marks obtained by the three candidates (out of 100) are given below :**

Subjects	Weights	Marks of Ram	Marks of Shyam	Marks of Manish
Maths	3	65	64	70
Statistics	4	63	60	65
Economics	1	70	80	52
English	2	58	56	63

WM of Ram = 63.3 Marks, WM marks of shyam = 82.4 WM of marks of manish 64.8 marks. Manish should get scholarship as his mean is maximum.

Q16. **Find the missing item.**

Marks	20	30	40	?	60	70
No. of Students	8	12	20	10	6	4

[Missing item = 50 marks]

Q17. Calculate the number of students against the class 30-40 of the following data, where $\bar{X} = 28$.

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	12	18	27	?	17	6

[Missing value = 80 marks]

Q18. Find the average marks from the following table :

Marks	No. of students
Below 10	25
Below 20	40
Below 30	60
Below 40	75
Below 50	95
Below 60	125
Below 70	190
Below 80	240

[Mean = 49.58 marks]

Q19. Calculate the mean from the following frequency distribution :

X	10 – 19	20 – 29	30 – 39	40 – 49	50 – 59	60 – 69	70 – 79	80 – 89
No. of persons	32	42	40	56	20	6	2	2

[Mean = 35.8]

Q20. Locate the missing frequency, if arithmetic mean of the following series is 44.8

X	20	30	40	50	60	70
f	5	?	15	10	8	5

[Missing frequency = 7]

Q21. The average of marks of 30 students in a class were 52. The top six students had an average of 31. What were the average marks of the other students ? [Average marks = 57.25]

Q22. A candidate obtained the following percentage of marks in an examination : English 60; Business Studies 75; Maths 63; Accouints 59; Economics 55. Find the candidate's weighted arithmetic mean, if weights 1, 2, 1, 3, 3 respectively are allotted to the subjects.

[Weighted arithmetic Mean = 61.5]

Q23. Why are measures of central value calculated ?

Ans. Measures of central values summarize the whole series of statistical data into a single number that describes the series.

Q24. Define an average. Why do we call an average to be a measure of central tendency ?

Ans. "An average is a single value within the range of the data that is used to represent all the values in the series."

Since an average is somewhere within the range of the data, it is sometimes called the measures of central tendency.

Q25. What are the different types of averages? Give names of each.

Ans. Types of Averages:

(a) Mathemical Averages. Arithmetic mean; Geometric mean and Harmonic mean.

(b) Positional Averages. Median and Mode.

Q26. Define arithmetic mean. In one sentence, differentiate a simple arithmetic mean from a weighted arithmetic mean.

Ans. Arithmetic mean is the value obtained by dividing the sum of the values of the individual' items of a series by the number of observations.

$$\bar{X} = \frac{\Sigma X}{n}$$

Simple arithmetic mean is obtained where there is equal importance to all the items of the series whereas weighted arithmetic mean is obtained after giving weights as figures to different values of data to indicate the relative importance of items.

Q27. Mention two merits and demerits of arithmetic mean.

Ans. Merits of arithmetic mean :

a. Simple to calculate and understand

b. Based on all observations

Demerits of arithmetic mean :

a. Affected by extreme values

b. Problem of calculation in case of missing value

Q28. Calculate arithmetic averages of the following information :

a. Marks obtained by 10 students :
30, 62, 47, 25, 52, 39, 56, 66, 12, 24

b. Income of 7 families (In Rs) : Also show $\Sigma(X - \bar{X}) = 0$
550, 490, 670, 890, 435, 590, 575

c. Height of 8 students (In cm) :
140, 145, 147, 152, 148, 144, 150, 151

[$\bar{X}$: a = 41.3 Marks, b = Rs 600, c = 147.12 cm]

Q29. Calculate mean of the following frequency distribution :

Values :	60	62	64	67	70	73	77	81	85	89
Frequency :	54	82	103	176	212	180	115	78	50	21

[$\bar{X}$: Rs 70.94]

Q30. **Calculate the arithmetic mean from the following data :**

Marks	No. of Students
Less than 10	5
Less than 20	15
Less than 30	55
Less than 40	75
Less than 50	100

$\bar{X} = 30$ Marks

Q31. **A candidate obtains 46% marks in English, 67% in Mathematics, 53% in Hindi 72% in History and 58% in Economics. It is agreed to give triple weights to marks in English and double weights to marks in Mathematics as compared to other subjects. Calculate Weighted Mean. Also compare it with simple Arithmetic Mean.**

[Marks : $\bar{X}w = 56.9$, $\bar{X} = 59.2$]

Q32. **The mean marks of 100 students were found to be 40. Later on it was discovered that a score of 53 was misread as 83. Find the corrected mean corresponding to the corrected score.** [Correct $\bar{X} = 39.7$ Marks]

Q33. **The mean weight of 25 boys in group A of a class is 61 kg and the mean weight of 35 boys in group B of the same class is 58 kg Find the mean weight of 60 boys.** [$\bar{X}_{1,2} = 59.25$ kg]

Q34. **Calculate Combined Mean :**

Section	Mean Marks	No. of students
A	75	50
B	60	60
C	55	50

[$\bar{X}_{1,2,3} = 63.125$ Marks]

Q35. **The average marks for Statistics in a class of 30 were 52. The top six students had an average of 31 marks. What were the average marks of the otherv students?** [$\bar{X}_2 = 57.25$ Marks]

Q36. **Find the average wage of a worker from the following data :**

Wages (in Rs) :	Above 300	310	320	330	340	350	360	370
No. of workers :	650	500	425	375	300	275	250	100

[$\bar{X}$ = Rs. 339.23]

Q37. **Calculate arithmetic measure from the following data :**

Temp. ^{o}C	–40 to – 30	–30 to – 20	–20 to –10	–10 to 0	0 to 10	10 to 20	20 to 30
No. of days	10	28	30	42	65	180	10

[$\bar{X} = 4.29$ °C]

Q38. **A candidate obtains the following percentage of marks : Sanskrit 75, Mathematics 84, Economics 56, English 78, Politics 57, History 54,**

Geography 47. It is agreed to give double weights to marks in English, Mathematics and Sanskrit. What : the weighted, and simple arithmetic mean?

[$\bar{X}$w = 68.8, $\bar{X}$ = 64.43 Marks]

Q39. **Calculate weighted mean by weighting each price by the quantity consumed :**

Food items	Quantity Consumed	Price in Rupees (Per Kg)
Flour	500 kg	1.25
Ghee	200 kg	20.00
Sugar	30 kg	4.50
Potato	15 kg	0.50
Oil	40 kg	5.50

[$\bar{X}$w = Rs. 6.35]

— Notes —

CHAPTER – 10 – Measure of Central Tendency Median and Mode

Mathematical averages, viz., simple arithmetic meand and weighted arithmetic mena. Now we will discuss the positional averages – median and mode as well as partition values – quartiles.

MEDIAN

The middle value in the data set when its elements are arranged in a sequential order, that is, in either ascending order, that is, in teither ascending or descending order of magnitude.

1. Individual Series

In order to calculate median in an individual series or ungrouped data, first of all the data is arranged in ascending order or descending order and then the following formula is used

$$M = \text{Size of } \left[\frac{N+1}{2}\right]^{th} \text{ item}$$

Where,

M = Median; N = Number of items

If data set contains an odd number of items, then the middl item of the distribution known as median. However, if there are even number of items, then median is the average of the two middle items.

Ex.1 Find tout the median from the following data :

120, 200, 170, 800, 620, 350, 375, 640, 750

Sol. Calculation of median

Ans : 375.

Ans. Median 375. This meant that 50% of the items are less than or equal to 375 and 50% of the items are more than or equal to 375.

Even Number Series

Ex.2 Given below is the age of some students. Find out the median of their age.

20, 16, 19, 14, 10, 22, 11, 9

Sol. Calculation of Median

Serial No.	Age arranged in ascending order
1	9
2	10
3	11
4	14
5	16
6	19
7	20
8	22
N = 8	

Ans. 15, Median = $\frac{N+1}{2}$

Discrete Series

Ex.3 Calculate the median from the following data :

Size (X)	Frequency(f)
3	2
4	1
5	3
6	7
7	4

Ans. Median = 6

Continuous Series

Ex.4 Find the median for the folloiwng data :

X	f
0 – 10	3
10 – 20	4
20 – 30	2
30 – 40	7
40 – 50	9

Ans. Median = 6

Ex.5 If the arithmetic mean of the data given below is 28, find : (a) Missing frequency; (b) Median of the series.

Profit per retail shop (in Rs.)	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Number or retail shops	12	18	27	?	17	6

Sol. (a) Let the missing frequency of the class 30-40 be f

Calculation of Missing Frequency

Profit per retail shop (X)	No. of retail Shops (f)	Mid – Value (m)	fm
0 – 10	12	5	60
10 – 20	18	15	270
20 – 30	27	25	675
30 – 40	f	35	35f
40 – 50	17	45	765
50 – 60	6	56	330
	$\Sigma f = 80 + f$		$\Sigma fm = 2{,}100 + 35f$

$$\bar{X} = \frac{\Sigma fm}{\Sigma f}$$

$$28 = \frac{2{,}100 + 35f}{80 + f}$$

$$2{,}240 + 28f = 2{,}100 + 35f$$

$$7f = 140$$

$$\therefore \quad f = 20$$

Missing frequency = 20 shops.

(b) **Calculation of Median**

Profit per retail Shop (X)	No. of retail Shops (f)	c.f.
0 – 10	12	12
10 – 20	18	30 (c.f)
(ℓ_1) 20 – 30	27 (f)	57 Median Class
30 – 40	20	77
40 – 50	17	94
50 – 60	6	100
	$N = \Sigma f = 100$	

MEDIAN IN SPECIAL CASES

Cumulative Series ('Less than' or 'More than')

Ex.6 Calculate the median from the following data :

Marks	No. of students
Less than 5	4
Less than 10	10
Less than 15	20
Less than 20	30
Less than 25	55
Less than 30	77
Less than 35	95
Less than 40	100

Ans. 24.

Sol. Since we given the cumulative frequencies, we first find the simple frequency.

Marks (X)	No. of students (f)	c.f.
0 – 5	4	4
5 – 10	6	10
10 – 15	10	20
15 – 20	10	30 (c.f)
(ℓ_1) 20 – 25	25 (f)	55 Median Class
25 – 30	22	77
30 – 35	18	95
35 – 40	5	100
	$N = \Sigma f = 100$	

$$Me = \frac{N}{2} = \frac{100}{2} = 50^{th}$$

50th item lies in the group 20-25

$\therefore$ ℓ_1 = 20, c.f. = 30, f = 25, i = 5

$$Me = \ell_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 20 + \frac{50 - 30}{25} \times 5 = 24$$

Ans. Median = 24 marks.

Ex.7 Find the median of the following data :

Age greater than (in yrs.)	0	10	20	30	40	50	60	70
No. of persons	230	218	200	165	123	73	28	8

Sol. Note that it is ‘more than’ type frquency distribution. We will first convert the cumulative frequencies into simple frequencies.

Age (in yrs.)	No. of persons (f)	c.f.
0 – 10	12	12
10 – 20	18	30
20 – 30	35	65
30 – 40	42	107 (c.f)
(ℓ_1) 40 – 50	50 (f)	157 Median Class
50 – 60	45	202
60 – 70	20	222
70 and above	8	230
	$N = \Sigma f = 230$	

Ans. Median = 41.6 years.

Mid-Values are given

Ex.8 Compute median from the following table :

Marks	115	125	135	145	155	165	175	185	195
Frequency	6	25	48	72	116	60	38	22	3

Sol. In the given exmple, we are given the mid-values of the class-intervals of a continuous frequency distribution. The difference between two mid-values is 10. Hence $\frac{10}{2}$ = 5, thus 110-120, 120-130,... and so on upto 190-200.

Computation of Median

Marks	f	c.f.	
110 – 120	6	6	
120 – 130	25	31	
130 – 140	48	79	
140 – 150	72	151	(c.f.)
(ℓ_1)150 – 160	116(f)	267	Median Class
160 – 170	60	327	
170 – 180	38	365	
180 – 190	22	387	
190 – 200	3	390	
	$N = \Sigma f = 390$		

Ans. Median = 153.79 marks

Inclusive Class-Intervals

Ex.9 Compute median from the following data :

Daily wages (Rs.)	110 – 119	100 – 109	90 – 99	80 – 89	70 – 79	60 – 69	50 – 59
Frequency	15	40	45	60	50	40	15

Sol. This is case of incluse class-intervals. To calculate median, it should be made exclusive and arranged in the ascending order, as follows :

Daily wages (Rs.)	f	c.f.	
49.5 – 59.5	15	15	
59.5 – 69.5	40	55	
69.5 – 79.5	50	105	c.f
(ℓ_1)79.5 – 89.5	60 (f)	165	Median Class
89.5 – 99.5	45	210	
99.5 – 109.5	40	250	
109.5 – 119.5	15	265	
	$N = \Sigma f = 265$		

Ans. Median = Rs. 84.08

Open-End Series

Ex.10 The size of land holdings of 380 families in a village is given below. Find the median size of land holdings.

Size of Land holding (in acres)	No. of Families
Less than 100	40
100 – 200	89
200 – 300	148
300 – 400	64
400 and above	39

Sol. The given data consist of open-end classes. However, there is no need to complete the class-interval, to calculate the median.

Size of Land holding (in acres)	No. of families (f)	c.f.	
Less than 100	40	40	
100 – 200	89	129	(c.f.)
(ℓ_1) 200 – 300	148 (f)	277	Median Class
300 – 400	64	341	
400 and above	39	380	
	$N = \Sigma f = 380$		

Ans. Median = 241.21 acres.

Unequal Class-Intervals

Ex.11 Calculate the median of the following distribution of data :

Class – interval	0 – 10	10 – 30	30 – 60	60 – 80	80 – 90
Frequency	5	15	30	8	2

Sol. In this question, the class intervals are enequal. However, to calculate median, there is no need to make class-intervals equal.

Class – interval (X)	Frequency (f)	c.f.	
0 – 10	5	5	
10 – 30	15	20	(c.f.)
(ℓ_1) 30 – 60	30 (f)	50	Median Class
60 – 80	8	58	
80 – 90	2	60	
	$N = \Sigma f = 60$		

Ans. Median = 40

Calculation of Missing Frequencies.

Ex.12 The following table gives the distribution of monthly salary of 900 employees. However, the frequencies of the classes 40-50 and 60-70 are missing. If the median of the distribution is Rs. 59.25, find the missing frequencies.

Salaries (Rs. in '000)	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
No. of employees	120	?	200	?	185

Sol. Let f_1 and f_2 be the frequencies of the classes 40-50 and 60-70 respectively

Salaries in Rs. '000(X)	No. of employees (f)	c.f.
30 – 40	120	120
40 – 50	f_1	$120 + f_1$
50 – 60	200	$320 + f_1$
60 – 70	f_2	$320 + f_1 + f_2$
70 – 80	185	900
	$N = \Sigma f = 900$	

$$Me = \frac{N}{2} = \frac{900}{2} = 450^{th}$$

450th item lies in the group 50-60 (Given median = 59.25)

$\therefore$ $\ell_1 = 50$, c.f. = 120 + f_1, f = 200, i = 10

$$Me = \ell_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$59.25 = 50 + \frac{450 - (120 - f_1)}{200} \times 10$$

$$59.25 - 50 = \frac{450 - (120 - f_1)}{200 \div 10}$$

$9.25 \times 20 = 300 - f_1$

$f_1 = 145$

From summation of frequencies, we have

120 + f_1 + 200 + f_2 + 185 = 900

Putting the values of f_1, we get

120 + 145 + 200 + f_2 + 185 = 900

i.e. f_2 = 250

Ans. Frequencies of class 40 - 50 (f_1) = 145; Frequency of class 60 - 70 (f_2) = 250.

GRAPHIC LOCATION OF MEDIAN

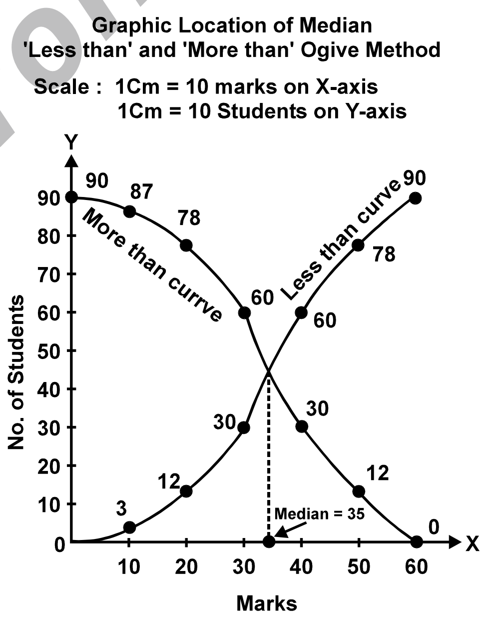
Ex.13 Determine the median graphically from the data given below :

Marks	0 – 10	10 – 20	20 – 30	30 – 40	50 – 60	60 – 70
No. of students	3	9	18	30	18	12

Sol. In order to calculate median by ‘Less than’ and ‘More than’ ogive method, we have to convert the series in cumulative frequency of ‘less than’ and ‘more than’ series.

Marks	No. of students	Marks	No. of students
Less than 10	3	More than 0	90
Less than 20	12	More than 10	87
Less than 30	30	More than 20	78
Less than 40	60	More than 30	60
Less than 50	78	More than 40	30
Less than 60	90	More than 50	12

On the basis of tables of ‘less than’ and ‘more than’, two Ogive curves and drawn.



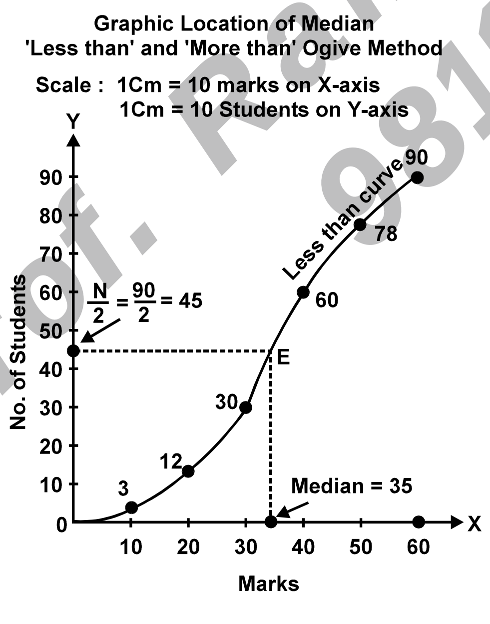
From the point of intersection (point E), a perpendicular (dotted line in the figure) is drawn on the X-axis. The dotted line cuts the X-axis at 35. Hence the median is 35 marks.

Ex.14 Determine the value of median graphically by 'less than' ogive the information given in **Ex.13.**

Sol. In order to culultive median by 'Less than' ogive method, we have to convert the series in cumulative of 'less than' series.

Marks	No. of students
Less than 10	3
Less than 20	12
Less than 30	30
Less than 40	60
Less than 50	78
Less than 60	90

On the basis of table of 'less than', one Ogive curve is drawn (see figure)



Median value = $\frac{N}{2} = \frac{90}{2} = 45$. Locating 45 on the Y-axis and a parallel line from 45 (dotted line in the figure) intersects the ogive at point E. Now, a perpendicular line drawn from point E cuts X-axis at 35. Hence the median is 35 marks.

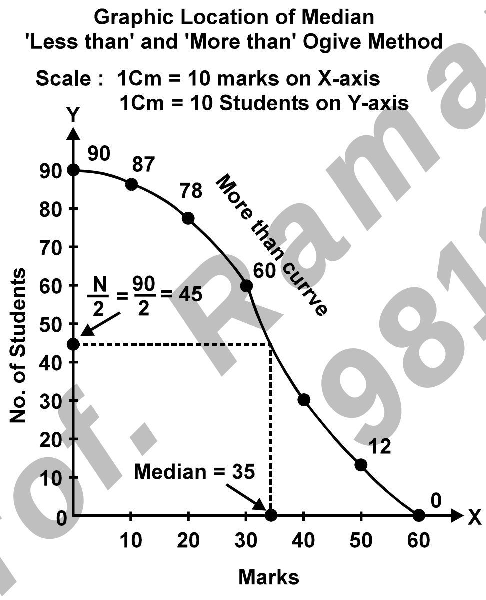
Ans. Median = 35 marks

Ex.15 Deterine the value of median graphically by 'more than' ogive with the information given in **Ex.13.**

Sol. In order to calculate median by 'More than' ogive method, we have to convert the series in cumulative frequency of 'more than' series.

Marks	No. of students
More than 0	90
More than 10	87
More than 20	78
More than 30	60
More than 40	30
More than 50	12

On the basis of table of 'more than', one Ogive curve is drawn (see figure)



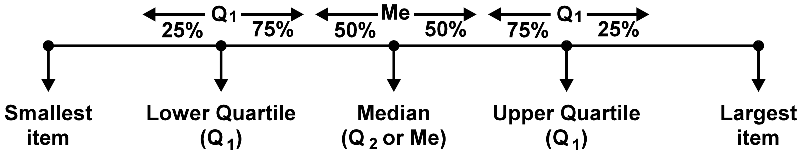
Median value $\frac{N}{2} = \frac{90}{2} = 45.$ Locating 45 on the Y-axis and a parallel line from 45 (dotted line in the figure) intersects the ogive at point E. Now, a perpendicular line drawn from point E cuts the X-axis at 35. Hence the median is 35 marks.

Ans. Median = 35 marks

APPLICATIONS OF MEDIAN

Quartiles

Median is a value which splits the series in two equal parts. Similarly, there are other positional values, which divide a series in a number of parts. The most common positional values besides median are Quartiles. Quartiles divide a series into four equal parts. For any series, there will be three quartiles as shown by the following figure :



COMPUTATION OF QUARTILES

While calculating Q_1 and Q_3, the series have to be arranged in ascending or descending order as in case of median.

Individual Series

In case of individual series, the values of lower quartile (Q_3) would be the size of

$\left[\frac{N+1}{4}\right]^{th}$ and $\left[\frac{N+1}{4}\right]^{th}$ items respectively.

Ex.16 From the data given below, calculate lower quartile (Q_1) and upper quartile (Q_3).

Pocket money (in Rs.) 46, 35, 28, 52, 54, 43, 35, 49, 46, 50, 41

Ans. Lower Quartile (Q_1) = Rs. 35; Upper Quartile (Q_3) = Rs. 50

Ex.17 Calculation first quartile and third quartile from the following data :

Marks of Students : 60, 38, 46, 43 50, 58, 65, 69

Sol. Arranging marks in ascending order, we get

38, 43, 46, 50, 58, 60, 65, 69

Calcuation of Lower Quartile (Q_1)

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

Ans. Lower Quartile (Q_1) = 43.75 marks; Upper Quartile (Q_3) = 63.75 marks.

Discrete Series

In case of discrete series also, the values of lower quartile (Q_1) and upper Quartile (Q_3) would be the size of $\left[\frac{N+1}{4}\right]^{th}$ and $3\left[\frac{N+1}{4}\right]^{th}$ items respectively. However, for value of N, the cumulative frequency is calculated.

Ex.17 From the following cumulative frequency :

X	10	20	30	40	50	60	70
f	2	3	5	10	5	3	2

Sol. We first calculate the cumulative frequency :

X	f	c.f.
10	2	2
20	3	5
30	5	10
40	10	20
50	5	25
60	3	28
70	2	30
	N = 30	

Calculation of Lower Quartile (Q_1)

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

Ans : $Q_1 = 30$; $Q_3 = 50$

Continuous Series

In case of continous series, the lower quartile (Q_1) is the $\left[\frac{N}{4}\right]^{th}$ item and the exact value of Q_1 is calculated by the following formula :

$$Q_1 = \ell_1 + \frac{\frac{N}{4} - c.f.}{f} \times i$$

$$Q_3 = \ell_1 + \frac{\frac{3N}{4} - c.f.}{f} \times i$$

Ex.18 With the help of following details calculate lower quartile, Median and upper quarterle.

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. of Students	16	14	23	17	7	3

Sol.

Marks (X)	No. of students (f)	c.f
0 – 10	16	16
10 – 20	14	30
20 – 30	23	53
30 – 40	17	70
40 – 50	7	77
50 – 60	3	80
	$N = \Sigma f = 80$	

Ans. Lower Quartile (Q_1)= 12.85 marks; Median = 24.35 marks; Upper Quartile (Q_3) = 34.12.

The following series relates to the daily income of workers employed in a firm. Compute

(a) highest income of lowest 50.% workers,

(b) minimum income earned by the top 25% workers, and

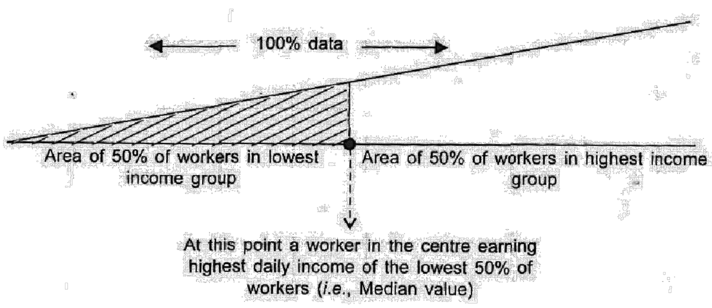
(c) maximum income earned by lowest 25% workers.

Daily Income (in Rs)	:	10-14	15-19	20-24	25-29	30-34	35-39
Number of workers	:	5	10	15	20	10	5

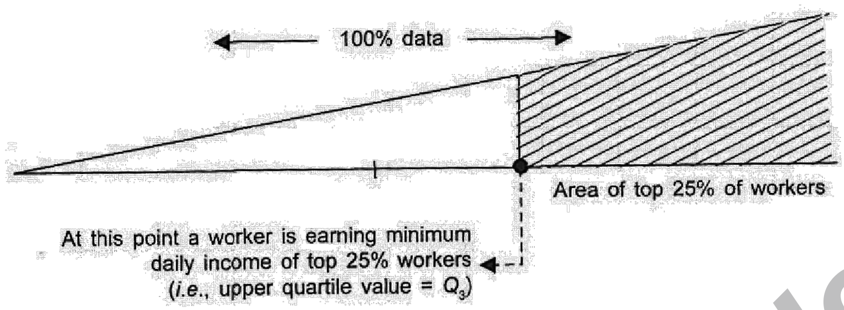
Before solving it let us understand the question.

1. As the data are of inclusive class intervals, we are required to convert the classes into class boundaries.

2. The data is arranged in ascending order, where



3. The data is arranged in ascending order, where



4. The data is arranged in ascending order, where

Therefore, we calculate median, Q_3 and Q_l to get the answers to the given questions.

Sol.

Daily Income (in Rs.) (X)	Numer of Workers (f)	Cumulative frequency (cf)
9.5 – 14.5	5	5
14.5 – 19.5	10	15
19.5 – 24.5	15	30
24.5 – 29.5	20	50
29.5 – 34.5	10	60
34.5 – 39.5	5	65

a. Computation of highest daily income in lowest 50% of workers. (Median)

Median is the value of $\left(\frac{N}{2}\right)^{th}$ item or $\left(\frac{65}{2}\right)^{th}$ or 32.5th item which lies in 224.5-29.5 class interval.

Applying suitable formula to get median value,

$$Me = l_1 + \frac{\frac{N}{2} - c.f}{f} \times i$$

$$= 24.5 + \frac{32.5 - 30}{20} \times 5$$

$$= 24.5 + \frac{2.5 \times 5}{20}$$

= 24.5 + 0.625 = 25.125

∴ Highest data income of lowest 50% workers is Rs 25.13.

b. Computation of minimum daily income earned by top 25% workers (Q_3)

Q_3 = Value of $3\left(\frac{N}{4}\right)^{th}$ item = $\frac{3 \times 65}{4}$ = 48.75^{th} value

Hence, Q_3 lies in class 24.5 - 29.5.

Applying suitable formula, we get

$$Q_3 l_1 + \frac{3\left(\frac{N}{4}\right)^{th} - c.f.}{f} \times i$$

$= 24.5 + \frac{48.75 - 30}{20} \times 5$

$= 24.5 + \frac{18.75 \times 5}{20}$

= 24.5 + 4.687 = 29.187

∴ Minimum daily income earned by top 25% workers is Rs 29.19.

c. Computation of maximum daily income earned by lowest 25% workers (Q_1)

Q_1 = Value of $\left(\frac{N}{4}\right)^{th}$ item = $\frac{65}{4}$ = 16.25^{th} value

Hence, Q_4 lies in class 19.5 - 24.5.

Applying suitable formula, we get

$$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i$$

$= 19.5 + \frac{16.25 - 15}{15} \times 5$

$= 19.5 + \frac{1.25 \times 5}{15}$

= 19.5 + 0.416 = 19.916

∴ Maximum daily income earned by lowest 25% workers is Rs 19.92.

Graphical Determination of Median and Quartiles

Q. Determine median and quartiles graphically from the following data :

Marks :	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Students :	7	10	20	13	17	10	14	9

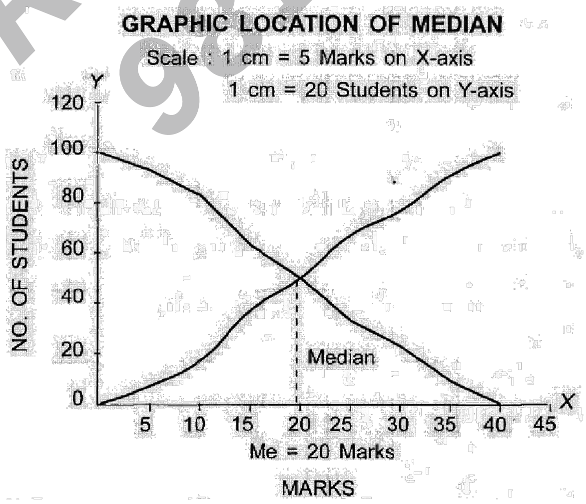
Sol.

Marks	f	Marks less than	Less than cumulative frequencies (c.f.)	Marks more than	More than cumulative frequencies (c.f.)
0 – 5	7	5	7	0	100
5 – 10	10	10	17	5	93
10 – 15	20	15	37	10	83
15 – 20	13	20	50	15	63
20 – 25	17	25	67	20	50
25 – 30	10	30	77	25	33
30 – 35	14	35	91	30	23
35 – 40	9	40	100	35	9
	N = 100				

First Method (only for Median).

Steps

1. Calculate ascending cumulative frequencies (less than) and descending cumulative frequencies (more than).
2. Draw two ogivesone by 'less than' and other by 'more than' methods.
3. From the intersecting point of two ogives, draw a perpendicular on X-axis.
4. The point where perpendicular touches X-axis, median value is determined.



Second Method (For Median and Quartiles).

Steps

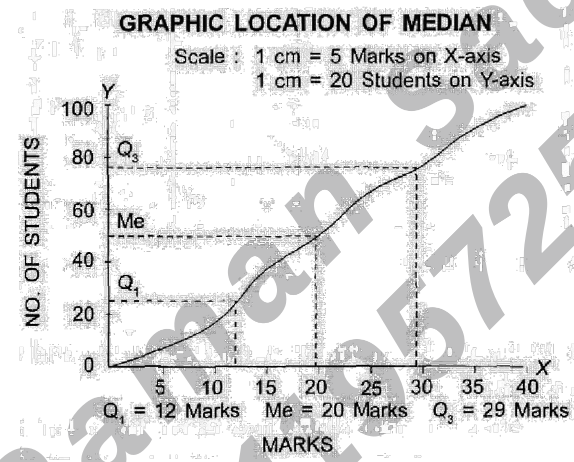
1. Calculate ascending cumulative frequency (less than).
2. Determine the value by the following formulae :

Me = size of $\left(\frac{N}{2}\right)^{th}$ item, i.e., $\frac{100}{2} = 50^{th}$ item

Q_1 = size of $\left(\frac{N}{4}\right)^{th}$ item, i.e., $\frac{100}{4} = 2.5^{th}$ item

Q_3 = size of $3\left(\frac{N}{4}\right)^{th}$ item, i.e., $3\left(\frac{100}{4}\right) = 7.5^{th}$ item

3. Locate 50, 25, 75 values on Y-axis and from them draw perpendiculars or cumulative frequency curve (ogive).
4. From these points where they meet the ogive draw another perpendicular touching X-axis.
5. The points where perpendicular touches X-axis, Q_1, Me and Q_3 are located.



Verification

Median Group 15 – 20

$$Me = l_1 + \frac{\left(\frac{N}{2}\right) - c.f.}{f} \times i$$

$$= 15 + \frac{50 - 37}{13} \times 5$$

$$= 15 + \frac{13 \times 5}{13} \times 20$$

Median = 20 Marks

Upper Quartile Group 25 – 30

$$Q_3 = l_1 + \frac{3\left(\frac{N}{4}\right) - c.f.}{f} \times i$$

$$= 25 + \frac{75 - 67}{10} \times 5$$

$$= 25 + \frac{8 \times 5}{10} \times 29$$

Q_3 = 29 Marks

Lower quartile group 10 – 15

$$Q_1 = l_1 + \frac{\left(\frac{N}{4}\right) - c.f.}{f} \times i$$

$$= 10 + \frac{25 - 17}{20} \times 5$$

$$= 10 + \frac{8 \times 5}{20} = 12$$

Q_1 = 12 Marks

Less than method' cumulative frequency curve is the reminder of the rule that at the first step of calculation of quartiles, the data is arranged in ascending order. However, median can be located on graph even by more than 'ogive' or calculated by arranging the data in descending order.

MODE

Mode is the value ocvcuring most frequently in a set of observations and around which other items of the set cluster most density.

CALCULASTION OIF MODE

1. Indiviual Series
2. Discrete Series and
3. Continuous Series.

Individual Series

Ex.20 From the heighs of 15 students, calculate the value of mode.

Height (in inches): 52 50 66 70 66 72 71 66 60 67 69 67 48 60 65

Sol. By arranging the series in an ascending order, we get :

48 50 52 60 60 65 66 66 66 67 67 69 70 71 72

By observation, height 66 inches occurs most, therefore, the mode (Z) is 66 inches.

Discrete Series

(i) Mode by observation, known as Inspection Method.

(ii) Mode by Grouping Method

Mode by Observation

Ex.21 Find out omode of the following series.

Daily Wages (in Rs.)	100	110	120	130	140	150
No. of persons	2	4	8	10	5	4

Sol. By inspection, we can see that 130 occurs most frequently in the series, hence modal daily wages = Rs. 130.

Mode by Grouping Method

1. **Grouping Table :** In this first table, groupings of frequencies are presented in six columns.
2. **Analysis Table :** In this second table, occurrence of frequencies or values in various grouping are written and added. Modal value is the value which occurs in the maximum number of grouping.

Ex.22 Calculate the value of Mode from the data given in **Ex.21** by grouping method.

Sol. First of all, grouping of the data is done.

In grouping method, values are first arranged in ascending order and the frequencies against each item are properly written. A grouping table normally consists of six columns. Frequencies are added in twos and threes and total are written between the values. If necessary, they can be added in fours and fives also.

Column 1. The maximum frequency is observed by putting a mark or a circle.

Column 2. Frequencies are grouped in twos.

Column 3. Leaving the first frequency, other frequencies are grouped in twos.

Column 4. Frequencies are grouped in threes.

Column 5. Leaving the first frequency, other frequencies are grouped in threes.

Column 6. Leaving the first two frequencies, other frequencies are grouped in threes.

After observing maximum total in each of these cases, put a mark, circle or make them bold on every total. An analysis table is prepared after completing grouping table in order to find out the item which is repeated the highest number of times. If the same procedure is adopted in continuous series, we shall be in a position to determine the modal class.

Wages in Rs. (X)	No. of Persons (f)	In Two's		In Three's		
	Column I	Column II	Column III	Column IV	Column V	Column VI
100	2	2+4=6		2+4+8=14		
110	4		4+8 =12		4+9+10=22	
120	8	8+10=18				8+10+5=23
130	10		10+5 =15	10+5+4=19		
140	5	5+4=9				
150	5					

After having prepared Grouping Table, we are required to prepare on Analysis Table. In this table, we enter the values having maximum frequencies in each column of Grouping Table by mean of ticks (T) as follows :

Table 2 : Analytic Table

Column No.	100	110	120	130	140	150
I				✓		
II			✓	✓		
III				✓	✓	
IV				✓	✓	✓
V		✓	✓	✓		
VI			✓	✓	✓	
Total	–	1	3	6	3	1

Since the value 130 has ccured the maximum number of times i.e., 6, the model income is Rs. 130.

Ans. Mode = Rs. 130.

Continuous Series

The value of mode is then determined by applying the following formula :

$$Mo = l_1 + \frac{f_1 - f_o}{2f_1 - f_o - f_2} \times i$$

or $$Mo = l_1 + \frac{f_1 - f_o}{(f_1 - f_o)(f_1 - f_2)} \times i$$

Where, Mo = Mode

l_1 = lower limit of modal class

f_1 = frequency of the modal class

f_o = frequency of the class preceding the modal class

f_2 = frequency of the class succeeding the modal class

i = class interval of the modal class

The above formula can also be expressed in the following way :

$$Mo = l_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

or $$Mo = l_1 + \frac{|f_1 - f_o|}{|f_1 - f_o| + |f_1 - f_2|} \times i$$

where, Mo = Mode

l_1 = lower limit of the modal class

D_1 = (Read delta 1), i.e., $|f_1 - f_o|$. The difference between the frequency of modal class and the frequency of the class before the modal class, i.e., preceding class (ignoring signs)

D_2 = (Read delta 2), i.e., $|f_1 - f_2|$. The difference between the frequency of the modal class and the frequency of the post modal class, i.e., succeeding class (ignoring signs)

Ex.23 From the following data of the ages of different persons, determine the modal age.

Size	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90
Frequency	4	10	25	15	23	22	12	9

Sol. Group Method

Ans. Mode = 58.59

MODE IN SPECIAL CASES

Cumulative Series ('Less than' or 'Mode than')

Ex.24 Find out the mode in the following series

Size (below)	5	10	15	20	25	30	35
Frequency	1	3	13	17	27	36	38

Sol. Here, we are given the data in the form of less than cumulative frequency distribution. To compute mode, we shall first arrange the data in the form of frequency distribution with continuous classes.

Calculation of Frequency Table

Size	c.f.	Frequency
0 – 5	1	1
5 – 10	3	2
10 – 15	13	10
15 – 20	17	4
20 – 25	27	10
25 – 30	36	9
30 – 35	38	2

In the given series, the distribution is irregular. Also the maximuim frequency (10) is repeated. Therefore, we will find mode by the method of grouping.

Ans. Mode = 24.28

Mid-Points are given

Ex.25 Calculate the mode from the following data :

Marks (Mid – values)	5	15	25	35	45	55	65	75
No. of Students	15	20	25	24	12	31	71	52

Sol. In the given example, mid-values of the series should be converted into a series with class-intervals.

Calculation of Class-Intervals

Marks	Frequency
0 – 10	15
10 – 20	20
20 – 30	25
30 – 40	24
40 – 50	12
50 – 60	31
60 – 70	71
70 – 80	52

In the given series, the distribution is irregular. Therefore, we will find mode by the method of grouping.

Ans. Mode = 66.78 marks.

Unequal Class-Intervals

Ex.26 Find the mode from the following data :

Class – interval	0 – 10	10 – 20	20 – 40	40 – 50	50 – 70	70 – 80
Frequency	10	14	40	35	42	10

Sol. In the given example, the class-interval are not equal. To calculate mode, the class intervals are made equal and frequencies are adjusted. We take the assumption that in this case frequencies are equally distributed.

Calculation of Frequency Table

Class – Intervals	Frequency	
0 – 10	10	
10 – 20	14	
20 – 30	20	
30 – 40	20	(f_0)
(ℓ_1) 40 – 50	35	(f_1) Modal Class
50 – 60	21	(f_2)
60 – 70	21	
70 – 80	10	

By inspection, it is clear that modal class is 40-50 as frequency of this class is maximum i.e. 35. To calculate mode, the following formula will be used.

$$\text{Mode (Z)} = \ell_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Ans. Mode = 45.17

Inclusive Class-Interval

Ex.27 Calculate the mode by Grouping Method in the following distribution.

Marks	40 – 49	50 – 59	60 – 69	70 – 79	80 – 89	90 – 99
No. of Students	12	30	24	20	12	2

Sol. **Ans.** Mode = 65.5 Marks.

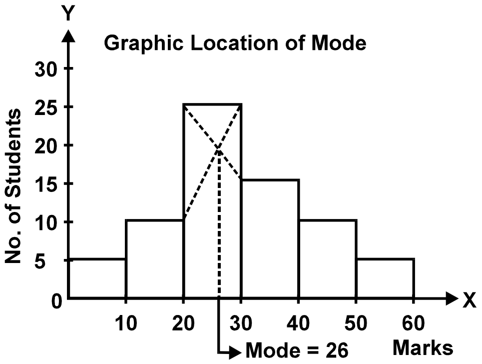
MODE BY GRAPHIC METHOD

Mode can be located graphically, with the hlep of histogram.

Ex.28 Find out the mode of the following series, using the Gaphic Method.

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. of students	5	10	25	15	10	5

Sol.



Verification : By Inspection, we find that modal class is 20-30. Applying the formula

$$M_0 = \ell_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Here $\ell_1 = 20,\ f_1 = 25,\ f_0 = 10,\ f_2 = 15,\ i = 10$

$$= 20 + \frac{25 - 10}{2 \times 25 - 10 - 15} \times 15$$

$$= 20 + \frac{15}{50 - 25} \times 10 = 20 + \frac{150}{25}$$

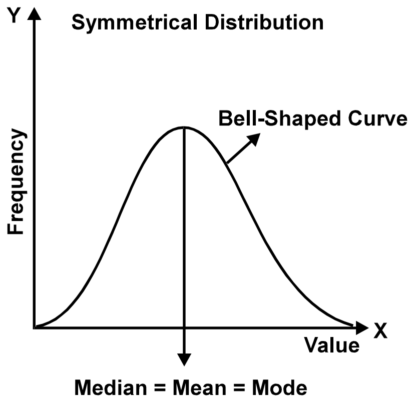
= 20 + 6 = 26

Ans. Mode = 26

RELATIONSHIP BETWEEN MEAN, MEDIAN AND MODE

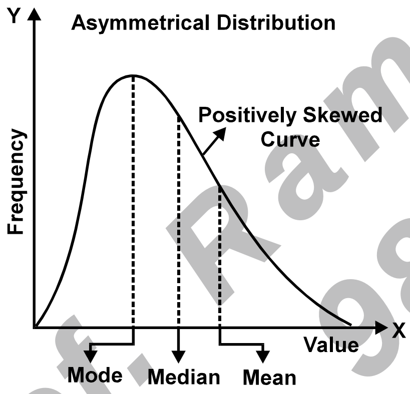
The relationship between mean median and mode depends upon the nature of distribution, which may be either symmetrical or asymmetrical.

1. **Symmetrical Distribution :** In case of symmetrical distribution, the values ofmean, median, and mode are equal, i.e., for symmetrical curves, Mean $(\bar{x})$ = Median (Me) = Mode (Z). The symmetrical distribution gives the shape of bell as seen in figure.

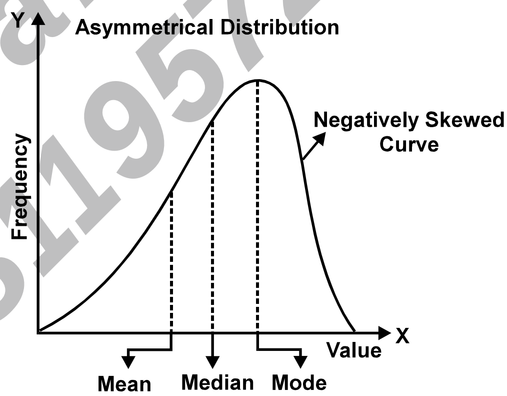


Symmetrical Distribution

2. **Asymmetrical Distribution :** In actual life, most of the distributions are not symmetrical. In an asymmetrical series, mean, median and mode have different values. The frequency curve is not bell shaped, i.e., height of the curve is not in the middle. An asymmetrical (skewed) distribution is either positively skewed or negatively skewed.



(Positive Skewed)



(Negatieve Skewed)

For a positively skewed distribution, most of the values of observations in a distribution fall to the right of the mode.

Mean > Median > Mode

For a negatively skewed distribution, values of lower magnitude are concentrated more to the left of the made.

Mean < Median < Mode

Accordingly to Karl Pearson, the relationship between mean, median and mode in an asymmetrical distribution is given by

Mode = Mean – 2(Mean – Median)

or Mode = 3 Median – 2 Mean

Ex.29 If the mean and median of moderately asymmetrical series are 26.8 and 27.9 respectively. Calculate the value of mode.

Sol. Using the empirical relationship, we know :

Mode = 3 Median – 2 Mean

= (3 × 27.9) – (2 × 26.8) = 88.7 – 53.6 = 30.1

Ans. Mode = 30.1

Bi-Modal Distribution

Ex.30 Calculate mode of the following series.

Marks	0 – 10	0 – 20	0 – 30	0 – 40	0 – 50	0 – 60	0 – 70	0 – 80	0 – 90
No. of students	4	6	24	46	67	86	96	99	100

Sol. Here, we are given the data in the form of less than cumulative frequency distribution. To compute mode, we shall arrange the data in the form of frequency distribution with continuous classes.

Table 1 : Grouping Table

Marks (X)	No. of Student (f)	In Two's		In Three's		
	Column I	Column II	Column III	Column IV	Column V	Column VI
0 - 10	4	6		24		
10 - 20	2		20		42	
20 - 30	18	40				61
30 - 40	22		43	62		
40 - 50	21	40			50	
50 - 60	19		29			32
60- 70	10	13		14		
70 - 80	3		4			
80 - 90	1					

Table 2 : Analysis Table

Column No.	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80	80 - 90
I				✓					
II			✓	✓	✓	✓			
III				✓	✓				
IV				✓	✓	✓			
V					✓	✓	✓		
VI			✓	✓	✓				
Total	–	–	2	5	5	3	1	–	–

Clearly, it is a bi-modal series and mode is ill defined. So, made in this case will be located with the help of formula :

Mode = 3 Median – 2 Mean

Calculation of Arithmetic Mean

Marks X	No. of students f	Mid – value m	$d = m - A$ (A=45)	$d' = \frac{m-A}{C}$ (C=10)	fd'
0 – 10	4	5	–10	–1	–16
10 – 20	2	15	–20	–2	–6
20 – 30	18	25	–30	–3	–36
30 – 40	22	35	–40	–4	–22
40 – 50	21	45 (A)	0	0	0
50 – 60	19	55	+10	+1	+19
60 – 70	10	65	+20	+2	+20
70 – 80	3	75	+30	+3	+9
80 – 90	1	85	+40	+4	+4
	$\Sigma f = 100$				$\Sigma fd' = -28$

$$\bar{X} = A + \frac{\Sigma fd'}{\Sigma f} \times C = 45 + \frac{-28}{100} \times 10 = 42.2$$

Mean = 42.2

Calcuation of Median

Marks X	No. of Students f	c.f.
0 – 10	4	4
10 – 20	2	6
20 – 30	18	24
30 – 40	22	46 (c.f.)
(ℓ_1) 40 – 50	21 (f)	67 Median Class
50 – 60	19	86
60 – 70	10	96
70 – 80	3	99
80 – 90	1	100
	$N = \Sigma f = 100$	

$$Me = \frac{N}{2} = \frac{100}{2} = 50^{th}$$

50th item lies in the group 40-50

$\therefore \quad \ell_1 = 40$, c.f. = 46, f = 21, i = 10

$$Me = \ell_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 40 + \frac{50 - 46}{21} \times 10$$

Median = 41.9

Mode = 3 Median – 2 Mean

= 3(41.9) – 2(42.2)

= 125.7 – 84.4

= 41.3

Ans. Mode = 41.3 Marks.

UNSOLVED PRACTICALS

1. Find out the median [Median = 18 Marks]

S.No.	1	2	3	4	5	6	7	8	9
Marks Obtained	10	12	14	17	18	20	21	30	32

2. Calculate the value of median. [Median = 19]

25, 20, 15, 45, 18, 7, 10, 64, 38, 12

3. Find out the median of the data given below :

X	160	150	152	161	156
f	5	8	6	3	7

[Median = 152]

4. Find the median of the following data : [Median = 40 years]

Age	20 – 25	25 – 30	30 – 35	35 – 40	40 – 45	45 – 50	50 – 55	55 – 60
No. of Persons	50	70	100	180	150	120	70	60

5. Calculate the median from the following data :

Marks	Frequency	Marks	Frequency
Less than 10	5	Less than 50	110
Less than 20	20	Less than 60	115
Less than 30	45	Less than 70	125
Less than 40	80	Less than 80	130

[Median = 35.71 Marks]

6. Find the median from the folloiwng : [Median = 28.24 Marks]

Marks (More than)	0	10	20	30	40	50	60	70
No. of Students	100	92	78	44	32	18	15	13

7. Compute median from the following data : [Median = 45]

Mid – values	37.5	42.5	47.5	52.5	57.5
No. of students	30	20	15	13	22

8. Calculate median from the following figures :

[Median = 39]

Class – interval	10 – 19	20 – 29	30 – 39	40 – 49	50 – 59	60 – 69
Frequency	12	19	20	21	15	13

9. Calculate median from following data : [Median = 28]

X	Below 10	10 – 20	20 – 30	30 – 40	40 – 50	50 and above
Frequency	3	7	15	9	6	4

10. From the followng data, compute median [Median = 40 Marks]

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. of Students	8	10	22	25	10	5

11. An incomplete distribution is given below : [F_1 = 34 ; F_2 = 45]

Variable	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	Total
Frequency	12	30	?	65	?	25	18	229

12. Detrmine the value of median from the following data with the help of : (i) 'Less than' and 'More than' Ogive Method; (ii) 'Less than' Ogive Method; (iii) 'More than' Ogive Method.

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. of Students	10	15	25	30	10	10

[Median = 30 Marks]

13. Calculate lower and upper quartiles.

S.No.	1	2	3	4	5	6	7	8	9	10
Marks	18	20	25	17	9	11	23	37	38	42

[Q_1 = 17 Marks ; Q_3 = 38 Marks]

14. From the following series, calculate lower quartile and upper quartile.

Variable	5	10	15	20	25	30	35	40
Frequency	16	18	22	21	24	14	11	9

[Q_1 = 10 ; Q_3 = 30]

15. Calculate upper and lower quartile from the following data :

Variable	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
Frequency	10	20	35	40	25	25	15

[Q_3 = 49 ; Q_1 = 23.57]

16. Find modal item of the following set of numbers :

[Mode = 5]

2, 5, 2, 3, 5, 5, 6, 4, 5, 3, 5, 2, 5, 7, 1

17. Compoute the mode from the following by : (i) Observation Method; (ii) Grouping Method

Height in inches	60	62	63	64	65	66	67	68	69	70
No. of Persons	5	13	18	20	21	30	23	12	4	2

[Mode = 30]

18. Calculate Mode of the following series [Mode = 13.33]

X	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30	30 – 35	35 – 40	40 – 45
f	20	24	32	28	20	16	34	10	8

19. Find out the mode value from the following data :

[Mode = 54.8]

Mid – Value	15	25	35	45	55	65	75	85
Frequency	5	8	12	16	28	15	3	2

20. Find out mode of the following frequency distribution :

Marks (More than)	0	10	20	30	40	50	60	70	80
No. of Students	40	38	33	25	15	7	5	2	0

[Mode = 35]

21. Calculate mode from the following data : [Mode = 44]

Variable	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of Students	5	12	40	32	28

22. Find out modal value from the following data :

Age	0 – 9	10 – 19	20 – 29	30 – 39	40 – 49	50 – 59
No. of Students	3	7	15	25	10	4

[Mode = 33.5]

23. The monthly profits in rupees (in thousands) of 100 shops are given below :

Marks	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. of Students	3	5	9	3	2

[Mode = 34]

24. Find lower quartile, median and upper quartile from the data given below

Class – Interval(More than)	10	20	30	40	50	60	70
Frequency	100	99	96	85	64	31	9

[Q_1 = 44.76 ; Q_3 = 62.72 ; Median = 54.24]

25. In the frequency distribution of 100 students given below. The number of students corresponding to marks groups 10-20 and 30-40 are missing from the table. However, the median is known to be 23. Find the missing frequencies

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of Students	8	?	40	?	10

[F_1 = 30 ; F_2 = 12]

26. Calculate the mean, median and mode of the following data :

Monthly Profit (Rs. '000)	Frequency
Less than 10	4
Less than 20	20
Less than 30	35
Less than 40	55
Less than 50	62
Less than 60	67

[Mean = 28.73 ; Median = 29 ; Mode = 32.78]

27. Calculate the median and mode from the following data :

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
No. of Students	2	18	30	45	35	20	6	3

[Median = 36.55 ; Mode = 36]

28. Calculate mode from the following data :

Income	15 – 24	25 – 34	35 – 44	45 – 54	55 – 64	65 – 74
No. of Workers	8	10	15	25	40	20

[Mode = 58.78]

29. Define median. What do you do with the data set before locating median ?

Ans. The median is that value of the variable which divides the group into two equal parts, one part comprising all values greater than the median value and the other part comprising all values smaller than the median value.

To calculate the median value, first of all the observations of the data series are arranged either in ascending or descending order of magnitudes and then the middlemost observation is called the median.

30. Mention any two merits and demerits of median

Ans. Merits of median:

a. Easy to calculate and easy to understand.

b. Not affected by extreme values.

Demerits of median:

a. Not based on all values.

b. Not a good representative.

31. What do you mean by mode ?

Ans. The mode of distribution is the value at the point around which the items tend to be most heavily concentrated.

32. When is the grouping method advised to locate the modal group in case of frequency distribution.

Ans. Grouping method is followed when the frequencies are more evenly spread over the different values of variable. If there are many observations having high frequency and the difference between the highest frequency and the next highest frequency is very small, for such cases grouping method is advised.

33. Mention two merits and demerits of mode.

Ans. Merits of mode:

a. The most representative value of a series.

b. Simple to calculate and understand.

Demerits of mode:

a. Not based on all items.

b. Not rigidly defined.

34. Of the different measures of central tendency, which is the most representative and why ?

Ans. If the purpose is statistical analysis of data set, then obviously arithmetic mean will be the best measure of central tendency because it can be used for further mathematical treatment.

35. Find out median of the following information :

Marks : 10, 70, 50, 20, 95, 55, 42, 60, 48, 80

[Me = 52.5 Marks]

36. We have the following frequency distribution of the size of 51 households. Calculate the arithmetic mean and the median.

Size	2	3	4	5	6	7	Total
No. of household	2	3	9	21	11	5	51

[X = 5, Me = 51]

37. Find out median, first quartile and third quartile of the following series :

Height (in inches)	58	59	60	61	62	63	64	65	66
No. of Persons	2	3	6	15	10	5	4	3	1

[Me = 61, Q_1 = 61 and Q_3 = 63]

38. The percentage of marks obtained by 68 students in an examination are given below. Compute the median.

Marks :	Below 20	20-40	40-60	60-80	Above 80
No. of Students :	0	5	22	25	16

[Me = 65.6]

39. Calculate the mean of the following distribution of daily wages of workers in a factory :

Daily Wages (in Rs):	100-120	140-160	160-180	180-200	Total
No. of Workers :	10	30	15	5	80

Also, calculate the median for the distribution of wages given above.

[X = 146.75, Me = 146.67]

40. An analysis for more efficiency in a factory, indicating the distribution of ages of orkers was as follows :

Age	16 – 19	20 – 29	30 – 39	40 – 49	50 – 59	60 – 64
Frequency	15	46	49	32	28	14

(a) Calculate the mean and median of the above data.

(b) Draw a histogram and indicate mean and mode therein.

[X = 37.48, Me = 36.32]

Hint. Get Mode on Histogram. Mean cannot be obtained on Histogram.

41. The percentage marks obtained by 100 students in an examination are given below. Compute the median, 1st and 3rd Quartiles.

Marks :	30-35	35-40	40-45	45-50	50-55	55-60	60-65
No. of Students :	14	16	18	23	18	8	3

[Me = 45.43, Q_1 = 38.4, Q_3 = 51.1]

42. Calculate Mean, Median and Mode from the following data :

Marks (mid – point)	59	61	63	65	67	69	71	73	Total
No. of students	1	2	9	48	131	102	40	17	350

[X= 67.9, Me = 67.75, Mo = 67.48]

43. Following is the distribution of marks of 50 students in a class :

Marks (more than) :	0	10	20	30	40	50
No. of students :	50	46	40	20	10	3

Calculate the Median Marks. If 60% of students pass this examination, find out the minimum marks obtained by a pass candidate.

[Me = 27.5, 25.5%]

44. The age in completed years of 50 persons is given below :

32 61 52 56 22 49 97 35 30 30 95 67 42
20 31 64 20 10 62 60 27 53 31 9 54 25
43 47 35 21 43 75 45 22 36 13 46 23 51
11 15 39 50 42 77 73 81 40 40 55

Prepare frequency table taking 10 - 19, 20 - 29,...as class intervals and calculate modal age. [Modal age = 42 years]

45. Determine the value of mode for the following data by using the formula :

Mode = 3, Median = 2 Mean.

Marks (Less than)	10	20	30	40	50	60	70	80	90
No. of students	5	15	98	242	367	405	425	438	439

46. Draw a 'less than' ogive from the following data and hence find out the value of Median.

Class	20 – 25	25 – 30	30 – 35	35 – 40	40 – 45	45 – 50	50 – 55	55 – 60
Frequency	6	9	13	23	19	15	9	6

47. Draw the histogram and estimate the value of mode from the following data:

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90
No. of students	0	2	3	7	13	11	9	2	1

48. Represent the following data by means of a histogram and find out mode.

Weekly wages	:	10-15	15-20	20-25	25-30	30-35	35-40	40-45
No. of workers	:	7	19	27	15	12	12	8

CHAPTER – 11 – MEASURES OF DISPERSION

Meaning of Dispersion

Dispersion is the extent to which values in a distribution differ from the average of the distribution.

"The degree to which numerical data tend to spread above an average value is called the variation or dispersion of the data'.

Dispersion : Absolute or Relative

The measure of dispersion can be either 'absolute' or 'relative'.

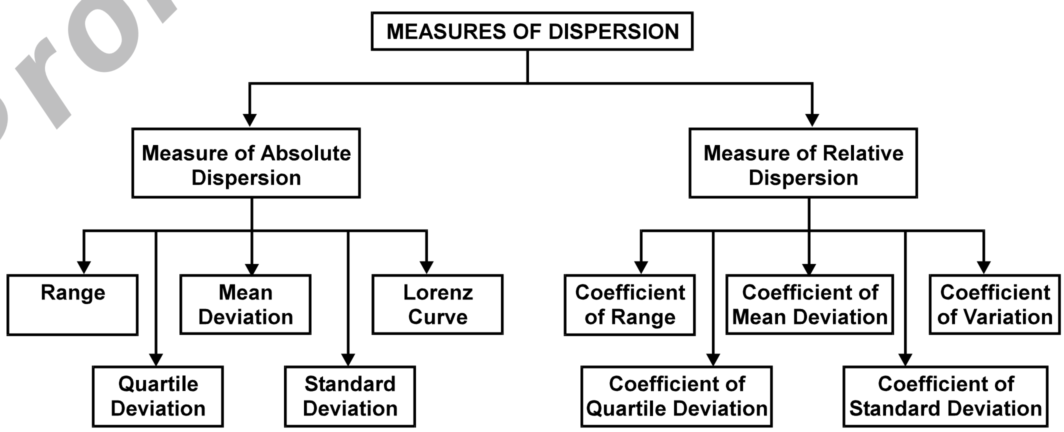
Absolute Measure

The measure of dispersion which are expressed in terms of the original units of a series are termed as Absolute Measures i.e., units in terms of which the data have been expressed like rupees, centimeters, kilograms etc.

Relative Measure

When the dispersion is measured as a percentage or ratio of the average , it is called relativ measures of dispersion.

Methods of Studying Dispersion



RANGE AND COEFFICIENT OF RANGE

Range

It is defined as the difference between the largest and the smallest item in a distribution.

$R = L - S$

Where,

R = Range

L = Largest item

S = Smallest item

If the items of a series be arrnaged in ascending (or descending) order, then the lowest and the highest values of the items are the two extreme items.

Coefficient of Range

The range is a measure of absolute dispersion and as such, cannot be usefully employed for comparing the variability of two distributions expressed in different units.

This relative measure, called the coefficient of range, refers to the ratio of the different between two extreme items (the largest and the smallest) of the distribution to thier sum.

$$\text{Coefficient of Range} = \frac{L - S}{L + S}$$

Where,

L = Largest item, and S = Smallest item

Individual Series

Ex.1 The folloiwng data represents the daily profits of a street vendor for 10 days. Calculate range and coefficient of range. Also calculate the percentage change in range, if the amount of maximum profit is omitted.

Profit (Rs.) 20, 24, 22, 40, 30, 35, 59, 55, 25, 70

Sol. In an ascending order, the profit s are : 20, 22, 24, 25, 30, 35, 40, 55, 59, 70

For the given values of profits, S = Rs. 20 and L = Rs. 70

Range = L – S = Rs. [70 – 20] = Rs. 50

$$\text{Coefficient of Range} = \frac{L - S}{L + S}$$

$$\text{Coefficient of Range} = \frac{70 + 20}{70 + 20} = \frac{50}{90} = 0.55$$

When maximum profit of Rs. 70 is omitted, then Range will be = L = 59 – 20 = 39.

Change in Range = Original Range – New Range = 50 – 39 = 11

$$\text{Percentage Change in Range} = \frac{11}{50} \times 100 = 22\%$$

Ans. Range = Rs. 45; Coefficent of range = 0.32; Percentage change in range (when maximum profit is omitted) = 22%.

Discrete Series

So, range is calculated by subtracting the smallest item from the largest item, without taking into account their frequencies.

Ex.2 Find the range and coefficient of range of the following distribution :

Items	3	4	5	6	7	8	9	10
Frequency	35	30	20	10	6	3	2	1

Ans. Range = 7; Coefficient of range = 0.54

Continuous Series

1. Difference between lower limit of the lowest class-interval and upper limit of the highest class-interval.
2. Difference between the mid-points of the lowest class-interval and the highest class-interval. Both the methods will give different answers. But, both the answers will be correct.

Ex.3 Calculate Range and its coefficient from the following data :

Marks (Below)	10	20	30	40	50	60	70	80	90	100
No. of Students	12	32	62	105	165	205	230	238	244	245

Sol. In the given example, we will first calculate the simple class-intervals and convert the series into non cumulative series, to determine the largest and smallest item. **Ans.** Range = 100 marks; Coefficient of range 1.

Marks	No. of students
0 – 10	12
10 – 20	20
20 – 30	30
30 – 40	43
40 – 50	60
50 – 60	22
60 – 70	40
70 – 80	55
80 – 90	10
90 – 100	6

Merits And Dismerits of Range

Merits

1. It is simplest to understand and easiest to compute.
2. It gives a quick measures of variability.
3. Range provides the broad picture of the data at a glance.

Demerits

1. Range is not based on all the observations. If all the items of a distribution are replaced except the smallest and the largest item, then the range of the distribution remains same.
2. Range is very much affected by fluctuations of sampling. Its value varies widely from sample to sample.
3. It does not give any idea about the pattern of the distribution. There can be two distributions with the same range but different patterns of distribution.
4. The range cannot be calculated in open-end distributions because of the absence of highest and lowest class boundaries.

QUARTILE DEVIATION

Interquartile Range

Intergquartile range refers to the difference between the values of two quartiles. symbolically,

Interequartile Range = $Q_3 - Q_1$

Quartile Deviation (Semi-Interquartile Range)

Quartile Deviation is known as the half of difference of upper quartile (Q_3) and the lower quartile (Q_1). It is half of the inter-quartile range (range among of the quartile). So, it is also known as the Semi Interquartile Range.

Symbolically,

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

Coefficient of Quartile Deviation

Quartile deviation is absolute meausre of dispersion. For comparative studies of variability of two distributions, we make use of relative measure, known as Coefficient of Quartile Deviation.

Symbolically,

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

Calcualtion of Quartile Deviation

Individual Series

Ex.4 Find the interquartile range, quartile deviation and coefficient of quartile deviation from the data given below :

200, 210, 208 160, 220, 250, 300

Sol. Calculation of lower quartile (Q_1) and upper quartile (Q_3).

S.No.	Items arrnaged in ascending order
1	160
2	200
3	208
4	210
5	220
6	250
7	300
N = 7	

$Q_1 = 200$

$Q_3 = 250$

Intrerquartile Range = 50

Intrerquartile Range = Q3 – Q1 = 250 – 200 = 50

Quartile Deviation = $\frac{Q_3 - Q_1}{2} = \frac{250 - 200}{2} = 25$

Coefficient of Quartile Deviation = $\frac{250 - 200}{250 + 200} = \frac{50}{450} = 0.11$

Ans. Interquartile range = 50; Quartile Deviation = 25; Coefficient of Quartile Deviation = 0.11.

Series of Individual Observations

Q. Find out interquartile range, quartile deviation and its coefficient frome following daily income of
workers in rupees :

145	130	200	210	198
234	159	160	178	257
260	300	345	360	390

Ans.

S.No.	Income (Rs.)	S.No.	Income (Rs.)
1	130	9	234
2	145	10	257
3	159	11	260
4	160	12	300
5	178	13	345
6	198	14	360
7	200	15	390
8	210		

Steps :

1. Arrange the data in ascending order to get the value of lower and upper quartiles.

2. Locate the value by finding out Q_1 = size of $\left(\frac{N+1}{4}\right)^{th}$ item and Q_3 = size of $3\left(\frac{N+1}{4}\right)^{th}$ item.

3. Apply the formulae to get interquartile range, quartile deviation and coefficient of artile deviation.

Thus, we get

Q_1 = size of $\left(\frac{15+1}{4}\right)^{th}$ item = 4th item

= Rs. 160

Q_3 = size of $3\left(\frac{15+1}{4}\right)^{th}$ item = 12th item

= Rs. 300

Interquartile Range = $Q_3 - Q_1$

Here, Q_3 = 300 and Q_1 = 160

= 300 – 160 = Rs. 140

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

Q.D. = $\frac{300-160}{2} = \text{Rs. }70$

Coefficient of Q.D. = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

= $\frac{300-160}{300+160} = \frac{140}{460} = 0.304$

Q. In a particular distribution, quartile deviation is 15 marks and coefficient quartile deviation is 0.6. Calculate lower and upper quartiles.

Ans. We are given

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

$\therefore \quad Q_3 - Q_1 = 30$ (equation 1)

Coefficient of Quartile deviation

$$\frac{Q_3 - Q_1}{Q_3 + Q_1} = 0.6$$

$$= \frac{30}{Q_3 + Q_1} = 0.6$$

$$= Q_3 + Q_1 = \frac{30}{0.6}$$

$\therefore \quad Q_3 + Q_1 = 50$ (equation 2)

Now equations (1) and (2) are solved,

$Q_3 - Q_1 = 30$

$Q_3 - Q_1 = 50$

$2Q_3 = 80$

$\therefore \quad Q_3 = \frac{80}{2} = 40$ Marks

Putting the value of Q_3 in equation (1)

$Q_3 - Q_1 = 30$

$40 - Q_1 = 30$

$\therefore \quad Q_1 = 40 - 30 = 10$ Marks

Thus, Upper Quartile = 40 Marks and Lower Quartile = 10 Marks

Discrete Series

Ex.6 From the following particulars, calculate the range of marks obtained by middle 50% of the students. Also calculate quartile deviation.

Marks	2	4	6	8	10	12
No. of students	3	5	10	12	6	4

Sol.

Marks(X)	No. of students (f)	c.f.
2	3	3
4	5	8
6	10	18
8	12	30
10	6	36
12	4	40

To calculate marks of middle 50% of students, we will have to calculate the difference between marks of 10th student (i.e. Q_1) and 30th student (i.e. Q_3), i.e., we have to calculate interquartile rane.

Ans. Q_1 = 6, Q_2 = 10, Interquartile Range = 4, Range of marks by middle 50% of students = 4

Continuous Series

Ex.7 Calculate the value of interquartile range, quartile deviation and coefficient of quartile deviation :

Marks	11 – 15	16 – 20	21 – 25	26 – 30	31 – 35	36 – 40	41 – 45	46 – 50	51 – 55
No. of students	10	17	22	31	42	32	26	19	14

Sol. This is a case of inclusive class-intervals. It should be first converted into exclusive series.

Marks (X)	Frequency (f)	c.f.
10.5 – 15.5	10	10
15.5 – 20.5	17	27
20.5 – 25.5	22	49
25.5 – 30.5	31	80
30.5 – 35.5	42	122
35.5 – 40.5	32	154
40.5 – 45.5	26	180
45.5 – 50.5	19	199
50.5 – 55.5	14	213
	$N = \Sigma f = 213$	

Ans. Q_1 = 26.18, Q_3 = 41.6, QD = 7.71,Coff. of Q.D. = 0.22, Inter Quartile Range = 15.42

Ex9. In a town, 25% of persons earned more than Rs 45,000 Where 75% earned more than Rs 18,000 Calculate the absolute and relative values of dispersion.

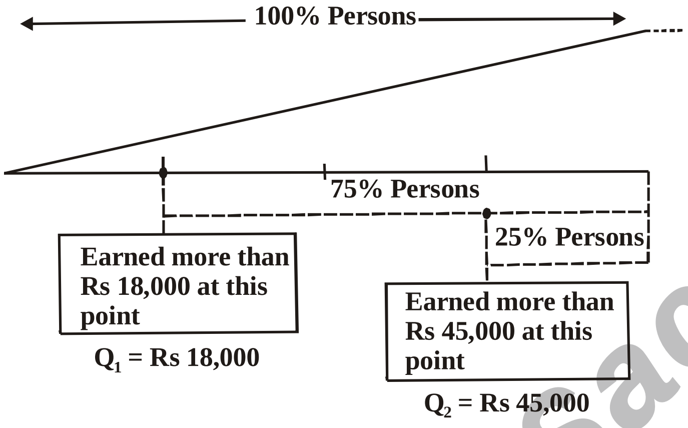
Solution.

Absolute value of dispersion

Quartile Deviation (Q.D.) = $\frac{Q_3 - Q_1}{2} = \frac{45,000 - 18,000}{2} = \text{Rs.}13,500$

Relative value of dispersion

$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{45,000 - 18,000}{45,000 + 18,000} = \frac{27,000}{63,000} = 0.428$$



MERITS AND DEMERITS OF QUARTILE DEVIATIONS

Merits

1. It is quite easy to understand and calculate.
2. It is the only measure of dispersion which can be used to deal with a distribution having open-end classes.
3. In comparision to range, it is less affected by extreme values.

Demerits

1. It is not based on all the observations
2. It is not capable of further algebraic treatment.
3. It is considerably affected by fluctuations in the sample.

MEAN DEVIATION

Meaning

"Mean deviation of a series is the arithmetic average of the deviations of various items from a measureof central tendency (mean, median or mod)."

Between mean and median, the latter is supposed to be better than the former, because the sum of the deviations from the median is less than the sum of the deviations from the mean.

Coefficient of Mean Deviation

Mean deviation or first moment of dispersion is an absolute measure of dispersion, expressed in the same units in which the original date are. It is divided by the average, from which it has been calculated. It is then known as the Coefficient of Mean Deviation.

Symbolically :

Coefficient of Mean Deviation from Mean $(\bar{X}) = \frac{MD_{\bar{X}}}{\bar{X}}$

Coefficient of Mean Deviation from Median (M) = $\frac{MD_M}{M}$

Calculation of Mean Deviation and Its Coefficient

Individual Series

Mean Deviation from Mean $(MD_{\bar{X}}) = \frac{\Sigma|X-\bar{X}|}{N} = \frac{\Sigma|D|}{N}$

Mean Deviation from Median $(MD_M) = \frac{\Sigma|X-M|}{N} = \frac{\Sigma|D|}{N}$

Ex.8 Which mean as the base, calculate the mean deviation and compare the variability of the two series A and B.

Series A	10	12	16	20	25	27	30
Series B	10	20	22	25	27	31	40

Sol. Calculation of Mean Deviation from Mean

Series A X_A	Deviations from Mean $X_A - \bar{X}_A$ |D|	Series B X_B	Deviations from Mean $X_B - \bar{X}_B$ |D|
10	10	10	15
12	8	20	5
16	4	22	3
20	0	25	0
25	5	27	2
27	7	31	6
30	10	40	15
$\Sigma X_A = 140$	$\Sigma\|D\| = 44$	$\Sigma X_B = 175$	$\Sigma\|D\| = 46$

Series A

$$\text{Mean}(\bar{X}_A) = \frac{\Sigma X_A}{N} = \frac{140}{7} = 20$$

Mean Deviation (Series A) = $\frac{\Sigma|D|}{2} = \frac{44}{7} = 6.28$

Coefficient of M.D. (Series A) = $\frac{M.D.}{\overline{X}_A} = \frac{6.28}{20} = 0.31$

Series B

Mean $(\overline{X}_B) = \frac{\Sigma X_B}{N} = \frac{175}{7} = 25$

Mean Deviation (Series B) = $\frac{\Sigma|D|}{N} = \frac{46}{7} = 6.57$

Coefficient of M.D. (Series B) = $\frac{M.D.}{\overline{X}_B} = \frac{6.57}{25} = 0.26$

4. Divide the total (Sf| D |) by the number of items to get mean deviation.

Mean Deviation from Mean $(MD_{\overline{X}}) = \frac{\Sigma f|X-\overline{X}|}{N} = \frac{\Sigma f|D|}{N}$

Mean Deviation from Median $(MD_M) = \frac{\Sigma f|X-M|}{N} = \frac{\Sigma f|D|}{N}$

Ex.9 Calculate mean deviation about arithmetic mean and coefficient of mean deviation.

Value (X)	10	11	12	13
Frequency (f)	3	12	18	12

Sol. Calculation of Mean Deviation about Arithmetic mean.

Values (X)	Frequency (f)	fX	$\|D\| = \|X-\overline{X}\|$ $\|X = 11.87\|$	$f\|D\|$
10	3	30	1.87	5.61
11	12	132	0.87	10.44
12	18	216	0.13	2.34
13	12	156	1.13	13.56
	$\Sigma f = 45$	$\Sigma fX = 534$		$\Sigma f\|D\| = 31.95$

Mean $(\overline{X}) = \frac{\Sigma fX}{\Sigma f} = \frac{534}{45} = 11.87$

Mean deviation from mean $(MD_{\overline{X}}) = \frac{\Sigma f|D|}{N} = \frac{31.95}{45} = 0.71$

Coefficient of Mean deviation = $\frac{MD_{\overline{X}}}{\overline{X}} = \frac{0.71}{11.87} = 0.06$

Ans. Mean deviation from mean = 0.71; Coefficient of mean deviation = 0.06.

Continuous Series

Mean Deviation from Mean $(MD_{\bar{X}}) = (MD_{\bar{X}} \frac{\Sigma f \mid m - \bar{X} \mid}{N} = \frac{\Sigma f \mid D \mid}{N}$

Mean Deviation from Median $(MD_M) = \frac{\Sigma f \mid m - M \mid}{N} = \frac{\Sigma f \mid D \mid}{N}$

where, m = Mid-point; $\bar{X}$ = Arithmetic Mean; M = Median.

Ex.10 Calculate mean deviation from median and Coefficient of mean deviation from the following data :

Marks(Mid – points)	10	30	50	70	90
No. of students	10	16	30	32	12

Sol. In the given example, we are given mid-points. Hence we will first convert the mid-points into class-intervals to calculate medina.

Calculation of Median

Mid – points m	Class – Interval X	f	c.f.
10	0 – 20	10	10
30	20 – 40	16	26
50	40 – 60	30	56
70	60 – 80	32	88
90	80 – 100	12	100
		$N = \Sigma f = 100$	

$$Me = \frac{N}{2} = \frac{100}{2} = 50^{th}$$

50th item lies in the group 40 - 60

$\therefore \quad \ell_1 = 40$, c.f. = 26, f = 30, i = 20

$$Me = \ell_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 40 + \frac{50 - 26}{30} \times 20$$

Median = 46

Calculation of Mean Deviation from Median

Mid – points (m)	f	$\|D\| = \|m - 56\|$	$f\|D\|$
10	10	46	460
30	16	26	416
50	30	6	180
70	32	14	448
90	12	34	408
	$N = \Sigma f = 100$		$\Sigma f\|D\| = 1,912$

Mean deviation from median (MD_M) = $\frac{\Sigma f|D|}{N} = \frac{1,912}{100}$ = 19.12 Marks

Coefficient of Mean deviation = $\frac{MD_M}{Median} = \frac{19.12}{56}$ = 0.34

Ans. Mean deviation from median 19.12 marks; Coefficient of mean deviation = 0.34.

MERIT AND DEMERITS OF MEAN DEVIATION

Merits

1. **Simplicity :** It is simple to calculate and easy to understand.
2. **Based on all observations :** Mean deviation is a more comprehensive measure of dispersion (compared to range and quartile deviation) as it is based upon all items of the series.
3. **Less effect of extreme values :** As compared with standard deviation, it is less affected by extreme observations.
4. **Rigidly defiend :** Mean deviation is rigidly defined and its value is precise and definite.
5. **Better Measure for comparison :** Mean deviation is based on the deviations from an average. So, it provides a better measure for comparison about formation of different distributions.

Demerits

1. **Not capable of algerabic treatment :** Mean deviation ignores the positive and negative signs of deviations. As a result, this method cannot lbe used for further algebraic treatment.
2. **Not well defined :** It is not well defined of dispersion since deviations can be taken from any measure of central tendency and mean deviation calculated from different averages (mean, median, mode) will not be same.

3. **Not suitable for open-end classes :** Mean deviation cannot be computed for distribution with open-end classes.
4. **Less reliable :** Meandeviation when calculated from mode is not reliable because in many cases mode has no fixed value.
5. **Difficult calculations :** It mean, mode and median are in fractions, then calculation of mean deviation becomes difficult.

STANDARD DEVIATION

Meaning

"Standard deviation is the square root of the arithmetic averageo of the squares of the deviations measured from the mean."

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

where

s = Standard Deviation

Σx^2 = Sum total of the squares of deviations from actual mean.

N = Number of pair of observations.

CALCULATION OF STANDARD DEVIATION

Individual Series

Actual Mean Method

In this method, deviations are taken from the actual mean. The steps involved in calculation of standard deviation are :

Step 1 Calculate the actual mean $(\bar{X})$ of the observation.

Step 2 Find out deviations of each item of the series from mean, i.e., calculate $(X - \bar{X})$ and denote the deviations by x.

Step 3 Square the deviations and obtaine total, i.e. Σx^2

Step 4 Apply the following formula

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}}$$

where

s = Standard Deviation

Σx^2 = Sum total of the squares of deviations from actual mean.

N = Number of pair of observations.

Ex.11 Calculate the standard deviation from the following data :

5 8 7 11 14

Sol. Calculation of Standard Deviation (Actual Mean Method)

Values (X)	$X - \bar{X}$ (x)	x^2
5	$5 - 9 = -4$	16
8	$8 - 9 = -1$	1
7	$7 - 9 = -2$	4
11	$11 + -9 = +2$	4
14	$14 - 19 = +5$	25
$\Sigma X = 45$		$\Sigma x^2 = 50$

Arithmetic mean$(\bar{X})$ = $\frac{\Sigma X}{N} = \frac{45}{5} = 9$

Standard deviation(s) = $\sqrt{\frac{\Sigma x^2}{N}}$

$$s = \sqrt{\frac{50}{5}} = \sqrt{10} = 3.16$$

Ans. Standard deviation = 3.16

Direct Method :

In this method, standard deviation without finding out the deviations from the actual mean. The steps involved in the Direct methe.

Step 1 Calculate the actual mean $(\bar{X})$ of the observation

Step 2 Square the observations and obtain the total, i.e., ΣX^2

Step 3 Apply the following formula

$$\sigma = \sqrt{\frac{\Sigma X^2}{N} - (\bar{X})^2}$$

where

s = Standard Deviation

ΣX^2 = Sum total of the square of observations

$\bar{X}$ = Actual Mean

N = Number of pair of observation.

Ex.12 Calculate the Standard Deviation of the data given in **Ex.11** by the Direct Method.

Sol. Calculation of Standard Deviation (Direct Method)

Values (X)	X^2
5	25
8	64
7	49
11	121
14	196
$\Sigma X = 45$	$\Sigma X^2 = 455$

Arithemtic mean$(\bar{X})$ = $\frac{\Sigma X}{N} = \frac{45}{5} = 9$

Standard Deviation(s) = $\sqrt{\frac{\Sigma X^2}{N} - (\bar{X})^2}$

$$s = \sqrt{\frac{455}{5} - (9)^2} = \sqrt{91 - 81}$$

$$= \sqrt{10} = 3.16$$

Ans. Standard Deviation = 3.16

Short-Cut Method (Assumed Mean Method)

$$s = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$$

where

s = Standard Deviation

Sd = Sum total of deviations from assumed mean

Σd^2 = Sum total of squares of deviations

N = Number of pair of observations

Ex.13 Calculate the Standard Deviation of the data given in **Ex.11** by the Short-Cut Method (Assumed Mean Method.

Sol. Calculation of Standard Deviation (Short-Cut-Method)

Values (X)	$d = X - A$ (A=7)	d^2
5	$5 - 7 = -2$	4
8	$8 - 7 = +1$	1
7(A)	$7 - 7 = 0$	0
11	$11 - 7 = +4$	16
14	$14 - 7 = +7$	49
$N = 5$	$\Sigma d = 10$	$\Sigma d^2 = 70$

Standard deviation (s) = $\sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$

$$s = \sqrt{\frac{70}{5} - \left(\frac{10}{5}\right)^2} = \sqrt{14 - 4}$$

$$= \sqrt{10} = 3.16$$

Ans. Standard deviation = 3.16

Discrete Series

Actual Mean Method

Apply the following formula

$$s = \sqrt{\frac{\Sigma fx^2}{N}}$$

Where : s = Standard Deviation

Σfx^2 = Sum total of the squared deviations multiplied by frequency

N = Number of pair of observations

Ex.15 Calculate standard deviation by the actual mean method.

Size	5	10	15	20
Frequency	2	1	4	3

Sol. Calculation of Standard Deviation (Actual Mean Method)

Size (X)	Frequency (f)	fX	$x = X - \bar{X}$	x^2	fx^2
5	3	10	–9	81	162
10	1	10	–4	16	16
15	4	60	+1	1	4
20	3	60	+6	36	108
	$N = \Sigma f = 10$	$\Sigma fX = 140$			$\Sigma fx^2 = 290$

Arthmetic mean$(\bar{X})$ = $\frac{\Sigma fX}{\Sigma f} = \frac{140}{10} = 14$

Standard Deviations = $\sqrt{\frac{\Sigma fx^2}{N}}$

Here, $\Sigma fx^2 = 290$ and $N = 10$

$$s = \sqrt{\frac{290}{10}} = \sqrt{29}$$

$$= 5.38$$

Ans. Standard Deviation = 5.38

Direct Method

Apply the following formula :

$$s = \sqrt{\frac{\Sigma fX^2}{N} - (\bar{X})^2}$$

Where

s = Standard Deviation

$\bar{X}$ = Actual Mean

Σfx^2 = Sum total of the squared observations multiplied by frequency.

N = Number of pair of observations

Ex.16 Calculate the Standard Deviation of the data given in **Ex.15** by the Direct method.

Sol. Calculation of Standard Deviation (Direct Method)

Size (X)	Frequency (f)	fX	X^2	fX^2
5	2	10	25	50
10	1	10	100	100
15	4	60	225	900
20	3	60	400	1,200
	$N = \Sigma f = 10$			$\Sigma fX^2 = 2,250$

Arithmetic mean$(\bar{X})$ = $\frac{\Sigma fX}{\Sigma f} = \frac{140}{10} = 14$

Standard deviations = $\sqrt{\frac{\Sigma fX^2}{N} - (\bar{X})^2}$

Here, $\Sigma fX^2 = 2,250$, $N = 10$, $\bar{X} = 14$

$$s = \sqrt{\frac{2,250}{10} - (14)^2} = \sqrt{225 - 196}$$

$$= \sqrt{29} = 5.38$$

Ans. Standard Deviation = 5.38

Short-Cut Method (Assumed Mean Method)

Apply the following formula :

$$s = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

Ex.17 Calculate the Standard Deviation of the data given in **Ex.15** by the Short-Cut Method (Assumed Mean Method).

Sol. Calculation of Standard Deviation (Short-Cut Method)

Size (X)	Frequency (f)	$d = X - A$ (A=10)	fd	d^2	fd^2
5	2	–5	–10	25	50
10(A)	1	0	0	0	0
15	4	+5	20	25	100
20	3	+10	30	100	300
	$N = \Sigma f = 10$		$\Sigma fd = 40$		$\Sigma fd^2 = 450$

Standard Deviation(s) = $\sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$

Here, $\Sigma fd^2 = 450, N = 10, \Sigma fd = 40$

$$s = \sqrt{\frac{450}{10} - \left(\frac{40}{10}\right)^2} = \sqrt{45 - 16}$$

$$= \sqrt{29} = 5.38$$

Ans. Standard deviation = 5.38

Step Deviation Method

Apply the following formula : $\sigma = \sqrt{\frac{\sum fd^{1_2}}{N} - \left(\frac{\sum fd^1}{N}\right)^2} \times \bar{C}$

Where :

s = Standard Deviation

$\Sigma fd'^2$ = Sum total of the squared step deviations multiplied by frequency

$\Sigma fd'$ = Sum total of step deviations mutliplied by frequency

C = Common Factor

N = Number of pair observations.

Ex.18 Calculate the Standard Deviation of the data given in **Ex.15** by the Step Deviation Method.

Sol. Calculation of Standard Deviation (Step deviation method)

Size (X)	Frequency(f)	$d = X - A$ A=10	$d' = \frac{X-A}{C}$ C=5	fd'	d'^2	fd'^2
5	2	–5	–2	–2	1	2
10 (A)	1	0	0	0	0	0
15	4	5	+1	4	1	4
20	3	10	+2	6	4	12
	$N = \Sigma f = 10$			$\Sigma fd' = 8$		$\Sigma fd'^2 = 18$

Standard deviation(s) = $\sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$

Here $\Sigma fd'^2 = 18,\ N = 10;\ \Sigma fd' = 8;\ C = 5$

$$s = \sqrt{\frac{18}{10} - \left(\frac{8}{10}\right)^2} \times 5 = \sqrt{1.8 - .64} \times 5$$

$$= 5.38$$

Ex.19 Calculate the standard deviation from the given data by : (i) Actual Mean Method; (ii) Direct Method; (iii) Short-Cut Method; (iv) Step Deviation Method.

X :	2	4	6	8
f :	3	1	4	2

Sol. Calcuation of Standard Deviation

Actual mean Method

X	f	fX	$x = X - \bar{X}$	x^2	fx^2
2	3	6	–3	9	27
4	1	4	–1	1	1
6	4	24	+1	1	4
8	2	16	+3	9	18
	$N = \Sigma f = 10$	$\Sigma fX = 50$			$\Sigma fx^2 = 50$

$$\bar{X} = \frac{\Sigma fX}{\Sigma f} = \frac{50}{10} = 5$$

$$s = \sqrt{\frac{\Sigma fx^2}{N}} = \sqrt{\frac{50}{10}}$$

$$s = \sqrt{5} = 2.236$$

Direct Method :

X	f	fX	X^2	fX^2
2	3	6	4	12
4	1	4	16	16
6	4	24	36	144
8	2	16	64	128
	$N = \Sigma f = 50$			$\Sigma fX^2 = 300$

$$\bar{X} = \frac{\Sigma fX}{\Sigma f} = \frac{50}{10} = 5$$

$$s = \sqrt{\frac{\Sigma fX^2}{N} - (\bar{X})^2} = \sqrt{\frac{300}{10} - (5)^2}$$

$$s = \sqrt{5} = 2.236$$

Ans. Standard Deviation = 2.236

Short Cut Method

X	f	$d = X - A$ (A=4)	fd	d^2	fd^2
2	3	–2	–6	4	12
4 (A)	1	0	0	0	0
6	4	+2	+8	4	16
8	2	+4	+8	16	32
	$N = 10$		$\Sigma fd = 10$		$\Sigma fd^2 = 60$

$$s = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

$$s = \sqrt{\frac{60}{10} - \left(\frac{10}{10}\right)^2}$$

$$s = \sqrt{5} = 2.236$$

Step Deviation Method

X	f	$d = X - A$ (A=4)	$d' = \frac{X - A}{C}$ (C=2)	fd'	d'^2	fd'^2
2	3	–2	–1	–3	1	3
4(A)	1	0	0	0	0	0
6	4	+2	+1	+4	1	4
8	2	+4	+2	+4	4	8
	$N = 10$			$\Sigma fd' = 5$		$\Sigma fd'^2 = 15$

$$s = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$$

$$s = \sqrt{\frac{15}{10} - \left(\frac{5}{10}\right)^2} \times 2$$

$$s = \sqrt{1.25} \times 2 = 2.236$$

Ans. Standard Deviation = 2.236

Continuous Series

Apply the following formula :

$$s = \sqrt{\frac{\Sigma fx^2}{N}}$$

Where s = Standard Deviation

Σfx^2 = Sum total of the squared deviations multiplied by frequency

N = Number of pair of observations.

Ex.20 Calculate standard deviation by the actual mean method :

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of students	4	3	6	5	2

Sol. Calculation of Standard Deviation (Actual Mean Method)

Marks X	No. of students f	Mid – point m	fm	$x = m - \bar{X}$	x^2	fx^2
0 – 10	4	5	20	–19	361	1,444
10 – 20	3	15	45	–9	81	243
20 – 30	6	25	150	+1	1	6
30 – 40	5	35	175	+11	221	605
40 – 50	2	45	90	+21	441	882
	$N = \Sigma f = 20$		$\Sigma fm = 480$			$\Sigma fx^2 = 3{,}180$

Arithmetic mean $(\bar{X})$ = $\frac{\Sigma fm}{\Sigma f} = \frac{480}{20} = 24$

Standard deviations = $\sqrt{\frac{\Sigma fx^2}{N}}$

Here, $\Sigma fx^2 = 3{,}180$ and $N = 20$

$$s = \sqrt{\frac{3{,}180}{20}} = \sqrt{159}$$

$$= 12.61 \text{ marks}$$

Ans. Standard Devition = 12.61 marks.

Direct Method

Apply the following formula

$$s = \sqrt{\frac{\Sigma fm^2}{N} - (\bar{X})^2}$$

Where : s = Standard Deviation

$\bar{X}$ = Actual Mean

Σfm^2 = Sum total of the squared mid-points multiplied by frequency

N = Number of pair of observations

Ex.21 Calculate of Standard Deviation (Direct Method)

Marks X	No. of students f	Mid – points m	fm	m^2	fm^2
0 – 10	4	5	20	25	100
10 – 20	3	15	45	225	675
20 – 30	6	25	150	625	3,750
30 – 40	5	35	175	1,225	6,125
40 – 50	2	45	90	2,025	4,050
	$N = \Sigma f = 20$		$\Sigma fm = 480$		$\Sigma fm^2 = 14{,}700$

Arithmetic mean $(\bar{X})$ = $\frac{\Sigma fm}{\Sigma f} = \frac{480}{20} = 24$

Standard deviations = $\sqrt{\frac{\Sigma fm^2}{N} - (\bar{X})^2}$

Here, $\Sigma fm^2 = 14,700;\ N = 20;\ \Sigma fm = 480$

$$s = \sqrt{\frac{14,700}{20} - (20)^2} = \sqrt{735 - 576}$$

$$= \sqrt{159} = 12.61$$

Ans. Standard Deviation = 12.61 Marks.

Short-Cut Method (Assumed Mean)

Apply the following formula :

$$s = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

Where : s = Standard Deviation

Σfd^2 = Sum total of the squared deviations multiplied by frequency

Sfd = Sum total of deviations multiplied by frequency

N = Number of pair of observations

Ex.22 Calculate of Standard Deviation (Shot-cut Method)

Marks X	No. of students f	Mid – point m	d = m – A A=25	fd	d^2	fd^2
0 – 10	4	5	–20	–80	400	1,600
10 – 20	3	15	–10	–30	100	300
20 – 30	6	25(A)	0	0	0	0
30 – 40	5	35	+10	+50	100	500
40 – 50	2	45	+20	+40	400	800
	$N = \Sigma f = 20$			$\Sigma fd = -20$		$\Sigma fd^2 = 3,200$

Standard deviation(s) = $\sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$

Here, $\Sigma fd^2 = 3,200;\ N = 20;\ \Sigma fd = -20$

$$s = \sqrt{\frac{3,200}{20} - \left(\frac{-20}{20}\right)^2} = \sqrt{160 - 1}$$

$$= \sqrt{159} = 12.61 \text{ Marks}$$

Ans. Standard Deviation = 12.61 Marks

Step Deviation Method

Apply the following formula :

$$s = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$$

Where : s = Standard Deviation

$\Sigma fd'^2$ = Sum total of the squared step deviations multiplied by frequency

$\Sigma fd'$ = Sum total of step deviations multiplied by frequency

C = Common Factor

N = Number of pair of observations

Ex.23 Calculate standard deviation by the actual mean method :

Class	1 – 3	3 – 5	5 – 7	7 – 9	9 – 11	11 – 13	13 – 15
Frequency	1	9	25	35	17	10	3

Sol. Calculation of Standard Deviation (Actual Mean Method)

Class X	Frequency f	Mid – point m	fm	$x = m - \bar{X}$	x^2	fx^2
1 – 3	1	2	2	–6	36	36
3 – 5	9	4	36	–4	16	144
5 – 7	25	6	150	–2	4	100
7 – 9	35	8	280	0	0	0
9 – 11	17	10	170	+2	4	68
11 – 13	10	12	120	+4	16	160
13 – 15	3	14	42	+6	36	108
	$N = \Sigma f = 100$		$\Sigma fm = 800$			$\Sigma fx^2 = 616$

Arithmetic mean $(\bar{X}) = \frac{\Sigma fm}{\Sigma f} = \frac{800}{100} = 8$

Standard Deviation $(s) = \sqrt{\frac{\Sigma fx^2}{N}}$

Here, $\Sigma fx^2 = 616$ and $N = 100$

$$s = \sqrt{\frac{616}{100}} = \sqrt{6.16} = 2.48$$

Ans. Standard deviation = 2.48

Ex.24 Calculate the standard deviation from the following data by : (i) Actual Mean Method (ii) Direct Method; (iii) Short-Cut Method; (iv) Step Deviation Method.

X :	0 – 10	10 – 20	20 – 30	30 – 40
f :	2	3	4	1

Sol. Calcuation of Standand Deviation

Actual Mean Method

X	f	m	fm	$x = m - \bar{X}$	x^2	fx^2
0 – 10	2	5	10	–14	196	392
10 – 20	3	15	45	–4	16	48
20 – 30	4	25	100	+6	36	144
30 – 40	1	35	35	+16	256	256
	N = 10		$\Sigma fm = 190$			$\Sigma fx^2 = 840$

$$\bar{X} = \frac{\Sigma fm}{\Sigma f} = \frac{190}{10} = 19$$

$$s = \sqrt{\frac{\Sigma fx^2}{N}} = \sqrt{\frac{840}{10}}$$

$$s = \sqrt{84} = 9.165$$

Direct Method

X	f	m	fm	m^2	fm^2
0 – 10	2	5	10	25	50
10 – 20	3	15	45	225	675
20 – 30	4	25	100	625	2,500
30 – 40	1	35	35	1,225	1,225
	N = 10		$\Sigma fm = 190$		$\Sigma fm^2 = 4,450$

$$s = \sqrt{\frac{\Sigma fm^2}{N} - (\bar{X})^2}$$

$$= \sqrt{\frac{4,450}{10} - (19)^2}$$

$$s = \sqrt{84} = 9.165$$

Ans. Standard Deviation = 9.165

Short-cut Method

X	f	m	d	fd	d^2	fd^2
0 – 10	2	5	–10	–20	100	200
10 – 20	3	15(A)	0	0	0	0
20 – 30	4	25	+10	+40	100	400
30 – 40	1	35	+20	+20	400	400
	N = 10			Σfd = 40		$\Sigma fd^2 = 1,000$

$$s = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} = \sqrt{\frac{1000}{10} - \left(\frac{40}{10}\right)^2}$$

$$s = \sqrt{84} = 9.165$$

Step Deviation Method

X	f	m	d	d′	fd′	d'^2	fd'^2
0 – 10	2	5	–10	–1	–2	1	2
10 – 20	3	15(A)	0	0	0	0	0
20 – 30	4	25	+10	+1	+4	1	4
30 – 40	1	35	+20	+2	+2	4	4
	N = 10				Σfd′ = 4		$\Sigma fd'^2 = 10$

$$s = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$$

$$s = \sqrt{\frac{10}{10} - \left(\frac{4}{10}\right)^2} \times 10$$

$$s = \sqrt{0.84} \times 10 = 9.165$$

Ex25. From the following information, find standard deviation of x and y variables :

$\sum x = 235,$ $\qquad \sum y = 250$

$\sum x^2 = 6750,$ $\qquad \sum y^2 = 6840$

N = 10

Solution. Calculation of standard deviation of x and y variables.

x	y
$\sigma x = \sqrt{\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2}$	$\sigma y = \sqrt{\frac{\sum y^2}{N} - \left(\frac{\sum y}{N}\right)^2}$
$= \sqrt{\frac{6750}{10} - \left(\frac{235}{10}\right)^2}$	$= \sqrt{\frac{6840}{10} - \left(\frac{250}{10}\right)^2}$
$= \sqrt{675 - (23.5)^2}$	$= \sqrt{684 - (25)^2}$
$= \sqrt{675 - 552.25}$	$= \sqrt{684 - 625}$
$= \sqrt{122.75}$	$= \sqrt{59}$
$= 11.079$	$= 7.68$

VARIANCE

Variance is another measure based on standard deviation. By variance, we mean the square of the standard deviation.

Variance $= \sigma^2$

Standard Deviation (s) = $\sqrt{\text{variance}}$

Smaller the value of variance (σ^2), lesser is the variability or greater the consistency and vice-versa.

RELATIVE MEASURES OF STANDARD DEVIATION

Coefficient of Standard Deviation

The standard deviation is an absolute measure of dispersion. To compare the variability in two series, a relative measure of standard is found out. It is known as 'coefficient of standard deviation' or 'standard coefficient of dispersion'.

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

Coefficient of Variation

When two or more groups of similar data are to be compared with respect to stability for uniformly or consistency or homogeneity), coefficient of variation the most appropriate measure.

It incidcates the relationship between the standard deviation and the arithmetic mean expressed in terms of percentage.

$$C.V = \frac{\sigma}{\overline{X}} \times 100$$

Where

C.V. = Coefficient of Variation, s = Standard Deviation, $\overline{X}$ = Arithmetic Mean.

MISCELLANEOUS PRACTICALES

Ex.25 Abatsman is to be selected for a cricket team. The choice is between X and Y on the basis of their five previous scores, which are :

X	25	85	40	80	120
Y	50	70	65	45	80

Which batsman should be selected, we want :

(i) a higher sum getter, or

(ii) a more reliable batsman, we will calculate average of both the batsman

Sol. (i) To determine Higher Run getter batsman, we will calculate average of both the batsman.

Batsman X

(X_x)	$X_x - (X_x)$	x_x^2
25	–45	2,025
85	+15	225
40	–30	900
80	+10	100
120	+50	2,500
$\Sigma X_x = 350$		$\Sigma x_x^2 = 5,750$

Batsman Y

(X_y)	$X_y - (X_y)$	x_y^2
50	–12	144
70	+8	64
65	+3	9
45	–17	286
80	–18	324
$\Sigma X_x = 310$		$\Sigma x_x^2 = 830$

Batsman X

Arithmetic Mean

$$\overline{X} = \frac{\Sigma X}{N}$$

SX = 350, N = 5

$$\therefore \quad \overline{X} = \frac{350}{5} = 70$$

Batsman Y

$$\overline{Y} = \frac{\Sigma Y}{N}$$

SX = 350, N = 5

$$\therefore \quad \overline{Y} = \frac{310}{5} = 62$$

∴ Average score = 70 runs

∴ Average score = 62 runs

Standard Deviation

$\sigma = \sqrt{\frac{\Sigma x^2}{N}}$

$Sx^2 = 5750$ and N = 5

∴ $\sigma = \sqrt{\frac{5750}{5}} = \sqrt{1150}$

= 33.91 Runs

$\sigma = \sqrt{\frac{\Sigma y^2}{N}}$

$Sy^2 = 830$ and N = 5

∴ $\sigma Y = \sqrt{\frac{830}{5}} = \sqrt{166}$

= 12.88 Runs

Coefficient of Standard Deviation

Coefficient of S.D. = $\frac{\sigma x}{\overline{X}}$

s = 33.91 and $\overline{X}$ = 70

∴ Coeff. of SD $= \frac{33.91}{70}$

= 0.484

Coefficient of S.D. = $\frac{\sigma y}{\overline{Y}}$

s = 12.88 and $\overline{Y}$ = 62

∴ Coeff. of S.D. = $\frac{12.88}{62}$

= 0.207

Variance

$sx^2 = (s^2) = \frac{\Sigma\left(X - \overline{X}\right)^2}{N} = \frac{\Sigma x^2}{N}$

$S(X - \overline{X})^2 = Sx^2 = 5750$

Here, $X - \overline{X} = x$

∴ $\sigma x^2 = \frac{5750}{5}$

= 1150 Runs

$sy^2 = \frac{\Sigma\left(Y - \overline{Y}\right)^2}{N} = \frac{\Sigma y^2}{N}$

$S(Y - \overline{Y})^2 = Sy^2 = 830$

Here, $Y - \overline{Y} = y$

∴ $\sigma y^2 = \frac{830}{5}$

= 166 Runs

Coefficient of Variation

$C.V._x = \frac{\sigma x}{\overline{X}} \times 100$

$s_x = 33.91$ and $\overline{X}$ = 70

∴ C.V. = $\frac{33.91}{70} \times 100$

= 48.44%

$C.V._y = \frac{\sigma y}{\overline{Y}} \times 100$

$s_y = 33.91$ and $\overline{Y}$ = 62

∴ C.V. = $\frac{12.88}{62} \times 100$

= 20.77 %

(i) Batsman X should be selected as a higher run scorer as his average score (70 runs) is greater than the average score of Y (62 runs).

(ii) Batsman Y is a more reliable batsman in the team because his coefficient variation (CV. = 20.77%) is less than that of batsman X (C.V. = 48.44%)

Ex.26 Calculate variance and coefficient of variation from the followig data :

Size	4	5	6	7	8	9	10
Frequency	3	7	22	60	85	32	8

Sol. Calculation of Variance and Coefficient of Variation

Size X	Frequency f	X – 7 d	fd	fd^2
4	3	–3	–9	27
5	7	–2	–14	28
6	22	–1	–22	22
7	60	0	0	0
8	85	+1	+85	85
9	32	+2	+64	128
10	8	+3	+24	71
	N = 217		$\Sigma fd = 128$	$\Sigma fd^2 = 362$

Variance

$$\text{Variance } (s^2) = \frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd^2}{N}\right)^2$$

Here, Sfd^2 = 362, Sfd = 128 and N = 217

∴ $\text{Variance } (s^2) = \frac{362}{217} - \left(\frac{128}{217}\right)^2$

= $1.668 - (0.589)^2$

= 1.668 – 0.347 = 1.32

Coefficient of Variation

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

Let us calculate s and $\overline{X}$,

$$\sigma = \sqrt{\text{Variance}} = \sqrt{1.32} = 1.15$$

$$\overline{X} = A + \frac{\Sigma fd}{N} = 7 + \frac{128}{217} = 7 + 0.59 = 7.59$$

Applying the formula, we get

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

Here, s = 1.15 and $\overline{X}$ = 7.59

∴ $C.V. = \frac{1.15}{7.59} \times 100 = 0.1515 \times 100$

= 15.15%

Continuous Series

Ex.27 To check the quality of two bulbs and their life in burning hours was estimated as under for 100 bulbs of each brand.

Life (in hrs.)	No. of bulbs Brand A	No. of bulbs Brand B
0 – 50	15	2
50 – 100	20	8
100 – 150	18	60
150 – 200	25	25
200 – 250	22	5
Total	100	100

(i) Which brand gives higher life?

(ii) Which brand is more dependable?

Sol.

Brand A

Life (in hrs.) X	No. of bulbs (f)	Mid – point m	m – 125 d	$\frac{m-125}{50}$ d′	fd′	fd′^{2}
0 – 50	15	25	–100	–2	–30	60
50 – 100	20	75	–50	–1	–20	20
100 – 150	18	125	0	0	0	0
150 – 200	25	175	+50	+1	+25	25
200 – 250	22	225	+100	+2	+44	88
	N = 100				Σfd′ = 19	Σfd′^{2} = 193

Brand B

Life (in hrs.) X	No. of bulbs (f)	Mid – point m	m – 125 d	$\frac{m-125}{50}$ d′	fd′	fd′^{2}
0 – 50	2	25	–100	–2	–4	8
50 – 100	8	75	–50	–1	–8	8
100 – 150	60	125	0	0	0	0
150 – 200	25	175	+50	+1	25	25
200 – 250	5	225	+100	+2	10	20
	N = 100				Σfd′ = 23	Σfd′^{2} = 61

Coefficient of Variation (Brand A)

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

Coefficient of Variation (Brand B)

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

Let us calculate first s and $\overline{X}$.

Standard Deviation

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$$

Here, Sfd´² = 193,

Sfd´² = 19

N = 100

and C = 50

$$\therefore \quad \sigma = \sqrt{\frac{193}{100} - \left(\frac{19}{100}\right)^2} \times 50$$

$$= \sqrt{1.93 - (0.19)^2} \times 50$$

$$= \sqrt{1.8939} \times 50$$

= 1.376 × 50

= 68.8 hrs.

Arithmetic Mean

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times C$$

Here, A = 125,

Sfd´ = 19,

N = 100

and C = 50

$$\therefore \quad \overline{X} = 125 + \frac{19}{100} \times 50$$

= 125 + (0.19) × 50

= 125 + 9.5 = 134.5 hrs.

Applying the formula, now we get get

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

where, s = 68.8

and $\overline{X}$ = 134.5

$$\therefore \quad C.V. = \frac{68.8}{134.5} \times 100 = 51.15\%$$

Standard Deviation

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$$

Here, Sfd´² = 61,

Sfd´² = 23

N = 100

and C = 50

$$\therefore \quad \sigma = \sqrt{\frac{61}{100} - \left(\frac{23}{100}\right)^2} \times 50$$

$$= \sqrt{0.61 - (0.23)^2} \times 50$$

$$= \sqrt{0.5571} \times 50$$

= 0.746 × 50

= 37.32 hrs.

Arithmetic Mean

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times C$$

Here, A = 125,

Sfd´ = 23,

N = 100

and C = 50

$$\therefore \quad \overline{X} = 125 + \frac{23}{100} \times 50$$

= 125 + (0.33) × 50

= 125 + 11.5 = 136.5 hrs.

Applying the formula, now we

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

where, s = 37.32

and $\overline{X}$ = 136.5

$$\therefore \quad \frac{37.32}{136.5} \times 100 = 27.34\%$$

a. Since the average life of bulbs of brand B (136.5 hrs) is greater than that of brand A (134.5 hrs), therefore the bulbs of brand B give a higher life.

b. Since CV. of bulbs of brand B (27.34%) is less than of brand A (51.15%), therefore the bulbs of B are more dependable.

Ex.28 The number of employees, wages per employee and the variance of wages per employee for two factories are given below :

	Factory A	Factory B
No. of Employees	50	100
Average wage per employee per day (Rs)	120	85
Variance of wages per employee per day (Rs)	9	16

a. In which factory is there greater variation in the distribution of wages per employee ?

b. Suppose in factory B, the wages of an employee are wrongly noted as Rs 120 instead of Rs 100.

What would be the corrected variance for factory B ?

Ans.

a. Calculation of Coefficient of Variation :

Factory A

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

Here, $\overline{X}$ = 120 and s = $\sqrt{9}$

$$\therefore \quad C.V. = \frac{3}{120} \times 100$$

= 2.5 %

Factory B

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

Here, $\overline{X}$ = 85 and s = $\sqrt{16}$

$$\therefore \quad C.V. = \frac{4}{85} \times 100$$

= 4.7 %

Since coefficient of variation is higher for factory B, there is greater variation in the distribution of wages in factory B,

b. Correcting Mean and Variation :

Mean

$$\overline{X} = \frac{\Sigma X}{N} \text{ or } N\overline{X} = \Sigma X$$

For Factory B :

N = 100 and $\overline{X}$ = 85

$\therefore$ 100 × 85 = 8500

It is not correct SX

Corrected SX = 8500 – 120 + 100 = Rs. 8480

$$\therefore \quad \text{Corrected } \overline{X} = \frac{\Sigma X}{N} = \frac{8480}{100} = 84.80$$

Variance = s^2

$$\sigma^2 = \frac{\Sigma X^2}{N} - \left(\overline{X}\right)^2$$

Here, s^2 = 16, $\overline{X}$ = 85 and N = 100

$$\therefore \quad 16 = \frac{\Sigma X^2}{100} - (85)^2$$

SX^2 = 1600 + 722500 = 724100

It is not correct SX^2

Corrected SX^2 = 724100 – $(120)^2$ + $(100)^2$

= 724100 – 14400 + 100000 = 719700

Corrected Variance = $\frac{\text{Corrected } \Sigma X^2}{N} - \left(\text{Corrected } \bar{X}\right)^2$

$= \frac{719700}{100} - (84.8)^2$

= 7197 - 7191.04 = Rs 5.96

Ex.29 The sum of 10 values is 100 and the sum of their squares is 1090. Find the coefficient of variation.

Ans. We are given,

N = 10, SX = 100, and SX^2 = 1090

Coefficient of variation (C.V.) = $\frac{\sigma}{X} \times 100$

Apply the following formula to get Mean ($\bar{X}$)

$\bar{X} = \frac{\Sigma X}{N} \quad \therefore \quad N\bar{X} = \Sigma X$

$10\bar{X} = 100$

$\therefore \quad \bar{X} = \frac{100}{10} = 10$

Apply the following formula to obtain standard deviation (s) by direct method.

$\sigma = \sqrt{\frac{\Sigma X^2}{N} - \left(\bar{X}\right)^2}$

$\sigma^2 = \frac{\Sigma X^2}{N} - \left(\bar{X}\right)^2$

$= \frac{1090}{10} - (10)^2$

= 109 – 100

= 9

$\therefore \quad \sigma = \sqrt{9} = 3$

Therefore, $\text{C.V.} = \frac{\sigma}{\bar{X}} \times 100$

$= \frac{3}{10} \times 100$

= 30

Thus, $\bar{X}$ = 10 and C.V. = 30

Ex.30 Calculate the standard deviation if coefficient of variation is 23.21, number of items is 110 and mean is 21.

Sol. Coefficient of Variation = $\frac{\sigma}{\bar{X}} \times 100$

$$23.21 = \frac{\sigma}{21} \times 100$$

$$s = \frac{23.21 \times 21}{100} = 48.74$$

Ans. Standard deviation (s) = 4.874

Ex.31 Particulars regarding the income of two villages are given below :

	Village X	Village Y
Number of People	500	600
Average Income	Rs. 186	Rs. 175
Standard Deviation	9	10

(i) What is the average income of the village X and Y taken together ?

(ii) Which village has a larger income ?

(iii) In which village, variation in income is greater ?

Sol. (i) Average income of the village X and Y taken together (Combined Mean)

Combined Mean

$$(\bar{X}_{X,Y}) = \frac{N_X \bar{X}_X + N_Y \bar{X}_Y}{N_X + N_Y}$$

Given $\bar{X}_X = 186$, $\bar{X}_Y = 175$, $N_X = 500$, $N_Y = 600$

$$(\bar{X}_{X,Y}) = \frac{(500 \times 186) + (600 \times 176)}{500 + 600} = \frac{1,98,000}{1,100} = \text{Rs. } 180$$

(ii) Income of village X = 500 × 186 = Rs. 93,000

Income of village Y = 600 × 175 = Rs. 1,05,000

Thus, Village Y has a larger income.

(iii) $$C.V._X = \frac{\sigma_X}{\bar{X}} \times 100 = \frac{9}{180} \times 100 = 4.84\%$$

$$C.V._Y = \frac{\sigma_Y}{\bar{Y}} \times 100 = \frac{10}{175} \times 100 = 5.71\%$$

There is more variability in Village Y.

Ans. (i) Average income of the village X and Y taken together = Rs. 180; (ii) Village Y has a larger income; (iii) In Village Y, variation in income is greater.

Ex.32 For a group of 200 candicates, the mean and standard were found to be 40 and 15. Later on it was discovered that the score 43 was misread as 53. Find the correct mean and standard deviations corresponding to the corrected figure.

Sol. Calculation of Correct Mean

$$\bar{X} = \frac{\Sigma X}{N}$$

or $$SX = \bar{X}N$$

So, $\Sigma X = 40 \times 20 = 8,000$

But 8,000 is a wrong value as one score was misread as 53 instead of 43.

So, Correct $\Sigma X = 8,000$ – incorrect item + correct item

= 8,000 – 53 + 43 = 7,990

$$\text{Correct } \bar{X} = \frac{\Sigma X}{N} = \frac{7,990}{200} = 39.95$$

Calcualtion of Correct Standard deviation

$$s = \sqrt{\frac{\Sigma X^2}{N} - (\bar{X})^2}$$

$$15 = \sqrt{\frac{\Sigma X^2}{200} - (40)^2}$$

$$15 = \sqrt{\frac{\Sigma X^2}{200} - 1600}$$

Squaring both the sides

$$225 = \frac{\Sigma X^2}{200} - 1600$$

$$225 \times 200 = \Sigma X^2 - 1,600 \times 200$$

$$\Sigma X^2 = 3,20,000 + 45,000 = 3,65,000$$

But, it is incorrect value

Correct ΣX^2 = Incorrect ΣX^2 – (Incorrect item)^2 + (Correct item)^2

= 3,65,000 – (53)^2 + (43)^2

= 3,65,000 – 2,809 + 1,849 = 3,64,040

$$\text{Correct } s = \sqrt{\frac{\text{correct } \Sigma X^2}{N} - (\text{Correct } \bar{X})^2} = \sqrt{\frac{3,64,040}{200} - (39.95)^2}$$

$= \sqrt{1,820.2 - 1,596} = \sqrt{224.2} = 14.97$

Ans. Correct Mean=39.95 marks; Correct Standard Deviation = 14.97 marks.

Combined Standard Deviation

As we can calculate mean of two or more than two series, we can also compute combined standard deviation of two or more than two series. The formula in case of two series :

$$\sigma_{1,2} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$$

Where $\sigma_{1,2}$ = Combined standard deviation of two groups

σ_1 = Standard deviation of first group

σ_2 = Standard deviation of second group

$d_1^2 = (\bar{X}_1 - \bar{X}_{1,2})^2$

$d_2^2 = (\bar{X}_2 - \bar{X}_{1,2})^2$

$\bar{X}_{1,2}$ = Arithmetic mean of first group

$\bar{X}_1$ = Arithmetic mean of first group

$\bar{X}_2$ = Arithmetic mean of second group

N_1 = Number of observations of first group

N_2 = Number of observations of second group

This formula can be extended upto N number of series. If there are three series, then the combined standard deviation is :

$$\sigma_{1,2,3} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^3 + N_1d_1^2 + N_2d_2^2 + N_3d_3^2}{N_1 + N_2 + N_3}}$$

Where, $d_1^2 = (\bar{X}_1 - \bar{X}_{1,2,3})^2$, $d_2^2 = (\bar{X}_2 - \bar{X}_{1,2,3})^2$ and $d_3^2 = (\bar{X}_3 - \bar{X}_{1,2,3})^2$

Ex.33 Find the combined standard deviation from the following data :

	Boys	Girls
Number	30	20
Mean	20	30
Standard deviation	4	5

Sol. To calculate combined standard deviation, we will have to first calculate combined mean

Combined Mena$(\bar{X}_{1,2})$= $\dfrac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$

Given, $\bar{X}_1 = 20,\ \bar{X}_2 = 30,\ N_1 = 30,\ N_2 = 20$

$$= \frac{(30 \times 20) + (20 \times 30)}{30 + 20}$$

$$= \frac{1,200}{50} = 24$$

$$X_{1,2} = 24$$

Calculatin of Combined Standard Deviation

$$\sigma_{1,2} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$$

$d_1 = \bar{X}_1 - \bar{X}_{1,2} = 20 - 24$

$d_1 = -4$

$d_2 = \bar{X}_2 - \bar{X}_{1,2} = 30 - 24$

$d_2 = 6$

Putting the values of $N_1 = 30,\ \sigma_1 = 4,\ N_2 = 20,\ \sigma_2 = 5,\ d_1 = -4,\ d_2 = 6$

$$= \sqrt{\frac{30(4)^2 + 20(5)^2 + 30(-4)^2 + 20(6)^2}{30 + 20}}$$

$$= \sqrt{\frac{480 + 500 + 480 + 720}{50}}$$

$$= \sqrt{43.6} = 6.6$$

Ans. Combined Standard Deviation = 6.6

Ex.34 Mean and standard deviations of two distributions of 100 and 150 items are 50, 5, and 40, 6 respectively. Find the standard deviation of all the 250 items taken together.

Sol. Combined Mean $(\bar{X}_{1,2})$ = $\dfrac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$

Given $\bar{X}_1 = 50,\ \bar{X}_2 = 40,\ N_1 = 100,\ N_2 = 150$

$$= \frac{(100 \times 50) + (150 \times 40)}{100 + 150}$$

$$= \frac{11,00}{250}$$

$$\bar{X}_{1,2} = 44$$

Calculation of Combined Standard Deviation

$$\sigma_{1,2} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$$

$$d_1 = \bar{X}_1 - \bar{X}_{1,2} = 50 - 44$$

$$d_1 = 6$$

$$d_2 = \bar{X}_2 - \bar{X}_{1,2} = 40 - 44$$

$$d_2 = -4$$

Putting the values of $N_1 = 100,\ \sigma_1 = 5,\ N_2 = 150,\ \sigma_2 = 6,\ d_1 = 6,\ d_2 = -4$

$$= \sqrt{\frac{100(5)^2 + 150(6)^2 + 100(6)^2 + 150(-4)^2}{100 + 150}}$$

$$= \sqrt{\frac{2,500 + 5,400 + 3,600 + 2,400}{250}}$$

$$= \sqrt{55.6} = 7.456$$

Ans. Combined Sandard Deviation = 7.456

MERITS, DEMERITS AND USES OF STANDARD DEVIATION

Merits

1. Based on all Values
2. Rigidly Defined
3. Less effect of fluctuation in sampling
4. Algebraic Treatment
5. Better mathematical process

Demerits

1. Difficult to Compute
2. More stress on extreme items

3. Depend upon units of measurement

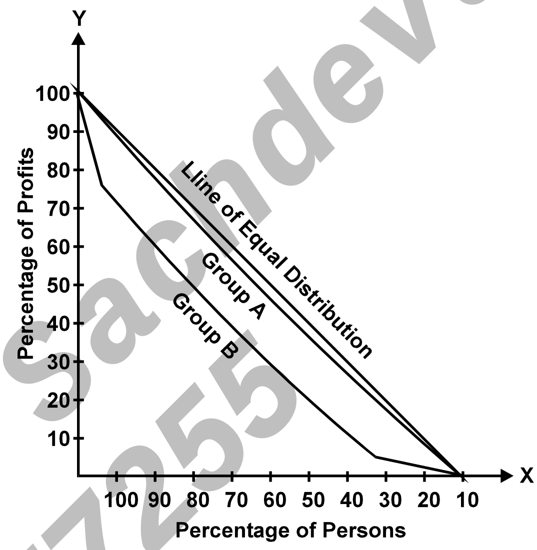
Comparision Between Mean Deviation and Standard Deviation.

S.No.	Mean Deviation	Standard Deviation
1.	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.
2.	Mean deviation can be computed from mean, median or mode and its value offer in thiese cases (unless the distribution is normal).	The standard deviation is always calculated from the arithmetic mean.
3.	Mean Deviation does not take into account the algebraic signs (plus of minus) in its calculation which is ilogical.	In calculation of standard deviation, the deviation, the deviations are square. So, the plus and minus signs need not be omitted.
4.	Mean Deviation is not capable of further algebraic treatment as it considers only the absolute values. So, it is not possible to compute combined mean deviation.	Standard deviation is capable of futher algebraic treatment, i.e., we can find the combined standard deviation of two or more series.

LORENZ CURVE

1. The size of items or mid-points of class-intervals are cumulated. The last cumulative total is considereed as equal to 100 and then percentage are obtained for these various cumulative values.
2. Frequencies are also made cumulative. Calculate the percentage for each cumulative frequency by taking the total equal to 100.
3. The percentage of cumulative frequencies are represented on the X-axis and the percentage of the cumulative totals of the values of the variable are represented on the Y-axis.
4. We start on the X-axis from 100 and reach 0. On the contrary, we start from 0 and reach 100 on the Y-axis.
5. Zero on X-axis and 100 on Y-axis are joined by a line. This line will form 45° angle. This line is known as "line of Equal Distribution" or "Equality line."

6. Plot the various points corresponding to the values of the variables X and Y and then join these points with a smooth free hand curve. The curve so obtained shows the actual distribution. This curve to known as Lorenz Curve.
7. The greater is the distance between the curve and line of equal distribution, the greater is the dispersion.



The Lorenz curve always lies below the line of equal distribution, unless the distribution is uniform. If the distribution is uniform, the Lorenz Curve will coincide with the line of equal distribution.

The curve taht is fartehst from the diagonal line represents greatest inequality.

Ex.31 From the folloiwng table, draw Lorenz curve for number of persons in Group A and B and interpret the results.

Profit Earned (Rs. in '000)	20	30	40	50	60
Group A	6	8	10	12	14
Group B	15	10	9	11	5

Sol.

Profit Earned (Rs. in '000)	Cumulative Profit (Rs.)	Cumulative %	Group A No. of Persons	Group A Cumulative (nos.)	Group A Cumulative (%)	Group B No. of Persons	Group B Cumulative No.	Group B Cumulative (%)
20	20	10	6	6	12	15	15	30
30	50	25	8	14	28	10	25	50
40	90	45	10	24	48	9	34	68
50	140	70	12	36	72	11	45	90
60	200	100	14	50	100	5	50	100

The Lorenz Curve of Group B is farthest from the line of equal distribuiton. So, Group B shows greater inequality as compared to Group A.

MERITS AND DEMERITS OF LORENZ CURVE

Merits

1. Lorenz Curve is attractive and it gives a rough idea of extent of dispersion.
2. With the help of Lorenz curve, it becomes easy to compare two or more series.

Demerits

1. Lorenz curve gives only a relative idea of the dispersion as compared with the line of equal distribution. It does not provide us any numerical value of the variability for the given distribution.
2. The method of drawing Lorenz Curve is very difficult.

UNSOLVED PRATICALS

Q1. Find range and coefficient of range from the weekly wage (in Rs.) of 10 workers of a factory:

310, 350, 420, 105, 115, 290, 245, 450, 300, 375

[Range Rs. 345 ; Coff. of average = 0.62]

Q2. From the following data calculate range and coefficient of range

Marks	10	20	30	40	50	60	70
No. of students	8	12	7	30	10	5	2

[R = 60 Marks ; Coff. of Range = 0.35]

Q3. Following are the marks obtained by 25 students of class XI in an exam. Find out range and coefficient of range of the marks :

Marks	5 – 9	10 – 14	15 – 19	20 – 24	25 – 29	30 – 34
No. of Students	4	6	3	2	6	4

[R = 30 Marks ; Coff. 0.369]

Q4. Find out the value of quartile deviation and its coefficient from the follwoing data :

Roll No.	1	2	3	4	5	6	7
Marks	20	28	40	12	30	15	50

[QB = 12.5 Marks ; Coff. 0.45]

Q5. Compute coefficient of Quartile deviation from the following data :

Marks	10	20	30	40	50	60
No. of students	4	7	15	8	7	2

[Coff. of QD = 0.333]

Q6. Estimate an appropriate measure of dispersion for the following data :

Wages(Rs.)	Less than 25	25 – 30	30 – 35	35 – 40	Above 40
No. of workers	2	10	26	16	7

[QD = Rs. 3.40 Coff. 0.1]

Q7. Calculate the mean deviaton form median and its coefficient by mean from the data given below

X 210 220 225 225 225 235 240 250 270 280

[MD = 17.6 ; Coff. 0.74]

Q8. Find ouit the mean deviation from the median and its coefficient.

Marks	10	11	12	13	14
No. of Students	3	12	18	12	3

[MD = 0.75 Marks]

Q9. Find the mean-deviation from mean and its coefficient for the given data :

X	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
F	3	5	7	2	9	4

[MD = 14.33 ; Coff. 0.44]

Q10. Find the standard deviation of the height of 10 men given below :

160, 160, 161l 162, 163, 163, 164, 164, 170 [SD = 2.72]

Q11. Measurements are made to the nearest cm. of the heights of 10 children. Calculate mean and standard deviation.

Height (cms)	60	61	62	63	64	65	66	67	68
No. of children	2	0	15	29	25	12	10	4	3

[Mean = 63.89 cms ; SD = 1.6 cms]

Q12. Calculate standard deviation from the following data :

Class	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
Frequency	17	11	8	5	4	3	2

[SD = 15.81]

Q13. The mean of two samples of sizes 50 and 100 respectively are 54.1 and 50.3 and the standard deviations are 8 and 7. Find the mean and the standard deviation of the sample of size 150 obtained by combining the two simples. [Mean = Rs. 39 ; Cors. SD = Rs. 9.16]

Q14. The given table shows the daily income of workers of two factories. Draw the Lorenz Curves for both the factories.

Daily Income(Rs.)	0 – 100	100 – 200	200 – 300	300 – 400	400 – 500
Factory A	8	7	5	3	2
Factory B	15	6	2	1	1

Q15. Find the mean deviation from mean and its coefficient for the given data :

Marks (more than)	0	10	20	30	40	50	60
No. of Students	200	180	150	100	40	15	5

[MD from mean = 9.05 ; Coff of MD = 0.306]

Q16. Calculate standard deviation of the following data :

Age in years (below)	10	20	30	40	50	60	70	80
No. of persons	15	30	53	75	100	110	115	125

[SD = 19.76 years]

Q17. Price of a particular item in 10 years in two cities are given below, which city has more stable prices?

City A	55	54	52	53	56	58	52	50	51	49
City B	108	107	105	105	106	107	104	103	104	101

[GV City A = 4.99% B = 1.90% ; City D has more stable prices]

Q18. The mean and standard deviation of a series of 20 items are 20 and 5 respectively. While calculating these measures, an item of 13 was wrongly read as 30. Find out the correct mean and standard deviation.

[Correct mean = 19.15 ; SD = 4.66]

Q19. In two Towns A and B, daily pocket money and standard deviations are given below :

Town	Average Daily Pocket	Standard Deviation	No. of Teenagers
A	34.5	5.0	476
B	28.5	4.5	524

(i) Which town A or B pays out the larger amount of daily pocket money ?

Town D = Rs. 16,422 ; Town B = Rs. 14,934

(ii) What is the average daily pocket money of all teenagers taken together ?

Combined pocket money Rs. 31.36

(iii) Calculate coefficient of variation of each town. Which town is more variable in terms of pocket money.

[CB (Town A) : 14.49% ; CV (Town B) = 15.79%]

Q20. Calculate the standard deviation and coefficient of dispersion from the data below :

Mid – Points	5	15	25	35	45	55	65	75
Frequency	5	8	7	12	28	20	10	10

[SD = 18.49 ; CV = 41.09%]

Q21. The following data shows the expected life of two models of T.V : A and B

Life (No. of years)	0 – 2	2 – 4	4 – 6	6 – 8	8 – 10	10 – 12
Model A	5	16	13	7	5	4
Model B	2	7	12	19	9	1

Which modal has greater uniformly ?

[CV of Model A = 54.91 % ; CV of Model B = 36.21%]

Q22. The sum of 10 values is 100 and the sum of their squares is 1090. Find the coefficient of variation.

Ans : $\overline{x} = 10,\ \sigma = 3,\ Cv = 30$

Q23. The means and standard deviations of two bulbs are given below :

	Brand I	Brand II
Mean	800 hours	770 hours
Standard deviation	100 hours	60 hours

Calculate a measure of relative dispersion for two brands and interpret the result.

Ans. For relative measure of dispersion from given information of two brand We will calculate Coefficient of Variation.

Brand I

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

Brand II

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

Ans: 12.5%

Given : $\overline{X} = 800$

$\sigma = 100$

$\therefore \quad \frac{100}{800} \times 100 = 12.5\%$

Given : $\overline{X} = 770$

$\sigma = 60$

$\therefore \quad \frac{60}{770} \times 100 = 7.79\%$

Hence, the bulbs of brand II are more consistent as compared to brand I.

ABSOLUTE AND RELATIVE MEASURES OF DISPERSION OR VARIATIONS

Types of Measures of Dispersion/Variation

Absolute Measure of Variation

- Range
- Inter-Quartile Range and
- Quartile Deviation
- Mean Deviation
- Standard Deviation

Relative Measure of Variation/ Coefficient of Variation

- Coefficient of Range
- Coefficient of Quartile Deviation
- Coefficient of Mean Deviation
- Coefficient of Standard Deviation

(A) ABSOLUTE MEASURE

Absolute dispersion is measured in the same units as the data. For instance, if the original data are in rupees, the absolute measure is also be in rupees, if the data are in kg, the measure will be in kg etc. For this reason absolute dispersion cannot be used to compare the scatter or variability in series where units of measure are different or when averages of one distribution than that in other distributions differ in size. The obsolute measures are range, interquartile, range, quartile deviation, mean deviation and standard deviation.

(B) RELATIVE MEASURE

For comparing two or more series where units of measure are different relative measures are used because they are calculated as the percentage or the coefficient of the absolute measure of dispersion. Therefore, it is also called coefficient of dispersion or coefficient of variation. The relative measures are coefficient of range, coefficient of quartile deviation, coefficient of mean deviation and coefficient of standard deviation. 100 times the coefficient of dispersion based on standard deviation is called coefficient of variation, abbrivated as CV.

Thus, CV. = $C.V. = \frac{\sigma}{\overline{X}} \times 100$

Thus, absolute measure is used only within a series and is measured in the same unit as that of series. Relative measure on the other hand, is used for comparing of variability of two or more series where units of measurement may be the same or different, it is obtained as percentages or coefficients.

Q1. What do you mean by dispersion? How is it linked to the measures of central tendency?

Ans. Measures of dispersion are the different statistically calculated figures that measure the degree of variation and dispersion of the data set.

Measures of dispersion indicate the scatteredness or the dispersion of the different observations of the data set around the measures of central tendency. This helps us in judging the compactness of the data set as well as the representativeness or Reliability of the measures of the central tendency.

Q2. Mention important measures of dispersion.

Ans. Important measures of dispersion:

(a) Absolute measures: Range, Mean deviation, Quartile deviation, Standard Deviation.

(b) Relative measures: Coefficient of range, Coefficient of quartile deviation, Coefficient of mean deviation, Coefficient of variation.

Q. Give two objectives of the measures of dispersion.

Ans. Objectives of measures of dispersion:

(a) Measures of dispersion help us in testing the reliability of an average. If the variation is small the degree of reliability will be low.

(b) Measures of dispersion are also useful in comparing two or more series with regard to disparities. A great degree of dispersion would mean lack of uniformity or consistency of the data.

Q. How the absolute measures of dispersion are different than the relative measures? Give two examples of each.

Ans. Absolute measures are those measures that express the dispersion in the same unit in which the data is collected like collected like metre, centimetre, rupees etc. Hence they cannot be used for comparison of the dispersion of different data sets. Measures like Range, Mean Deviation, Quartile Deviation, Standard Deviation etc. come in this category.

Relative measures are free from the units of measurement and are used for the comparison of the different data sets. Measures like coefficient of quartile deviation, coefficient of mean deviation, coefficient of variations etc. come under this category,

Q. Define range. Mention any two uses of range.

Ans. Range is defined as the difference between the largest and the smallest values in the data set.

Range is being widely used in defining the environmental standards for different pollutants and to study the fluctuations in stock markets.

Q. What is mean deviation? Define it.

Ans. Mean deviation is defined as the arithmetic mean of all the absolute deviations ignoring ± signs of the different values from any one of the measures of central tendency particularly mean and median.

Q. Mention any two merits and two demerits of mean deviation.

Ans. Merits of Mean Deviation:

(a) Mean deviation is easy to calculate and simple to understand.

(b) It is based on all observations of the series.

Demerits of Mean Deviation:

(a) It is not a popular measure as the ± signs are ignored in calculation of mean deviation.

(b) If extreme values are present, then mean deviation from mean could be very different from mean deviation from median.

Q. What is a line of equal distribution? What is its relationship with the Lorenz Curve?

Ans. The diagonal (OA) of the box in regard to Lorenz Curve is called the line of equal distribution as the abscissa and ordinate of any point from this line are equal to I each other. Hence the same percentage of income is enjoyed by the same percentage I of households or the income distribution is fair. Any departure from this line I indicates inequality. The area between the equal distribution line and Lorenz Curve is a measure of inequality or dispersion in income distribution.

Q. The daily wages of ten workers are given below. Find out range and its coefficient.

No.of Workers	A	B	C	D	E	F	G	H	I	J
Wages in (Rs.)	175	50	50	55	100	90	125	145	70	60

[Range = 125; Coefficient of Range = 0.55]

Q. Find range and coefficient of range of the following :

(a) Per, day earning of seven agricultural labourers in Rs :

60, 72, 36, 85, 35, 52, 72

(b) Mean temperature deviation from normal (2002).

Jan.	Feb.	Mar.	Apr.	May	June
+1.5	+2.4	+3.1	–1.5	–0.4	+3.3
July	Aug.	Sep.	Oct.	Nov.	Dec.
– 0.1	– 0.6	– 1.5	– 0.6	– 1.9	– 6.1

(a) Range = 50, Coeff. of Range = 0.42

(b) Range = +9.4, Coeff. of Range = - 3.36

Q. Calculate range and coefficient of range of the following data :

X	:	10	15	20	30	40	50
f	:	4	12	7	3	5	2

[Range = 40; Coefficient of Range = 0.67]

Q. Find out Quartile Deviation, Interquartile Range and Coefficient of Quartile Deviation of the following series :

Height (in inches)	:	58	59	60	61	62	63	64	65	66
No. of Persons	:	2	3	6	15	10	-5	4	3	1

[Q.D. = 1, Interquartile Range = 2, Coefficient of Q.D. = 0.016]

Q. Calculate Quartile and Coefficient of Quartile Deviation of the following data :

Marks	:	5-9	10-14	15-19	20-24	25-29	30-34	35-39
Students	:	1	3	8	5	4	2	2

[Q_1 = 15.906, Q_2 = 20, Q_3 = 26.687 and Coefficient of Q.D. = 0.25]

Q. Calculate lower and upper Quartiles, Quartile Deviation and Coefficient of Quartile Deviation of the following series :

Values	:	5-6	6-7	7-8	8-9	9-10	10-11
Frequency	:	5	8	12	15	6	2

[Q_1 = 6.875, Q_3 = 8.733, Coefficient of Q.D. = 0.119]

Q. Calculate the Interquartile Range for the data given below :

Class	:	35-40	30-35	25-30	20-25	15-20	10-15	5-10	0-5
Frequency	:	1	4	9	11	10	6	5	4

Q. Calculate Mean Deviation from mean and median of the following data :

X	:	54	71	57	52	49	45	72	57	47

[M.D. from $\bar{X}$ = 7.33, from Me = 7.11]

Q. Calculate (a) Median Coefficient of dispersion, and (b) Mean Coefficient of dispersion from the following data :

Size of item	:	4	6	8	10	12	14	16
Frequency	:	2	4	5	3	2	1	4

[M.D. from $\bar{X}$ = 3.32 and Coefficient of M.D. = 0.342
M.D. from Me = 3.24 and Coefficient of M.D. = 0.405]

Q. Calculate Mean and Mean Deviation and coefficient of M.D. for the following distribution:

Weekly wages	:	20-40	40-60	60-80	80-100
Workers	:	20	40	30	10

[Mean = 56, M.D. = 15.2 and Coefficient of M.D. = 0.27]

Q. Find Mean Deviation from median of the marks secured by 100 students in a class-test as given below:

Marks	:	60-63	63-66	66-69	69-72	72-75
No. of students	:	5	18	42	27	8

[M.D. = 2.26]

Q. Calculate Standard Deviation of the following two series. Which series has more variability :

A	:	58	59	60	65	66	52	75	31	46	48
B	:	56	87	89	46	93	65	44	54	78	68

[A : $\bar{X}$ = 56, S.D. = 11.7, CV. = 20.89%
B : $\bar{X}$ = 68, S.D. = 17.1, CV. = 25.14%

Series B has more variability]

Q. Calculate Mean and Standard Deviation from the following data :

Marks (Above) :	0	10	20	30	40	50	60	70	80
No. of Students :	150	140	100	80	80	70	30	14	0

[$\bar{X}$ = 39.27, S.D. = 22.85]

Q. Calculate Mean, Standard Deviation and Variance :

Variable :	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Frequency:	2	5	7	13	21	16	8	3

[$\bar{X}$ = 21.9, S.D. = 7.95, s^2 = 63.20]

Q. Calculate the Coefficient of Variation for the following distribution of the wages of 200 workers in a factory :

Wages (in Rs) :	40-49	50-59	60-69	70-79	80-89	90-99	100-109	110-119
No. of Workers :	11	23	40	60	35	16	9	6

[$\bar{X}$ = 74.45, S.D. = 15.921, CV. = 21.38%]

Q. Calculate the arithmetic mean and standard deviation and variance from the following distribution :

Class :	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Frequency :	2	5	7	13	21	16	8	3

[$\bar{X}$ = 21.9, S.D. = 7.99, s^2 = 63.97]

Q. You are given the following data about height of boys and girls :

	Boys	Girls
Number	72	38
Average height (in inches)	68	61
Variance of distribution (in inches)	9	4

a. Calculate Coefficient of Variation.

b. Decide whose height is more variable.

[(a) Boys : CV. = 4.41%, Girls : CV. = 3.27%] [(b) Boys]

Q. From the data given below state which series is more consistent :

Variable :	10-20	20-30	30-40	40-50	50-60	60-70
Series A :	10	18	32	40	22	18
Series B :	18	22	40	32	18	10

[Series A : $\bar{X}$ = 42.14, S.D. = 14.06, CV = 33.36%

Series B : $\bar{X}$ = 37.86, S.D. = 14.06, CV = 37.14%]

Q. The following table gives the distribution of wages in the two branches of a factory :

Monthly wages (In Rs) :	100-150	150-200	200-250	250-300	300-350	Total
Number of workers :						
Branch A :	167	207	253	205	168	1000
Branch B :	63	93	157	105	82	500

Find the mean and standard deviation for the two branches for the wages separately.

a. Which branch pays higher average wages ?

b. Which branch has greater variability in wages in relation to the average wages ?

c. What is the average monthly wage for the factory as a whole ?

d. What is the variance of wages of all the workers in the two branches A and B taken together ?

[Branch A : Mean = Rs 225, S.D. = Rs 66.20, CV. = 29.42%
Branch B : Mean = Rs 230, S.D. = Rs 62.15, CV. = 27.02%
(a) Branch B pays higher average monthly wages.
(b) Branch A has greater variability.
(c) Combined Mean = Rs 226.67
(d) Combined Variance = Rs 4215]

Q. For a group of 50 male workers, the mean and standard deviation of their weekly wages are Rs 63 and Rs 9 respectively. For a group of 40 female workers, these are Rs 54 and Rs 6 respectively. Find mean and standard deviation for acombined group of 90 workers.

[Combined : $\bar{X}_{1,2} = 59$, $s_{1,2} = 9$]

Q. Coefficient of variation of two series are 58% and 69% and their standard deviation are 21.2 and 15.6. What are their means ?

[Mean $\bar{X}$ = 36.55 and 22.608]

Q. Draw a Lorenz curve from the following :

Income (Rs '000)	:	20	40	60	100	160
No. of Persons ('000)						
Class A	:	10	20	40	50	80
Class B	:	16	14	10	6	4

Q. Following is the frequency distribution of marks obtained by students in Economics and Statistics. Analyse the data by drawing a Lorenz Curve.

Marks (mid-values)	:	5	15	25	35	45	55	65	75	85	95
No. of students (Eco.)	:	10	12	13	14	22	27	20	12	11	9
No. of students (Stat.)	:	1	2	26	50	59	40	10	8	3	1

— Notes —

CHAPTER – 12 – MEASURES OF CORRELATION

Q. Define Correlation

Correlation analysis help us in determining the degree of relationship between two or more variables. If there is correlation between two variables, it may be due to the following reasons :

1. **Both variables being influenced by a third variables**

   It is possible that a high degree of correlation in the analysis. For example, a high degree of correlation between the yield per acre rice and of jute may be due to the fact that both are related to the amount of rainfall.

2. **Mutual Depdence (Cause and Effect)**

   It may happen that the two variables shows a high degree of correlation but it is difficult to designate one as cause and other as effect. For examole, increase in price of a commodity decreases it's demand.

3. **Pure Chance**

   The correlation between the two variables may be obtained due to sheer coincidence (or pure chance). Such a correlation is known as spurious. We may get a high degree of correlation between two variables in a sample (say, size of shoes and income of people in a locality) when in fact there does not exist any relationship.

Importance or Significance of correlation

1. The correlation coefficient helps in measuring the extent of relationship between two variables in one figure.
2. Correlation anyalsiss facilitates understanding of economic behaviour and helps in locating the critically important variables on which others depend.
3. When two variables are correlated, then value of one variable can be estimated, given the value of another. This is one with the help of regression equations.
4. Correlation facilitates the decision-making in the business world. It reduces the range of uncertainty as predictions based on correlation are likely to be more reliable and near to reality.

TYPES OF CORRELATION

Positive and negative correlation

Positive Correlation

When two variables move in the same direction i.e. when one increases the other also increases and when one decreases the other also decreases, then such a relation is called positive correlation. For example : Relationship between height and weight, income and expenditure age of husband and age of wife etc.

Positive Correlation

Increase in both the variables (X & Y)

X	20	25	28	30	35
Y	10	12	15	18	20

Decrease in both the variables (X & Y)

X	30	25	20	15	10
Y	20	18	16	14	12

Positive correation is said to occur when both the variables vary in the same direction.

Negative Correlation

When two variables move in opposite directions, i.e., when one increases the other decreases and when one decreases the other increases, then such a relation is called negative correlation. For exmaple : Relationship between price and demand, day temperature and sale of woollen garments etc.

Negative Correlation

Rise in value of one variable (X) and fall in over (Y)

X	10	12	14	16	18
Y	30	25	20	15	10

Rise in value of one variable (Y) and fall in other (X)

X	50	40	30	20	10
Y	5	10	15	20	25

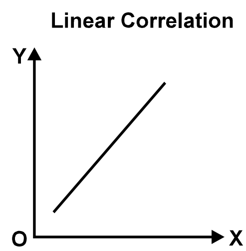
Negative correlation is said to occur when both the variables vary in the opposite direction.

Linear and Non-Linear (Curvilinera) Correlation

Linear Correlation

Linear correlation is said to exist if the amount of change in one variable tends to bear a constant ratio to the amount of change in the other variable. The graph of variables having such a relationship will form a straight line.

X	10	20	30	40
Y	5	10	15	20

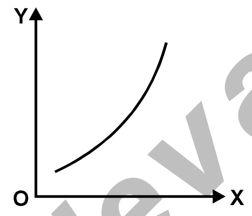


Non-Linear (Curvilinear) Correlation

In non-linear or curvilinear correlation, the amount of change in one variable does not bear a constant to the amount of change in the other related variable.

For example, when we double the use of fertilzers, the production of rice would not necessarily.

Non-Linear Correlation



X	10	20	30	40
Y	5	7	12	20

DEGREE OF CORRELATION

Perfeect Correlation

If the relationship between the two variables is such that the values of the two variables change (increase or decrease) in the same proportion, correlation between them is said to be perfect. Perfect correlation may be positive or negative. If the proportionate chage in the values of the two variables is in the same direction, it is called perfect positive correlation and the value is described as + 1. Changes are in the reverse direction, then the relationship is known as perfect negative correlation and described as –1.

Zero Correlation

When there is no relationship between the two variables, we say that there is absence of correlation. So, a change in the value of one variable has no particular effect on the value of the other variable. In this case, the value of coefficient will be zero.

Limited Degree of Correlation

Economic data do not indicate perfect positive or negative correlation. At the same time, the case of absence of correlation of economic data are, indeed, very limited. The value of correlation coefficient (r) normally lies in between +1 and –1. The limited degree of correlation can be high, moderate or low.

Degree of Correlation

Degree of Correlation	Positive Correlation	Negative Correlation
Perfect Correlation	+1	–1
High degree of Correlation	Between +0.75 and 1	Between – 0.75 and –1
Moderate degree of correlation	Between + 0.5 & 0.75	Between – 0.5 and – 0.75
Low degree of Correlation	Between + 0.25 & 0.5	Between – 0.25 and – 0.5
Very low degree of Correlation	Between 0 and 0.25	Between 0 and – 0.25
No Correlation	0	0

METHODS OF MEASUREMENTS OF CORRELATION

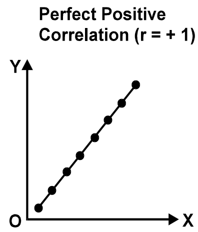
1. Scatter Diagram
2. Karl Pearson's Coefficient of Correlation
3. Spearman's Rank Correlation Coefficient

Scatter Diagram

The simplest method of studying the relationship between two variables is the scatter diagram. In this method, the values of the variables X and Y are plotted along the X-axis and Y-axis respectively. The values of two variables are shown by dots on the graph. This graphical representation of the values of the variables is known as scatter diagram or dot diagram. Inspection of the scatter diagram gives an idea of the nature and intensity of the relationship. The scatter diagram can be interpreted in the positive ways.

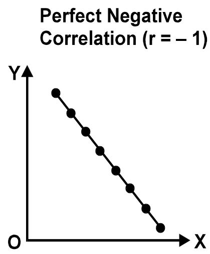
1. **Perfect Positive Correlation**

   If all the points of scatter diagram fall on a straight line with positive slope, then the correlation is said to be perfectly positive (r = + 1).



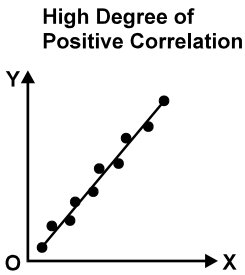
2. **Perfect Negative Correlation**

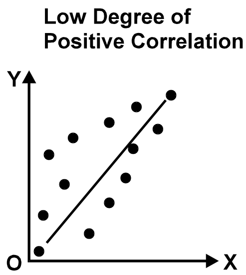
   If all the points of scatter diagram fall on a straight line with negative slope, then the correlation is said eo be perfectly negative (r = – 1).



3. **Positive Correlation**

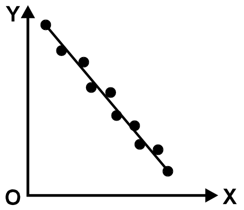
   When all the points of scatter diagram cluster around a straight line going upwards from left to right, the correlation is positive correlation.

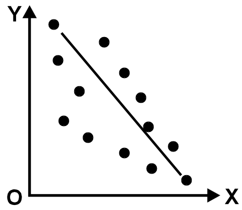




4. **Negative Correlation**

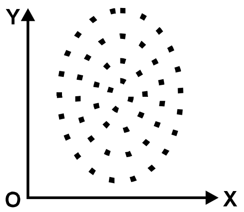
   When all the points of scatter diagram cluster around a straight line with negative slope, the correlation is said to be negative as shown in figure.

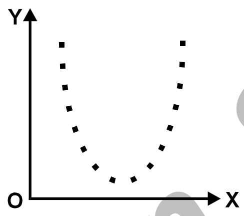




5. **No Correlation**

If the points are scattered in a haphazard manner, then it is a case of zero or no correlation.





Particals on Scatter Diagram

Ex.1 Make a scatter diagram for the following data and state the type of correlation between X and Y.

X	10	20	30	40	50
Y	70	140	210	280	350

Sol. The scatter diagram is obtained by plotting the values of Series X on the X-axis and values of Series Y on the Y-axis. Plotting the values (10, 70), (20, 140),... (50, 350) on the graph paper, we get the scatter diagram (see figure).

It is obvious from the scatter diagram that there is perfect positive correlation between the values of Sereis X and Series Y.

Ex.2 Draw a scatter diagram to represent the following values of X and Y variables. Comment on the type and degree of correlation.

X	15	18	30	27	25
Y	7	10	17	16	12

Sol. Plot the values of variable X on the X-axis and variable Y on the Y-axis.

A glance at the above scatter diagram shows that there is an upward trend of the dots from lower left-hand corner to the upper right-hand corner. It means, there is positive correlation between values of X and Y variables.

Merits and Demerits of Scatter Diagram

Merits

1. **Simplicity :** It is a simple and a non-mathematical method of studying correlation between two variables.
2. **Easily understandable :** It can be easily understood and interpreted. It enables us to know the presence or absence of correlation at a single glance of the diagram.
3. **Not affected by extreme items :** It is not influenced by the size of extreme values, whereas most of the mathematical methods lack this quality.
4. **First step :** It is a first step in investingating the relationship between two variables.

Demerits

1. **Non-mathematical method :** This method does not indicate the exact numerical valuke of correlation which is possible by other mathematical methods of correlation.
2. **Rough Measure :** It gives only a broad and rought idea of the degree and nature of correlation between two variables. Thus, it is only a qualitative expression rather than a quantitative expression.
3. **Unsuitable for large observations :** It is not possible to draw a scatter diagram on a graph paper in case of more than two variables.

KARL PERSON'S COEFFICIENT OF CORRELATION

According to Karl Pearson, Coefficient of correlation is determined by dividing the sum of products of deviatons from their respective means by their number of pairs and their standard deviations." The Karl Pearson's coefficient of correlation also known as 'Product Moment of Correlation' of 'Simple Correlation Coefficient'. It is denoted by the symbol 'r'.

$$\text{Coefficient of Correlation (r)} = \frac{\Sigma xy}{N \times \sigma_x \times \sigma_y}$$

Where,

N = Number of pair of observation

x = Deviation of X series from mean $(X - \bar{X})$

y = Deviation of Y series from mean $(Y - \bar{Y})$

σ_x = Standard deviation of X series, i.e., $\sqrt{\frac{\Sigma x^2}{N}}$

σ_y = Standard deviation of Y series, i.e., $\sqrt{\frac{\Sigma y^2}{N}}$

r = Coefficient of Correlation

Karl Pearson's method of calculating coefficient of correlation is based on covariance of the concerned variables. Let us illustrate this :

$$\text{Covarience (X, Y)} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{N} = \frac{\Sigma xy}{N}$$

$$r = \frac{\Sigma xy}{N \times \sigma_x \times \sigma_y}$$

or $$r = \frac{\Sigma xy}{N \times \sqrt{\frac{\Sigma x^2}{N}} \times \sqrt{\frac{\Sigma y^2}{N}}}$$

or $$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

This method is to be applied only when the deviation of items are taken from actual means and not from assumed means.

Ex.3 Find the coefficient of correlation between X and Y.

	X series	Y series
No. of items	30	30
Standard deviation	3	2

Summation of product of deviation of X and Y series from their respective means = 200.

Sol. Given, N = 30, $\sigma_x = 3$, $\sigma_y = 2$ and $\Sigma xy = 150$

$$r = \frac{\Sigma xy}{N \times \sigma_x \times \sigma_y} = \frac{150}{30 \times 3 \times 2}$$

$$= \frac{150}{180} = 0.833$$

Ans. Coefficient of correlation = 0.833. There is high degree of positive correlation between X and Y.

Ex.4 Find the standard deviation of Y series, if :

Coefficient of correlation = 0.25; Covariance between X and Y = 10; Variance of X = 64.

Sol. Covarience between X and Y = $\frac{\Sigma xy}{N} = 10$

Variance of X = $\sigma_x^2 = 64$

Coefficient of Correlation (r) = $\frac{\Sigma xy}{N \times \sigma_x \times \sigma_y}$

or $r = \frac{\Sigma xy}{N} \times \frac{1}{\sigma_x} \times \frac{1}{\sigma_y}$

$0.25 = 10 \times \frac{1}{8} \times \frac{1}{\sigma_y}$

$\sigma_y = \frac{10}{8 \times 0.25}$

Ans. Standard deviation of Y series = 5.

Ex.5 If covariancle between two variables X and Y is 9.4 and variance of Series X and Series Y are 10.6 and 12.5 respectively. Calculate the coefficient of correation.

Sol. Covariance between X and Y = $\frac{\Sigma xy}{N} = 9.4$

Variance of X = $\sigma_x^2 = 10.6$

$\sigma_x = \sqrt{10.6} = 3.25$

Variance of Y = $\sigma_y^2 = 12.5$

$\sigma_y = \sqrt{12.5} = 3.53$

Coefficient of Correlation (r) = $\frac{\Sigma xy}{N \times \sigma_x \times \sigma_y}$

or $r = \frac{\Sigma xy}{N} \times \frac{1}{\sigma_x} \times \frac{1}{\sigma_y}$

$= 9.4 \times \frac{1}{3.25} \times \frac{1}{3.53}$

$= 9.4 \times 0.307 \times 0.283 = 0.816$

Ans. Coefficient of correlation = 0.816. There is a high degree of positive correlation between X and Y.

CALCULATION OF KARL PEARSON'S COEFFICIENT OF CORRELATION

1. Actual Mean Method
2. Direct Method
3. Short-cut Method (Assumed Mean method or Indirect Method)

4. Step Deviation Method

Actual Mean Method :

Steps for calculation

Step 1 Calculat the means of the two sedries (X and Y) i..e, calculate $\bar{X}$ and $\bar{Y}$.

Step 2 Take the deviation of X series from $\bar{X}$ (mean of X) and denote the deviations by x.

Step 3 Square these deviations and obtain the total, i.e. Σx^2.

Step 4 Take the deviations of Y series from $\bar{Y}$ (mean of Y) and denote the deviation by y.

Step 5 Square these deviations and obtain the total, i.e. Σy^2.

Step 6 Multiply the respective deviations of X and Y series and obtain their total, i.e., Σxy

Step 7 Substitute the values of Σxy, Σx^2, Σy^2 in the following formula

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

Ex.6 Calculate the coefficient of correlation for the following data by the Actual Mean

X	12	15	18	21	24	27	30
Y	6	8	10	12	14	16	18

Sol. Calculation of Correlation (Actual Mean Method

X – series			Y – series			
X	$x = X - \bar{X}$	x^2	Y	$y = Y - \bar{Y}$	y^2	xy
12	–9	81	6	–6	36	54
15	–6	36	8	–4	16	24
18	–3	9	10	–2	4	6
21	0	0	12	0	0	0
24	+3	9	14	+2	4	6
27	+6	36	16	+4	16	24
30	+9	81	18	+6	36	54
$\Sigma X = 147$		$\Sigma x^2 = 252$	$\Sigma Y = 84$		$\Sigma y^2 = 112$	$\Sigma xy = 168$

$\bar{X} = \frac{\Sigma X}{N} = \frac{147}{7} = 21$ $\bar{Y} = \frac{\Sigma Y}{N} = \frac{84}{7} = 12$

Coefficient of Correlation (r) = $\frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$

Here, $\Sigma xy = 168$; $\Sigma x^2 = 252$; $\Sigma y^2 = 112$

$$= \frac{168}{\sqrt{252 \times 112}} = \frac{168}{\sqrt{28,224}} = \frac{168}{168} = 1$$

Ex.8 Calculate the number of number of items, when standard deviation of series X is 6 and coefficient of correlation is 0.75. Also given that sum of the product of deviations of X and Y from actual means is 180 and sum of squares of deviations of Y from actual means is 160.

Sol.

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

$$0.75 = \frac{180}{\sqrt{\Sigma x^2 \times 160}}$$

$$(0.75)^2 = \frac{(180)^2}{\Sigma x^2 \times 160}$$

$$0.5625 = \frac{32,400}{\Sigma x^2 \times 160}$$

$$\Sigma x^2 = \frac{32,400}{160 \times 0.5625} = \frac{32,400}{90}$$

$$\Sigma x^2 = 360$$

Standard Deviation $(\sigma_x) = \sqrt{\frac{\Sigma x^2}{N}}$

$$6 = \sqrt{\frac{360}{N}}$$

$$(6)^2 = \frac{360}{N}$$

$$N = \frac{360}{60} = 10$$

Ans. Number of items = 10

Direct Method

$$\text{Coefficient of Correlation (r)} = \frac{N\Sigma XY - \Sigma X.\Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2} \times \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

Ex.9 Calculate the coefficient of correlation of the data given in **Ex.6** by the Direct Method.

Sol. Calculation of Coefficient of Correlation (Direct Method)

X – Series		Y – Series		
X	X^2	Y	Y^2	XY
12	144	6	36	72
15	225	8	64	120
18	324	10	100	180
21	441	12	144	252
24	576	14	196	336
27	729	16	256	432
30	900	18	324	540
$\Sigma X = 147$	$\Sigma X^2 = 3,339$	$\Sigma Y = 84$	$\Sigma Y^2 = 1,120$	$\Sigma XY = 1,932$

$$\text{Coefficient of Correlation (r)} = \frac{N\Sigma XY - \Sigma X.\Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2} \times \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

Here, $\Sigma X = 147$, $\Sigma Y = 84$, $\Sigma X^2 = 3,339$; $\Sigma Y^2 = 1,120$; $\Sigma XY = 1,932$ and $N = 7$

$$= \frac{7 \times 1,932 - 147 \times 84}{\sqrt{7 \times 3,339 - (147)^2} \times \sqrt{7 \times 1,120 - (84)^2}}$$

$$= \frac{13,524 - 12,348}{\sqrt{23,373 - 21,609} \times \sqrt{7,840 - 7,056}}$$

$$= \frac{1,176}{\sqrt{1,764} \times \sqrt{784}}$$

$$= \frac{1,176}{42 \times 28} = \frac{1,176}{1,176}$$

$$= 1$$

Ans. Coefficient of Correlation = 1. There is perfect positive correlation between the values of Series X and Series Y.

Short-Cut Method (Assumed Mean Method)

$$r = \frac{\Sigma dxdy - \frac{\Sigma dx \times \Sigma dy}{N}}{\sqrt{\Sigma dx^2 - \frac{(\Sigma dx)^2}{N}} \times \sqrt{\Sigma dy^2 - \frac{(\Sigma dy)^2}{N}}}$$

Where

N = Number of pair of observations

Sdx= Sum of deviations of X values from assumed mean.

Sdx= Sum of deviations of Y values from assumed mean.

Sdx**2** = Sum of squared deviation of X values from assumed mean

Sdy**2** = Sum of squared deviations of Y values from assumed mean.

Sdxdy = Sum of the products of deviations dx and dy.

Sol. Calculation of Correlation (Short-Cut Method)

X – Series			Y – Series			
X	dx = X – A (A=18)	dx^2	Y	dy = Y – A (A=10)	dy^2	dxdy
12	–6	36	6	–4	16	24
15	–3	9	8	–2	4	6
18(A)	0	0	10(A)	0	0	0
21	+3	9	12	+2	4	6
24	+6	36	14	+4	16	24
27	+9	81	16	+6	36	54
30	+12	144	18	+8	64	96
	$\Sigma dx = 21$	$\Sigma dx^2 = 315$		$\Sigma dy = 14$	$\Sigma dy^2 = 140$	$\Sigma dxdy = 210$

$$\text{Coefficient of Correlation (r)} = \frac{\Sigma dxdy - \frac{\Sigma dx \times \Sigma dy}{N}}{\sqrt{\Sigma dx^2 - \frac{(\Sigma dx)^2}{N}} \times \sqrt{\Sigma dy^2 - \frac{(\Sigma dy)^2}{N}}}$$

Here, $\Sigma dxdy = 210, \Sigma dx = 21, \Sigma dy = 14, N = 7, \Sigma dx^2 = 315, \Sigma dy^2 = 140$

$$= \frac{210 - \frac{(21)(14)}{7}}{\sqrt{315 - \frac{(21)^2}{7}} \times \sqrt{140 - \frac{(14)^2}{7}}}$$

$$= \frac{210 - 42}{\sqrt{252} \times \sqrt{112}}$$

$$= \frac{168}{168} = 1$$

Ans. Coefficient of Correlation = 1. There is perfect positive correlation between the values of Series X and Series Y.

Ex.11 Calculate the Karl Pearson's Coefficient of correlation fro the following data :

(i) Sum of deviation of X values (Sdx) = 5

(ii) Sum of deviation of Y values (Sdy) = 4

(iii) Sum of squares of deviations of X values (Sdx^2) = 40

(iv) Sum of squares of deviation of Y values (Sdy^2) = 50

(v) Sum of the product of deviation of X and Y values (Sdxdy) = 32

(vi) No. of pairs of observation (N) = 10

Sol. Coefficient of Correlation (r) = $\dfrac{\Sigma dxdy - \dfrac{\Sigma dx \times \Sigma dy}{N}}{\sqrt{\Sigma dx^2 - \dfrac{(\Sigma dx)^2}{N}} \times \sqrt{\Sigma dy^2 - \dfrac{(\Sigma dy)^2}{N}}}$

Here, Sdxdy = 32, Sdx = 5, Sdy = 4, N = 10, Sdx^2 = 40, Sdy^2 = 50

$$= \frac{32 - \dfrac{(5)(4)}{10}}{\sqrt{40 - \dfrac{(5)^2}{10}} \times \sqrt{50 - \dfrac{(4)^2}{10}}}$$

$$= \frac{32 - 2}{\sqrt{37.5 \times 48.4}} = \frac{30}{42.60}$$

= 0.704

Ans. Coefficient of correlation = 0.704. There is a high degree of positive correlation between X and Y.

Coefficient of Correlation (r) = $\dfrac{\Sigma dx'dy' - \dfrac{\Sigma dx' \times \Sigma dy'}{N}}{\sqrt{\Sigma dx'^2 - \dfrac{(\Sigma dx')^2}{N}} \times \sqrt{\Sigma dy'^2 - \dfrac{(\Sigma dy')^2}{N}}}$

Ex.12 Calculate the coefficient of correlation of the data by the Step Deviation

X	12	15	18	21	24	27	30
Y	6	8	10	12	14	16	18

Sol. Calcualtion of Coefficient of Correlation (Step Deviation Method)

X – Series				Y – Series				
X	$dx = X - A$ (A=18)	$dx = \frac{dx}{C}$ (C=3)	dx'^2	Y	$dy = Y - A$ (A=10)	$dy' = \frac{dy}{C}$ (C=2)	dy'^2	$dx'dy'$
12	–6	–2	4	6	–4	–2	4	4
15	–3	–1	1	8	–2	–1	1	1
18(A)	0	0	0	10(A)	0	0	0	0
21	+3	+1	1	12	+2	+1	1	1
24	+6	+2	4	14	+4	+2	4	4
27	+9	+3	9	16	+6	+3	9	9
30	+12	+4	16	18	+8	+4	16	16
		$\Sigma dx' = 7$	$\Sigma dx'^2 = 35$			$\Sigma dy' = 7$	$\Sigma dy'^2 = 35$	$\Sigma dx'dy' = 35$

$$r = \frac{\Sigma dx'dy' - \frac{\Sigma dx' \times \Sigma dy'}{N}}{\sqrt{\Sigma dx'^2 - \frac{(\Sigma dx')^2}{N}} \times \sqrt{\Sigma dy'^2 - \frac{(\Sigma dy')^2}{N}}}$$

Here, $\Sigma dx'dy' = 35, \Sigma dx' = 7, \Sigma dy' = 7, N = 7$

$$= \frac{35 - \frac{(7)(7)}{7}}{\sqrt{35 - \frac{(7)^2}{7}} \times \sqrt{35 - \frac{(7)^2}{7}}} = \frac{28}{\sqrt{28} \times \sqrt{28}} = \frac{28}{28}$$

$= 1$

Ans. Coefficient of Correlation = 1. There is perfect positive correlation between the values of Series X and Series Y.

Ex.13 Calculate the coefficient of between X and Y and comment on their relationship.

X	–3	–2	–1	1	2	3
Y	9	4	1	1	4	9

Sol. Calculation of Coefficient of Correlation (Direct Method)

X – Series		Y – Series		
X	X^2	Y	Y^2	XY
–3	9	9	81	–27
–2	4	4	16	–8
–1	1	1	1	–1
1	1	1	1	1
2	4	4	16	8
3	9	9	81	27
$\Sigma X = 0$	$\Sigma X^2 = 28$	$\Sigma Y = 28$	$\Sigma Y^2 = 196$	$\Sigma XY = 0$

Coefficient of Correlation (r) = $\dfrac{N\Sigma XY - N - \Sigma X.\Sigma XY}{\sqrt{N\Sigma X^2 - (\Sigma X)^2} \times \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$

Here, SX = 0; SY = 28; SX^2 = 28; SY^2 = 196; SXY = 0 and N = 6

$$= \frac{6 \times 0 - 0 \times 28}{\sqrt{6 \times 28 - (0)^2} \times \sqrt{6 \times 196 - (28)^2}}$$

$$= \frac{0 - 0}{\sqrt{168 - 0} \times \sqrt{1176 - 784}}$$

$$= \frac{0}{\sqrt{168} \times \sqrt{392}} = 0$$

Ans. Coefficient of Correlation = 0. There no lienar relation between X and Y.

SPEARMAN'S RANK CORRELATION

*The coefficient of correlation obtained on the basis of ranks is called **'Spearman's Rank Correlation'***.

$$r_k = 1 - \frac{6\Sigma D^2}{N^3 - N}$$

where

r_k = Coefficient of rank correlation

SD^2 = Sum of square of Rank Difference

N = Number of pairs of observation.

COMPUTATION OF RANK CORRELATION

1. When Ranks are given.
2. When Ranks are not given.
3. When Rans are equal or repeated.
4. Apply the following formula

$$r_k = 1 - \frac{6\Sigma D^2}{N^3 - N}$$

Ex.14 Calculate Spearman's Rank Correlation of coefficient from the ranks given below :

X	2	1	4	3	5	7	6
Y	1	3	2	4	5	6	7

Sol. Calcualtion of Rank Correlation

X (R_1)	Y (R_2)	$(R_1 - R_2)$ D	D^2
2	1	+1	1
1	3	–2	4
4	2	+2	4
3	4	–1	1
5	5	0	0
7	6	+1	1
6	7	–1	1
N = 7			$\Sigma D^2 = 12$

Rank Coefficient of Correlation $(r_k) = 1 - \frac{6\Sigma D^2}{N^3 - N}$

Here, $\Sigma D^2 = 12; N = 7$

$$r_k = 1 - \frac{6(12)}{(7)^3 - 7} = 1 - \frac{72}{336}$$

$$= 0.786$$

Ans. Rank coefficient of correlation = 0.786. There is high degree of positive correlation.

Ex.15 Ten competitiors in the singing competition are ranked by three judges in the following ordre :

Judge 1	1	5	4	8	9	6	10	7	3	2
Judge 2	4	8	7	6	5	9	10	3	2	1
Judge 3	6	7	8	1	5	10	9	2	3	4

Use rank correlation coefficient to discus which pair of judges have the nearest approach to common testes in singing sscompetition.

Sol. In order to determine which pair of judges has the nearest approach to common taste in beauty we shall have to calcluate the rank coefficient of correlation between the rankings of :

(i) First and Second Judge

(ii) Second and Third Judge

(iii) First and Third Judge

(i) First and Second Judge

Rank Corelation between First and Second Judge

Rank by Judge 1 (R_1)	Rank by Judge 2 (R_2)	$(R_1 - R_2)$ D	D^2
1	4	–3	9
5	8	–3	9
4	7	–3	9
8	6	+2	4
9	5	+4	16
6	9	–3	9
10	10	0	0
7	3	+4	16
3	2	+1	1
2	1	+1	1
$N = 10$			$\Sigma D^2 = 74$

Rank Cofficient of correlation $(r_k) = 1 - \frac{6\Sigma D^2}{N^3 - N}$

Here $\Sigma D^2 = 74$; $N = 10$

$$r_k = 1 - \frac{6(74)}{(10)^3 - 10}$$

$$= 1 - \frac{444}{990}$$

= 0.552

(ii) Second and Third Judge

Rank Correlation between Second Judge and Third Judge

Rank by Judge 2 (R_2)	Rank by Judge 3 (R_3)	$(R_2 - R_3)$ D	D^2
4	6	–2	4
8	7	+1	1
7	8	–1	1
6	1	+5	25
5	5	0	0
9	10	–1	1
10	9	+1	1
3	2	+1	1
2	3	–1	1
1	4	–3	9
$N = 10$			$\Sigma D^2 = 44$

Rank Coefficient Correlation $(r_k) = 1 - \frac{6\Sigma D^2}{N^3 - N}$

Here $\Sigma D^2 = 44; N = 10$

$$r_k = \frac{6(44)}{(10)^3 - 10} = 1 - \frac{264}{990}$$

= 0.733

(iii) First and Third Judge

Rank Correlation between First and Third Judge

Rank by Judge 1 (R_1)	Rank by Judge 3 (R_3)	$(R_1 - R_3)$ D	D^2
1	6	–5	25
5	7	–2	4
4	8	–4	16
8	1	+7	49
9	5	+4	16
6	10	–4	16
10	9	+1	1
7	2	+5	25
3	3	0	0
2	4	–2	4
$N = 10$			$\Sigma D^2 = 156$

Rank Coefficient Correlation $(r_k) = 1 - \dfrac{6\Sigma D^2}{N^3 - N}$

Here $\Sigma D^2 = 156$; $N = 10$

$$r_k = 1 - \frac{6(156)}{(10)^2 - 10}$$

$$= 1 - \frac{936}{990} = 0.0545$$

The second and third judge have the nearest approach in common testes in singing competition, because the rank coefficinet of correlation is highest (0.733) between them.

When Ranks are not given

Ex.16 The following tables gives the marks obtained by 10 students in Economics and Accounts. Calculate the rank correlation.

Marks in Economics	35	90	70	40	95	45	60	85	80	50
Marks in Accounts	45	70	65	30	90	40	50	75	85	60

Sol. We are given actual ranks and not the ranks. So, assigning ranks from the lowest to the highest, we get

Calculation of Rank Correlation

Marks in Economics	Marks in Accounts	Ranks(R_1) in Economics	Ranks (R_2) in Accounts	$(R_1 - R_2)$ D	D^2
35	45	1	3	–2	4
90	70	9	7	+2	4
70	65	6	6	0	0
40	30	2	1	+1	1
95	90	10	10	0	0
45	40	3	2	+1	1
60	50	5	4	+1	1
85	75	8	8	0	0
80	85	7	9	–2	4
50	60	4	5	–1	1
$N = 10$					$\Sigma D^2 = 16$

Rank Coefficient Correlation $(r_k) = 1 - \frac{6\Sigma D^2}{N^3 - N}$

Here $\Sigma D^2 = 16; N = 10$

$$r_k = 1 - \frac{6(16)}{(10)^3 - 10}$$

$$= 1 - \frac{96}{990} = 0.903$$

Ans. Rank Coefficient of corelation = 0.903. There is high degree of positive correlation.

When Ranks are Equal or Repeated

Ex.17 Calculate the coefficient of correlation of the following data by the Spearman's Rank Correlation method.

X	19	24	12	23	19	16
Y	9	22	20	14	22	18

Sol. Calculation of Rank Coefficient

X	Y	R_1	R_2	$(R_1 - R_2)$ D	D^2
19	9	3.5	1	+2.5	6.25
24	22	6	5.5	+0.5	0.25
12	20	1	4	–3	9
23	14	5	2	+3	9
19	22	3.5	5.5	–2	4
16	18	2	3	–1	1
N = 6					$\Sigma D^2 = 29.5$

Here, number 19 is repeated twice in series X and number 22 is repeated twice in series Y. Therefore, in X, m = 2 and in Y, m = 2.

Rank Coefficient Correlation

$$(r_k) = 1 - \frac{6\left(\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + ...\right)}{N^3 - N}$$

Here $\Sigma D^2 = 29.50;\ N = 6$

$$= 1 - \frac{6\left(29.50 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)\right)}{6^3 - 6}$$

$$= 1 - \frac{6(29.50 + 0.5 + 0.5)}{210}$$

$$= 1 - \frac{6 \times 30.5}{210} = 1 - \frac{183}{210}$$

$= 1 - 0.871$

$= 0.128$

Ans. Rank's Coefficient of correlation = 0.128. There is very low degree of positive correlation.

Ex.18 The coefficinet of rank correlation of the ranks obtained by 10 students in Accounts and Maths ws found to be 0.8. It was later discovered that that difference in ranks in the two subjects obtained by one of the students was wrongly taken as 7 instead of 9. Find the correct coefficient of rank correlation.

Sol. We are given : n = 10, r_k = 0.8

We have $r_k = 1 - \frac{6\Sigma D^2}{(N^3 - N)}$

$$0.8 = 1 - \frac{6\Sigma D^2}{(10)^3 - 10}$$

$$\frac{6\Sigma D^2}{990} = 1 - 0.8$$

$$SD^2 = \frac{990 \times 0.2}{6} = 33$$

Since one difference is wrongly taken as 7 instead of 9, the corected value of SD^2 is given

Corrected $SD^2 = 33 - 7^2 + 9^2$

$= 33 - 49 + 81$

$= 65$

$\therefore$ Corrected $(r_k) = 1 - \frac{6 \times 65}{990} = 1 - \frac{390}{990}$

$= 1 - 0.394$

$= 0.606$

Ans. Correct coefficinet of rank correlation = 0.606

Ex.19 A person while calculating coefficinet of correlation between two variables X and Y, obtained the following results .

$$N = 8,\ \Sigma X = 120,\ \Sigma X^2 = 600,\ \Sigma Y = 90,\ \Sigma Y^2 = 250 \text{ and } \Sigma XY = 356$$

However, at the time of checking the calculations, he discovered that two paris of observations (8, 10) and (12, 7) were wrongly entered instead of (8, 12) and (10, 8). Determine the correct value of coefficient of correlation.

Sol. Corrected SX = 120 – 8 – 12 + 8 + 10 = 118

Corrected $SX^2 = 600 - 8^2 - 12^2 + 8^2 + 10^2 = 556$

Corrected SY = 90 – 10 – 7 + 12 = 8 = 93

Corrected $SY^2 = 250 - 10^2 - 7^2 + 12^2 + 8^2 = 309$

Corrected SXY = 356 – (8 × 10) – (12 × 7) + (8 × 12) + (10 × 8) = 368

Coefficient of Correlation (r) = $\frac{N\Sigma XY - N - \Sigma X.\Sigma XY}{\sqrt{N\Sigma X^2 - (\Sigma X)^2} \times \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$

$$= \frac{30 \times 368 - 118 \times 93}{\sqrt{30 \times 556 - (118)^2} \times \sqrt{30 \times 309 - (93)^2}}$$

$$= \frac{11,040 - 10,974}{\sqrt{16,680 - 13,924} \times \sqrt{9,270 - 8,649}}$$

$$= \frac{66}{\sqrt{2,756} \times \sqrt{621}} = \frac{66}{52.50 \times 24.92} = \frac{66}{1,308.3} = 0.05$$

Ans. Coefficient of Correlation = 0.05

UNSOLVED PRACTILES

Q1. **Represent correlation between the following figures through scatter diagram.**

X	8	16	24	31	42	50
Y	70	58	50	32	26	12

[There is high degree of negative correlation between X and Y]

Q2. **Given the following pair of values of the variables X and Y :**

X	8	10	12	11	9	7	13	14	15
Y	5	7	9	8	6	4	10	11	12

[The perfect positive correlation between X and Y]

Q3. **The data on price and quantity purchased relating to a commodity for 10 months are given below: Calculate coefficient of correlation between price and quantity.**

Pr ice (Rs.)	10	14	12	11	9	7	15	16	18	20
Quantity (kg)	25	20	30	32	35	40	19	16	12	10

[Coff. of Correlation is -0.958]

Q4. **Find out the correlation between the marks in Statistics and marks in Accountancy :**

No. of students	1	2	3	4	5	6	7	8	9	10
Marks in Statistics	20	35	15	40	10	35	30	25	45	30
Marks in Accountancy	25	30	20	35	20	25	25	35	35	30

[Coff. of Correlation is 8.76]

Q5. **The data on price and demand for commodity is given below :**

Pr ice (Rs.)	14	16	17	18	190	20	21	22	23
Demand (kg)	84	78	70	75	66	67	62	58	60

[Coff. of Correlation is -0.954]

Q6. **Find coefficient of correlation from the following figures :**

X	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5
Y	1,000	2,000	3,000	4,000	5,000	6,000	7,000	8,000	9,000

[Coff. of Correlation is Y]

Q7. **Calculate coefficient of correlation from the following data :**

(i) Sum of deviation of X values = – 6

(ii) Sum of deviation of Y values = 1

(iii) Sum of squares of deviations of X values = 196

(iv) Sum of squares of deviation of Y values = 87

(v) Sum of the product of deviation of X and Y values = 124

(vi) No of pairs of observations = 6

[Coff. of Correlation is 0.973]

Q8. **Ten competitors in a debate contest are ranked by three judges in the following order :**

Competitors	A	B	C	D	E	F	G	H	I	J
Ranks by 1st Judge	7	4	10	5	9	8	6	2	1	3
Ranks by 2nd Judge	4	1	9	10	7	3	2	5	6	8
Rank by 3rd Judge	10	2	8	5	7	6	9	1	4	3

[1sth and 3rd Judge pair has neariest approach]

Use the ranking correlation method and state which pair of judge have the nearest approach.

[Between 1st and 2nd year 0.103 between Ist and IIIrd is 0.73 ;
Correlation between 2nd and 3rd = 0.006]

Q9. **A group of 8 students got the following marks in a test in Maths and Accountancy.**

Marks in Maths	50	60	65	70	75	40	80	85
Marks in Accountancy	80	71	60	75	90	82	70	50

[Coff. of Mark Correlation is -0.5]

Q10. **Calculate the coefficient of rank correlation from the following data :**

X	48	33	40	9	16	16	65	24	16	57
Y	13	13	24	6	15	4	20	9	6	19

[Coff. of rank Correlation is 0.73]

Q11. **The coefficient of rank correlation of the marks obtained by 10 students in two particular subjects was found to be 0.5. It was later discovered that the difference in ranks in two subjects obtained by one of the students was wrongly taken as 3 instead of 7. What should be the correct value of coefficient of rank correlation ?**

[Coff. of Correlation is 0.257]

Q12. **Find the standard deviation of X series, if coefficient of correlation between two series X and Y is 0.35 and their covariance in 10.5 and variance of Y series is 56.25.** [SD = 4]

Q13. **Find the coefficient of rank correlation between the marks obtained in Mathematics and those in statistics by 10 students of a class.**

Marks in Mathematics	12	18	32	18	25	24	25	40	38	22
Marks in Statistics	16	15	28	16	24	22	25	36	34	19

[Coff. of rank Correlation is 0.95]

Q14. **Find Karl Pearson's coefficient for the following data :**

Marksin English	45	70	65	30	90	40	50	75	85	60
Marks in Maths	35	90	70	40	95	40	60	80	80	50

[Coff. of rank Correlation is 0.9031]

Q15. **From the following data compute product moment correlation between X and Y :**

	X series	Y series
No of items	15	15
Arithmetic Mean	25	18
Square of deviation from Mean	136	138

Summation of product of deviations of X and Y series from their respective means = 122.

Ans: 0.891 (High Degree of Positive Correlation b/w X & Y)

Q16. **Two series X and Y with 50 items each have standard deviations 4.5 and 3.5 respectively. If the summation of product of deviations of X and Y series from their respective arithmetic means be 420. Find the Coefficient of correlation between and Y.**

Ans: 0.531 (Moderate Degree)

Q17. **If the covariance between X and Y variables is +12.3 and variances of X and Y are respectively 13.8 and 16.4. Find the Karl Pearson's coefficient of correlation between them.**

Ans. We are given,

$$\text{Covariance of X and Y} = \frac{\sum xy}{N} = 12.3$$

Variance of X (σx^2) = 13.8

∴ $\sigma x = \sqrt{13.8} = 3.71$

Variance of Y $(\sigma y^2) = 16.4$

$\sigma y = \sqrt{16.4} = 4.05$

Applying formula, $r = \frac{\sum xy}{N.\sigma x.\sigma y} = \frac{\sum xy}{N} \times \frac{1}{\sigma x} \times \frac{1}{\sigma y}$

Now, we get $r = 12.3 \times \frac{1}{3.71} \times \frac{1}{4.05}$

$= 12.3 \times 0.27 \times 0.25 = 0.83$

Hence, there is high degree of positive correlation between X and Y

Q18. **Find the standard deviation of X series if coefficient of correlation between two series X and Y is = 0.28 and their covariance is 76 and variance of Y series is 81.90.**

Ans. Given, coefficient of correlation (r) = 0.28

Covariance of X and Y = $\frac{\sum xy}{N} = 7.6$

Variance of Y (σy^2) = 81.90

∴ $\sigma y = \sqrt{81.90} = 9.05$

Applying formula, $r = \frac{\sum xy}{N.\sigma x.\sigma y} = \frac{\sum xy}{N} \times \frac{1}{\sigma x} \times \frac{1}{\sigma y}$

Now, we get $0.28 = 7.6 \times \frac{1}{\sigma x} \times \frac{1}{9.05}$

or $0.28 \times 9.05\ \sigma x = 7.6$ or $2.534\ \sigma x = 7.6$

$\sigma x = \frac{7.6}{2.534} = 2.99$ approx 3

Therefore, variance of X $(\sigma x)^2 = (3)^2 = 9$

Q19. **Calculate the number of items for which r + 0.8, $\sum xy = 200$ standard deviation of Y = 5; and $\sum x^2 = 100$, where x and y denotes deviation of items from actual mean.**

Ans. Applying formula,

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \times y^2}} \text{ or } 8.0 = \frac{200}{\sqrt{100 \times \sum y^2}}$$

Now, we get $(0.8)^2 = \frac{(200)^2}{100 \times \sum y^2}$ or $0.64 = \frac{40000}{100 \times \sum y^2}$

or $0.64 \times 100 \times \sum y^2 = 40000$

$64 \sum y^2 = 40000$

$\therefore \quad \sum y^2 = \frac{40000}{64} = 625$

Now, $\sigma y = \sqrt{\frac{\sum y^2}{N}}$ or $5 = \sqrt{\frac{625}{N}}$

or $(5)^2 = \frac{625}{N}$ or $25 = \frac{625}{N}$

or $25\ N = 625 \quad \therefore \quad N = \frac{625}{25} = 25$

Hence, number of items is 25.

Q20. **Draw a scatter diagram and calculate Karl Pearson's coefficient of correlation between X and Y. Interprete the result and comment on their relationship.**

X	:	1	3	4	5	7	8
Y	:	2	6	8	10	14	16

Q21. **Distinguish between univariate and bivariate analysis with examples.**

Ans. Univariate analysis is one which is related to the observations of a single variable only. The measures of central tendency and dispersion etc. are single univariate analysis because they take up only a single variable at a time. Bivariate analysis is one when data in two different variables are studied together. The technique of correlation comes under this category.

Q22. **What are the different types of correlation? Name them.**

Ans. Correlation is classified into four broad categories :

(a) Positive and Negative Correlation

(b) Simple and Multiple Correlation

(c) Partial and Multiple Correlation

(d) Linear and Non-linear Correlation.

Q23. **Briefly explain the meaning of four types of correlation.**

Ans.(a) Positive and Negative Correlation. Under positive correlation, the direction of change is same for both variables like increase in X leading to increase in Y. The correlation between fertiliser use and yield per hectare is positive correlation. Under negative correlation, the direction of change is opposite, i.e., the increase in one variable leads to decrease in the other. The correlation between price and quantity demanded is an example of negative correlation.

(b) Simple and Multiple correlation. Simple correlation is the relationship between two variables. When three or more variables are studied the relationship can be multiple.

(c) Partial and Multiple Correlation. Under partial correlation, the relationship between two or more variables is examined keeping all other variables as constant, e.g. under law of demand we study the relationship between price and demand keeping all other factors as constant or amount of rainfall and yield of wheat in a certain constant temprature.

The multiple correlation is based on the study of the relationship between three or more relevant variables.

(d) Linear and Non-linear Correlation. Linear correlation is one in which the change on X is related to change in Y in a constant proportion such that for every 1 % of change on X, Y will change by some fixed per cent.

Under non-linear correlation, the change in X does not bear a constant proportion to change on Y.

Q24. **How do you calculate the correlation between qualitative variables?**

Ans. Correlation between qualitative variables like beauty, intelligence, good behaviour, honesty etc. can be calculated by Spearman's Rank Correlation method by using the ranks of the individual.

This is done by using the following formula :

$$rk = 1 - \frac{6\Sigma D^2}{N^3 - N}$$

Q25. **Mention two merits and two demerits of Rank Correlation.**

Ans. Merits :

(a) Rank correlation can be conveniently calculated and is easy to understand.

(b) Qualitative variables like beauty, intelligence, honesty etc. can be measured only by the Spearmen's Rank Correlation.

Demerits :

(a) Rank Correlation cannot be used for large series.

(b) This method lacks precision as compared to Karl Pearson's method.

Q26. **Positive correlation is found between number of children born and export over last decade. Yes or No.**

Ans. No. It is a non-sense or spurious correlation. There is no cause and effect relationship between children born and export increase.

Q27. **If the coefficient of correlation between X and Y is zero, does it mean that there is the absence of any type of relation between them ?**

Ans. The value of the coefficient of correlation may be zero. It .means that there is zero correlation. It does not mean the absence of any type of relation between them.

Two values are uncorrelated. However, other type of relation may be there and there is no linear relationship between them.

Q28. **From the following data compute the product moment correlation between X and Y :**

	X Series	Y Series
Arithmetic Mean	25	18
Square of deviations from Arithmetic Mean	136	138

Summation of products of deviations of X and Y series from their respective means = 122 Number of points of values = 15 [r = 0.89]

Q29. **Calculate Karl Pearson's Coefficient of Correlation on the following data :**

X :	15	18	21	24	27	30	36	39	42	48
y :	25	25	27	27	31	33	35	41	41	45

[r = 0.98]

Q30. **The deviations from their means of two series (X and Y) are given below :**

X :	- 4	-3	-2	-1	0	+1	+2	+3	+4
y :	+3	-3	-4	0	+4	+1	+2	-2	-1

Calculate Karl Pearson's coefficient of correlation and interpret the result.

[r = 0]

Q31. **Find the product moment correlation of the following data :**

X :	1	2	3	4	5	6	7	8	9
y :	9	8	10	12	11	13	14	16	15

Q32. **Calculate the coefficient of correlation for the following ages of husbands and wives in years at the time of their marriage.**

Age of husbands :	23	27	28	28	29	30	31	33	35	36
Age of wives :	18	20	22	27	21	29	27	29	28	29

[r = +0.82]

Q33. **Find suitable coefficient of correlation for the following data :**

Fertilizers used (in tons) :	15	18	20	24	30	35	40	50
Productivity (in tons) :	85	93	95	105	120	130	150	160

[r = +0.99]

Q34. **The total of the multiplication deviation of X and Y = 3044**

Number of pairs of observations = 10

Total of the deviation of X = -170

Total of the deviation of Y = -20

Total of squares of deviation of X = 2264

Total of the squares of deviation of Y = 8288

Find out Karl Pearson's coefficient of correlation when assumed mean of X and Y are 82 and 68 respectively. [r = + 0.78]

Q35. **Number of pairs of observations of X and Y series = 10.**

X series : Arithmetic Average = 65
: Standard deviation = 23.33
Y series : Arithmetic Average = 66
Standard deviation = 14.9

Summation of products of corresponding deviations of X and Y series = +2704. Calculate product moment correlation of X and Y series.
[r = +0.78]

Q36. **Calculate Spearman's rank correlation from the following data :**

X	:	10	12	8	15	20	25	40
y	:	15	10	6	25	16	12	8

[r = +0.143]

Q37. **Calculate rank coefficient of correlation of the following data :**

X	:	80	78	75	75	68	67	60	59
Y	:	12	13	14	14	14	16	15	17

[r - 0.93]

Q38. **Twelve entries were submitted in a flower show competition. They were ranked by two judges as under :**

Entires	1	2	3	4	5	6	7	8	9	10	11	12
Judge A	7	8	2	1	9	3	12	11	4	10	6	5
Judge B	6	4	1	3	11	2	12	10	5	9	7	8

Calculate Spearman's rank correlation. [r = +0.86]

Q39. **From the following data calculate coefficient of correlation by the method of rank differences.**

X	:	75	68	95	70	60	80	81	50
Y	:	120	134	150	115	110	140	142	100

[r = +0.93]

CHAPTER – 13 – INDEX NUMBERS

Meaning of Index Numbers

Index numbers are used to measure changes in group of items that are related to each other. The changes are measured from time to time or from place to place or from one commodity to another. They are usually expressed in percentages.

Features / Characteristics of Index Numbers

1. **Index number are specialised averges**
2. **Index numbers are expressed in percentages**
3. **Index numbers measure the effect of changes in relation to time or place**
4. **Index Numbers measures the change not capable of direct measurement**

METHODS OF CONSTRUCTING PRICE INDEX NUMBERS

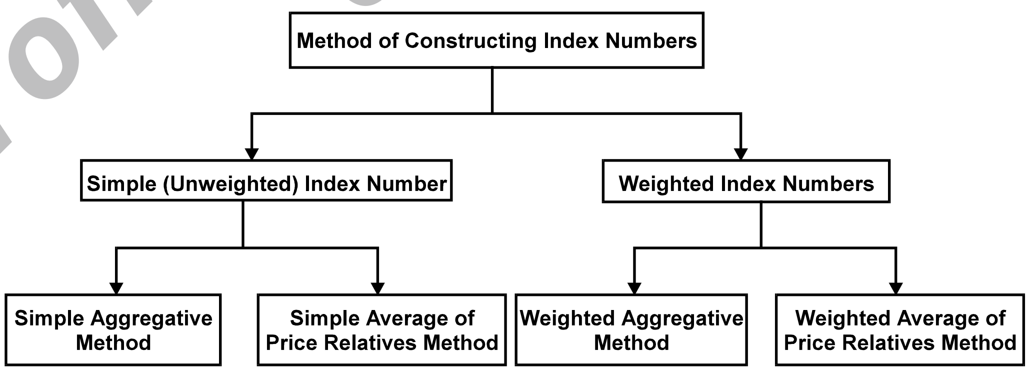
The various methods of constructing price index numberc can be grouped under two heads :

1. Unweighted Index Numbers;
2. Weighted Index Numbers.

Both of these methods of constructing index numbers are further classified as :

(i) Aggregative Method;

(ii) Average of Relatives Method



Unweighted Index Numbers

In the unweighted index numbers, each item is supposed to have the same weight as no weight is expressly assigned to any item.

Ex.1 Conctruction index numbers for 2008 taking 1994 as the base year from the following data by Simple Aggregate Method :

Commodity	Price in 1994	Price in 2008
Wheat	Rs. 10 / kg	Rs. 15 / kg
Rice	Rs. 12 / kg	Rs. 30 / kg
Pulses	Rs. 20 / kg	Rs. 50 / kg
Sugar	Rs. 18 / kg	Rs. 20 / kg

Sol. Construction of Price Index

Year 1994 as the base year

Commodity	Price in 1994 (Rs.)	Price in 2008 (Rs.)
Wheat	10	15
Rice	12	30
Pulses	20	50
Sugar	18	20
	$\Sigma p_0 = 60$	$\Sigma p_1 = 115$

Price Index for year 2008 with year 1994 as base

$$P_{01} = \frac{\Sigma p_1}{\Sigma p_0} \times 100$$

$$= \frac{115}{60} \times 100$$

$$= 191.67$$

The price index number (191.67) reveals that there is a net increase of 91.67% in prices in the year 2008 compared to the prices in the year 1994.

Ans. Price Index Number = 191.67

Ex.2 Calculate the price index by first taking 2000 as base year and then 2002 as base

Year	Price (Rs.)
2000	40
2001	50
2002	60
2003	70
2004	80
2005	90
2006	95

Sol. Calcualtion of Price Index :

Year	Price (in Rs.)	Index Numbers BaseYear = 2000 $\left(\frac{P_1}{P_0}\times100\right)$	Index Number Base Year = 2002 $\left(\frac{P_1}{P_0}\times100\right)$
2000	40	$\frac{40}{40}\times100=100$	$\frac{40}{50}\times100=66.66$
2001	50	$\frac{50}{40}\times100=125$	$\frac{50}{60}\times100=83.33$
2002	60	$\frac{60}{40}\times100=150$	$\frac{60}{60}\times100=100$
2003	70	$\frac{70}{40}\times100=175$	$\frac{70}{60}\times100=116.67$
2004	80	$\frac{80}{40}\times100=200$	$\frac{80}{60}\times100=133.33$
2005	90	$\frac{90}{40}\times100=225$	$\frac{90}{60}\times100=150.00$
2006	95	$\frac{95}{40}\times100=237.5$	$\frac{95}{60}\times100=158.33$

Ex.3 From the following data, construct on index for 2003 taking 1997 as base by the simple average of relatives method :

Commodities	A	B	C	D	E
Prices (1997)	50	40	80	100	20
Price (2007)	70	60	100	120	20

Sol. Price Index Number by Simple Average of Relative Method :

Commodity	Prices in 1997(Rs.) p_0	Prices in 2003 (Rs.) p_1	Price Relatives $\frac{p_1}{p_0}\times100$
A	50	70	$\frac{70}{50}\times100=140$
B	40	60	$\frac{60}{40}\times100=150$
C	80	100	$\frac{100}{80}\times100=125$
D	100	120	$\frac{120}{100}\times100=120$
E	20	20	$\frac{20}{10}\times100=100$
$N=5$			$\Sigma\left(\frac{p_1}{p_0}\times100\right)=635$

$$P_{01} = \frac{\Sigma\left(\frac{p_1}{p_0} \times 100\right)}{N} \qquad = \frac{635}{5} = 127$$

The prince index number of 127 shows the increase of 27% in prices in the year 2003 as compared to year 1997.

Ans. Price Index Number = 127.

Merit and Demerits of Average Price Relative Index

Meris.

(i) The value of this index is not affected by the units in which prices of commodities are quoted.

(ii) Equal importance is given to each commodity and extreme commodities do not influnce the index number.

Demerits

(i) As it is an unweighted index, each price relative is given equal importance.

(ii) Difficulty is faced with regard to the selection of an appropriate average.

WEIGHTED INDEX NUMBERS

Such weights indicate the relative importance of items or commodities included in the calculation of an index Weighted index numbers can be constructed by two methods :

(i) Weighted Aggregative method and

(ii) Weighted Average of Price Relatives Method.

Weighted Aggregative Method

Some of the important methods of constructing Weighted Aggregate Indices are :

1. Laspeyre's Method
2. Paasche's Method
3. Fisher's Ideal Method

Laspeyre's Method

$$p_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

Paasche's Method

$$p_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_1 q_1} \times 100$$

Fisher's Method

$$p_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

$$\text{Value Index Number} = \frac{\text{Current values}}{\text{Base year values}} \times 100$$

$$V_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_o q_o} \times 100$$

or $$V_{01} = \frac{\Sigma V_1}{\Sigma V_o} \times 100$$

where p_{01} = Price index number

q_{01} = Quantity index number

V_{01} = Value index number

p_1 = Current year price

p_0 =Base year price

q_1 = Current year quantity

q_0 = Base year quantity

V_1 = Current year value (Sp_1q_1)

V_0 = Base year value (Sp_0q_0)

Q. Construct Price Index and Quantity Index of 2010 from the following data by (i) Laspeyre's method, and (ii) Paasche's method. Also calculate Value Index for 2010.

Commodities	2005 Base year Price	2005 Base year Quantity	2010 Base year Price	2010 Base year Quantity
A	10	30	12	50
B	8	15	10	25
C	6	20	6	30
D	4	10	6	20

Ans. Construction of Price Index Numbers

Commodities	Base year (2005) Price (P_o)	Base year (2005) quantity (q_o)	Current year (2010) Price (P_o)	Current year (2010) Quantity (P_o)	p_oq_o	p_oq_1	p_1q_o	p_1q_1
A	10	30	12	50	300	500	360	600
B	8	15	10	25	120	200	150	250
C	6	20	6	30	120	180	120	180
D	4	10	6	20	40	80	60	120
					$\Sigma p_oq_o = 580$	$\Sigma p_oq_1 = 960$	$\Sigma p_1q_o = 690$	$\Sigma p_1q_1 = 1150$

Laspeyre's Price Index Number

Symbolically,

$$p_{o1} = \frac{\Sigma p_1 q_o}{\Sigma p_o q_o} \times 100$$

$$= \frac{690}{580} \times 100 = 118.96$$

Laspeyre's Quantity Index :

$$q_{o1} = \frac{\Sigma q_1 p_o}{\Sigma q_o p_o} \times 100$$

$$= \frac{960}{580} \times 100 = 165.52$$

Paasche's Method :

$$p_{o1} = \frac{\Sigma q_1 p_1}{\Sigma q_o p_1} \times 100$$

$$= \frac{1150}{960} \times 100 = 119.79$$

Thus, the price index number of 2010 is 119.79. In other words, there is net increase in prices of commodities in the year 2010 to the extent of 19.79% as compared to 2005.

Paasche's Quantity Index :

$$q_{o1} = \frac{\Sigma q_1 p_1}{\Sigma q_o p_1} \times 100$$

$$= \frac{1150}{690} \times 100 = 166.67$$

Value Index Number :

$$V_{o1} = \frac{\Sigma q_1 p_1}{\Sigma q_o p_1} \times 100$$

$$= \frac{1150}{580} \times 100 = 198.28$$

Ex.4 For the data given int he following table compoute index numbers by : (i) Laspeyre's method (ii) Fisher's ideal method :

	Base year		Current Year	
Commodity	Price (Rs.)	Quantity	Price (Rs.)	Quantity
	p_0	q_0	p_1	q_1
A	10	30	12	50
B	8	15	10	25
C	6	20	6	30
D	4	10	6	20

Sol. Construction of Price Index Numbers

	Base year		Current Year					
Commodity	Price (Rs.)	Qty.	Price (Rs.)	Qty.	p_0q_0	p_0q_1	p_1q_0	p_1q_1
	p_0	q_0	p_1	q_1				
A	10	30	12	50	300	500	360	600
B	8	15	10	25	120	200	150	250
C	6	20	6	30	120	180	120	180
D	4	10	6	20	40	80	60	120
					Σp_0q_0 = 580	Σp_0q_1 = 960	Σp_1q_0 = 690	Σp_1q_1 = 1150

(i) Laspeyre's Method

$$p_{01} = \frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times 100$$

$$= \frac{690}{580} \times 100$$

$= 1.18965 \times 100 = 118.965$

(ii) Paache's Method

$$p_{01} = \frac{\Sigma p_1q_1}{\Sigma p_0q_1} \times 100$$

$$= \frac{1{,}150}{960} \times 100$$

$= 1.1979 \times 100 = 119.79$

(iii) Fisher's Method

$$p_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100$$

$= \sqrt{\frac{690}{580} \times \frac{1,150}{960}} \times 100$

$= \sqrt{1.18965 \times 1.1979} \times 100$

$= \sqrt{1.4250} \times 100 = 1.1937 \times 100$

= 119.37

Ans. (i) Laspeyre's = 118.965; (ii) Paasche's = 119.79; (iii) Fisher's = 119.37

Ex.5 Calculate the price index number by (i) Laspeyre's method (ii) Paasche's Method, (iii) Fisher's Method.

	Base year (1999)		Current Year (2008)	
Commodity	Price (Rs.) p_0	Value (Total Expenditure)	Price (Rs.) p_1	Value (Total Expenditure)
A	2	200	3	300
B	8	82	10	100
C	12	60	15	90
D	7	49	10	80

Sol. In case of both the year,s total value is given. So, we will have to first calculate the quantity by applying the following formula : Quanity = $\frac{\text{Value}}{\text{Price}}$

Construction of Price Index Numbers

	Base year (1999)		Current Year (2008)					
Commodity	Price (Rs.) p_0	Qty. q_0	Price (Rs.) p_1	Qty. q_1	p_0q_0	p_0q_1	p_1q_0	p_1q_1
A	2	100	3	100	200	200	300	300
B	8	9	10	10	72	80	90	100
C	12	5	15	6	60	72	75	90
D	7	7	10	8	49	56	70	80
					Σp_0q_0 = 381	Σp_0q_1 = 408	Σp_1q_0 = 535	Σp_1q_1 = 570

(i) Laspeyre's Method

$p_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$

$= \frac{535}{381} \times 100$

= 140.42

(ii) Paasche's Method

$$p_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

$$= \frac{570}{408} \times 100$$

$$= 139.70$$

(iii) Fisher's Method

$$p_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

$$= \sqrt{\frac{535}{381} \times \frac{570}{408}} \times 100$$

$$= 140.05$$

Ans. (i) Laspeyre's = 140.42 (ii) Paache's = 139.70; (iii) Fisher's = 140.05

Weighted Average of Price Relatives Method

Apply the formula :

$$p_{01} = \frac{\Sigma RW}{\Sigma W}$$

Ex.6 Calculate the index number by weihted relatives method from the following data for the year 2007 with 2000 as the base year.

Commodity	Quantity in 2000 (Units)	Price in 2000 (Rs.)	Price in 2007 (Rs.)
A	80	5	8
B	65	8	14
C	42	12	18
D	37	4	5
E	31	4	5
F	15	2	4

Sol. Construction of Weighted Index Numbers

Commodity	Weight q_0	Price (2000) p_0	Price (2005) p_1	Value Weight $(p_0q_0)W$	$\frac{p_1}{p_0}\times 100$ R	RW
A	80	5	8	400	160	64,000
B	65	8	14	520	175	91,000
C	42	12	18	504	150	75,600
D	37	4	5	148	125	18,500
E	31	4	5	124	125	15,500
F	15	2	4	30	200	6,000
				$\Sigma W = 1,726$		$\Sigma RW =$ 2,70,600

Weighted Average of Price Relatives

$$p_{01} = \frac{\Sigma RW}{\Sigma W}$$

$$= \frac{2,70,600}{1,726} = 156.77$$

The index number of 156.77 shows the increases of 56.77% in prices in the year 2005 as compared to your 2000.

CONSUMER PRICE INDEX (CPI)

Meaning

Consumer Price Inde reflects the average increases in the cot of the commodities consumed by a class so that they can maintain the same standard of living in the current year as in the base year.

Methods of Constructing CPI

(i) Aggregate Expenditgure Method or Weighted Aggregate Method;

(ii) Family Budget Method or Method of Weighted Average of Price Relatives.

Aggregate Expenditure Method

The method is similar to the laspeyre's method of constructing weighted index.

$$\text{Consumer Price Index} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

Family Budget Method

In this method, the family budgets of a large number of people, for whom the index is meant, are carefully studied.

Consumer Price Index = $\frac{\Sigma RW}{\Sigma W}$ $\qquad R = \frac{P_1}{P_o} \times 100$

Ex.7 Calculate cost of living index, for the following data, using aggregate expenditure and family budget method :

Commodity	Price (in Rs.)		Qantity in units
	1994	2009	1994
A	10	15	15
B	8	12	20
C	20	24	10
D	32	40	5
E	15	20	6
F	12	18	2
G	8	10	1

Sol. Aggregate Expenditure Method :

	Price (in Rs.)			Aggregate Expenditure	
Commodity	1994 p_0	2009 p_1	Quantity in units (1994) q_0	p_0q_0	p_1q_0
A	10	15	15	150	225
B	8	12	20	160	240
C	20	24	10	200	240
D	32	40	5	160	200
E	15	20	6	90	120
F	12	18	2	40	36
G	8	10	1	8	10
				$\Sigma p_0q_0 = 792$	$\Sigma p_1q_0 = 1071$

Concumer Price Index for the year 2009

$$= \frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times 100$$

$$= \frac{1,071}{792} \times 100$$

= 135.22

Family Budget Method

Commodity	Price (in Rs.) 1994 p_0	2009 p_1	Price Relatives $R = \frac{p_1}{p_0} \times 100$ R	Quantity (1994) q_0	p_0q_0 W	RW
A	10	15	150	15	150	22,500
B	8	12	150	20	160	24,000
C	20	24	120	10	200	24,000
D	32	40	125	5	160	20,000
E	15	20	133.33	6	90	12,000
F	12	18	150	2	24	3,600
G	8	10	125	1	9	1,000
					$\Sigma W = 792$	$\Sigma RW = 1,07,100$

Consumer Price Index for the year 2009

$$= \frac{\Sigma RW}{\Sigma W}$$

$$= \frac{1,07,100}{792} = 135.22$$

It show that there is an increase of 35.22% in prices in the year 2009 as compared to year 1994.

Ans. Consumer Price Index by Aggregate Expenditure and Family Budget Method = 135.22

Ex.8 An enquiry into the budget of middle class families in a certain city gave the following information :

Expenses	Food	Fuel	Clothing	Rent	Misc.
	35%	10%	20%	15%	20%
Price in 2004 (Rs.)	1500	250	750	300	400
Price in 1995 (Rs.)	1400	200	500	200	250

What is the cost of living index of 2004 as compared with 1995 ?

Sol. Calculation of Cost of Living Index (Family Budget Method)

Commodity	Price (in Rs.) 1995 p_0	2004 p_1	Price Relatives $R = \frac{p_1}{p_0} \times 100$ R	Weights W	RW
Food	1400	1500	107.14	35	3,749.9
Fuel	200	250	125	10	1,250
Clothing	500	750	150	20	3,000
Rent	200	300	150	15	2,250
Misc.	250	400	160	20	3,200
				$\Sigma W = 100$	$\Sigma RW = 13,449.9$

Cost of Living Index of the year 2004

$$= \frac{\Sigma RW}{\Sigma W} = \frac{13,449.9}{100}$$

$$= 134.499$$

Ans. Cost of Living Index = 134.499. It means there is an increase of 34.499% in prices in 2004 as compared to 1995.

Ex.14 The consumer price index for June 2005 was 125. The food index was 120 and that of other items 135. What is the percentage of the total weight given to food ?

Sol. Price Relatives (Index) denoted by R is given for food (120) and other items (135). Let the weight food items be W_1 and that of other items be W_2. Now, the given information can be summarised as under :

Items	Index (R)	Weight (W)	RW
Food	120	W_1	$120W_1$
Other Items	135	W_2	$135W_2$
		$\Sigma W = 100$	$\Sigma RW = 120W_1 + 135W_2$

Consumer Price Index = $\frac{\Sigma RW}{EW}$

$$125 = \frac{120W_1 + 135W_2}{100}$$

$$12500 = 120W_1 + 135W_2 \quad ...(1)$$

We also have

$$W_1 + W_2 = 100 \quad ...(2)$$

Solving (1) and (2), we get

$$W_1 = 66.67 \text{ and } W_2 = 33.33$$

Ans. Percentage of total weight given to food (W_1) = 66.67%

Use of Consumer Price Index (CPI) Number

1. Consumer price index numbers helps in wage negotiations, formulation of wage policy, price policy, rent control , taxation and general economic policy formulation.
2. The government and business units use the consumer price index numbers to reglate the Dearness allowance (DA).
3. The CPI are used to measure purchasing power of the consumer in rupees. The formula for calculating the purchasing power of the ruppe is

$$\text{Purchasing Power} = \frac{1}{\text{Consumer Price Index}} \times 100$$

4. Index numbers tell us the change in real wages. Real wages can also be determined, in the following manner :

$$\text{Real Wages} = \frac{\text{Money Wages}}{\text{Consumer Price Index}} \times 100$$

5. Consumer price index numbers are also used for analysing markets for particular kinds of goods and service.

Ex.15 During a certain period, the cost of living index goes up from 110 to 200 and the daily wages of a worker was also raised from Rs. 80 to Rs. 125. Has the worker really gained., and if so, by how much in real terms.

Sol. With increase in cost of living index from 110 to 200, the daily wages of the worker should be increased from Rs. 80 to $\frac{80 \times 200}{110}$ = Rs. 145.45. However, the daily wages have gone up only to Rs. 125.

Hence, the worker has not gained. In fact, his real wages have gone down.

The real wage of the worker is $\frac{125 \times 110}{200}$ = Rs. 68.75 as compared to Rs. 80 before the price rise.

Ex.16 If the salary of a person in the base year is Rs. 4,000 per annum and the current year salary is Rs. 6,000, by how much should his salary rise to maintain the same standard of living, if the CPI is 400?

Sol. With increase in cost of living index from 100 to 400l the annual salary of the person should be : $\frac{400 \times 4,000}{100}$ = Rs. 16,00.

However, the annual salary has increased to only Rs. 6,000.

To maintain the same standard of living, the salary of the person should rise by Rs. 10,000 (Rs. 16,000 – s. 6,000 = Rs. 10,000).

INDEX OF INDUSTRIAL PRODUCTION

The index number of industrial production measure changes in the level of industrial production comprising many industries.

$$\text{Index Number of Industrial Production} = \frac{\Sigma\left(\frac{q_1}{q_0}\times 100\right)W}{\Sigma W}$$

where,

q_1 = Level of production in the current year

q_0 = Level of production in the base year

W = Weight or relative importance of industrial output.

Ex.17 From the following data, construct index of industrial production :

Industry	Output (in units) 2000	2007	Weights
Mineral Prod.	125	190	35
Chemical Prod.	80	140	40
Electrical Prod.	170	272	10
Textile	220	308	15

Sol. Calcuation of index of Industrial Production

Industry	Output (in units) 2000	2007	Weights	$\frac{q_1}{q_0}\times 100$	$\left(\frac{q_1}{q_0}\times 100\right)W$
Mineral Prod.	125	190	35	152	5,320
Chemical Prod.	80	140	40	175	7,000
Electrical Prod.	170	272	10	160	1,600
Textile	220	308	15	140	2,100
				SW = 100	16,020

$$\text{Index Number of Industrial Production} = \frac{\Sigma\left(\frac{q_1}{q_0}\times 100\right)W}{\Sigma W}$$

$$= \frac{16,020}{100} = 160.20$$

Industrial production has increased by 60.20% in year 2007

Ans. Index of Industrial Production = 160.20

USES OF INDEX NUMBERS

Index numbers are the signs and guide posts along the business highway that indicate to the businessman how he should derive or manage his affiars.

1. **Helps in Policy Formulation**

   Index numbers are indispensable tools for the management of any governent or non-government organisation.

2. **Index numbers act as Economic Barometers**

   Index numbers are used to feel the pressure of the economic and business behaviour, as well as to measure ups and downs in the general economic condition of a country.

3. **Help in studying trends**

   Index numbers are very useful in studying the trend or tendency of a series spread over a period of time. It also helps in forecasting the future trends.

4. **To measure and compare changes**

   Index numbers helps in comparative changes in two variables.

5. **Index numbers help to measure purchaing power**

   The value of money depends on its purchasing power but the purchasing power of money depends on the prices of the commodities.Index numbers are in helpful in finding out the intrinsic worth of money as contrasted with its nominal worth. This helps in formualating the wage policy of th ecountry.

6. **Index numbers in help in deflating various values**

   The price index number helps to adjust monetary figures of various periods for changes in prices.

LIMITATIONS OF INDEX NUMBERS

1. **Provides relative changes only**

   Index numbers are only estimates of relative chasnges in various events, which is obtained on the basis of average of all the items. Hence, it does no apply to individual units.

2. **Lack of perfect accuracy**

   Quite often, index numbers are based on sample items, each and every item is not considered.

3. **Difference between purpose and method of contruction**

When an index number is constructed for a special purpose by a specific method, then such index number will not be appropriate for all other purposes and situations.

4. **Ignorees qualitative changes**

While constructing the price or production index numbers no attention is paid to the changes in quality of the product.

5. **Manilpuations are possible**

Index numbers can be consturcted in such a manner so that the desired result can be obtained.

Wholesale price index numbers

Wholesale price index numbers are those price index numbers which measure the general chagnes in the wholesale prices of goods in a country.

UNSOVLVED PRACTICALS

Q1. **Construct index numbers by aggregative method (based on the price of 1998) from the following figures :** [I.N. 140.235]

Items	A	B	C	D	E	F
Price (1998)	200	60	350	100	60	80
Price (2005)	240	90	600	110	62	90

Q2. **Calculate index number for 2010 on the base prices for 2005 from the following by average of price relative method.**

Items	:	Bricks	Timber	Plaster Board	Sand	Cement
Prices (2005)	:	10	20	21	2	7
Prices (2010)	:	16	21	6	3	14

[Index No. = 147]

Q3. **Construct the index number for 2010 taking 2000 as base by price relative method using arithmetic mean.**

Commodities	:	A	B	C	D
Prices (2000)	:	10	20	30	40
Prices (2010)	:	13	17	60	70

[Index No. = 147.5]

Q4. **Calculate the index numbers from the following data using : (i) Laspeyre's method, (ii) Paasche's method (ii) Fisher's ideal method.**

[Laspeyer's 124.44 ; Paasche's 124.42 ; Fisher's 124.23]

Commodity	Base year		Current year	
	Price (in Rs.) p_0	Quantity q_0	Price (in Rs.) p_1	Quantity q_1
A	8	100	10	120
B	4	60	5	80
C	10	20	12	25
D	12	25	15	30
E	3	5	4	6

Q5. **The follwoing table contains information from the raw material purchase records of a small factory for the year 2000 and 2005 :**

[Fisher's I.N. 121.75]

Commodity	2000		2005	
	Price (Rs. / Unit) p_0	Total Value q_0	Price (Rs. / Unit) p_1	Total Value q_1
A	5	50	6	72
B	7	84	10	80
C	10	80	12	96
D	4	20	5	30
E	8	56	8	64

Q6. **Calculate weighted average of price relative index number of prices for 2003 on the basis of 1999 from the following data :**

[I.N. 149.715]

Commodity	Quantity in 1999	Price (in Rs.)1999	Price (in Rs.)2003
A	20	20	35
B	12	12	18
C	3	10	11
D	4	5	5
E	6	4	5

Q7. **Calculate consumer price index number for the following data by aggregate expenditure and family budget Method.**

[Consumer price I.N. is 116.34]

Expenses	Weights	Price in Rs. Base Year	Price in Rs. Current year
Food	45	300	350
Rent	20	200	225
Fuel	8	100	150
Clothing	10	150	175
Misc.	17	250	300

Q8. **Construct the index of industrial production from the following data:**

Industry	Output (in tonnes)		Weight
	2002	2005	
Mining	120	180	25
Electrical Products	200	290	45
Manufactured Goods	150	220	30

[Index of industrial production is 146.75]

Q9. **Calculate the cost of living index number for 2005 taking 2000 as base year from the following data by family budget method :**

[Cost of Living index is 120.18]

Items	Quantity (in kg)	Price in 1990 (in Rs./kg)	Price in 1995 (in Rs./kg)
A	15	10.00	12.00
B	20	16.50	20.00
C	8	6.00	7.50
D	12	15.00	16.00
E	10	8.00	11.50

Q10. **The following data relate to the prices and quantities of 4 commodities in the years 1995 and 2005. Construct the index numbers of price for the year 2005 by using 1995 the base year by :**

(i) Laspeyre's method (ii) Paasche's method (iii) Fisher's ideal method

Commodity	1995		2005	
	Price (in Rs.) p_0	Quantity q_0	Price (in Rs.) p_1	Quantity q_1
A	5	100	6	150
B	4	80	5	100
C	2.5	60	5	72
D	12	30	9	33

[Laspeyer's 118.05 ; Paasche's -119.18 ; Fisher's Ideal 118.61]

Q11. **Calculate the index number for 2007 with 1999 as base using the weighted average of price relative method for the following data :**

Commodity	Quantity in 1999 (in units)	Price 1999	Price 2007
A	2	12	24
B	8	8	12
C	4	15	27
D	5	6	18
E	1	10	12

[I.N. is 188.297]

Q12. **From the data given below construct the consumer price index number :**

Commodity	Price Relatives	Weights
Food	250	45
Rent	150	15
Clothing	320	20
Fuel and Lighting	190	5
Miscellaneous	300	15

[Index Number = 253.5]

Measuring rate of inflation :

WPI is used to measure the rate of inflation. The rate of inflation is useful to know the real value of income, savings and wealth, etc. Using WPI of 2003-04 and 2004-05 for all the commodities from the table given above, the rate of inflation can be calculated as under :

$$\text{Rate of inflation} = \left[\frac{\text{WPI of current year}}{\text{WPI of previous year}} \times 100\right] - 100$$

$$= \left[\frac{189.5}{180.3} \times 100\right] = 5.1\%$$

or

$$\text{Rate of inflation} = \left[\frac{\text{WPI of current year} - \text{WPI of previous year}}{\text{WPI of previous year}}\right] \times 100$$

$$= \left[\frac{189.5 - 180.3}{180.3}\right] \times 100$$

$= 5.1\%$

Thus, the annual inflation rate during 2004-05 was 5.1% in case of all commodities. One can also calculate inflation rates for different commodities or commodity groups as required for policy purposes.

LIST OF FORMULAS

1. Unweighted Simple Aggregative Method

Price Index $p_{o1} = \frac{\Sigma p_1}{\Sigma p_o} \times 100$

Quantity Index $q_{o1} = \frac{\Sigma q_1}{\Sigma q_o} \times 100$

Value Index $V_{o1} = \frac{\Sigma V_1}{\Sigma V_o} \times 100$

$$= \frac{\Sigma p_1 q_1}{\Sigma p_o q_o} \times 100$$

2. Unweighted Simple Average of Price Relative Method

$$p_{o1} = \frac{\Sigma\left(\frac{p_1}{p_o} \times 100\right)}{N}$$

3. Weighted Aggregative Method : Price Index Quantity Index

A. Laspeyre's Method $p_{o1} = \frac{\Sigma p_1 q_o}{\Sigma p_o q_o} \times 100$ $q_{o1} = \frac{\Sigma q_1 p_o}{\Sigma q_o p_o} \times 100$

B. Paasche's Method $p_{o1} = \frac{\Sigma p_1 q_1}{\Sigma p_o q_o} \times 100$ $q_{o1} = \frac{\Sigma q_1 p_1}{\Sigma q_o p_1} \times 100$

4. Weighted Average of Price Relative Method

$$p_{o1} = \frac{\Sigma\left[\left(\frac{p_1}{p_o} \times 100\right) \times (p_o q_o)\right]}{\Sigma p_o q_o} \text{ or } \frac{\Sigma PV}{\Sigma V}$$

Consumer Price Index

A. Aggregative Expenditure Method or Aggregative Method

$$CPI = \frac{\Sigma p_1 q_o}{\Sigma p_o q_o} \times 100$$

B. Family Budget Method

$$CPI = \left[\frac{\Sigma WR}{\Sigma W}\right] \text{ or } \left[\frac{\Sigma WI}{\Sigma W}\right]$$

Index of Industrial Production

Industrial Production Index No. $= \frac{\Sigma\left(\frac{q_1}{q_o}\right)W}{\Sigma W}$

Q1. **Clearly define the index number.**

OR

What is meant by index number?

Ans. Index number is a statistical device which measures the relative changes in the magnitude of related variables with respect to time or place.

Q2. **Mention two important uses of index numbers.**

Ans. Uses of Index Numbers :

a. Index numbers act as economic barometer and are used to feel the pulse of economy.

b. Helpful in the formulation of policies.

Q3. **Give two important limitations of index numbers.**

Ans. Limitations of Index Numbers :

a. Index members are only estimatesthey are true only on an average.

b. Index numbers prepared for one purpose cannot be effectively used for other purposes.

Q4. **Point out the steps and problems involved in the construction of index numbers.**

Ans. Steps and Problems in the Construction of Index Numbers :

a. Purpose of index number should be clearly defined.

b. Proper selection of commodities.

c. Collection of price quotations.

d. Selection of base period.

e. Choice of average.

f. Selection of weights.

Q5. **What care should be taken in the selection of commodities for the index numbers?**

Ans. Following care should be taken in the selection of commodities for the index numbers :

a. Commodities must be representative.

b. Commodities should be adequate.

c. Commodities must satisfy the criterion of comparability.

Q6. **What are the two methods of price quotations? Tell their meaning also.**

Ans. Methods of price quotations :

a. Money Prices. Quoting prices in terms of money per unit of commodity. For instance, potato at Rs. 10 per kg.

b. Quantity Prices. Quoting prices in terms of quantity per unit of money. For instance, $\frac{1}{2}$ kg. potato per 5 rupee.

Q7. **What are the two methods to determine the base period?**

Ans. Methods to determine the base period :

a. Fixed base period. One particular year may be taken as base = 100.

b. Chain base method. Prices in one year serve as base for the price relative in the succeeding year.

Q8. **What do you mean by weight in the context of index number ?**

Ans. The term 'weight' refers to relative importance of different items in the construction of index numbers.

Q9. **Construct the Index Number for 2010 with 2005 as base from the following prices of commodities by simple (Unweighted) aggregative method.** [Index Number = 164.48]

Commodities :	A	B	C	D	E
Price in Rs. 2005 :	50	40	10	5	2
Price in Rs. 2010 :	80	60	20	10	6

Q10. Using the following data and 2008 as the base period, compute simple aggregative price indices for the two fuels.

Item	Producer's Price (2008)	Producer's Price (2009)	Producer's Price (2010)
Coal (Rs.)	5	3	4
Crude oil (Rs.)	2	3	4

[Index Number : 2009 = 85.71, 2010 = 114.28]

Q11. Calculate index number for 2010 on the base prices for 2005 from the following by average of price relative method.

Items	Bricks	Timber	Plasater Board	Sand	Cement
Prices(2005)	10	20	5	2	7
Prices(2010)	16	21	6	3	14

[Index No. = 147]

Q12. Construct the index number for 2010 taking 2000 as base by price relative method using arithmetic mean.

Commodities :	A	B	C	D
Price (2000) :	10	20	30	70
Price (2010) :	13	17	60	70

[Index No. = 147.5]

Q13. Calculate price index number for 2008 of following data by weighted aggregative method using :

(a) Laspeyre's method (b) Paasche's method.

Commodity	Price(2001)	Quantity(2001)	Price(2008)	Quantity(2008)
A	4	20	6	10
B	3	15	5	23
C	2	25	3	15
D	5	10	4	40

[Laspeyre's : 137.77, Paasche's : 158.99]

Q14. From the data given below, construct Laspeyre's and Paasche's price index and quantity index numbers with base 2009 and interpret.

Commodity	Price(Rs.) 2009	Quantity(kg) 2009	Price(Rs.) 2010	Quantity(kg) 2010
A	4	2	6	3
B	3	5	2	1
C	8	2	4	6

[Laspeyre's : Price Index = 76.92, Quantity Index = 143.18;
Paasche's : Price Index = 69.84, Quantity Index = 130]

Q15. **Calculate weighted aggregative of actual price index number and quantity index number from the following data using (i) Laspeyre's Method, and (ii) Paasche's Method. Also calculate value index number and interpret them.**

Commodity	Base year Quantity lbs.	Base year Price per lb.	Current year Quantity lbs.	Current year Price per lb.
Bread	6	40 paise	7	30 paise
Meat	4	45 paise	5	50 paise
Tea	0.5	90 paise	1.5	40 paise

[Index Number = (i) 86.02, (ii) 81.25]

Q16. **Calculate price index number by weighted average of price relative method.**

Commodity	Price base year (in Rs.)	Price Current year (in Rs.)	Quantity base year (in kg.)
A	6.0	8.0	40
B	3.0	3.2	80
C	2.0	3.0	20

[Index Number = 122.3]

Q17. **From the data given below construct the consumer price index number :**

Commodity	Food	Rent	Clothing	Fuel and Lighting	Miscellaneous
Price relatives	250	150	320	190	300
Weights	45	15	20	5	15

[Index Number = 253.5]