INTRODUCTION

The word statistics is derived from the Latin word status.

The word statistics comes from the Italian word statista

INTRODUCTION • For a layman, 'Statistics' means numerical information expressed in quantitative terms. • This information may relate to objects, subjects, activities, phenomena, or regions of space. As a matter of fact, data have no limits as to their reference, coverage, and scope. • At the macro level, these are data on gross national product and shares of agriculture, manufacturing, and services in GDP (Gross Domestic Product) • At the micro level, individual firms, howsoever small or large, produce extensive statistics on their operations. • The annual reports of companies contain variety of data on sales, production, expenditure, inventories, capital employed, and other activities. These data are often field data, collected by employing scientific survey techniques. • . It is a discipline, which scientifically deals with data, and is often described as the science of data. In dealing with statistics as data, statistics has developed appropriate methods of collecting, presenting, summarizing, and analysing data, and thus consists of a body of these methods.

MEANING

In the early years "statistics means collection of facts about the state or the people in the state for political purpose. Data was collected about economic and social conditions of the people living in different parts of the country. For proper functioning of the state, it is essential to know the conditions under which people live and work, earn income and spend the wealth

MEANING

•Gottfried Achenwall in 1749 was the first to use the term statistics to refer a subject matter as a whole. •With the passage of time, the scope of statistics widened. Initially , statistics was regarded as a branch of Economics but now it has become full fledged independent subject

Definition

" Statistics is defined as collection, organising (Classification)

presentation, analysis and interpretation of

numerical data."

Acc. Croxton & Cowden

Statistics is the sciences and art of dealing with figures and facts.

PLURAL SENSE OF STATISTICS

What is statistics in plural sense ?

In plural sense, it means a systematic collection of numerical facts

"Statistics are numerical statements of facts in any department of enquiry placed in relation to each other " - BOWLEY

According to Webster " The classified facts respecting the condition of people in a state especially those facts which can be stated in numbers or in any tabular or classified arrangement

Plural sense of statistics

Professor Horace Secrist defined Statistics as

" Aggregate of facts Affected to marked extent by multiplicity of causes numerically expressed, enumerated or estimated according to reasonable standards of accuracy collected in a systematic manner for a pre-determined purpose and placed in relation to each other."

Characteristics Of Statistics in Plural Sense

- **Aggregate of facts.**
- **Affected to a marked extent by multiplicity of causes.**
- **Numerically Expressed.**
- **Enumerated or estimated according to reasonable standard of accuracy.**
- **Collected in a systematic manner.**
- **For a predetermined purpose.**
- **Placed in relation to each other**

THE CHARACTERISTICS WHICH NUMERICAL DATA MUST POSSESS IN PLURAL SENSE ARE :

1. AGGREGATES OF FACT

Single and isolated numerical expressions are not called statistics as such facts cannot be put in comparison or in relation to others

For e g : Adel 's marks in Management are 88 is not statistics

But marks of 10 students in Management 80,86,68,81,82,88,66,60,62,61 are statistics.

2. AFFECTED BY MULTIPLICITY OF CAUSES

Facts and figures are affected to a considerate extent by a number of forces acting together .

For e.g. : Change in demand is affected by so many causes such as change in price of the product , income of the consumer etc.

3. NUMERICALLY EXPRESSED

All statistics are numerical statements of facts that is expressed in numbers .

Qualitative expressions like young , old , good, bad etc are not recognized. So, a numerical value like 932 females per 1000 males in India and 1170 females per 1000 males in USSR.

Eg. Miss Aishwarya Rai is Miss World is not Statistics, but in 1994 Miss Aishwarya Rai is Miss World is Statistics

4. ESTIMATED ACCORDING TO A RESONABLE STANDARD OF ACCURACY

Facts and figures about any statement or phenomenon can be divided in two ways

First way : By actual counting and measurement

Second way : By estimation

Eg. Small gathering can be counted like no. of Students in 1 IT

But large gathering cannot be counted. Like any event in college where other college students are also invited.

5.COLLECTED IN A SYSTEMATIC MANNER

For accuracy and reliability of data, the figures should be collected in a systematic manner. Like unrelated data shd not be collected. But only data which is related to purpose has to collected.

In survey steps have followed in systamatic Manner.

6.COLLECTED FOR PREDETERMINED PURPOSE

Data collected with some objective in mind should be concrete and specific

For e.g. : If we want to collect data on prices, then we must be clear whether we have to collect wholesale or retail prices.

7.PLACED IN RELATION TO EACH OTHER

Figures(Numbers) data should be homogeneous but not heterogeneous.

For e.g. : Height of a person should not be compared with the rent he pays.

Statistics was originated when man started counting and keeping records in one way or other .It is applied in all

divisions of knowledge .It is used as a valuable tool in analyzing of tool of problems in physical science,commerce,medical,technologyetc

Statistics deals with numerical statements of statistics or statistical data. Statistics stands for presentation , analysis and interpretation of numerical data

STATISTICS IN SINGULAR SENSE OR STAGES IN STATISTICAL ENQUIRY / INVESTIGATION
OR
STATISTICAL METHODS

Large volume of Numerical Information gives rise to the need for Systamatic Methods which can be used for organising, analysing, and interpreting the information Effectively

STATISTICS IN SINGULAR SENSE

According to ''CROXTON & COURDEN''statistics is

''COLLECTION,CLASSIFICATION/ORGANISATION,

PESENTATION,ANALYSING AND INTERPRETATION OF NUMERICAL DATA''

CHARACTERISTICS /STAGES OF STATISTICS IN SINGULAR SENSE

<u>FEATURES IN SINGULAR SENSE</u>
1.COLLECTION OF DATA
2.CLASSIFICATION OF DATA
3.PRESENTATION OF DATA
4. ANALYSIS
5.INTERPRETATION

STAGE 1 COLLECTION OF DATA:- Collection of data is the first step in statistical investigation atmost care should be taken in collecting the data because they form foundation to statistical analysis .These should be such as to fulfill the basic objective of proposed study

STAGE 2 :- CLASSIFICATION AND ORGANISATION:- Data collection has to be organized properly hence it should be edited ,classified and then tabulated . Editing is done carefully so that omissions ,inconsistences may be corrected . Classification means arrangement of data according to some characteristic. After classification , tabulation of data is done

STAGE : 3:- PRESENTATION:- After the data has been collected and organized and is presented in a systematic manner which facilitates statistical analysis . Presentation of data can be done in diagram or graph

STAGE:4:- ANALYSIS:- Analysis means presentation of data by applying some techniques .The purpose is to digout the hidden information which is useful for decision making . Analysis helps in presenting the data in such a way that it draws some meaningful conclusion

STAGE:5:- INTERPRETATION:- Interpretation helps in drawing conclusion from the data collected and analyzed. Interpretation includes statistical thinking , skill and experience .If the data has been analyzed and not properly interpretate d the whole object of investigation fails

Difference between Singular and Plural sense of statistics

Singular	Plural
1. It is singular.	1. It is plural.
2. If refers to statistical methods.	2. It refers to a series of data.
3. It is always of scientific nature.	3. It may be of primary or secondary nature.
4. It is the form of tools applied to process the material.	4. It is the form of raw material.

Scope of statistics

Subject Matter Of Statistics

Statistical methods	Applied statistics
• Collection of data	• Descriptive applied statistics

• Classification of Data	• Scientific applied statistics
• Presentation of data	• Inductive or Inferential statistics
• Analysis of data	• Analytical statistics
• Interpretation of data	

Nature Of Statistics

As science as an art

Subject matter of statistics

Statistical methods or Stages in statistical enquiry

According to 'Croxton and Cowden' statistics is collection,classification/organisation, presentation,analysing and interpretation of numerical data.

Five stages of statistics

Stage 1: Collection of Data is the first step in statistics investigation , at most care should be taken in collecting the data because they form foundation to statistical analysis. These should be such as to fulfill the basic objective of proposed study.

Stage 2: Classification and organisation of data collected has to be organised properly. Hence, it should be edited, Classification and the tabular. Editing is done very carefully so that omissions, inconsistencies may be corrected. Classification means arrangement of data according to some characteristics. After classification tabulation of data is done.

Stage 3 : Presentation

After the data has been collected and organised and is presented in a systematic manner which facilitates statistical analysis. Presentation of data can be done in form of diagrams or graphs.

Stage 4 : Analysis

Analysis means presentation of data by applying Statistical techniques. The purpose is to dig out the hidden information which is useful for decision making., to draw some meaningful conclusions.

EXAMPLE : Measures of central tendency, dispersion,

Correlation, regression, index numbers, time series etc.

Analysis helps in presenting the data in such a

way that it draws some meaningful conclusion.

Stage 5 : Interpretation

Interpretation helps in drawing conclusions from the data collected and analyzed. It includes statistical thinking skill and experience. If the data has been analyzed and not properly interpreteded the whole object of investigation fails.

- **Applied statistics**

It deals with application of statistical problems like population, National income, wages, price, etc.

EXAMPLE: If we have to estimate the investments or savings of a country then special techniques has to be followed.

CLASSIFICATION OF APPLIED STATISTICS

1. DESCRIPTIVE APPLIED STATISTICS

2. SCIENTIFIC APPLIED STATISTICS

3. INDUCTIVE OR INFERENTIAL STATISTICS

4. ANALYTICAL STATISTICS

Classification of APPLIED STATISTICS

1. **Descriptive Applied Statistics :**

Descriptive applied statistics relates to data

which may be past or present. The business statistics

are descriptive applied statistics as they

are concerned with analysis measurement

and presentation of business facts.

EXAMPLE : The data collected in the census helps us in framing different political and social policies, similarly statistics concerned with prices, wages, industrial and agriculture production are useful in framing economic policies.

2. Scientific Applied Statistics

Scientific APPLIED STATISTICS deals with the formulation of physical and psychological laws on the basis of quantitative data for descriptive purpose by the use of established definite laws.

EXAMPLE: Income elasticity of demand to find the level of income.

3. Inductive or Inferential Statistics

Inductive statistics refers to forecasting, prediction, estimates or judgment made for a universe on the basis of random sampling techniques.

4.Analytical Statistics

All the statistical techniques based on mathematical theories such as association of attributes, correlation and regression analysis and probability theories are known as analytical statistics. Application of analytical tools of statistics enables to compare the results of two or more phenomena and establish the relationship.

- **Nature of Statistics**

In nature of statistics we study whether statistics is a science or an art.

1.Statistics as a Science

Science, by definition, is a systematic body of knowledge which studies the cause and effect relationship and endeavors to find out generalization with help of certain principles or laws. In simpler terms, it explains the facts. Further, the primary features of science are:

It is a systematic study of any subject

It takes a fact and tries to establish the relationship between cause and effect

Also, the laws of science are universal in nature

Tippet defines statistics as, "*A science, where the statistical method is a part of the general scientific methods and based on the some fundamental ideas and processes.*"

2.Statistics as an Art

Art is action or the actual application of science. While Science teaches us to know, Art teaches us to do. Further, Art has the following characteristics:

It is a group of actions which solve a problem

It does not describe the facts but examines the merits and demerits and suggests ways to achieve the objective

According to Tippet, "*Statistic is both a science and an art.* It is a science in its methods are basically systematic and have general application ,and an art because their successful application depends, to a considerable degree, on the skill and special experience of the statistician, and on his knowledge of the field of application."

Conclusion

Hence it can be said that statistic is a both science as well as an art. In work of professor tippet "statistics is both science and an art".

It is a science because its methods are basically systematic and have general application and art because there successful application depends to considerable degree on a skill and special experience .

Functions and limitation of statistics

Definition

According to croxton and cowden,statistics may be defind as a science of collection, organisation , presentation and interpretation of numerical data

Functions Of Statistics

1) DEFINTENESS :- Statistics presents facts in a precise and definite form. Thus helping in proper comprehension of what is stated. The statement like there is a lot of unemployement in India is not in definte form. But statement like there 217 unemployed for every 1000 is in definite form or there is 21.7% of unemployement in India.
2) CONDENSATION:-.The facts should be given in a definite form numericals are more convincing and the data can be condensed in precised form. It condenses mass of figures which helps in condensing the complex data into a few significant to make them understandable.

For Eg. From individual price level of all goods price situation cannot be determined . But if general price index is given it is easy to understand the price situation.

3) COMPARISION:- Statistics facilitates comparison and correlation . The relation between two groups is best represented by a certain mathematical quantity like average grades,ratio,co – efficients etc.
   For Eg. The per capita income of underdeveloped country is of much use if we know the per capita income of advanced countries.

4) TESTING HYPOTHESIS:- Statistical Methods are extremely useful in formulating and testing of hypothesis and to develop new theories.with the help of statistical techniques.
   For Eg The effects of imposing tax on impact of petrol on its consumptions of a country can be known. OR whether students were benefited from extra coaching or not.
5) PREDICTION OR FORECasting:- Statistical Methods helps in forecasting future events and plans..
   For Eg. If a business man has to decide how much he should producein 2020, then he would analyse the sales data of the previous year to forecast the maximum production.

6) FORMULATINF POLICIES. Statistical Methods provide the best material for framing suitable policies. Policies of organisations are formulated in advance before their implementation. A perfect knowledge of future trends with the help of statistical techniques , is usefull informing suitable policies and plans.
   For eg. The decision regarding the import of oil dependsupon the total consumption and internal production.

   Conclusion: In a net shell the important functions of statistics are definiteness , condensation , comparison predicition framing and testing of theories and policy making.

Limitation Of Statistics

Science of statistics has a universal application but suffers from a certain limitation unless the data is properly collected and critically interapted .There is every possibility that there may be wrong conclusion. There is also posibility of misuse of statistics. Following are the limitations of statistics.

1). Qualitative aspects are ignored : the phenomeon which cannot be expressed quantitatively cannot be a part of statistics.

2)Statistics dose not deal with individuals : statistics deal only with the aggregate , rather then individual items.

3)Results are there only on an average : in statistics ,it is not possible to study the effects of each factor seperately, as it is done in other sciencs. It is same for all.

4)Statistics may be misused : the statistical data can be misused by the ignorence or intentionally.

5)It can be used only by experts : the techniques of statistics are not simple and cannot be used by a lay man.only the experts having the knowledge of statistical methodology and its application can handle the data carefully.

6)Interferance may be wrong without any refrence and may lead to wrong conclusion.For Eg. In cloth business profits earned during 3 years are Rs.1000, Rs.2000, are Rs.3000 respectively, and in Paper business are Rs.3000 Rs.2000, Rs.1000. Thus the average profit in

both business in Rs.2000 year . It may state that status of both business is same which may not be true.

7)Statistics is only a means but not an end : the tools of statistics are only a means to understand a given problem , but does not give a method to solve a problem. For Eg. With the help of data % of unemployed can be know but not the method to solve it.

UNIT-2:CLASSIFICATION AND PRESENTATION OF DATA

PART1

CLASSIFICATION OF DATA.

- MEANING AND TYPES OF STATISTICAL INVESTIGATION OR ENQUIRY.

An investigation is inquiry or survey which means search for knowledge.Statistical inquiry or investigation means search for knowledge with the help of statistical methods.Statistical investigation is a comprehensive process which lays structure for decision making.It is a technical job which requires specialized knowledge and skill

- It provides answer to various management problems that the decision makers face.It gives a set of recommendations that helps business organisation to develop and implement most viable strategies
- According to Griffin,"Statistical enquiries have always require considerable skill on the part of statistician,Rooted in a broad knowledge of the subject matter area combined with considerable ingenuity in overcoming practical difficulties."
- A Statistical inquiry or investigationis nothing but search for the TRUTH. It attempts to find the best possible solution to a problem and supply Answers to all the questions emerging out of it.

If a Problem relates a physical or natural sciences , then data can be obtained by experiments. But if the problem relates to social sciences like business , Industry, Management, Economics, politics, etc . the relevant data can be gathered only through Statistical enquiry.

- **TYPES OF STATISTICAL INVESTIGATION:**
- 1.Special purpose inquiry.
- 2.General purpose inquiry.
- **1.SPECIAL PURPOSE INQUIRY.**
- It relates to the field in which we have special mission to fulfill.
- Eg:Use of holding a particular function at a particular place
- **2.GENERAL PURPOSE INQUIRY**
- It relates to the fulfillment of any objective under consideration for which data is collected.
- Eg:Total number of students in a college, how many of them are males and females etc.
- ***STAGES OF STATISTICAL INVESTIGATION.***
- The following are the stages of Statistical investigation:

- 1.Planning the statistical Investigation.
- 2.Execution of the investigation.
- **1.PLANNING THE STATISTICAL INVESTIGATION.**

Careful planning of statistical investigation is essential to get best results at minimum cost and time. Planning should proceed execution.The following are the points to be considered while planning a statistical investigation:

- Objectives of the investigation.
- Scope of investigation.
- Nature of information to be collected should be decided.
- Units of data collection should be defined.
- Source of data collection(i.e.)primary or secondary data should be decided.
- Methods of data collection should be(i.e)census method or sampling method should be decided before hand
- Choice of frame should be made.
- Reasonable standard of accuracy should be fixed (estimate or actual)
- Other considerations should include whether the investigation is:
- a. Official or semi official or private. b. Initial or repetitive.
- c. Confidential or open.
- d. Direct or Indirect.
- e. Regular or ad-hoc
- **2.EXECUTION OF THE INVESTIGATION.**

1.COLLECTION OF DATA
2.CLASSIFICATION OF DATA
3.PRESENTATION OF DATA
4. ANALYSIS
5.INTERPRETATION

COLLECTION OF DATA.

- The first step in the conduct of a statistical investigation or inquiry is collection of data.

The person who conducts the investigation/inquiry is known as an **INVESTIGATOR**

- The person from whom the information is collected are known as **RESPONDENTS/INFORMANTS**.
- The persons who help the investigator in collecting the data are called **ENUMERATORS (agent/assistant)**.

SOURCES OF DATA

INTERNAL SOURCES

- National Sample Survey (NSS) by the Government
- Individual Institutions
- Research bodies
- Economical and statistical departments
- Business organisations:- Profit, sales, expenditure etc

EXTERNAL SOURCES

PRIMARY DATA

SECONDARY DATA

- SOURCES SRCONDARY OF DATA

The following are the sources of Data:

1.Internal Source.

2.External Source.

a) Primary Data.

b) External Data.

- **INTERNAL SOURCE:**

When the data is collected from within the organization then it is said to be collected from internal sources.

INTERNAL DATA:

Data collected through internal sources is called internal data.

It comes from Government departments, business organisations which generate them in the form of production,purchase,expenditure,accounting system

Sales records ,customer feed back, Market Activity, Distribtor reports, profit etc., of any business organisation.

PRIMARY DATA

- MERITS
- **Orignal in nature!**
- **More reliable, authentic and accurate!**
  No scope of misinterpretation!
- **Can be used with confidence**
- **Matches the needs**
- **Free from Bias**
- DEMERITS
- **Expensive**
- **Time consuming**
- **Raw Data needs processing**
- **Easy to manipulate**
- **Creating new methods like questionaire**
- **Skilled Labour**

<u>Methods of collecting primary data</u>

1. Direct personal interview/investigation.
2. Indirect personal interview/investigation.
3. Information through correspondents.
4. Questionnaire method.

(a) Questionnaire sent through post/mail

(b) Questionnaire sent through Enumerators/**Schedule**

<u>Direct personal interview</u>:

According to this method, the **investigator** obtains the data by a personal interview. he contacts the source of information directly and personally. He will contact each and every possible source of information.**It is done in following situations. For eg. If investigator wants to collect data about wages of a factory, then he would go to factory and contact the workers to obatin the desired information**

- When the scope of enquiry is small

- When high degree of accuracy is requird
- When result of investigation to kept secret
- When area of investigation is hetrogeneous

Merits:

(a) The information collected will be more **accurate** , original and reliable

(b) The **response** will be more.

(c) There is **uniformity** in the collection of data.

(d) The **communication gap** can be filled by choosing the right words for the questions which are to be asked by the investigator from the respondents.

Demerits:

(a) It is very **costly** method.

(b) It is very **time consuming**.

(c) The chances of **personal bias** are greater.

(d) It requires **extra personal qualities of the interviewer** like tactfulness, courage, courtesy, etc.

(e) It can give **wrong results**.

(f) It can be used only when **area of investigation is small.**

2. Indirect personal Interview/investigation:

The investigator **contacts third parties or witnesses**, who are concerned with the information directly or indirectly and are capable of supplying the necessary information. This method is used by Goverrnment committees to get views of the people relating to enquiry. For eg. Number of students taking durgs

It is done only in the following situations:

(i) When the direct sources do no exist.

(ii) When the direct sources cannot be relied upon.

(iii) When the direct sources are indifferent on their part.

(iv) When the area of investigation is large.

Merits:

1.The investigator can **take the help of expert enumerators** to collect the data.

2. It is **simple and convenient** method of investigation.

3. It is **economical.**

Demerits:

1. If the **enumerators are not expert** in their field, then **wrong data** may be collected.
2. The chances of **personal bias** are greater.

3.Information through correspondents:

Under this method the investigator does not collect the information from the persons concerned directly. **He appoints generally local agents in different parts of the area** under investigation. These local agents are called **correspondents.** These correspondents collect the information and pass it on to the investigator from time to time.

It is done in the following situations:

(i) When the area of investigation is very large.

(ii) When regular information is required. For eg Newspaper require information regulary

Merits:

1. It is cheap or economical.
2. It covers large area.
3. It is regular

Demerits:

1. It is not very reliable because of the biaseness of the correspondents
2. The data so collected may not- be uniform

Questionnaire method:

In this method, the necessary information is collected from the respondents through a questionnaire. **A questionnaire is a set of questions relating to the enquiry**.

The information can be collected through questionnaires in two ways.

1 Questionnaire sent through post/Mailed Questionnaire method

2.Questionnaire sent through enumerators/**Schedule**

1.Questionnaires sent through post/mail:

In this case the questionnaire is sent to the informant and the **informant**

himself fills the answers to the various questions asked in it. And send it back.

This method is applicable:

(a) When the informant is literate.

(b) When the area of coverage is very wide.

Merits:

1. **Wide coverage** is possible.
2. It is **economical** because no enumerators are required.
3. It **saves time.**
4. **Errors** in the investigation **are very small** because the information is obtained directly from the persons concerned.

Demerits:

1. Information can be collected from the **literates only.**

2 There is very likelihood of **biasedness.**

3. **Supplementary questions** cannot be asked.

4. If questionnaire is **complex, it may not be responded**.
5. There are **chances of non-response**. It means the respondent may throw it as a **waste paper**.

Questionnaires sent through enumerators/Schedule

Under this method, the enumerators are appointed by the investigators and the **enumerators contact the informants**, **get replies** to the questionnaire and **fill them in their own hand writing** in the questionnaire form.

Merits:

1. Information can be collected even from the **illiterate persons**.

2. Information got through it, will be more **reliable and accurate**.

3. Because of personal contact by the field workers, the **defects** if any **in the questionnaire can be detected**

4. It covers **wide area.**

5. There are **less chances of non-response.**

Demerits:

1. It is **costly** because Enumerators has to be paid.
2. It is **time consuming**.
3. It requires **trained enumerators.**
4. The **personal bias** of enumerator may lead to false conclusions.
5. It can be **employed only by big organisations.**

Drafting the Questionnaire/Features of Good Questionnaire

1. The questionnaire should **not be too lengthy.**
2. A **decent paper** and printing is to be chosen.
3. The questions asked should be **well worded** and should not be ambiguous.
4. The questions asked should be in **proper order**.
5. **Irrelevant questions** should be avoided.
6. Questions **should not injure the feelings** of the respondents.
7. Questions should be such that the respondents has to do **minimum writing work**.
8. Necessary **instructions should be given**.
9. Questions involving **mathematical calculations** should be avoided.
10. There should be **guarantee** to keep the **answers secret.**
11. A **request for cooperation** should be made.
12. Some incentives should be given like **gifts or presents**.
13. If a questionnaire is been sent by post then a **self addressed and stamped envelope** should also be supplied.
14. Attractive questionaire impress the respondent.
15. Questionaire should be properly examined by the investigator before sending
16. The questionaire drafted should be tested with a small group of persons.

- **Differences between Questionnaire and Schedule**:

Questionnaire

1.It is mailed to respondents

2.Answers filled by respondents

3.Only Educated people can answer it

4.Response is low

5.It is cost effective/cheap

6.Confidentiality is maintained

7.Suitable for large areas

Schedule

1.It is taken to respondents

2.Answers filled by Enumerators

3.Can be used for uneducated people also

4.Response is high

5.It is Expensive

6.Confidentiality is not maintained

7.Suitable for small areas

SOURCES OF SECONDARY DATA

Sometimes it is not possible to collect the first hand data as it is time consuming and money consuming in such situations secondary data can be used . for ex: **D**ata published from newspapers, magazines, bulletins, reports, journals etc.

It refers to the data which is not collected by the investigator himself. ***It is second hand data.***

Secondary data can be classified into two types:

- PUBLISHED(information from various

Local, National, International Publications)

- UNPUBLISHED(data which is not published but used by someone)

SOURCES OF PUBLISHED DATA

1. International publications: Certain international institutes like world bank ,UNO ,IMF, etc publish reports from time to time regarding economic matters which are of great significance.

Ex: Annual reports of international labour organisation or by world bank, International Monetory Fund, UNO etc.

2. Official publications of central and state government: state and central government collects information regarding important economic variable like national income(NI),savings ,investments, employment, currency, etc.

3. Committee reports: sometimes the government appoints survey and enquiry commission to get the expert view on the matters of great importance .

ex : reports of public accounts committee of lok sabha, Tarrif Commission, National Agricultural Commission etc.

4. News papers and magazines: news papers like economic times, magazines like India today etc.Publishes the information regarding economic variables.

5. Individual research scholar: individually various scholars do the research and collects the information by some research institutions. Private Publications

2.Unpublished data:

When the data collected by someone but which are not published but used

by someone else. Is called as secondary data from unpublished sources.

Ex: Reports prepared by private investigation companies., reports of trade union, cooperative societies etc.

PRECAUTIONS TO BE TAKENWHILE USING THE SECONDARY DATA:-

While using the secondary data proper scrutinity of the secondary data must be done to have accurate and reliable conclusion

FOLLOWING POINTS HAVE TO BE CONSIDERED WHILE USING SECONDARY DATA:-

- *The investigator has to assure whether the data which is collected by qualified person or not.*
- *Whether the data is reliable or not.*
- *Whether the information is related to distanced post or near post.*
- *The purpose of collection of data is to be determined.*
- *Is the information sufficient or not.*

Difference between primary data and secondary data

PRIM ARY DATA	SECONDARY DATA

more	Comparatively less
Investigation agency collects the data on its own	Some other agencies collects the data for its own use which is used by others
It is time consuming	Takes less time to collect the data
More reliable as data is collected by the investigator himself	Less reliable as data is collected from someone
Requires big organization	Don't need of any organizational setup
Requires less precautions	Requires more precautions
The first hand data as it is directly collected from the source Direct personal interview Indirect personal interview	Second hand data 1 published 2 unpublished

Through correspondents Questionaire Schedule	

Characteristics of Primary data and Secondary data.

The investigator must decide whether he needs to use primary data and secondary data for his investigation

1. Nature and scope of enquiry.
2. Availability of financial resources.
3. Availability of time.
4. Degree of accuracy required.
5. Status of the investigation in either he is an individual investigator or an organization.

Statistical Units of Collection.

1. **Simple or Simple Unit**: It is expressed in terms of single determining units like liter, kilometer etc.
2. **Composite Unit**: It consists of two or more single units and is based upon more than one determining characteristics.
3. **Hypothetical Unit**: For comparison purpose sometimes hypothetical units like horsepower are

used. Hypothetical units have various definitions and different applications.

Methods of Collection of Data

Before hand and before starting a research the investigator has to decide whether the survey has to be conducted for the entire population or for a sample.

- Population: It means the number of people in a particular geographical area, but in Statistics population means all the units which are understudy or the number of the observations collected with the study.

Ex: If we want to know the average marks of 250 students, then the population of is 250 students.

- Sample: It means a small representative of the whole population.

ex: To know the average consumption of expenditure of people having income between 10,000-15,000,it is not necessary to contact each and every person but few people can be surveyed at random.

There are two methods of data collection:

i. Census Method: When the entire area of population is contacted, it is known as Census method. It is very

rarely used. ex: To know the total population of India the census of population is conducted after every ten years. With the help of census method, the data collected will give perfect results. But it is time consuming as well as expensive.

ii. Sample Method: It is used when small group is taken into account to represent the whole population.

Statistical Error:

The difference between the data collected and the actual value of fact is known as statistical error.

Reasons of Statistical Error:

i. Selection of wrong sample.
ii. Wrong information supplied by informant.
iii. Collection of data by estimation.
iv. Bias attitude of enumerator and investigator.
v. Collection of data by inexperienced enumerator.
vi. Selection of unused and unsuitable unit.

Biased Error: The errors are dangerous for investigation. These arises because of the biased opinion of the enumerator with the help of biased data, the investigator may prove anything which indicates the misuse of statistics. Hence, biased errors are intentional errors.

Unbiased Errors: These are not of serious type. These errors arise because of statistical techniques like approximation etc. They do not affect the results adversely. Unbiased error are unintentional errors.

Organization of Data:

The data is unorganized when it is collected either through or primary source or secondary source. Converting unorganized data and raw data into the required the form is called as organization of data.

**** **There are three Stages of organization*****

1. Editing

2.Clasification

3.Tabulation

1.Editing of the data : It refers to the detection of possible errors and irregularities committed during the process of collection of data.The data collected must be edited very carefully so that any omissions,inconsistancy, mistakes can be corrected or adjusted. If data is not edited then it leads to the wrong conclusion . According to Wed bailey & john Cummings the purpose of editing is .

a. Completeness
b. Uniformity
c. Consistency
d. Accuracy

2.Classification of Data : According to 'Horace Secrist ' classification is the process of arranging the data into sequences and groups according to their common characteristic but related parts .

Objects of classification:

1. To make the characteristics of **similarities and dissimilarities clear.**

2. To **present the complex**, scattered data in a concise, logical and **understandable form.**
3. To enable one to form mental pictures of objects of perfection and conception.
4. To make a **comparative study** possible.
5. To **remove irrelevant details** from the data and to make possible the tabulation of the data and its further analysis.
6. To make **clear** the underlying unity amongst different items.
7. To enable a statistical treatment of the material collected.
8. To make possible the **analysis** and generalisation of the data.
9. To pinpoint the most significant features of the data **at a glance.**

Merits of classification:

1. clarity: It presents data in simple and clear form. It removes complexity of data and makes it simple.
2. Stability: It should be stable to make comparison easier.
3. Flexibility: It should be flexible as to incorporate new ideas in future.
4. Homogeneity: Similar information is distinguished from dissimilar information.
5. Comparison: It facilitates comparison.
6. Condensation: It helps in shortening the huge mass of data.
7. Presentation: It is helpful in the presentation of data in the shape of tables, diagrams, graphs etc.
8. Analysis of data: Classification of data is essential for the application of statistical tools of analysis.

METHODS OR TYPES OF CLASSIFICATION
1. GEOGRAPHICAL
2. CHRONOLOGICAL
3.CONDITIONAL
4.QUALITATIVE: TWO FOLDED
MANI FOLDED
5. QUANTITATIVE

1.Geographical classification :

when data are classified on the basis of geographical areas like villages , city, state , zone, country etc
Example : statement wise population in india in 2020

2. Chronological classification :

when the data is classified on the basis of time like a date, hours , daily, weekly, monthly , quarterly , yearly , decade , etc

Example ; population of india since 1951

3.Conditional classification:

when data are classified on the basis of Gender, quality, honesty etc.,

4.Qualitative classification :

Data is classified **on the basis of** some attributes or **qualities** like literacy, honesty, intelligence , creativity, artistic skills etc.

(a) Simple or Two-fold or Dichotomous classification:

When the **data is classified into two categories** on the basis of presence or absence of an attribute, it is called two-fold classification.

Eg: Population can be divided into **Males & Females**

b) Manifold classification:

When the **Data is classified** on the basis of **more than one attribute**, it is called manifold classification.

Ex: Population may be classified according to gender, marital status, and literacy.

5) Quantitative classification :

when the data is classified on the basis of characteristic which can be measured is called as quantitative classification

example :

Marks	No. of students
0 – 10	5
10 -20	15
20 -30	17

30 -40	6
40-50	5

PROCESS OF CLASSIFICATION:
METHODS OF QUANTITATIVE CLASSIFICATION

(a) RAW DATA
When the Investigator has collected the data but not arranged the data in a systamatic manner is called as UnOrganised data or Raw data

Weight measurement of 50 workers

80 75 68 75 67 78 98 45 56 67 85 67 91 94 95 96 96 100 74 75 76 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 65 65 67 78 89 87 67 98 76 54 67 78

(c) Statistical DATA: DATA Presented in some order and sequence
(A) INDIVIDUAL SERIES
(B) FREQUENCY DISTRIBUTION SERIES :
1.DISCRETE SERIES
2. CONTINOUS SERIES
Hence we have three types of serries: IndividuAl, Discrete, and Continous

INDIVIDUAL SERIES The values of all units are shown Separetely either in ascending order or in descending order. Eg: Marks of students Roll No. wise for a Particular Subject can be arranged according to their performance i.e, highest to lowest
Eg marks of 10 Students of a class.
8 4 6 9 10 5 3 11 15 13
Arraning in assecending order
3
4
5
6
8
9
10
11
13
15
NOTE 1ST STEP IS TO ARRANGE IN ORDER

FREQUENCY DISTRIBUTION TERMINOLOGY

(a) VARIABLE:Any characteristic which varies in magnitute
It is that quantity which varies from an individual to another
Quantitative characterstic such as marks, age, height, weight etc
2 types
DISCRETE VARIABLE:These are the exact or finite, but cannot be expressed in either in decimals or in fractions.
Eg. No. of children to a family, watches, pen, books etc.
Continous variable:These variables can be expressed in fractions, and decimals

Eg. Height weight income etc

Continous series, Terminologies,and its types etc.

Continous series

The series which deals with continous variables is called as continous series.
It can be expressed in fractions or decimals with of all possibilities such as height, weight,income,marks etc.
Continous series is applicable when the observations are more.
EXAMPLE OF CONTIOUS SERIES

10-15	**5**
15-20	**17**
20-25	**22**
25-30	**6**
	N=50

MARKS	**NO.OF STUDENTS**
10-15	5
15-20	17
20-25	22
25-30	6
	N=50

1.CLASS : Each stated interval such as 10-15 and so on is called as class.
2.CLASS LIMIT : There are two limits for every class interval upper & lower limit.
3.CLASS INTERVAL : The difference between

the both the limits is called as class interval. It Is represented by I or c or h.

∴ Class interval = upper limit – lower limit

4. MID VALUES : Mid value is the average of both the class limits.

M= (lower limit + upper limit)/2

example:10-20 is the class interval then the M will be (10+20)/2 = 15

5. CLASS FREQUENCY : The number of observation which belong to a particular class is known as frequency of that class.

TYPES OF CONTINOUS SERIES

1. **INCLUSIVE SERIES : In inclusive series the upper limit of the precedding class is less than the lower limit of the succesdding class.**

EXAMPLE:

MARKS	**NO.OF STUDENTS**
10-19	5
20-29	17
30-39	22
40-49	6
	N=50

2.EXCLUSIVE SERIES : under these method the upper limit is excluded where as the lower limit included of the precedding class Is same as the lower limit of the suceededing series.

EXAMPLE

MARKS	**NO.OF STUDENTS**
10-20	5
20-30	17
30-40	22
40-50	6
	N=50

OPEN AND SERIES : If the value of any series of class are not given then it is called as open series.

MARKS	**NO.OF**

	STUDENTS
Below 20	5
20-30	17
30-40	22
Above 40	6
	N=50

Weight measurement of 50 workers

80 75 68 75 67 78 98 45 56 67 85 67 91 94
95 96 96 100 74 75 76 78 79 80 81 82 83 84
85 86 87 88 89 90 91 92 93 94 95 96 97 98
65 65 67 78 89 87 67 98 76 54 67 78

weight	Tally bars	No. Of workers f
45	I	1
54	I	1
56	I	1
65	II	2
67	𝍸 I	6
68	I	1
74	I	1
75	III	3
76	II	2

78	IIII	4
79	I	1
80	II	2
81	I	1
82	I	1
83	I	1
84	I	1
85	II	2
86	I	1
87	II	2
88	I	1
89	II	2
90	I	1
91	II	2
92	I	1
93	I	1
94	II	2
95	II	2
96	II	2
97	I	1
98	III	3
100	I	1

N=54

Conversation of inclusive into exclusive

The conversation factor is deducted from lower limit and added to upper limit

∴ CF = (upper limit of succedding – lower limit previous) /2

6 5 2 8 3 1 7 8 7 0 2 4 6 7 2 5 2 3 4 2 3 5 5 3 4 3 4 4 6 3

Individual series

0

1

2

2

2

2

3

3

3

3

3

3

4

4

4

4

4

5

5

5

5

6

6

6

7

7

7

8

8 N= 30

Discrete series

MARKS	TALLY BARS	NO. OF STUDENTS f
0	I	1
1	I	1
2	~~IIII~~	5
3	~~IIII~~ I	6
4	~~IIII~~	5
5	IIII	4
6	III	3
7	III	3
8	II	2

N= 30

Continous series EXCLUSIVE TYPE(UPPER LIMIT IS EXCLUDED)

CLASS (MARKS)	TALLY BARS	NO. OF STUDENTS F
0-2 (0,1)	II	2
2-4 (2,3)	~~IIII~~~~IIII~~ I	11
4-6 (4,5)	~~IIII~~IIII	9
6-8 (6,7)	~~IIII~~ I	6

8-10 (8,9)	II	2
		N=30

Continuos series

Inclusive type (both lower and upper limits are included)

Marks	Tally bars	No. of students f
0-1(0,1)	II	2
2-3(2,3)	~~IIII~~ ~~IIII~~ I	11
4-5(4,5)	~~IIII~~ IIII	9
6-7(6,7)	~~IIII~~ I	6
8-9(8,9)	II	2
		N=30

~~10~~ 17 15 22 11 16 19 24 29 18 25 26 23 27 30
12 15 18 15 21 28 30 28 14 13 10 16 20 22 29
19 23 11 14 30 14 17 20 24 16

DISCRETE SERIES

MARKS	TALLY BARS	NO. OF STUDENTS f
10	II	2
11	II	2

12	I	1
13	I	1
14	III	3
15	III	3
16	III	3
17	II	2
18	II	2
19	II	2
20	II	2
21	I	1
22	II	2
23	II	2
24	II	2
25	I	1
26	I	1
27	I	1
28	II	2
29	II	2
30	III	3
		N= 40

CONTINUOUS SERIES EXCLUSIVE SERIES

~~10~~ ~~17~~ ~~15~~ ~~22~~ 11 16 19 24 29 18 25 26 23 27 30

12 15 18 15 21 28 30 28 14 13 10 16 20 22 29

19 23 11 14 30 14 17 20 24 16

CLASS	TALLY BARS	NO. OF STUDENTS f
10-15 (10,11,12,13,14)	~~IIII~~ IIII	9
15-20	~~IIII~~ ~~IIII~~ II	12
20-25	~~IIII~~ ~~IIII~~	9
25-30	~~IIII~~ II	7
30-35	III	3
		N=40

~~10~~ ~~17~~ ~~15~~ ~~22~~ 11 16 19 24 29 18 25 26 23 27 30
12 15 18 15 21 28 30 28 14 13 10 16 20 22 29
19 23 11 14 30 14 17 20 24 16

CONTINUOUS INCLUSIVE TYPE (both lower and upper limits are included)

CLASS	TALLY BARS	NO. OF STUDENTS f
10-14 (10,11,12,13,14)	~~IIII~~ IIII	9
15-19	~~IIII~~ ~~IIII~~ II	12
20-24	~~IIII~~ ~~IIII~~	9
25-29	~~IIII~~ II	7

30-34	III	3
		N=40

56 71 18 31 56 40 81 64 59 49 42 12 67 7 39 76
48 7 80 84 34 58 61 28 45 0 63 61 11 81 12 24
75 64 68 43 58 79 36 83 9 64 38 30 32 60 27 18
54 63 37 58 83 41 47 38 83 45 56 28

EXCLUSIVE TYPE

CLASS	TALLY BARS	f
0-10	IIII	4
10-20	𝍸	5
20-30	IIII	4
30-40	𝍸 IIII	9
40-50	𝍸 III	8
50-60	𝍸 III	8
60-70	𝍸 𝍸	10
70-80	IIII	4
80-90	𝍸 II	7
		N=60

CONVERSION OF INCLUSIVE SERIES INTO EXCLUSIVE SERIES

GIVEN INCLUSIVE TYPE

CLASS	NO. OF STUDENTS f
10-14 (10,11,12,13,14)	9
15-19	12
20-24	9
25-29	7
30-34	3
	N=40

FIND CORRECTION FACTOR(CF)

CF=(L L of SUCCEDING CLASS – U L PRECEEDING CLASS)/2

=(15-14)/2=1/2=0.5

(25-24)/2=1/2=0.5

New Lower limit = old lower limit-CF

New Upper limit = old upper limit +CF

CLASS	NO. OF STUDENTS

	f
(10-0.5) –(14+0.5)	9
(15-0.5) –(19+0.5)	12
(20-0.5)-(24=0.5)	9
(25-0.5)-(29+0.5)	7
(30-0.5)-(34+0.5)	3
	N=40

Exclusive series

CLASS	NO. OF STUDENTS F
9.5-14.5	9
14.5-19.5	12
19.5-24.5	9
24.5-29.5	7
29.5-34.5	3
	N=40

OPEN END SERIES

MARKS	NO.OF STUDENTS f
BELOW 10	5
10-20	6
20-30	8

30-40	12
ABOVE 40	6

CONVERSION OF OPEN END SERIES INTO EXCLUSIVE

MARKS	NO.OF STUDENTS f
0-10	5
10-20	6
20-30	8
30-40	12
40-50	6

Coversion into Less than cumulative frequency distribution

Marks	No. of students f
Less than 10 (0-10)	5
Less than 20 (0-20)	5+6=11
Less than30(0-30)	5+6+8=19
Less than 40 (0-40)	5+6+8+12=31
Less than 50 (0-50)	5+6+8+12+6=37

CONVERT THE FOLLOWING INTO LESS THAN CUMULATIVE FREQUENCY DISTRIBUTION

MARKS	**NO. OF STUDENTS** **f**
0-10	**4**
10-20	**16**
20-30	**13**
30-40	**5**

40-50	12
50-60	10
60-70	9

N= 69

MARKS	NO. OF STUDENTS F =LESS THAN CF	(OR) CF
LESS THAN10	4=4	4
LESS THAN 20	4+16=20	4+16=20
LESS THAN 30	4+16+13=33	20+13=33
LESS THAN 40	4+16+13+5=38	33+5=38
LESS THAN 50	4+16+13+5+12=50	38+12+50
LESS THAN 60	4+16+13+5+12+10=60	50+10=60
LESS THAN 70	4+16+13+5+12+10+9=69	60+9=69

MARKS	NO. OF STUDENTS F =LESS THAN CF
LESS THAN10	4
LESS THAN 20	20
LESS THAN 30	33
LESS THAN 40	38
LESS THAN 50	50
LESS THAN 60	60
LESS THAN 70	69

CONVERSION OF LESS THAN CUMULATIVE INTO EXCLUSIVE TYPE

MARKS	F
0-10	4
10-20	20-4=16
20-30	33-20=13
30-40	38-33=5
40-50	50-38=12
50-60	60-50=10
60-70	69-60=9

Marks	cf
Less than 5	5
Less than 10	9
Less than 15	17
Less than 20	37
Less than 25	45
Less than 30	69
Less than 35	75

EXCLUSIVE TYPE

MARKS	F
0-5	5 =5
5-10	9-5 =4
10-15	17-9=8
15-20	37-17=20
20-25	45-37=8
25-30	69-45=24
30-35	75-69=6

CONVERT THE FOLLOWING INTO MORE THAN CUMULATIVE FREQUENCY DISTRIBUTION

EXCLUSIVE TYPE

MARKS	NO.OF STUDENTS f
0-10	5
10-20	6
20-30	8

30-40	12
40-50	6

N=37

MORE than cumulative frequency distribution

MARKS	**MORE THAN CF**	**(OR) CF**
MORE THAN 0 (0-50)	**5+6+8+12+6=37**	**37**
MORE THAN 10(10-50)	**6+8+12+6=32**	**37-5=32**
MORE THAN 20(20-50)	**8+12+6=26**	**32-6=26**
MORE THAN 30(30-50)	**12+6=18**	**26-8=18**
MORE THAN40(40-50)	**6**	**18-12=6**

CONVERSION OF MORE THAN CUMULATIVE INTO EXCLUSIVE TYPE

MARKS	**F**
0-10	**37-32=5**
10-20	**32-26=6**
20-30	**26-18=8**
30-40	**18-612**
40-50	**6**

CONVERT THE FOLLOWING INTO MORE THAN CUMULATIVE FREQUENCY DISTRIBUTION

MARKS	**NO. OF STUDENTS** **f**

0-10	4
10-20	16
20-30	13
30-40	5
40-50	12
50-60	10
60-70	9

MARKS	CF
MORE THAN 0	69
MORE THAN 10	69-4=65
MORE THAN 20	65-16=49
MORE THAN 30	49-13=36
MORE THAN 40	36-5=31
MORE THAN 50	31-12=19
MORE THAN 60	19-10=9

Conversion into exclusive when mid values are given

Mid value	f
5	6
15	5
25	8
35	2
45	4
55	8

MID VALUE = (LOWER LIMIT + UPPER LIMIT)/2

Range= 15-5=10= HEIGHT OF CLASS (UPPER LIMIT – LOWER LIMIT)

LL= MV-(R/2)=5-(10/2)=5-5=0

UL=MV+(R/2)=5+(10/2)=5+5=10 or LL + Range= 0 +10=10

CLASS	
0-10	
10-20	
20-30	
30-40	
40-50	
50-60	

MID VALUE	F
3	6
6	5
9	8
12	2
15	4
18	8
21	2

RANGE= 6-3=3
LL=3-(3/2)=3-1.5=1.5
UL=LL+R=1.5+3=4.5

CLASS	F
1.5-4.5	6
4.5-7.5	5
7.5-10.5	8
10.5-13.5	2
13.5-16.5	4
16.5-19.5	8
19.5-22.5	2

Tabulation

DIAGRAMS

ACCORDING TO M. J Moraney diagram represents and register a meaningful impression of facts and give a clear picture before we think.

Importance or utility of Diagram

1. **Gives a clear picture of the data**
2. **Makes comparision easy**
3. **Universally accepted**
4. **They have a impressive value**
5. **Saves times energy and money**
6. **Can be used only for numerical statistical analysis**
7. **Gives more information than a table**
8. **Data can be condensed with diagram and can be easily remembered.**

Rules of making a diagram

1. **Heading must be clear which conveys the purpose of diagram**
2. **Size not be too small nor too big.**
3. **Diagram must be attractive**
4. **Must be proportionate in height and width.**
5. **Scale must be presented along with the diagram.**
6. **Suitable title and footnote must be given**
7. **Vertical diagram should be preferred, to the horizontal diagram.**
8. **The date should be accurate**
9. **It must be self explanatory.**
10. **It must be neat and clean.**

TYPES OF DIAGRAM

Diagrams are mainly divided into four main category.

- **One dimensional :-line diagram, simple bar diagram, sub divided diagram, cumulative bar diagram ,% bar diagram, multiple bar diagram, etc**
- **Two dimensional :-rectangles, squares, circles etc**
- **Three dimensional :- cubes, cylinders, spheres, prisms etc**
- **Pictograms and cartograms.**

1.One dimensional diagram.

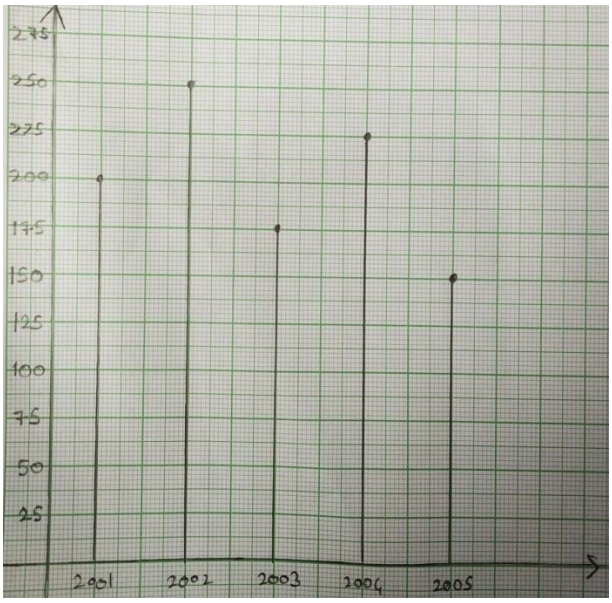
Width of each bar should be same. Length dimesion is given more importance as it represents the size

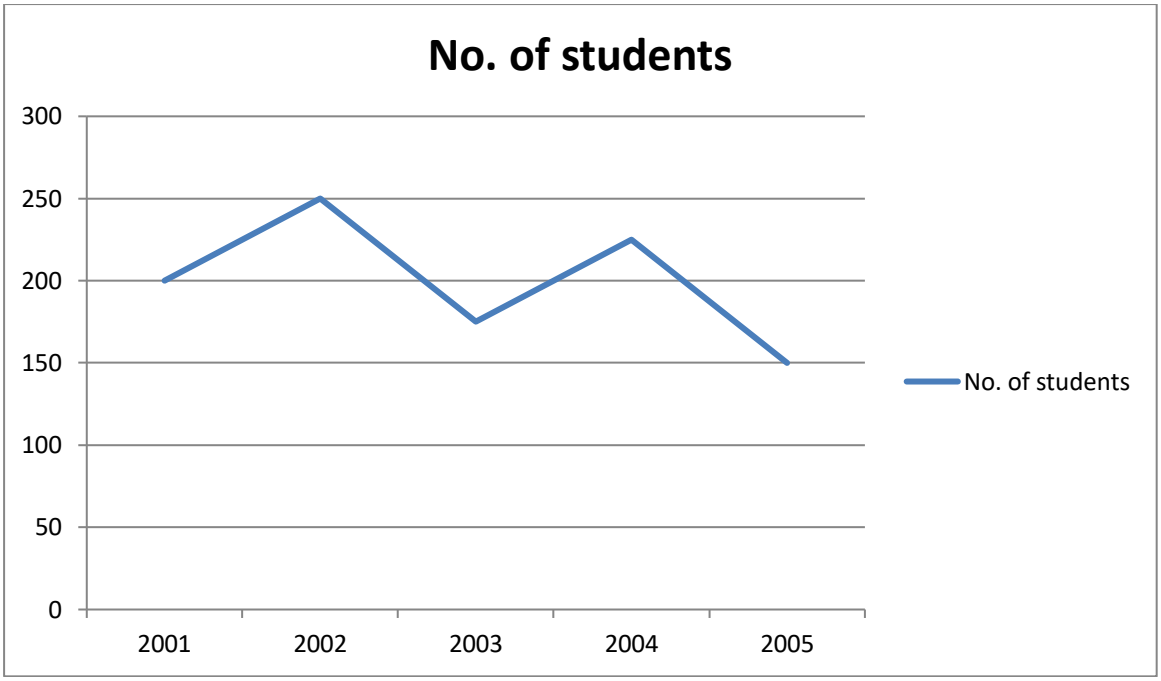
- **It is further classified into two-**
- **Line diagram**
- **simple bar diagram,**
- **sub divided diagram,**
- **cumulative bar diagram ,**
- **% bar diagram,**
- **multiple bar diagram, etc**

Line diagram:-one line is drawn only variable. Vertical or horizontal

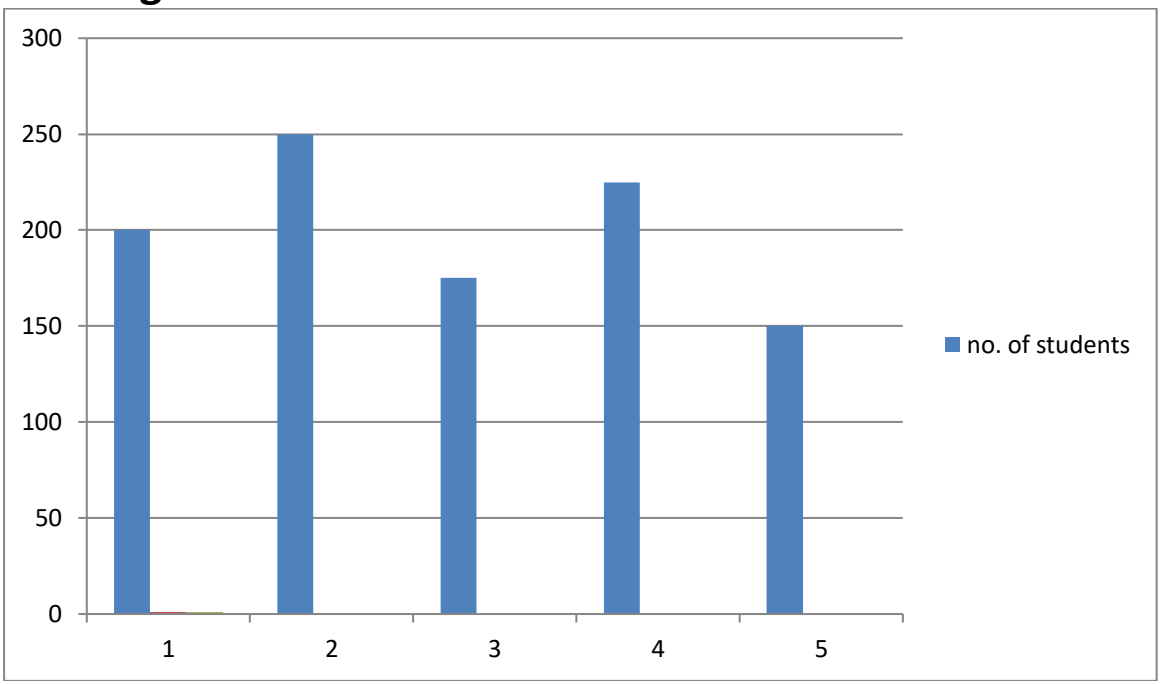
year	No. of students
2001	200
2002	250
2003	175
2004	225
2005	150

Line diagram scale on x-axis 1cm = 1 year, on y-axis 1cm = 25 students



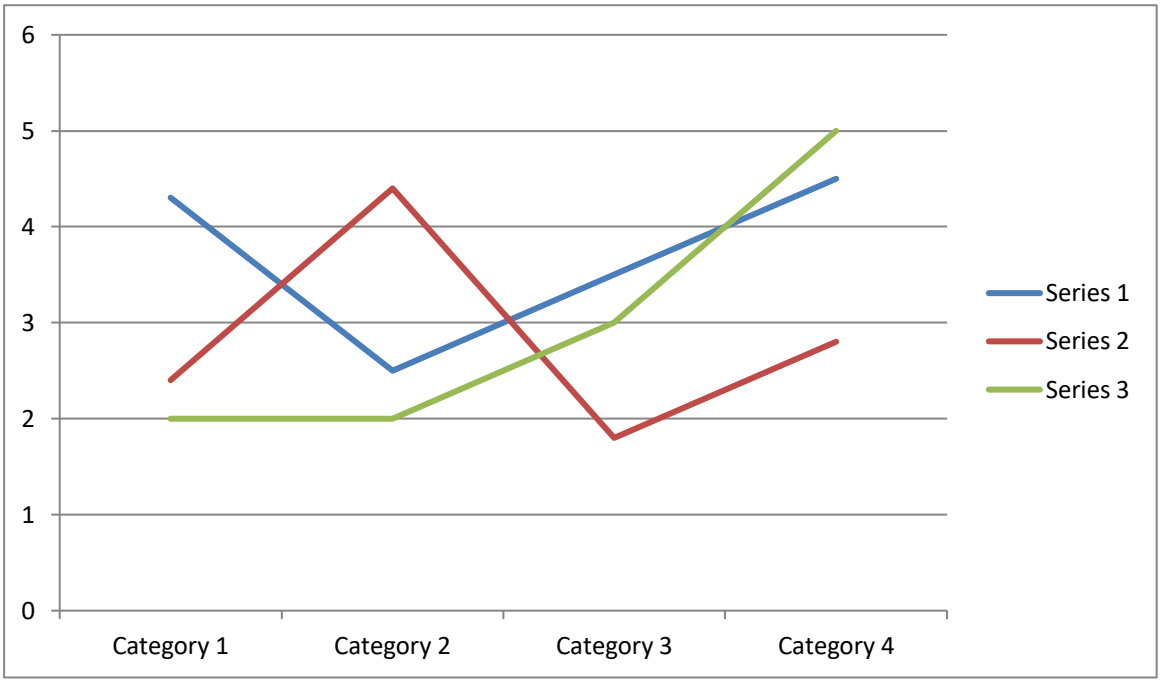


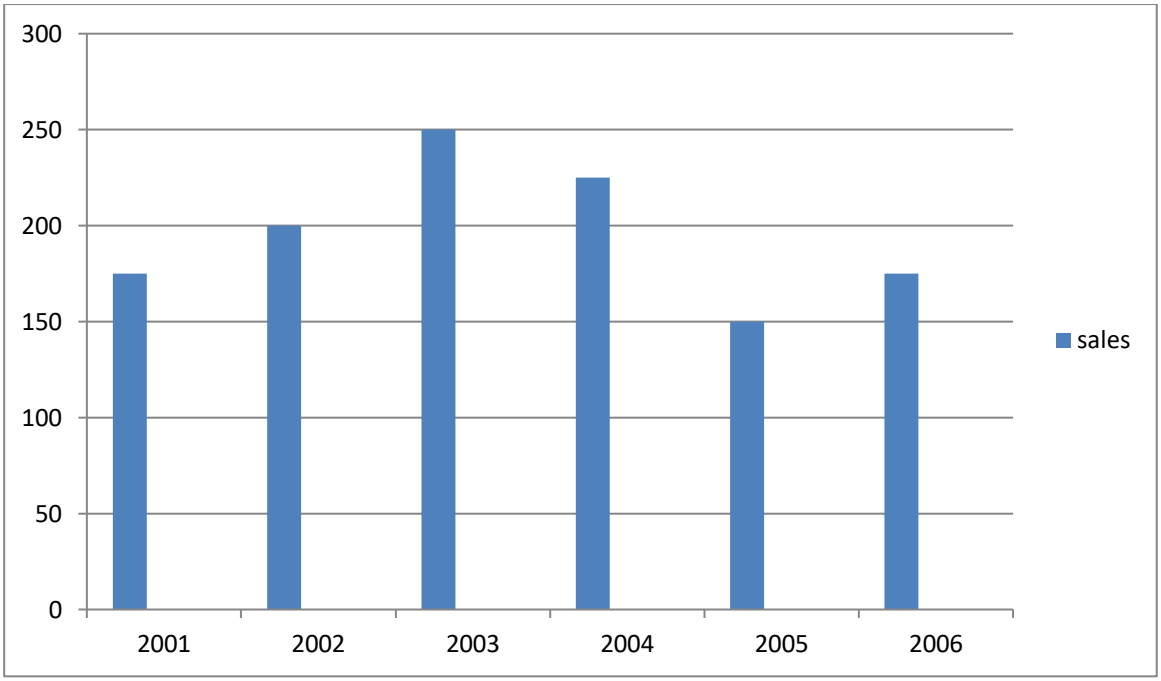
Bar diagram



year	Sales (in thousand Rs)
2001	175

2002	**200**
2003	**250**
2004	**225**
2005	**150**
2006	**175**



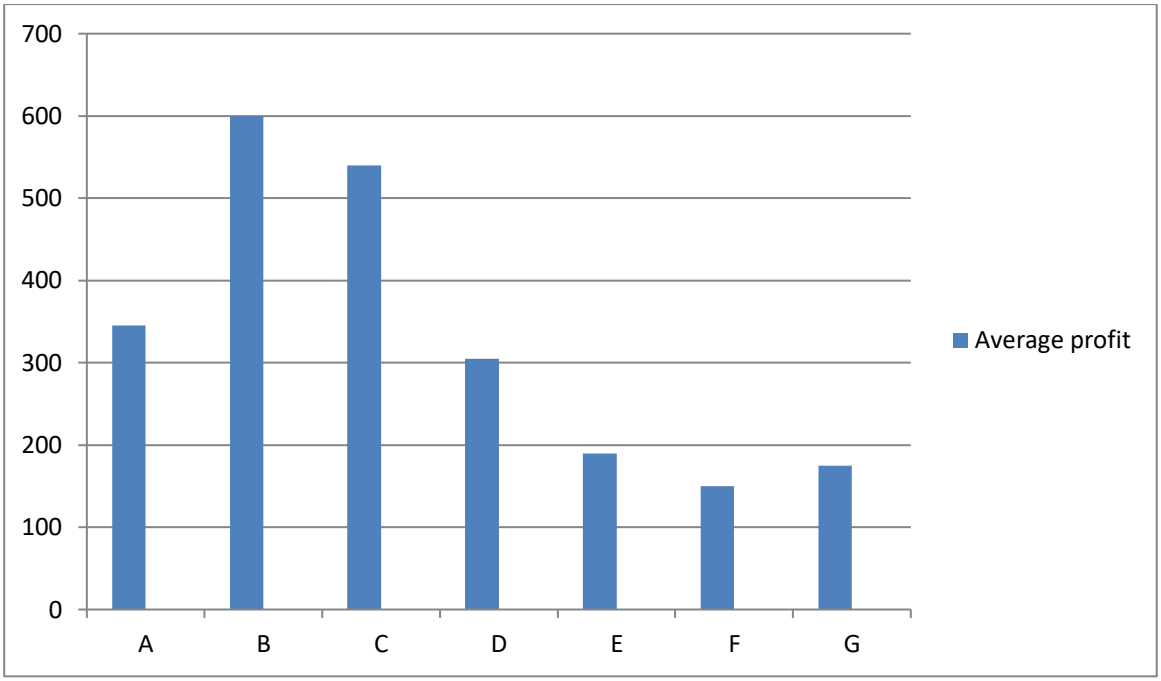


Line and bar diagram

Scale on y-axis 1cm = rs 100

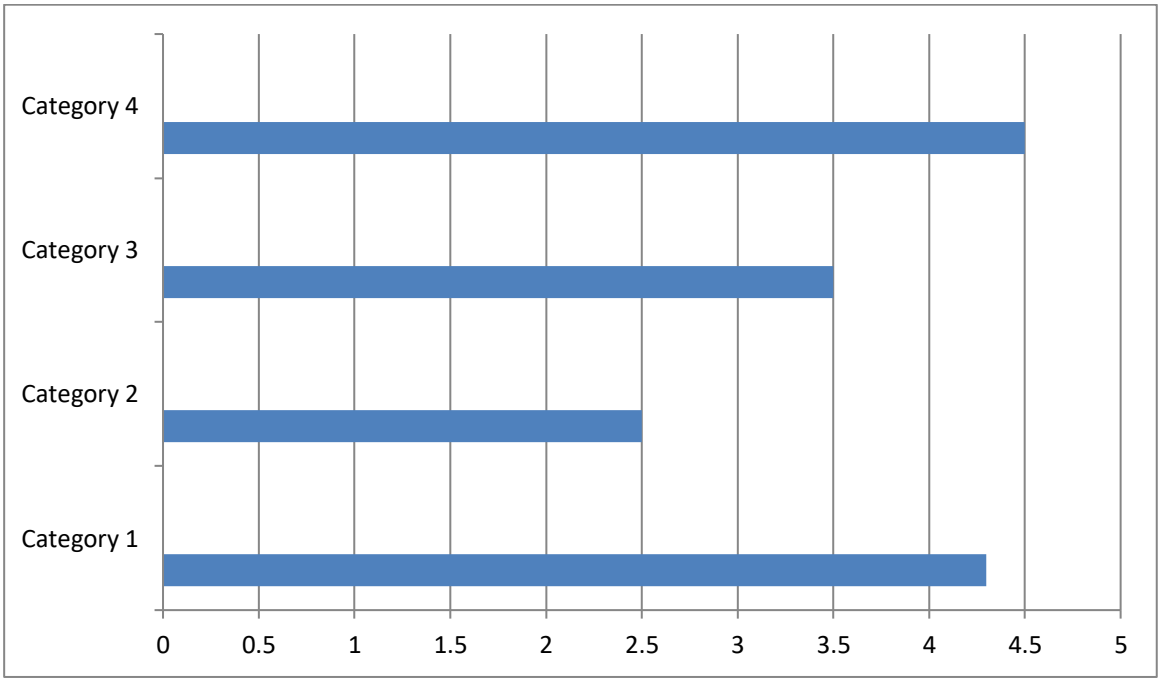
On x-axis 1 cm = 1 firm

Firm	Average profit
A	345
B	600
C	540
D	305
E	190
F	150
G	175



Horizontal bar diagram.

A horizontal bar graph is a graph whose dependent variable is shown on the horizontal scale. The length of each bar is proportional to the quantity to be presented and all bars go across from left to right.



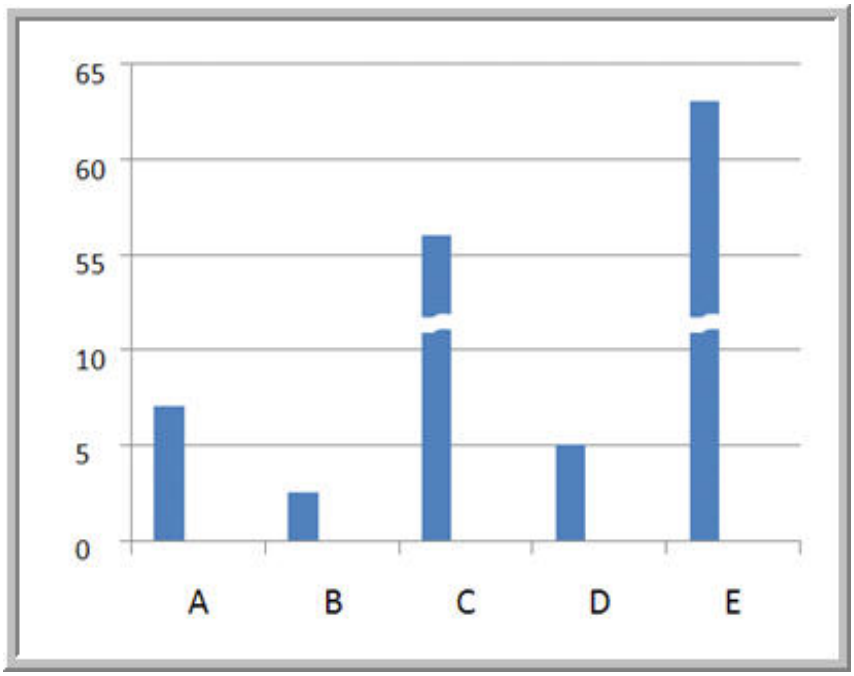
What and where is horizontal bar diagram used for?

Bar graphs usually present categorial and numeric variables grouped in class intervals. They consist of an axis and a series or labelled horizontal or vertical bars. The bars depict frequencies of different values of a variable or simply the different values themselves. The numbers on the x-axis of a bar graph or the y-axis of a column graph are called the scale.

Broken bar diagram.

This diagram is used when value of some variable is very high or low as compared to others. In this case the bars with bigger terms may be shown broken.

Name	A	B	C	D	E
Marks	7	3	58	5	64



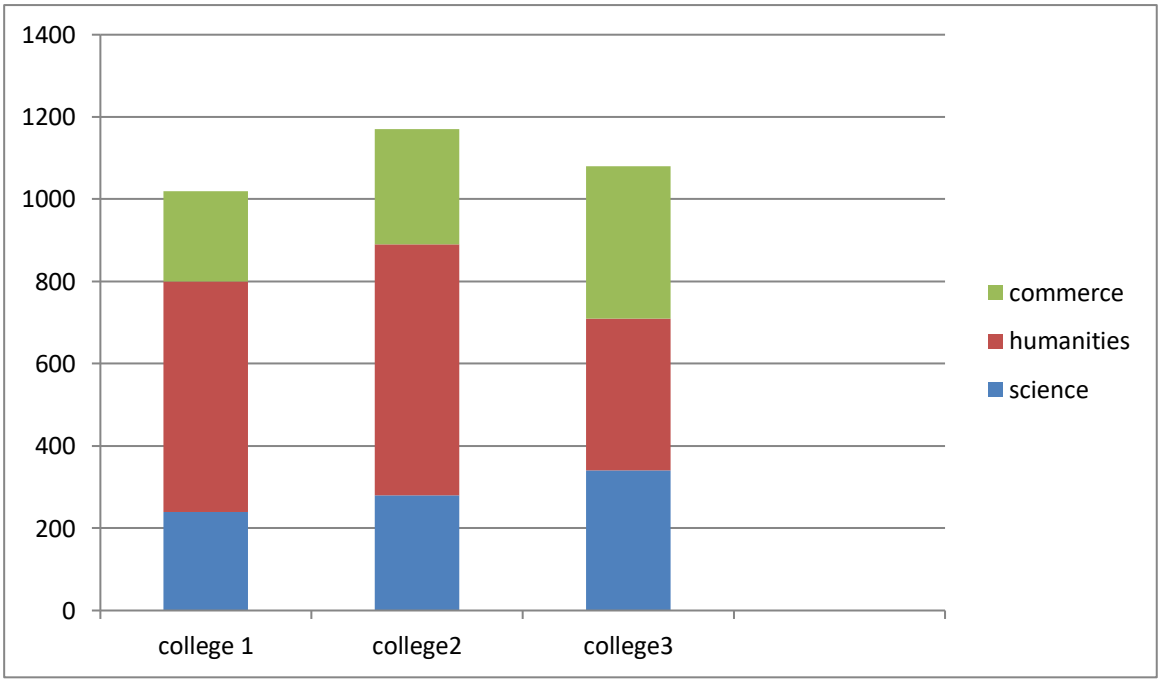
MULTIPLE BAR DIAGRAMS

Subdivided bar diagram

Presented by multible bar diagramsWe add different variables for a particular period and draw it in single bar

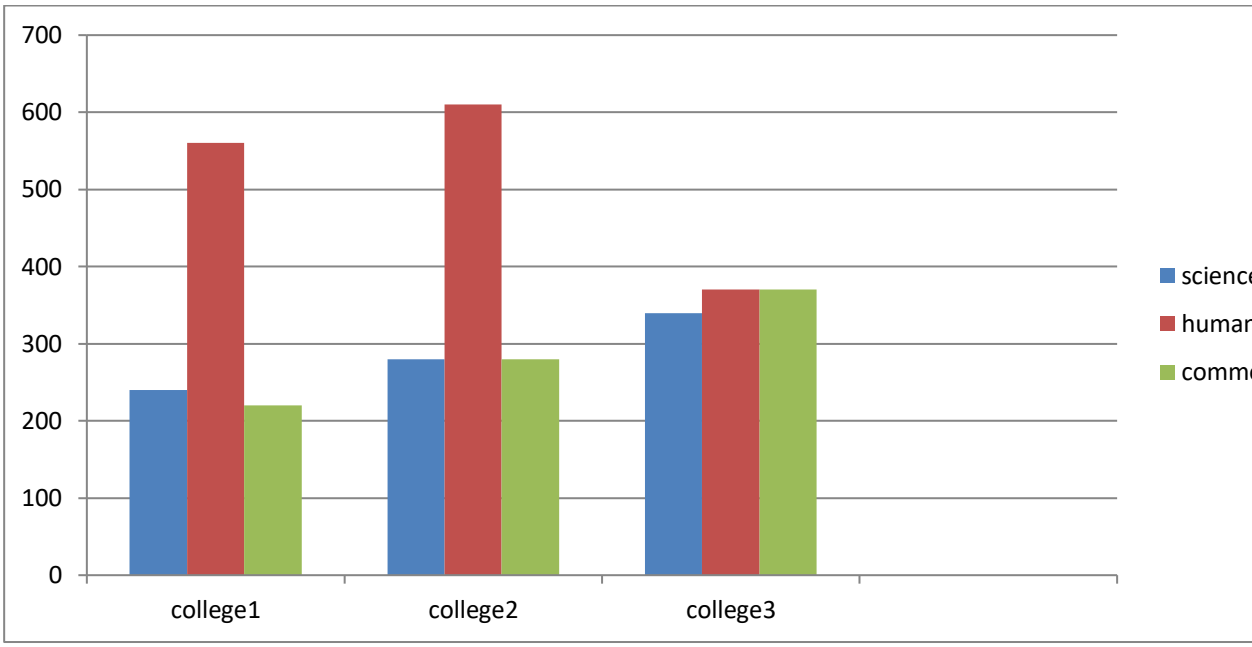
	science	humanities	commerce
college1	240	560	220
college2	280	610	280
college3	340	370	370

Scale on y axis 1cm=200students



MULTIPLE BAR DIAGRAM

Scale on y axis 1cm = 100 students

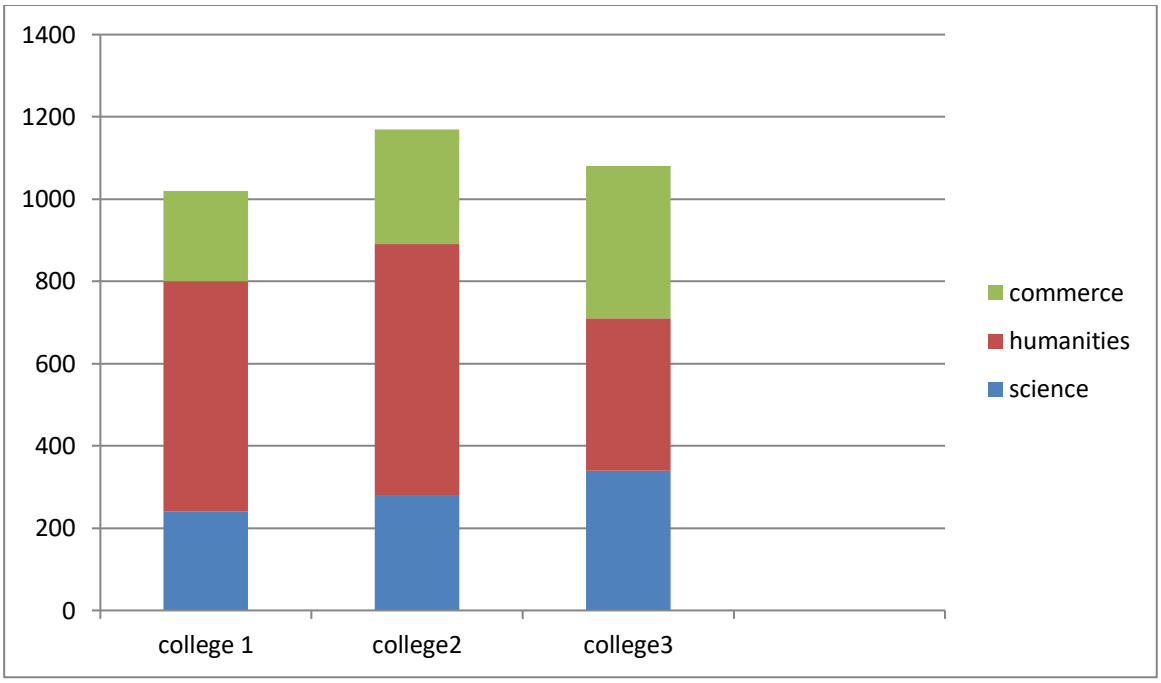


CUMULATIVE BAR DIAGRAM

SUBJECT\ college	COLLEGE 1		COLLEGE 2		COLLEGE 3	
	NO.	CUMU LATIVE	NO.	CUMUL ATIVE	No.	CUMU LATIVE
science	240	240	280	280	340	340
humani ties	560	800	610	890	370	710
commer ce	220	1020	280	1170	370	1080

Scale on x-axis 1cm = 1 college

On y-axis 1cm =200 students



EXAMPLE 11. Production of grains in Punjab is as follows. Present the data by a suitable diagram

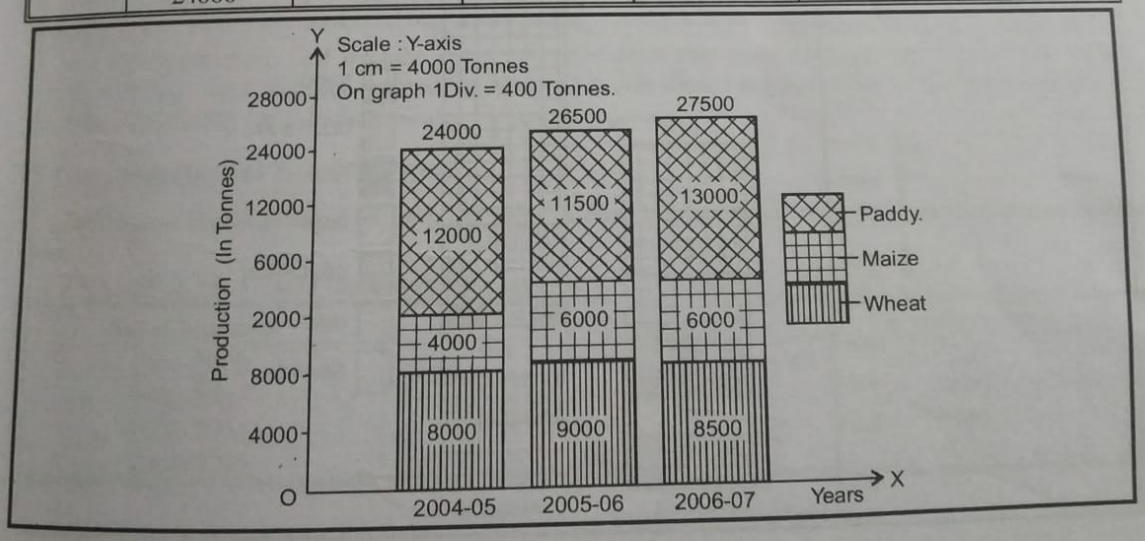
Production in Tonnes	Wheat	Maize	Paddy
2004-05	8000	4000	12000
2005-06	9000	6000	11500
2006-07	8500	6000	13000

Solution

This figure is also like the last one of example 8. Taking time period on X-axis and Production on Y–axis, we get the bar chart as below. On Y- axis ; 1 Div. = $\frac{4000}{10}$ = 400 Tonnes.

This figure can also be prepared by taking the variables, Wheat, Maize and Paddy on X-axis.

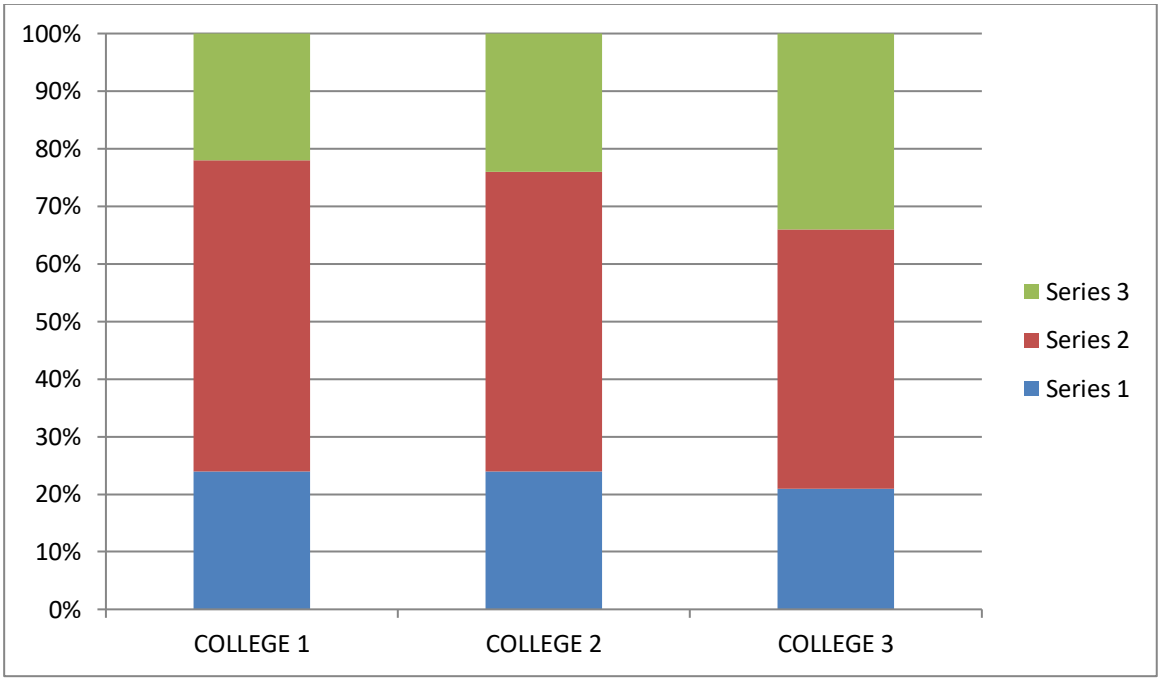
	2004-05		2005-06		2006-07	
	Production	Cumulative	Production	Cumulative	Production	Cumulative
Wheat	8000	8000	9000	9000	8500	8500
Maize	4000	12000	6000	15000	6000	14500
Paddy	12000	24000	11500	26500	13000	27500
	24000		26500		27500	



PERCENTAGE BAR DIAGRAM

SUBJECT	COLLEGE 1			COLLEGE 2			COLLEGE 3		
	NO.	CUMULATIVE	%	NO.	CUMULATIVE	%	NO.	CUMULATIVE	%
science	240	240	24	280	280	24	340	340	31
humanities	560	800	78	610	890	76	370	710	66
commerce	220	1020	100	280	1170	100	370	1080	100

FOR eg 240/1020 X100=23.5=24%



Two dimentional:- both length and width are taken proportionately

A. Rectangle

Year\No. of students	Honors	IT	IFA	Total
2017	200	250	150	600

2018	**250**	**300**	**250**	**800**
2019	**300**	**350**	**350**	**1000**

No. Of students\year	**2017**	**2018**	**2019**
Honors	**200**	**250**	**300**
IT	**250**	**300**	**350**
IFA	**150**	**250**	**350**
TOTAL	**600**	**800**	**1000**

Ratio of total students year wise

600:800:1000

3:4:5

Hence width will be in the ratio of 3:4:5 i,e 3CM, 4CM, 5CM

Scale on y axis 1cm = 100 students,

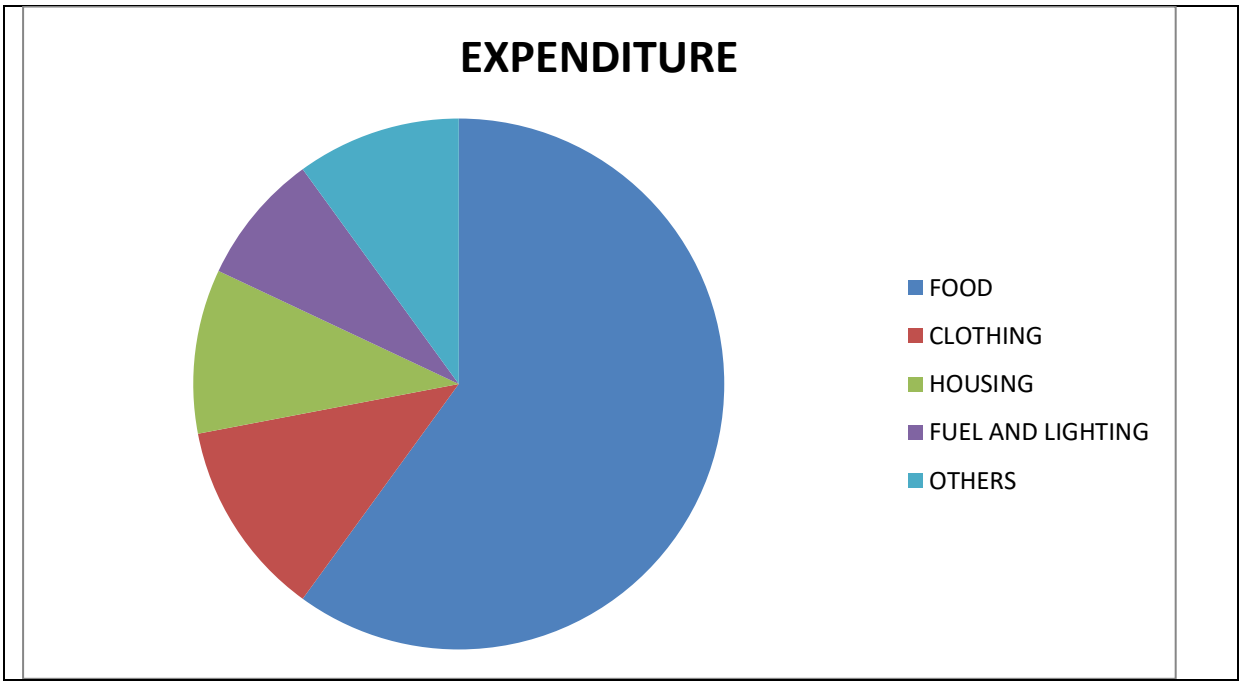
On x- axis year 2017 width 3cms

Year 2018 width 4 cms

Year 2019 width 5 cms

PIE CHART OR SUBDIVIDED CIRCULAR DIAGRAM OR ANGULAR DIAGRAM

ITEMS	EXPENDITURE	degree
FOOD	60	60/100 x360=216
CLOTHING	12	12/100x360=43
HOUSING	10	10/100x36036
FUEL AND LIGHTING	8	8/100x360=29
OTHERS	10	10/100x360=36
TOTAL	100	360



Source	Revenue in millions of Rs.	Degree
Customs	80	53
Excise	190	127
Income tax	160	107
Corporate tax	75	50
miscellaneous	35	23

total	540	360

TOPIC: TYPES/KINDS OF FREQUENCY DISTRIBUTION GRAPHS

I.Histogram

II. Frequency Polygon

III. Smoothed Frequency Curve

IV. Ogive Curves or Cumulative Frequency Curves

IMPORTANT NOTE

FOR CONSTRUCTION OF HISTOGRAM, FREQUENCY POLYGON AND FREQUENCY CURVE, THE CLASSES MUST BE EQUAL AND GIVEN IN EXCLUSIVE SERIES AND FREQUENCY MUST BE SIMPLE FREQUENCY. IF CLASSES ARE NOT EQUAL AND NOTGIVEN IN EXCLUSIVE SERIES AND FREQUENCY AS SIMPLE FREQUENCY THEN CHANGE THE QUESTION AND THEN DRAW GRAPHS

I. HISTOGRAM

Histogram consists of rectangular bars attached to each other.

Histogram can be constructed under 2 situations

1. **When Classes are equal**
2. **When Classes are Unequal**

Important Points for drawing Histrogram:

1. **Take Variable on X-Axis and Simple Frequency on Y-Axis**
2. **Take the variable as it is on X-Axis and for Simple frequency take desired scale (measurements) on Y-Axis**

TOPIC: CONSTRUCTION OF HISTOGRAM WHEN CLASSES ARE EQUAL

Problem 1: Construct Histogram from the given data

Marks	0-20	20-40	40-60	60-80	80-100
No .of Students	20	50	90	38	15

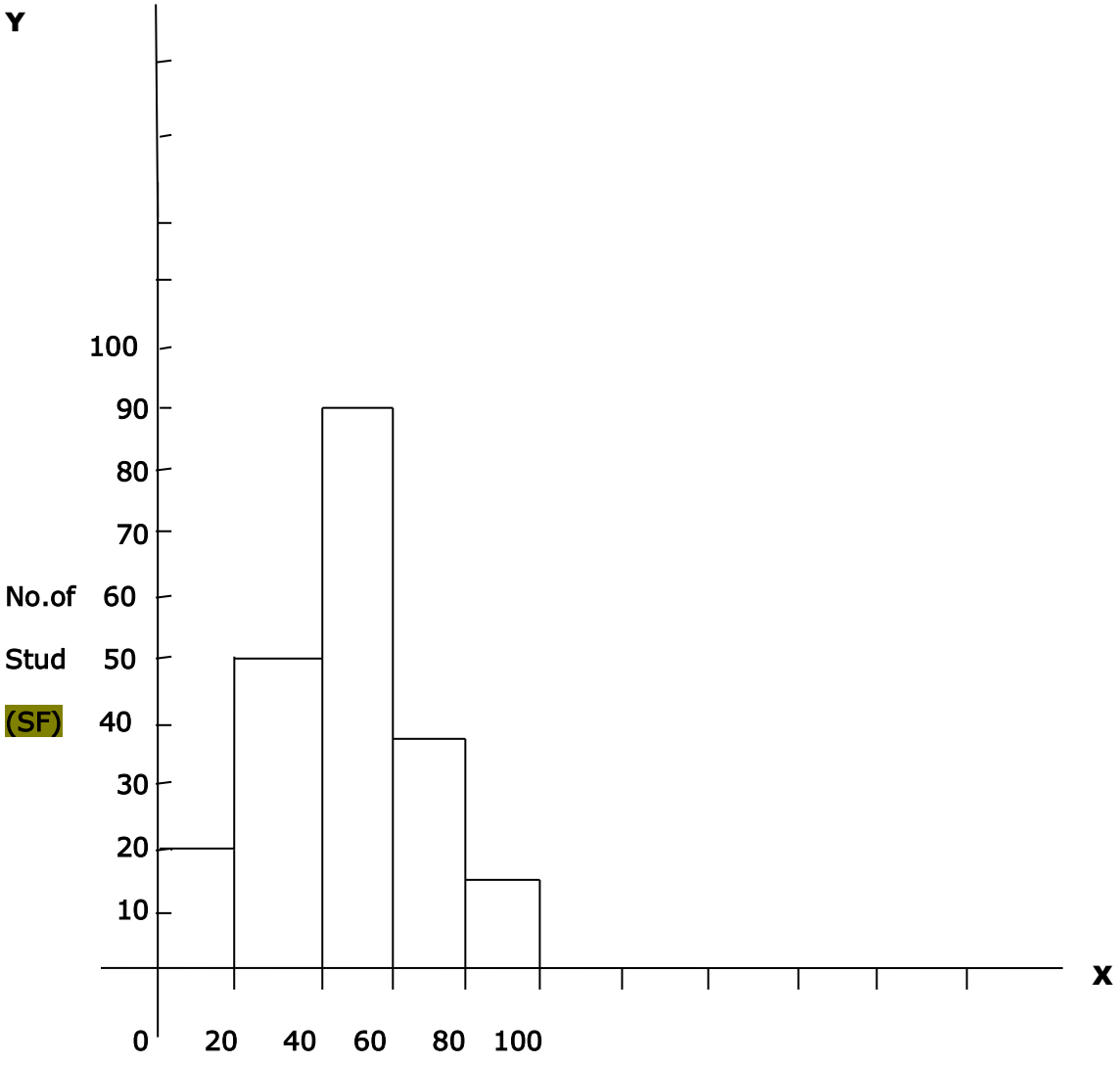
Note: Take Marks on X-Axis and Simple Frequency on Y-Axis

Histogram

Scale (Measurements)

X-Axis: 1cm = 20 marks

Y-Axis: 1cm = 10 Students

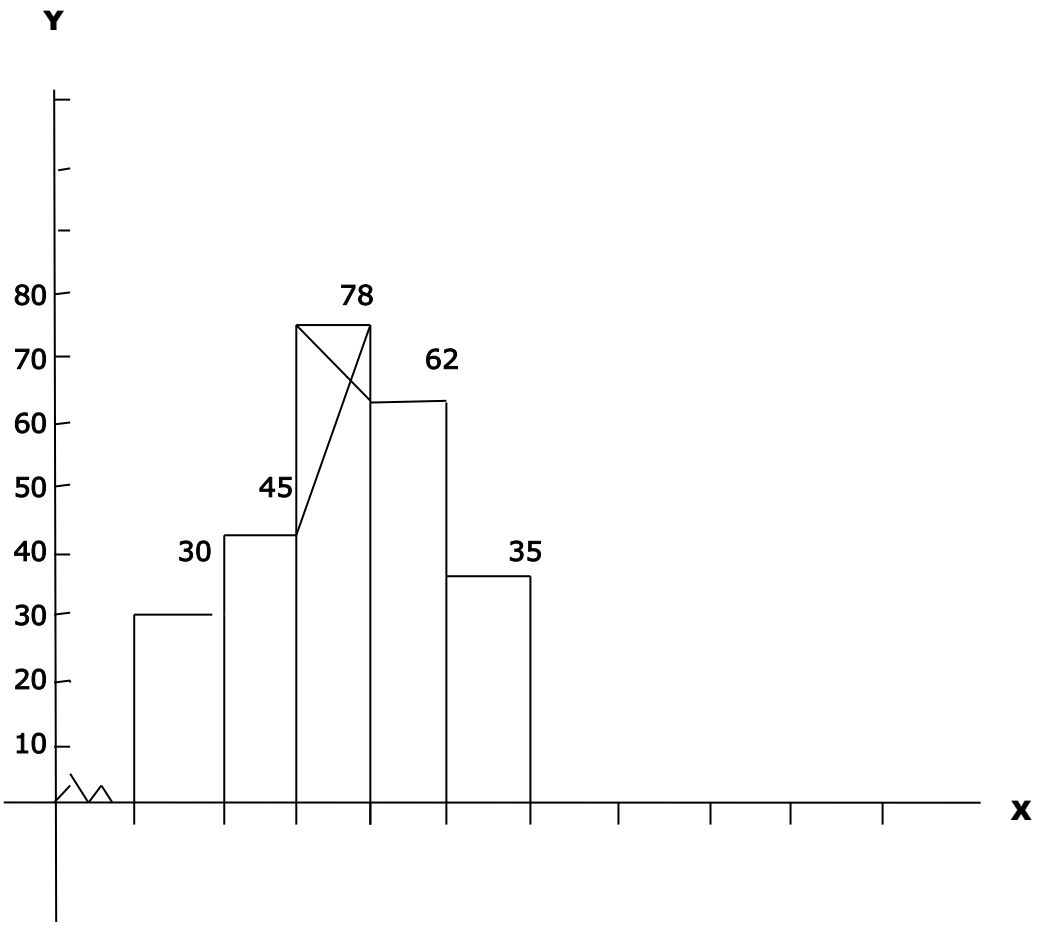


Marks

WAGES(RS)	325-350	350-375	375-400	400-425	425-450	
NO. OF WORKERS	30	45	75	60	35	

HISTOGRAM

Scale (Measurements)
X-Axis: 1cm = Rs.25
Y-Axis: 1cm = 10 workers



0 325 350 375 400 425 450

Wages

TOPIC: PROBLEMS ON CONSTRUCTION OF HISTOGRAM WHEN THE SIZE OF CLASSES IS UNEQUAL

Important Point:

when the size ofthe classes is unequal then compulsorily the classes should be made equal either by adding or dividing and then only histogram or frequency polygon or frequency curve should be drawn.

Problem 3:

Marks	0-10	10-20	20-30	30-60	60-80	80-85	85-90	90-100
No.of students	5	10	4	18	4	3	5	9

Solution:

Working notes:

1.lets us take the desired size of each class as 10

2. The class 30-60 can be divided into 3 classes of size 10 as 30-40,40-50,50-60 and the frequency will be 18/3 = 6 each.

3.The class 60-80 can be divided into 2 classes of size 10 as 60-70,70-80 and the frequency will be 4/2 = 2 each.

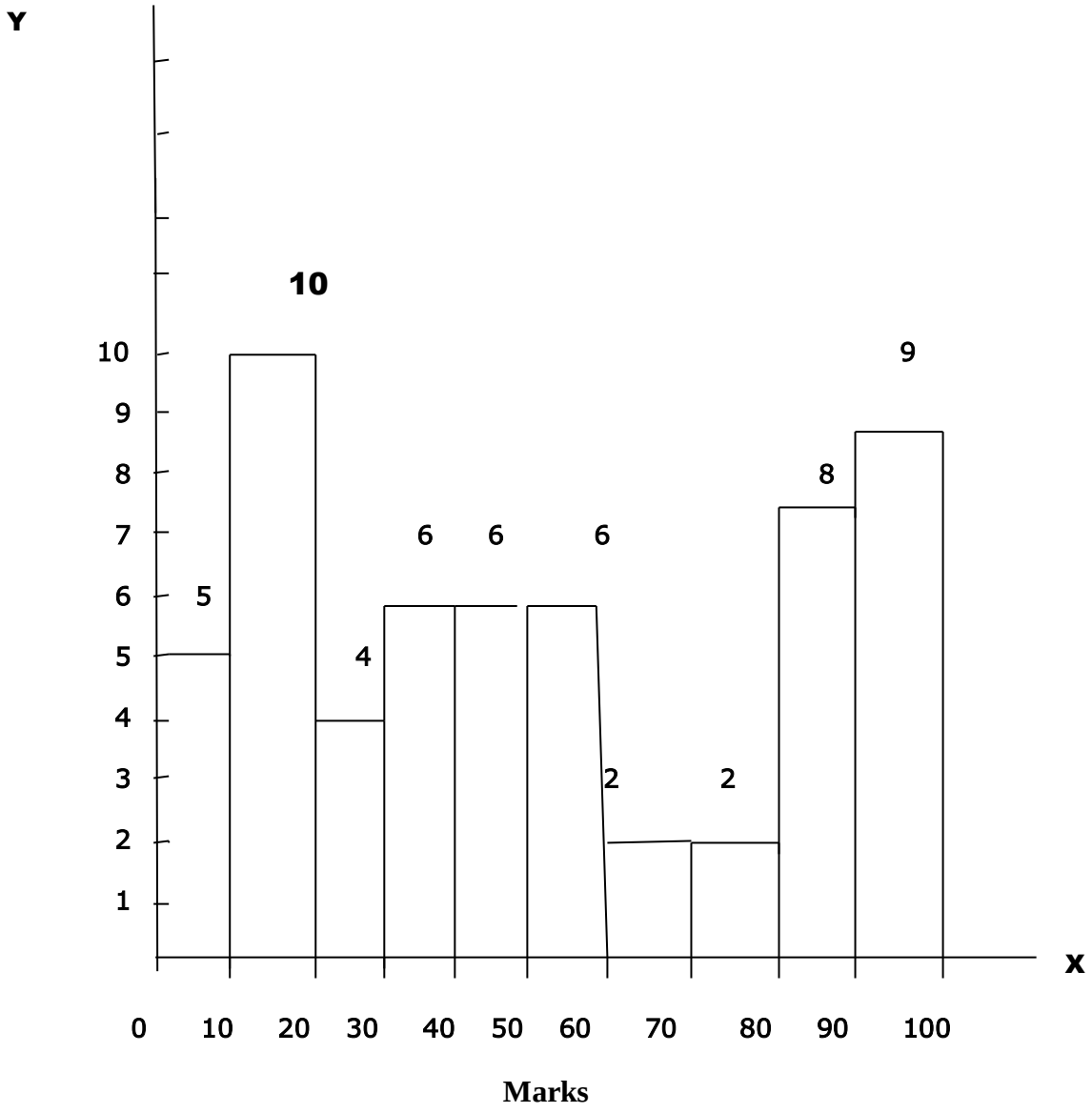
4.The classes 80-85 and 85-90 can be clubbed or added as 80-90 and the frequency will be 3+5=8.

Therefore the new question is

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No.of students	5	10	4	6	6	6	2	2	8	9

HISTOGRAM

Scale (Measurements)
X-Axis: 1cm = 10 marks
Y-Axis: 1cm = 1 Student



Problem: 4

Draw an Histogram from the following data

Marks	10-15	15-20	20-25	25-30	30-40	40-60
No.of students	7	19	27	15	12	12

Solution:

Working Notes:

1.let us take desired size of each class as 5

2.The class 30-40 can be divided into 2 classes of size 5 as 30-35, 35-40 and the frequency will be 12/2= 6 each.

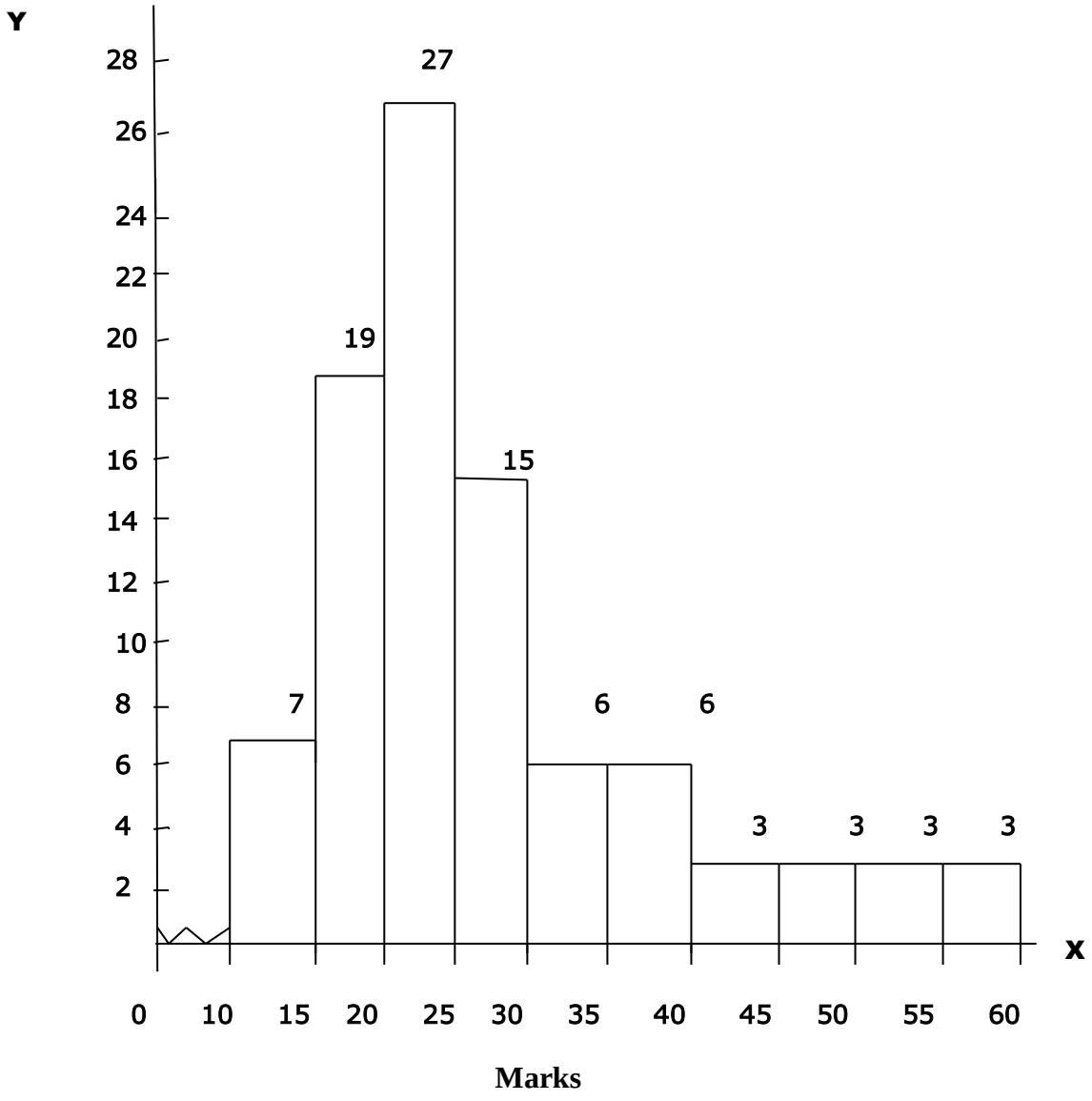
3.The class 40-60 can be divided into 4 classes of size 5 as 40-45, 45-50,50-55 and 55-60 and the frequency will be 12/4 = 3 each.

Therefore new question is :

Marks	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
No.of students	7	19	27	15	6	6	3	3	3	3

HISTOGRAM

Scale (Measurements)
X-Axis: 1cm = 5 marks
Y-Axis: 1cm = 2 Students



TOPIC: CONSTRUCTION OF FREQUENCY POLYGON & FREQUENCY CURVE

Frequency Polygon: It can be constructed in two ways

a) using histogram b) without using histogram

Steps for construction of frequency polygon using histogram

1.Take the classes on the X-axis and simple frequency on Y-axis

2. Draw a histogram

3.Mark the middle points of the rectangle bars and join all the points with scale. extend the two lines to touch the X-axis at the middle of the beside boxes on both the sides.

Frequency Curve: it is also known as smoothed frequency curve .It is drawn with free hand.It is used to remove the Ruggedness (pointed ends) of the frequency polygon.

Note:

1. **Important Point given for histogram is applicable for frequency polygon & Frequency curve also.**
2. **if the problem is silent then draw frequency polygon and frequency curve using histogram.**

Problem 5:

Draw a frequency polygon and frequency curve using histogram

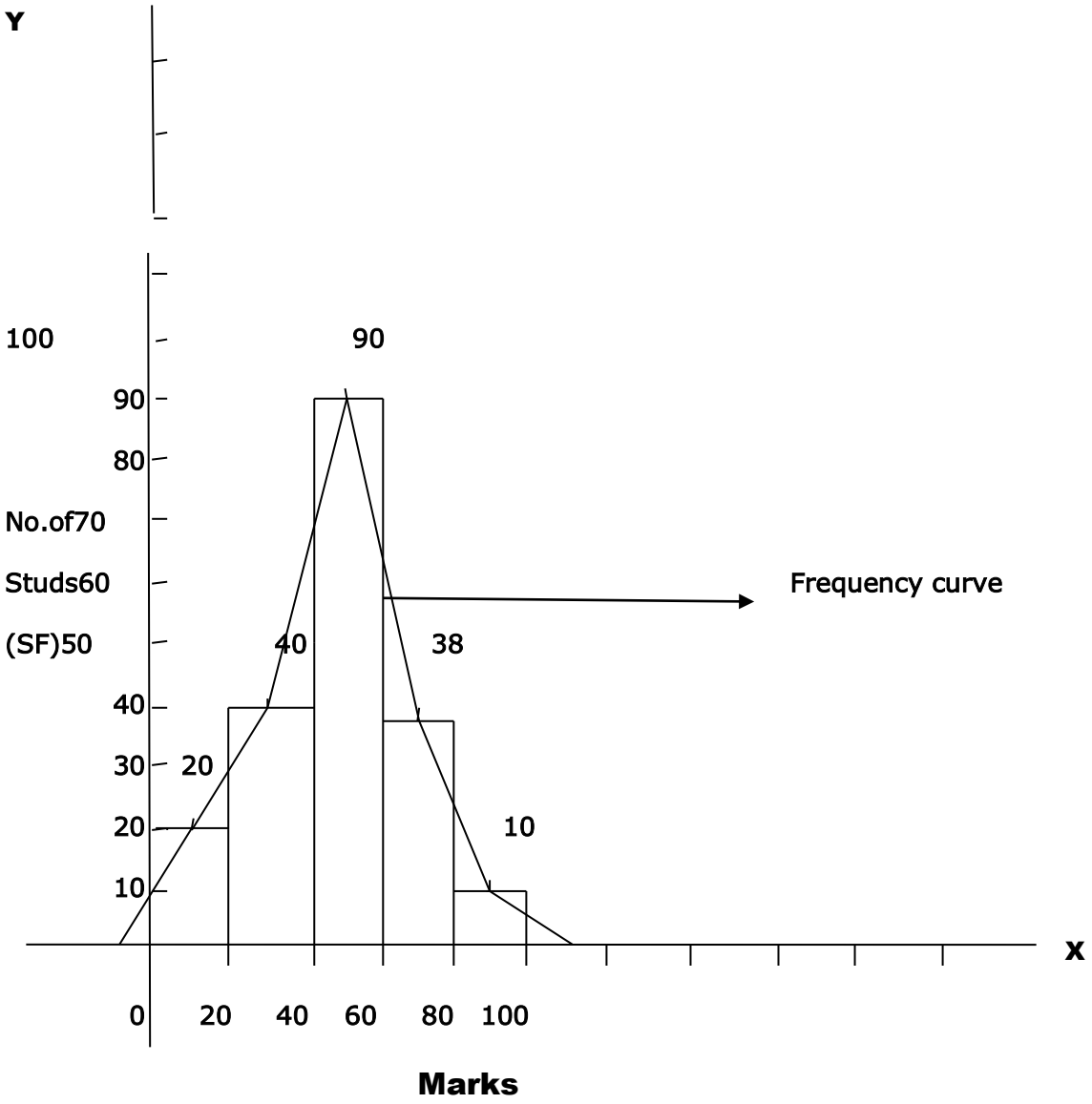
Marks	0-20	20-40	40-60	60-80	80-100
No.of students	20	40	90	38	15

Frequency Polygon & Frequency curve using Histogram

Scale (Measurements)

X-Axis: 1cm = 20 marks

Y-Axis: 1cm = 10 Students



Problem: 6

Draw Histogram, Frequency Polygon and Frequency Curve

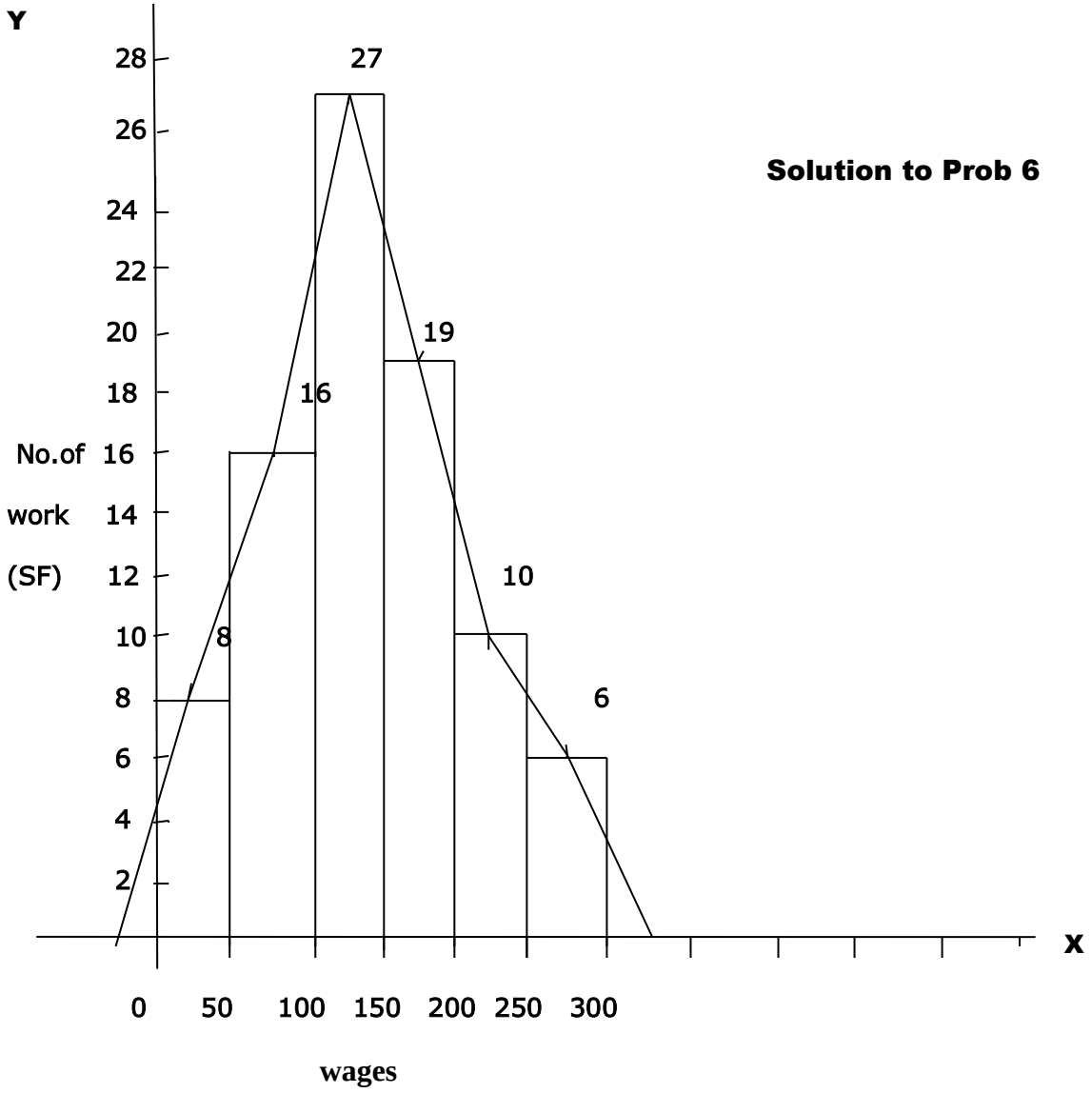
Wages	0-50	50-100	100-150	150-200	200-250	250-300
No.of workers	8	16	27	19	10	6

Frequency Polygon & Frequency Curve using Histogram

Scale (Measurements)

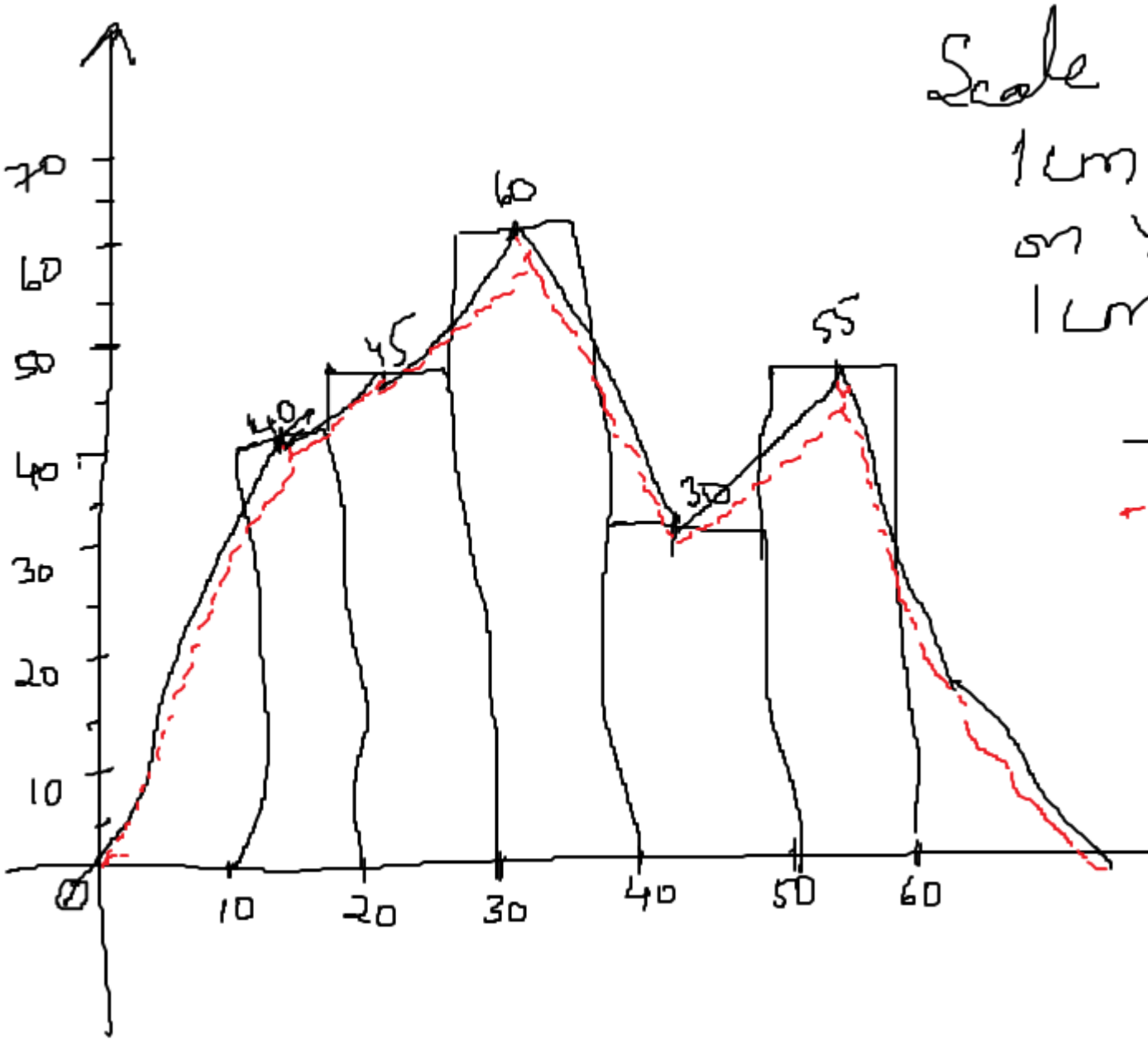
X-Axis: 1cm = Rs.50

Y-Axis: 1cm = 2 workers



Solution to Prob 6

MARKS	10-20	20-30	30-40	40-50	50-60
NO. OF STUDENTS	40	45	60	30	55



Problem No: 7

Draw a Frequency Polygon and Frequency Curve using Histogram

Marks	10-19	20-29	30-39	40-49	50-59
No.of students	20	40	80	55	35

Solution:

Note: In the problem classes are given under Inclusive series therefore convert into Exclusive series and then draw Histogram, Frequency Polygon & Frequency Curve

Steps to convert Inclusive Series into Exclusive Series

1. **Correction Factor = $\frac{\text{lower limit of 2}^{nd}\text{ class - upper limit of 1}^{st}\text{ class}}{2}$**

$$= \frac{20 - 19}{2} = \frac{1}{2} = 0.5$$

2. **The above Correction Factor of 0.5 should be deducted from all lower limits and added to all upper limits.**

Lower limits	Upper limits
10 - 0.5= 9.5	19+0.5=19.5
20 - 0.5=19.5	29+0.5=29.5
30 - 0.5=29.5	39+0.5=39.5
40 - 0.5=39.5	49+0.5=49.5
50 - 0.5=49.5	59+0.5=59.5

Therefore the new question under Exclusive Series is:

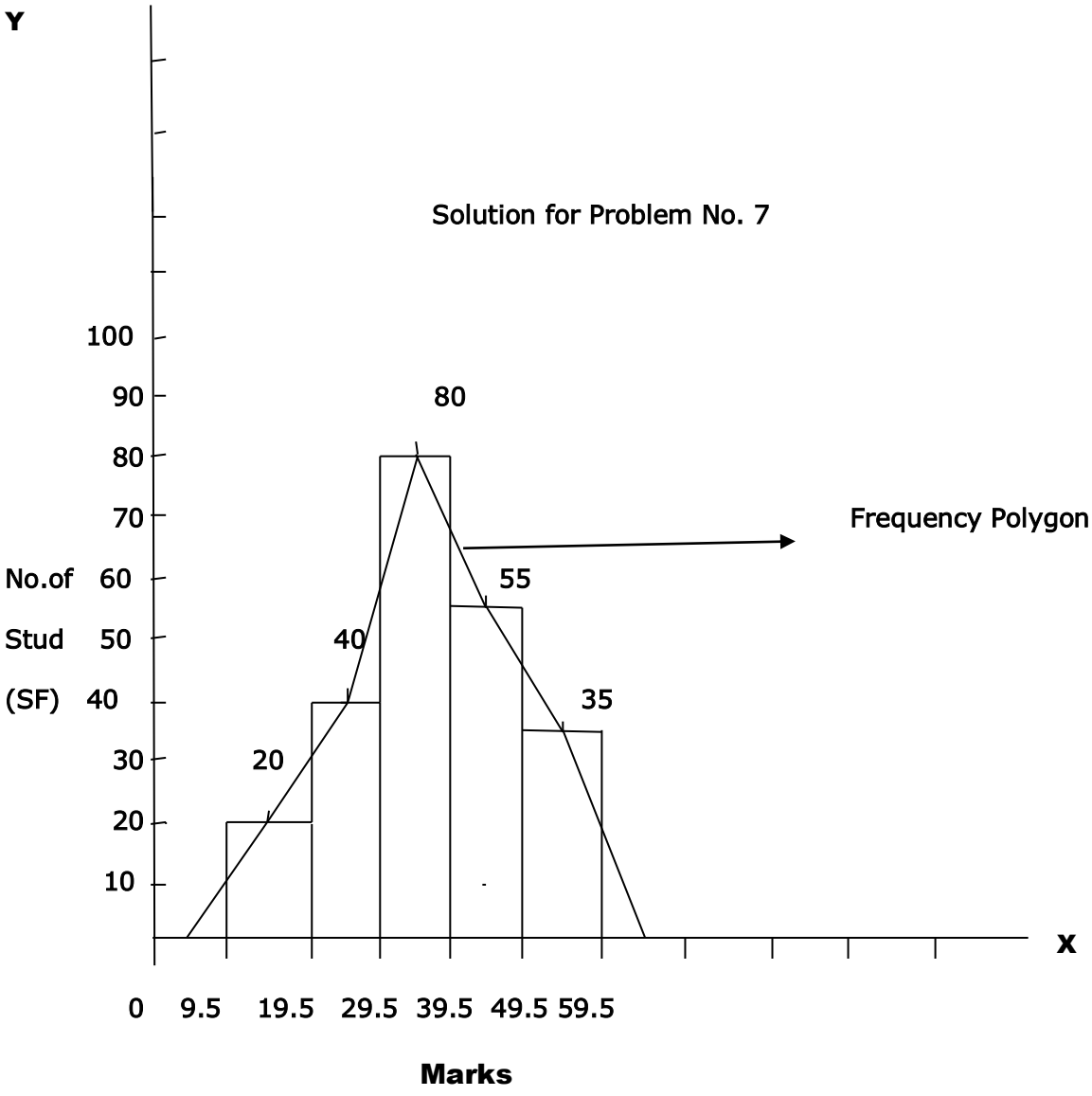
Marks	9.5-19.5	19.5-29.5	29.5-39.5	39.5-49.5	49.5-59.5
No.of students	20	40	80	55	35

Frequency Polygon & Frequency Curve using Histogram

Scale (Measurements)

X-Axis: 1cm = 10 marks

Y-Axis: 1cm = 10 Students



TOPIC: CONSTRUCTION OF FREQUENCY POLYGON AND FREQUENCY CURVE WITHOUT USINGHISTOGRAM

Steps

1.find out mid points/midvaluesof all the classes by using the formula

Midpoints =$\frac{\text{lower limit + upper limit}}{2}$

2. Take the midpoints on the X-axis as it is and the simple frequency on Y-axis using

measurements

3.mark the midpoints with its respective simple frequencies and join all the points with scale fordrawing Frequency polygon and use free hand for frequency curve.

Problem No: 8

Draw frequency polygon and frequency curve without using histogram

Marks	0-20	20-40	40-60	60-80	80-100
No.of students	30	40	90	60	20

Solution:

Calculation of Midpoints/Midvalues

Midpoint of 1st class = $\frac{0 + 20}{2} = \frac{20}{2} = 10$

Midpoint of 2nd class = $\frac{20 + 40}{2} = \frac{60}{2} = 30$

Midpoint of 3rd class = $\frac{40 + 60}{2} = \frac{100}{2} = 50$

Midpoint of 4th class = $\frac{60 + 80}{2} = \frac{140}{2} = 70$

Midpoint of 5th class = $\frac{80 + 100}{2} = \frac{180}{2} = 90$

Therefore the new question is

Marks	Midpoints	No.of students (Simple Frequency –SF)
0-20	10	30
20-40	30	40
40-60	50	90
60-80	70	60
80-100	90	20

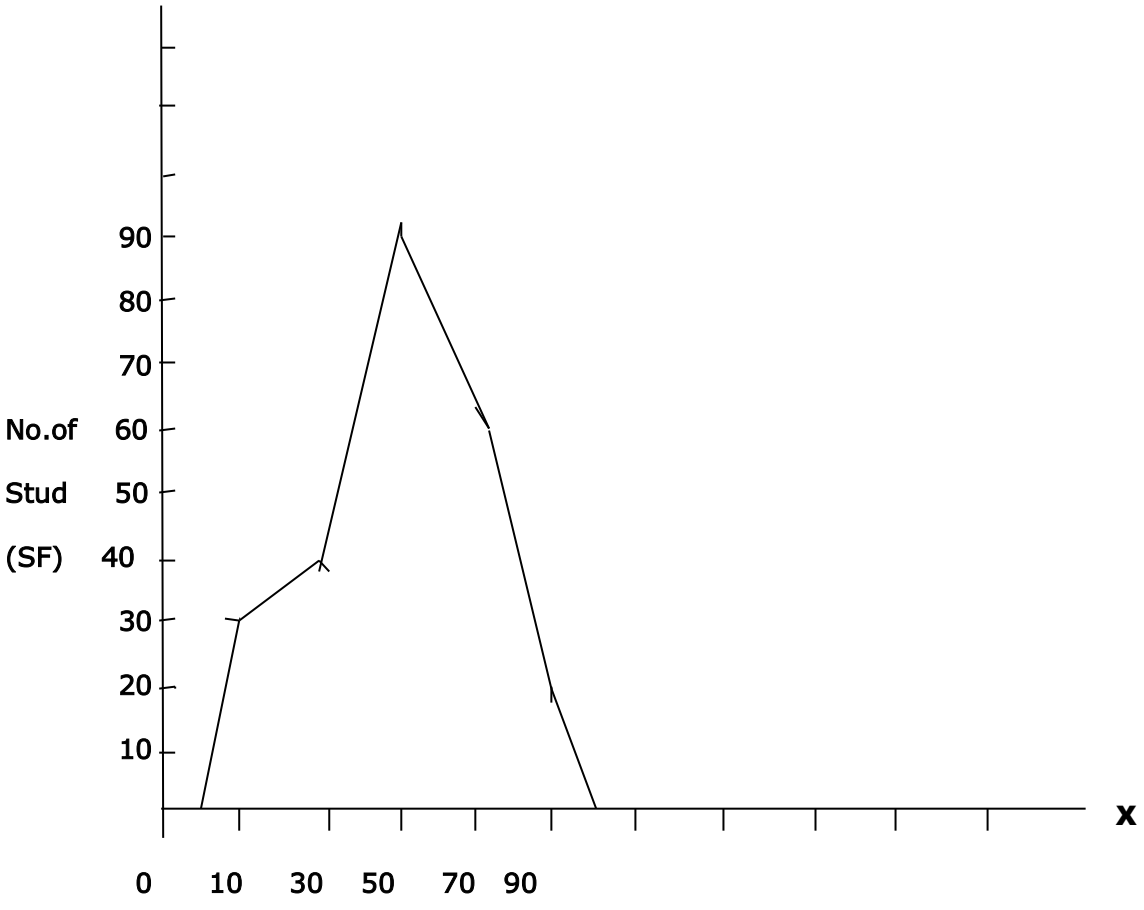
Frequency Polygon & Frequency Curve without using Histogram

Scale (Measurements)

X-Axis: 1cm = 20 marks

Y-Axis: 1cm = 10 Students

Y



Problem No: 9

Draw frequency polygon and frequency curve without using histogram

Marks	0-10	10-20	20-30	30-40	40-50
No.of students	40	30	10	12	15

Solution:

Calculation of Midpoints/Midvalues

Midpoint of 1st class = $\frac{0 + 10}{2} = \frac{10}{2} = 5$

Midpoint of 2nd class = $\frac{10 + 20}{2} = \frac{30}{2} = 15$

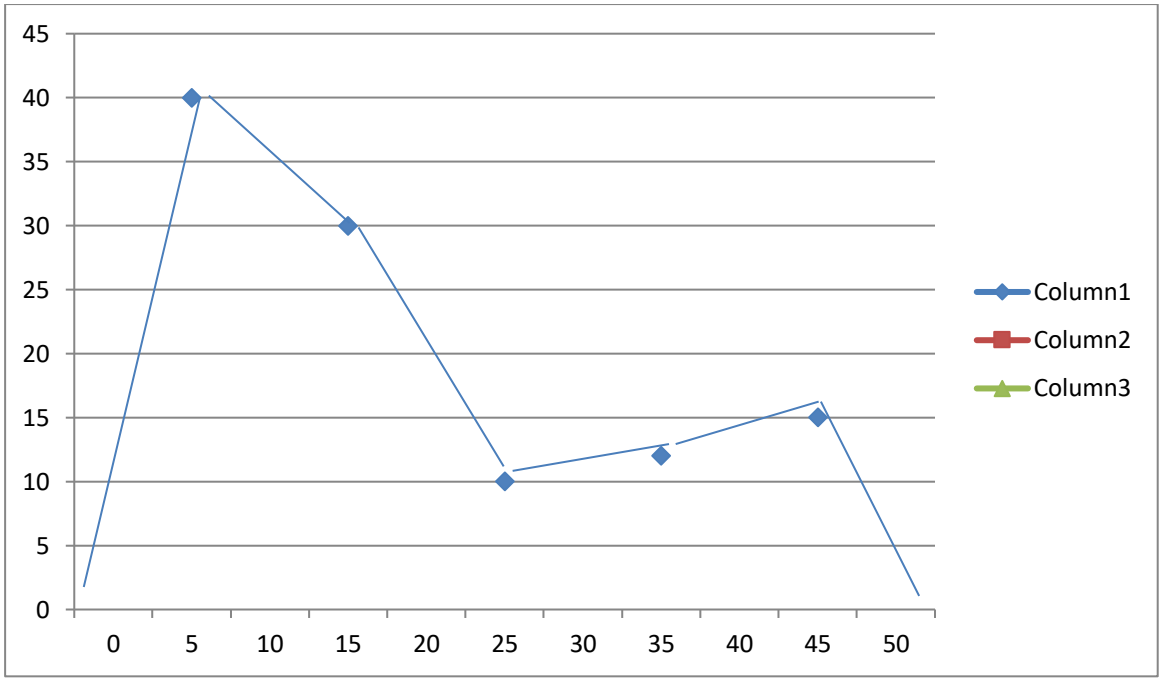
Midpoint of 3rd class = $\frac{20 + 30}{2} = \frac{50}{2} = 25$

Midpoint of 4th class = $\frac{30 + 40}{2} = \frac{70}{2} = 35$

Midpoint of 5th class = $\frac{40 + 50}{2} = \frac{90}{2} = 45$

Therefore the new question is

Marks	Midpoints	No.of students (Simple Frequency –SF)
0-10	5	40
10-20	15	30
20-30	25	10
30-40	35	12
40-50	45	15



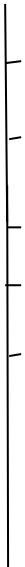
Frequency Polygon & Frequency Curve without using Histogram

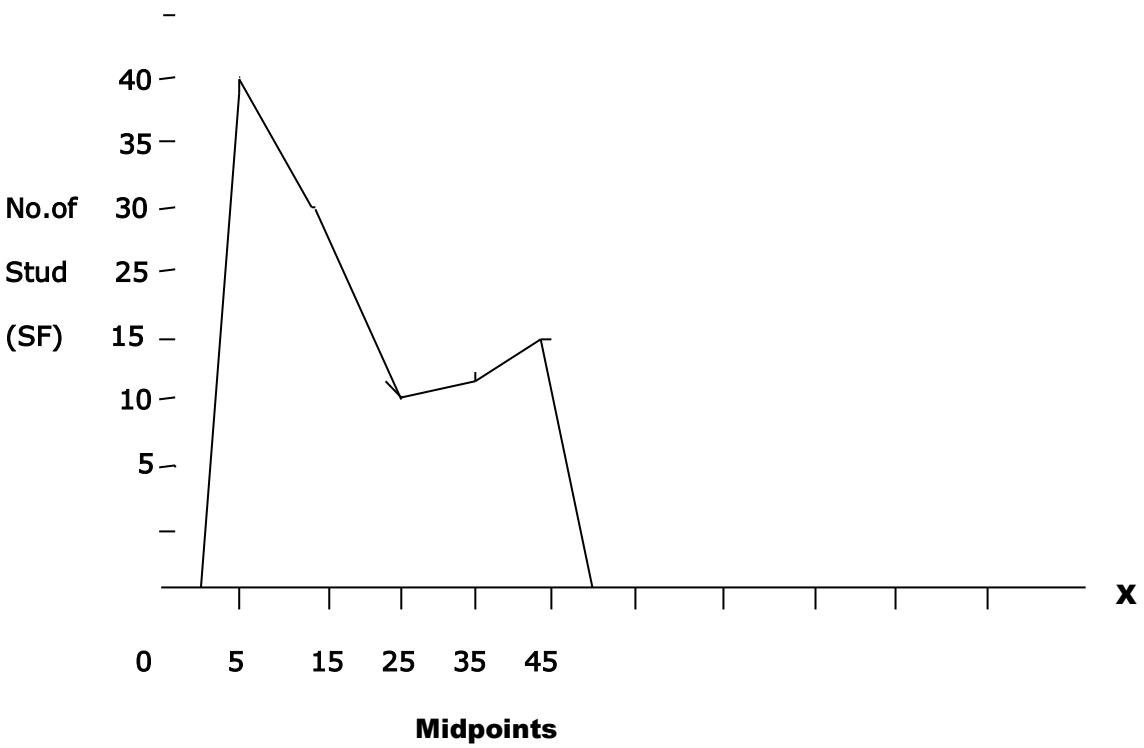
Scale (Measurements)

X-Axis: 1cm = 10 marks

Y-Axis: 1cm = 5 Students

Y





TOPIC: CALCULATION OF MODE GRAPHICALLY:

mode is that value of the variable with is repeated highest number of times. By the method of inspection ,mode is that value of the variable which has highest frequency.

Important note

For calculation of mode graphically, the classes must be equal and given in exclusive series and frequency must be simple frequency. If classes are not equal and not given in exclusive series and frequency as simple frequency then change the question and then find out mode

Steps to calculate Mode Graphically:

1. Draw a Histogram.

2.Decide the Modal class (the class with the highest simple frequency)

3. Draw two lines diagonally inside the modal class rectangle bar to the upper corner of adjacent rectangles.

4. Draw a straight line from the point of intersection of two diagonals touching the X-axis andfind out the value of the mode (MODE IS THAT VALUE OF THE VARIABLE WHERE THE STRAIGHT LINE TOUCHES THE X AXIS)

Problem No: 10

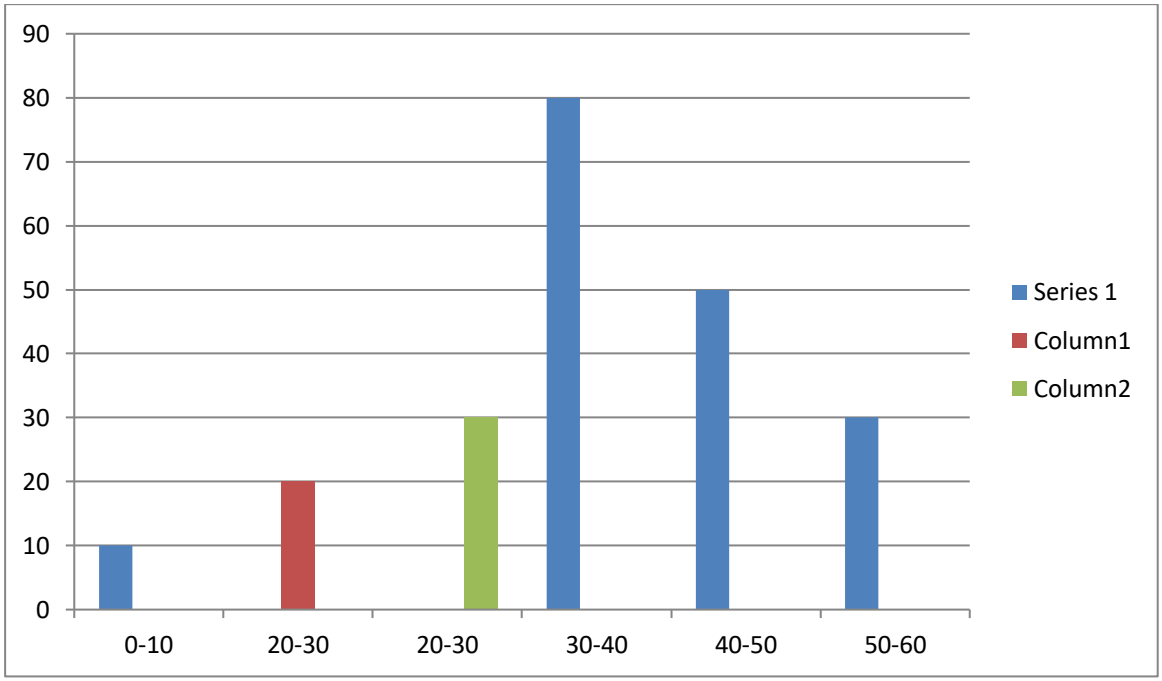
Calculate mode graphically

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No of students	10	20	30	80	50	30

Solution:

Working notes:

1. **By the method of inspection, mode lies in 30-40 class because it has got highest simple frequency of 80 students.**
2. **The value of mode is 36 marks because the line of intersection is touching the X axis at 36 marks.**

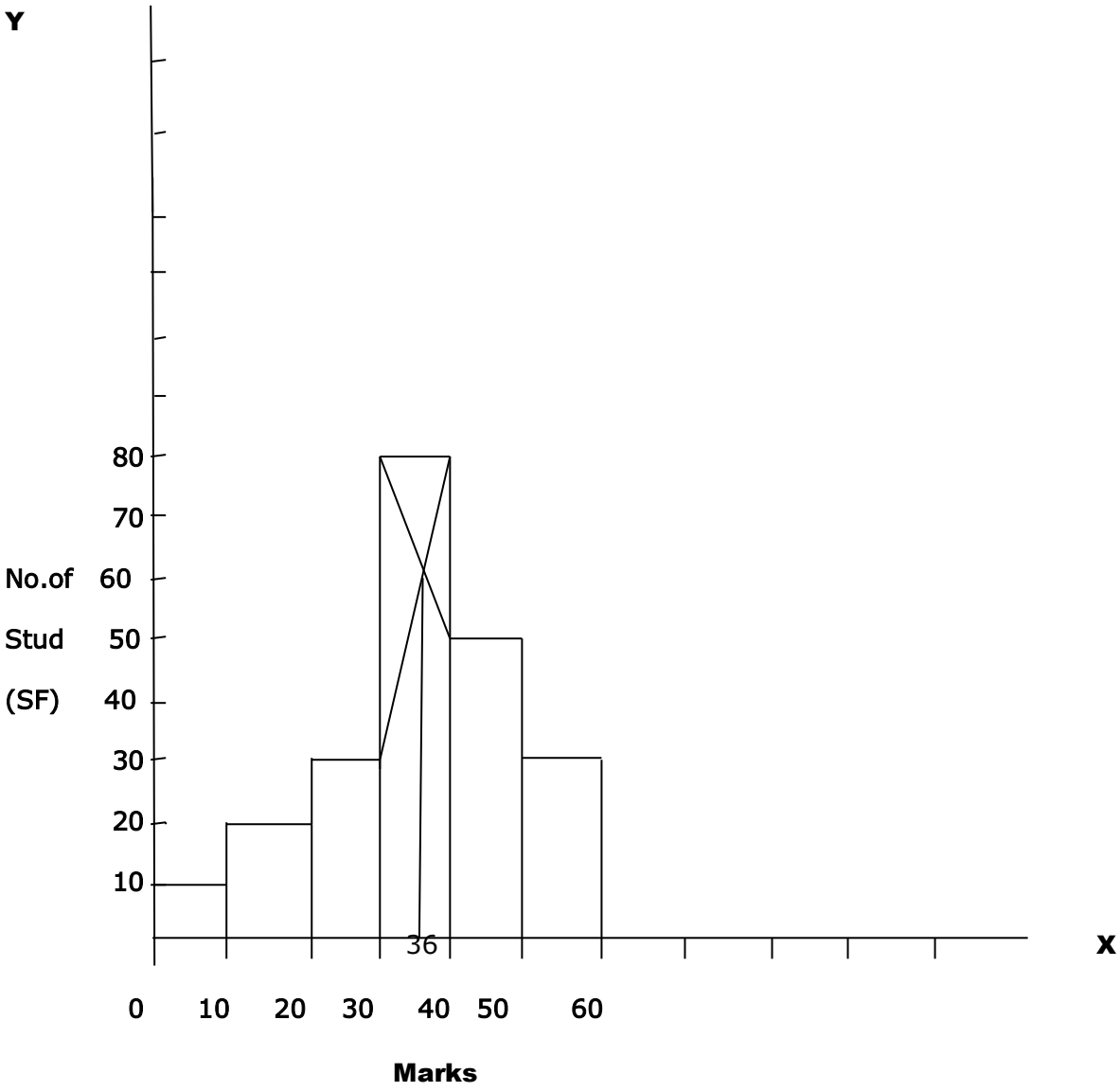


Calculation of Mode Graphically

Scale (Measurements)

X-Axis: 1cm = 10 marks

Y-Axis: 1cm = 10 Students



Problem No: 11

Calculate mode graphically

Marks	0-20	20-40	40-60	60-80	80-100
No of students	20	50	90	45	15

Solution:

Working notes:

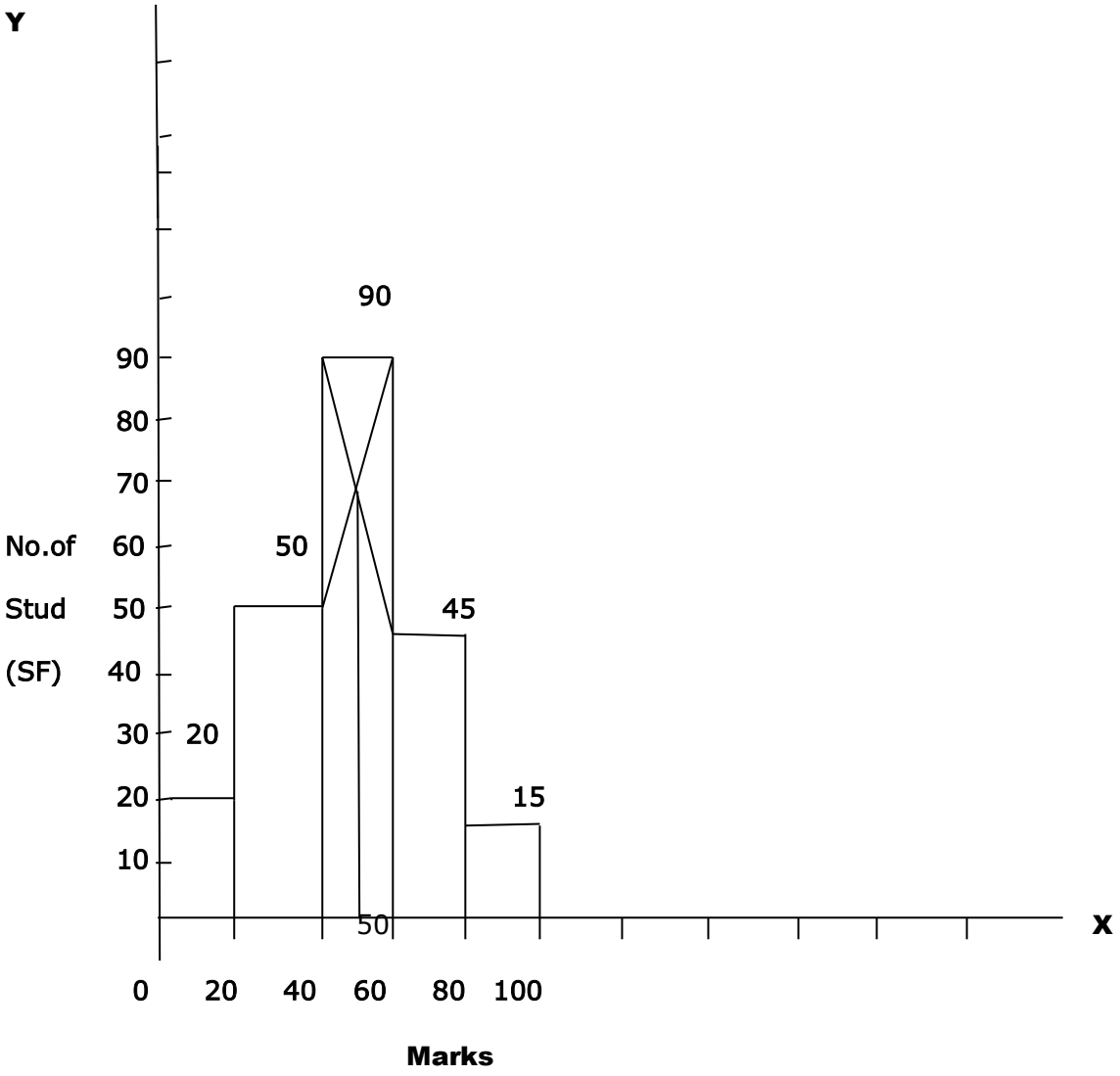
1. **By the method of inspection, mode lies in 40-60 class because it has got highest simple frequency of 90 students.**
2. **The value of mode is 50 marks because the line of intersection is touching the X axis at 50 marks.**

Calculation of Mode Graphically

Scale (Measurements)

X-Axis: 1cm = 20 marks

Y-Axis: 1cm = 10 Students



TOPIC: CUMULATIVE FREQUENCY CURVES OR OGIVE CURVES:

Sometimes it is necessary to know the number of items whose values are less than or more than a certain amount. The graphic presentation of cumulative frequency distribution is known as Ogive curves or cumulative frequency curves. These are 2 types of Ogivecurves (1). less than ogive curve (2). more than ogive curve

Steps to draw Less than Ogive Curves and More than Ogive Curves

1. **convert the given simple frequency into Less than Cumulative (LCF) and More than cumulative (MCF)**
2. **Take the classes on X-axis and LCF & MCF values on Y-axis**
3. **for drawing less than ogive curve we need upper limitsso mark these upper limits with its respective LCF values and join all points with scale**
4. **for drawing more than ogive curve we need lower limitsso mark these lower limits with its respective MCF valuesand join all points with scale**

Note:LESS THAN OGIVE CURVE WILL BE RISING UPWARDS AND MORE THAN OGIVE CURVE WILL BE FALLING DOWNWARDS.

TOPIC: CALCULATION OF MEDIAN GRAPHICALLY:

Median is the middle value of the given variables. Median divides adistribution into 2 equal parts

steps for calculation of median graphically

1. **Draw less than ogivecurve and more than ogive curve (refer above steps)**
2. **Draw a straight line from the point of intersection of both the ogive curves touching the X-axis. THE VALUE OF THE VARIABLE AT WHICH THE STRAIGHT LINE TOUCHES THE X-AXIS WILL BE THE ANSWER OF MEDIAN.**

Problem No: 12

Draw less than ogive curve and more than ogive curve and also locate median graphically

Marks	0-50	50-100	100-150	150-200	200-250	250-300	300-350
No.of students	7	16	31	48	26	14	8

Solution

converting simple frequency into LCF for drawing Less than Ogive Curve

Marks (upper limits)	No.of students (Simple Frequency – SF)	Less than Cumulative Frequency (LCF)
50	7	7
100	16	7+16=23
150	31	23+31=54
200	48	54+48=102
250	26	102+26=128
300	14	128+14=142
350	8	142+8=150

converting simple frequency into MCF for drawing More than Ogive Curve

Marks (Lower limits)	No.of students (simple frequency - SF)	More than Cumulative Frequency (MCF)
0	7	150
50	16	150-7=143
100	31	143-16=127
150	48	127-31=96
200	26	96-48=48
250	14	48-26=22
300	8	22-14=8

Note :

1. **Take the classes on X-axis and LCF and MCF values on the Y-axis**
2. **For drawing less than ogive curve, mark the upper limits with its respective LCF values and join all the points with scale.**
3. **For drawing more than ogive curve, mark the lower limits with its respective MCF values and join all the points with scale.**

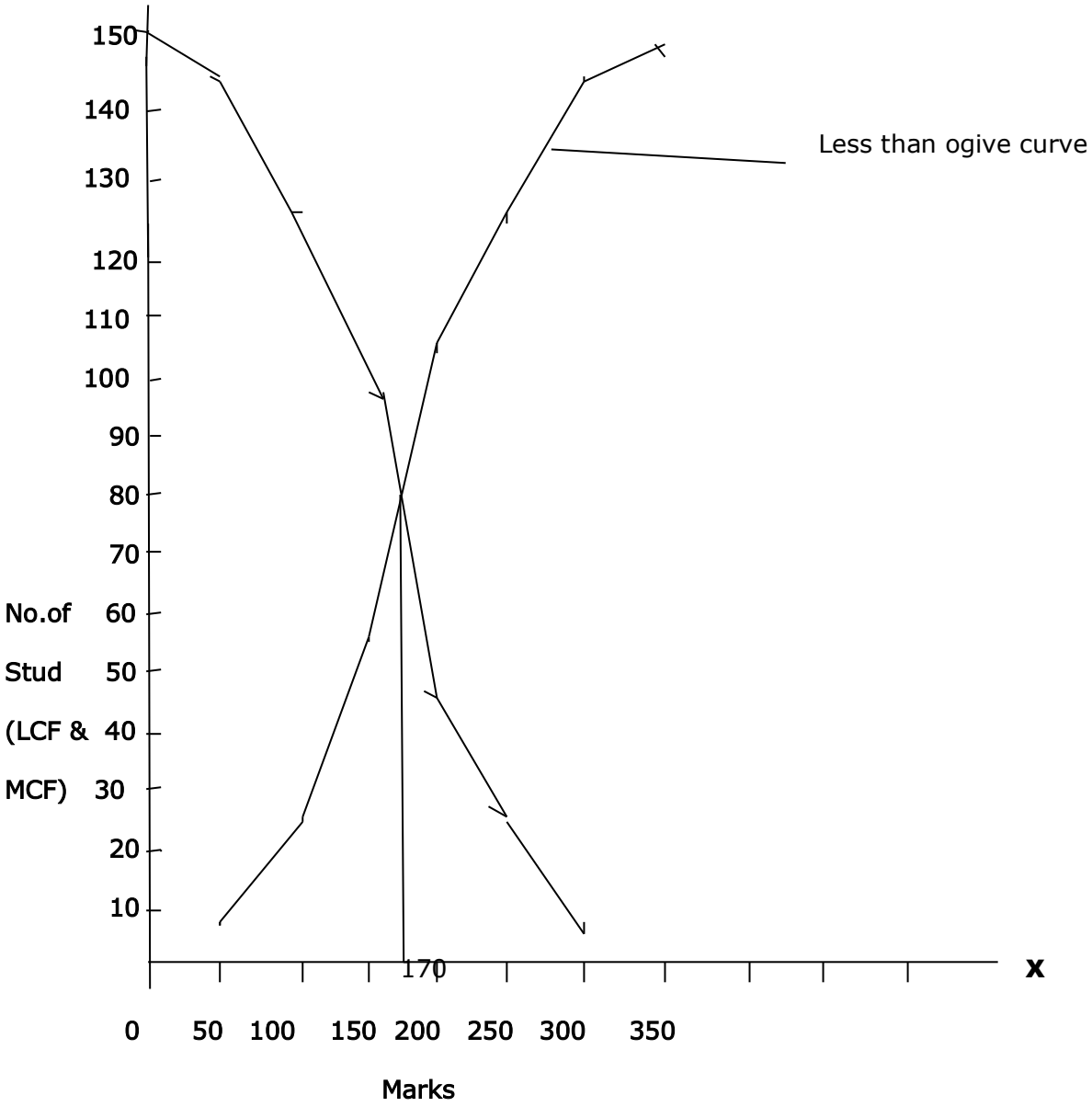
The line of intersection is touching X-axis at 170 marks, therefore value of median is 170 marks.

Calculation of Median Graphically using Ogive Curves

Scale (Measurements)

X-Axis: 1cm = 50 marks

Y-Axis: 1cm = 10 Students

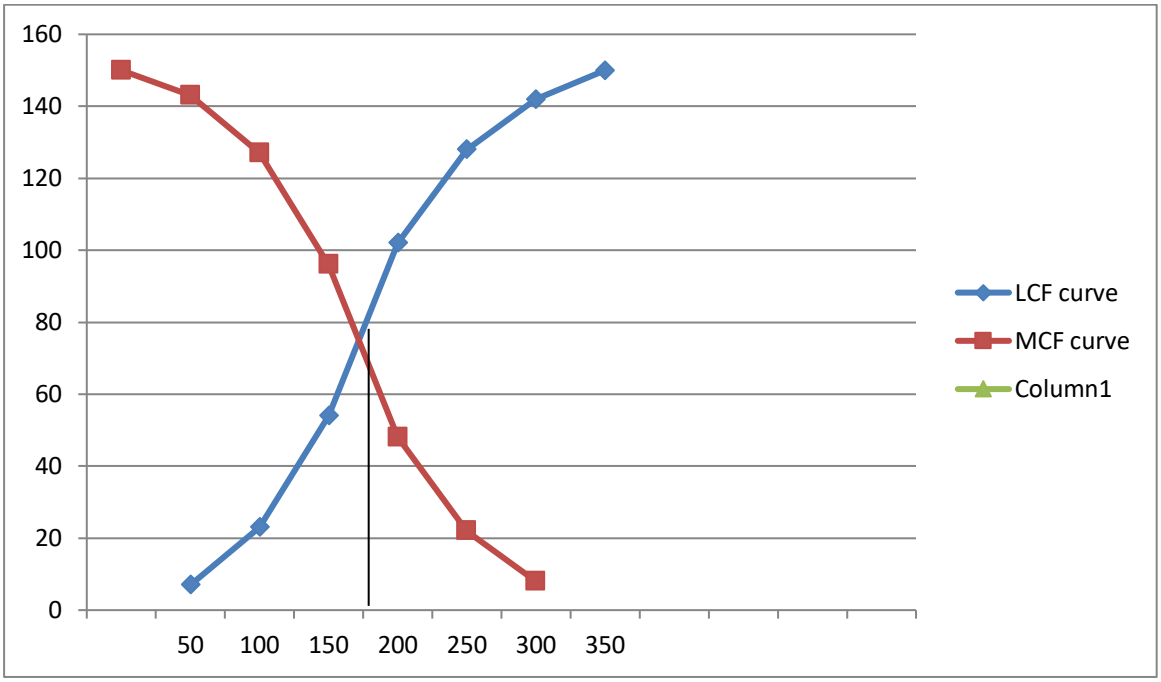


Marks	0-50	50-100	100-150	150-200	200-250	250-300	300-350
No.of students	7	16	31	48	26	14	8

marks	LCF curve	MCF curve
0	-	150
50	7	143
100	23	127
150	54	96
200	102	48
250	128	22
300	142	8
350	150	-

Scale on x-axis 1 cm = 50 marks

On y-axis 1cm = 20 students



Problem No:13

Draw less than ogive curve and more than ogive curve and also locate median graphically

wages	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45
No.of workers	3	2	12	13	20	18	15	10	7

Solution

Converting simple frequency into LCF for drawing less than Ogive Curve

Marks (upper limits)	No.of students (Simple Frequency – SF)	Less than Cumulative Frequency (LCF)
5	3	3
10	2	3+2=5
15	12	5+12=17
20	13	17+13=30
25	20	30+20=50
30	18	50+18=68
35	15	68+15=83
40	10	83+10=93
45	7	93+7=100

Converting simple frequency into MCF for drawing more than Ogive Curve

Marks (Lower limits)	No.of students (simple frequency - SF)	More than Cumulative Frequency (MCF)
0	3	97+3=100
5	2	95+2=97
10	12	83+12=95
15	13	70+13=83
20	20	50+20=70
25	18	32+18=50
30	15	17+15=32
35	10	7+10= 17
40	7	7

Note :

1. **Take the classes on X-axis and LCF and MCF values on the Y-axis**
2. **For drawing less than ogive curve, mark the upper limits with its respective LCF values and join all the points with scale.**
3. **For drawing more than ogive curve, mark the lower limits with its respective MCF values and join all the points with scale.**

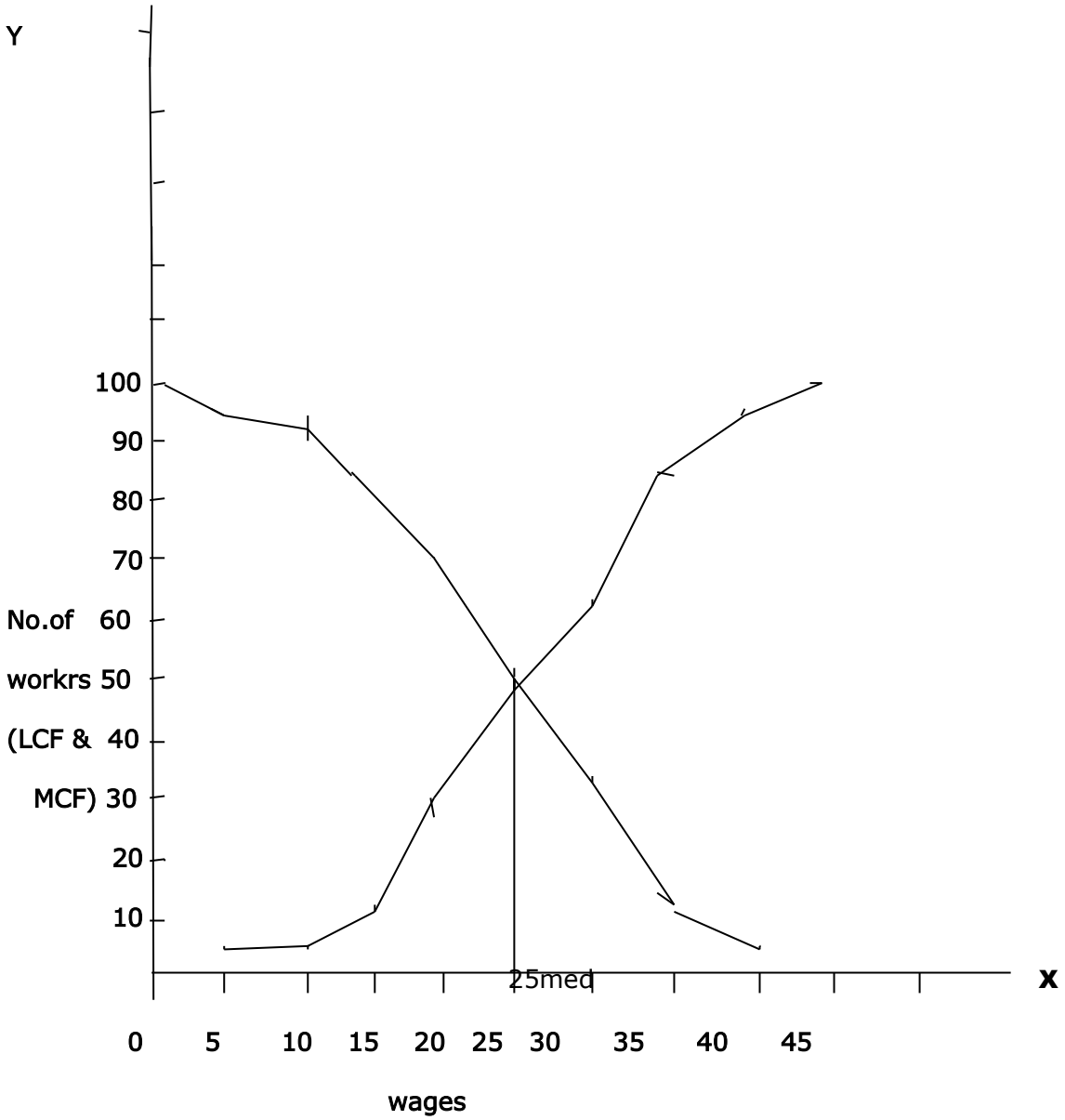
The line of intersection is touching X-axis at Rs. 25 (wages),therefore value of median is Rs.25 wages.

Calculation of Median Graphically using less than & more than Ogive Curves

Scale (Measurements)

X-Axis: 1cm = Rs.5 wages

Y-Axis: 1cm = 10 workers

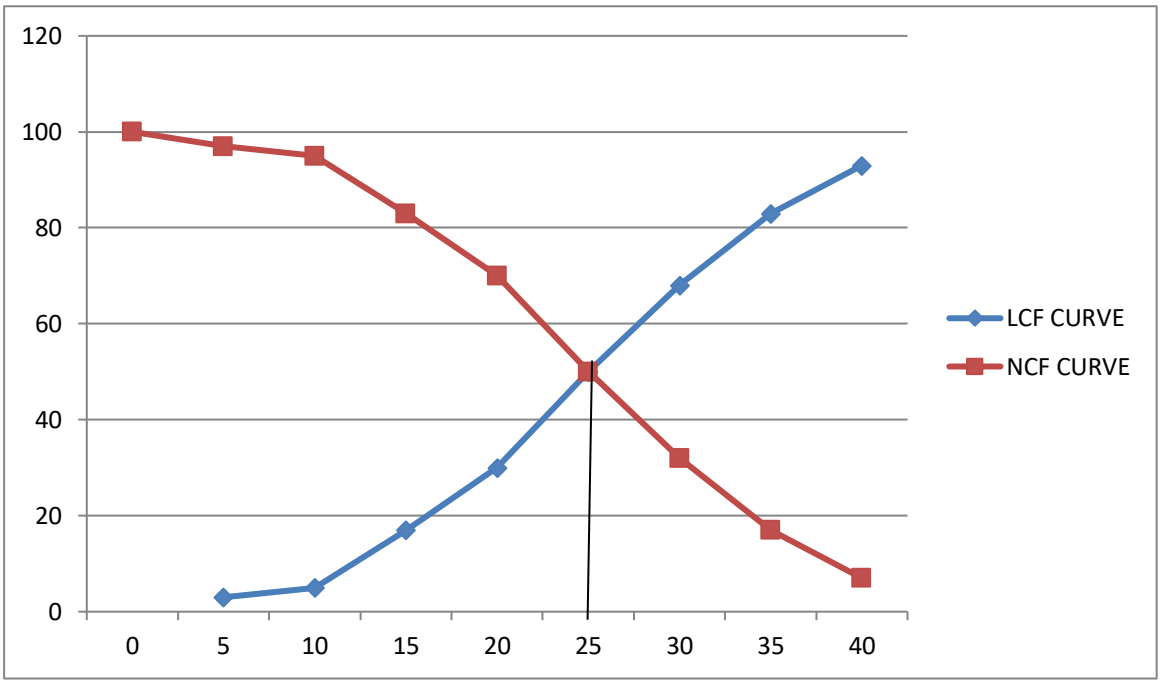


SCALE on x-axis 1cm = Rs. 5 wages

On y axis 1xm = 20 workers

marks	LCF CURVE	NCF CURVE
0	-	100
5	3	97
10	5	95
15	17	83
20	30	70
25	50	50
30	68	32
35	83	17

40	93	7
45	100	-



UNIT 3

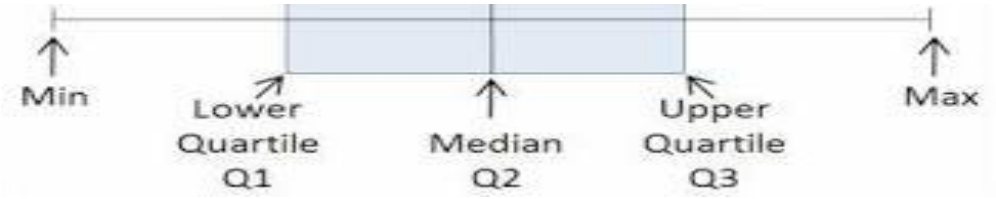
MEASURES OF CENTRAL TENDENCY -II

Median:

Median is the middle most value. It divides complete data into two equal parts.

Arrange the data in ascending order

Quartile: 4 parts



Q_1 divides the complete data into the first 25% whereas Q_3 divides data into the last 75%.

Interquartile Range $= Q_3 - Q_1$

Semi Interquartile Range or Quartile deviation $= \frac{Q_3 - Q_1}{2}$

Coefficient of Quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

Individual Series

Quartile = 4

$Q_1 = \left(\frac{N+1}{4}\right)^{th}$ **term or observation**

Q_2 **or Median** $= 2\left(\frac{N+1}{4}\right)^{th}$ **term** $= \left(\frac{N+1}{2}\right)^{th}$ **term**

$Q_3 = \left(\frac{3(N+1)}{4}\right)^{th}$ **term**

Octiles =8

$O_r = \left(\frac{r(N+1)}{8}\right)^{th}$ **term**

Deciles =10

$\mathbf{D_k = \left(\frac{k(N+1)}{10}\right)^{th}}$ **term**

Percentiles = 100

$\mathbf{P_p = \left(\frac{p(N+1)}{100}\right)^{th}}$ **term**

STEPS: .

Question 1: Calculate the value of median from the following figures
X = 250, 75, 85,210, 42,101,300,175, 145,
Also calculate O_2, D_3 , P_{55} , Interquartile deviations, Semi interquartile deviations and coefficient of Quartile deviation.
Solution:Arrange data in ascending order.

X	ORDER
42	**1ST**
75	**2ND**
85	**3RD**
101	**4TH**
145	**5TH = Me**
175	**6TH**
210	**7TH**
250	**8TH**
300	**9TH**

N = 9

Calculation of Q1

Q1 = Size of $\left(\frac{\mathbf{N+1}}{\mathbf{4}}\right)^{\mathbf{th}}$ term

= Size of $\left(\frac{9+1}{4}\right)^{th}$ term

= Size of $\left(\frac{10}{4}\right)^{th}$ term

= Size of 2.5^{TH} TERM

Q1= 2^{ND} TERM + 0.5(3^{RD} TERM – 2^{ND} TERM)

Q1 = 75 +0.5(85-75)

Q1 = 75 +0.5X10=75+5=80

Calculation of median = Q2

M = Size of $\left(\frac{\mathbf{N+1}}{\mathbf{2}}\right)^{\mathbf{th}}$ term

= Size of $\left(\frac{9+1}{2}\right)^{th}$ term

= Size of $\left(\frac{10}{2}\right)^{th}$ term

M = Size of 5^{th} term = 145

$\therefore M = 145$

Calculation of Q3

Q3 = Size of $3\left(\frac{\mathbf{N+1}}{\mathbf{4}}\right)^{\mathbf{th}}$ term

= Size of $3\left(\frac{9+1}{4}\right)^{th}$ term

= Size of $3\left(\frac{10}{4}\right)^{th}$ term

Size of 3x2.5 th term

Q3 = size of 7.5^{th} term

Q3 = 7^{th} term + 0.5(8^{th} term – 7^{th} term)

210 + 0.5(250-210)

=210 + 0.5x40

210 + 20

=230

Calculation of Interquartile nge , Semi interquartile range and Coefficient of Quartile deviation

Interquartile deviations Range = $Q_3 - Q_1$

Given $Q_1 = 80$, $Q_3 = 230$

= 230 – 80

= 150

Semi Interquartile Range or Qurtile deviation = $\frac{\mathbf{Q_3 - Q_1}}{\mathbf{2}}$

Given $Q_1 = 80$, $Q_3 = 230$

= $\frac{230 - 80}{2}$

$= \frac{150}{2}$

$= 75$

Coefficient of Quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

Given $Q_1 = 80$, $Q_3 = 230$

$= \frac{230 - 80}{230 + 80}$

$= \frac{150}{310}$

$= 0.484$

Calculation of O_2 , D_3 , P_{55}

- **$O_r = \left(\frac{r(N+1)}{8}\right)^{th}$ term**

  Given r = 2, N = 9

  $O_2 = \left(\frac{2(9+1)}{8}\right)^{th}$ term

  $= \left(\frac{2(10)}{8}\right)^{th}$ term

  $= \left(\frac{20}{8}\right)^{th}$ term

  = Size of 2.5 th term

  = 2 nd term + 0.5 (3 rd term - 2 nd term)

  = 75 + 0.5 (85-75)

  = 75 + 0.5(10) = 75 + 5

  $\therefore O_2 = 80$

- **$D_k = \left(\frac{k(N+1)}{10}\right)^{th}$ term**

  Given k = 3, N = 9

  $D_3 = \left(\frac{3(9+1)}{10}\right)^{th}$ term

  $= \left(\frac{3(10)}{10}\right)^{th}$ term

  $= \left(\frac{30}{10}\right)^{th}$ term

  = Size of 3 th term = 85

  $\therefore D_3 = 85$

- **$P_p = \left(\frac{p(N+1)}{100}\right)^{th}$ term**

  Given p = 55, N = 9

  $P_{55} = \left(\frac{55(9+1)}{100}\right)^{th}$ term

  $= \left(\frac{55(10)}{100}\right)^{th}$ term

  $= \left(\frac{550}{100}\right)^{th}$ term

= Size of 5.5 th term

= 5 th term + 0.5 (6 th term -5 th term)

= 145 + 0.5 (175 - 145) = 145 + 0.5 (30)

= 145 + 15 = 160

$\therefore P_{55} = 160$

PROBLEMS

1. Calculate Median, Quartiles, Interquartile deviations, Semi interquartile deviations, coefficient of Quartile deviation, 8 Decile (D_8) , 3 Octile (O_3), 61 Percentile (P_{61}) for the following 22,26,14,30,18,11,35,41,12,32

11 1ST
12 2ND
14 3RD
18 4TH
22 5 TH
26 6TH
30 7TH
32 8TH
35 9TH
41 10TH

Answers: Q1= 13. 5

2. Calculate Median, Quartiles, Interquartile deviations, Semi interquartile deviations, coefficient of Quartile deviation, 5 Decile (D_5) , 7 Octile (O_7), 77 Percentile (P_{77}) for the following 5,7,9,12,8,7,15,25,11

3. Calculate Median, Quartiles, Interquartile deviations, Semi interquartile deviations, coefficient of Quartile deviation, 6 Decile (D_6) , 5 Octile (O_5), 87 Percentile (P_{87}) for the following age of students 18,16,14,11,13,10,9,20,15

4. Calculate Median, Quartiles, Interquartile deviations, Semi interquartile deviations, coefficient of Quartile deviation, 8 Decile (D_8) , 2 Octile (O_2), 50 Percentile (P_{50}) for the following 30,45,75,65,50,52,28,40,39,35,42

22 26 14 30 18 11 35 41 12 32

ARRANGE IN ORDER

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
11	**12**	**14**	**18**	**22**	**26**	**30**	**32**	**35**	**41**

Q1 = (N+1)/4TH TERM

Q1 = (10+1)/4TH TERM

Q1 =2.75TH TERM

Q1 = 2ND TERM +0.75 (3RD TERM - 2ND)

12+ 0.75(14-12)

12+ 1.5=13.5

Q3 = 3(N+1)/4TH TERM

Q3 = 3(10+1)/4TH TERM

Q3= 3X2.75 TH TERM

Q3 = 8.25TH TERM

8TH TERM + 0.25(9TH TERM - 8TH TERM)

Q3 = 32 + 0.25(35-32)

Q3 = 32.75

MEDIAN OF DICRETE SERIES

X	8	10	52	25	37
F	5	3	4	2	1

8 , 8, 8, 8, 8,10,10,10,25, 25, 37, 52,52,52,52

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

ARRANGE THE DATA IN ORDER

X	F	CF
8 = Q1	5	5 Q1 LIES HERE
10 = Q2 = ME = O3	3	8 ME LIES HERE , O3 LIES HERE
25	2	10
37 =P 65	1	11 P65 LIES HERE
52= Q3= D7	4	15 Q3 LIES HERE, D7 LIES HERE
	N=∑F = 15	

Q1 = (N+1)/4TH TERM
Q1 = (15+1)/4TH TERM
Q1 =4TH TERM
Q1= 4th term
Q1 = 8
ME = Q2= (N+1)/2TH TERM
Q2 = (15+1)/2TH TERM
ME= 8TH TERM
ME = 10
Q3=3(N+1)/4TH TERM
Q3=3(15+1)/4TH TERM
Q3=12TH TERM
Q3=52
O3= 3(N+1)/8 TH TERM
O3= 3(15+1)/8TH TERM
O3 = 6TH TERM
O3= 10
D7= 7(N+1)/10TH TERM
D7= 11.2TH TERM
D7=52
P65=65(N+1)/100 TH TERM
P65 = 10.4TH TERM
P65= 37

HEIGHT IN CMS	165	156	163	149	168	159	155	170	16 4
NO. OF STUDENTS	5	8	9	6	7	3	2	1	6

Q1, Q2, Q3, O5, D9. P45

Step 1 arrange the data in order

Marks	35	23	19	22	42	46	12	50	18
No. Of students	2	3	8	9	4	5	7	8	3

marks	No. Of students = f	Cf
12	7	7
18	3	10
19 = Q1	8	18 Q1 lies here
22 = Q2	9	27 Q2 lies here
23	3	30
35=O5	2	32 O5 lies here
42 = D7	4	36 D7 lies here, P67 lies here
46 = Q3	5	41 Q3 lies here
50	8	49
	N=∑f= 49	

Q1 = (N+1)/4th term
Q1 = (49+1)/4= 12.5th term
Q1 is 12.5th term
Q1 = 19
Q2 =Me= (N+1)/2th term
Q2 = (49+1)/ 2 th term= 25th term
Q2= 22
Q3 = 3(N+1)/4th term
Q3 = 3(49+1)/4 th term
Q3 = 37.5th term
Q3 = 46

O5= 5(N+1)/8th term =31.25th term
O5= 35
D7 = 7(N+1)/10th term
D7 = 7(49+1)/10th term =35th term
D7 = 42
P67=67(N+1)/100th term
P67 = 67(49+1)/100 th term =33.55th term
P67 = 42

Continous series

Marks	No. Of students = f	Cf
0-10	3	3
10-20	5	8
20-30	8	16 Q1 lies here
30-40	15	31 Me lies here
40-50	6	37 Q3 lies here
50-60	5	42
	N=∑F=42	

Q1 lies in N/4th term

Q1 lies in 42/4th term

Q1 lies in 10.5th term

Q1 lies in 20- 30 CI

L= 20, h= 30-20=10, f= 8, N/4= 10.5, cf= 8 (cf <n/4)

Q1 = $L + \frac{\left(\frac{N}{4}\right) - cf}{f} xh$

Q1 = 20 + $\frac{10.5-8}{8}$x10

Q1= 20 + $\frac{2.5x10}{8}$ = 20 +3.125

Q1= 23.125

Me = Q2= $L + \frac{\left(\frac{n}{2}\right) - cf}{f} xh$

Me Lies in N/2th term

Me Lies in 42/2th term

Me lies in 21st term

Me lies in 30-40 CI

L= 30, h=40-30=10,f=15, n/2= 21, cf= 16 (cf<N/2)

Me = Q2= $L + \frac{\left(\frac{n}{2}\right) - cf}{f} xh$

Me =33.33

Q3 lies in (3N/4) th term

Q3 = (3x42/4) th term

Q3 lies in 31.5th term

Q3 lies in 40-50 C I

L = 40, h=50-40=10, f= 6, 3n/4=31.5, cf = 31

Q3 = $L + \frac{\left(\frac{3N}{4}\right) - cf}{f} xh$

Q3 = 40 + $\frac{31.5-31}{6} x10$

40.83

O4=$L + \frac{\left(\frac{4n}{8}\right) - cf}{f} xh$

D7=$L + \frac{\left(\frac{7n}{10}\right) - cf}{f} xh$

P65= $L + \frac{\left(\frac{65n}{100}\right) - cf}{f} xh$

mARKS	NO. OF STUDENTS
0-10	8
10-20	10
20-30	22
30-40	25
40-50	10
50-60	5

Marks	NO. OF STUDENTS=f	cf
0-10	8	8
10-20	10	18
20-40	22	40 Q1 lies here, Me=Q2 lies here
40-60	25	65 Q 3 lies here
60-90	10	75
90-100	5	80
	N = ∑f = 80	

Q1 lies in n/4th term

Q1 lies in 80/4th term

Q1 lies in 20th term

Q1 lies in20-40 CI

L=20, h=40-20=20, f =22, n/4=20,cf= 18

Q1 = $L + \frac{\left(\frac{N}{4}\right) - cf}{f} xh$

Q1 = $20 + \frac{20 - 18}{22} x20$

Q1= 21.81

Q2=Me lies in n/2th term

Q2 =Me lies in 80/2th term

Q2 =Me lies in 40th term

Q2=Me lies in 20-40 C I

L = 20, h = 40-20=20, f =22, n/2=40, cf=18

Q2 = =Me= $L + \frac{\left(\frac{N}{2}\right) - cf}{f} xh$

Me = $20 + \frac{40-18}{22} x20$

Me = 40

Q3 lies in 3N/4th term

Q3 lies in 3x80/4 th term

Q3 lies in 60th term

Q3 lies in 40-60 C I

L = 40, h = 60-40=20, f=25, 3n/4 =60, cf= 40

Q3 = L + $\frac{\left(\frac{3N}{4}\right)- cf}{f} xh$

Q3 =56

Interquartile Range = Q3-Q1

Semi Interquartile Range or Quartile deviation = $\frac{Q_3 - Q_1}{2}$

Coefficient of Quartile deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

CALCULATION OF MEDIAN WHEN MID POINTS ARE GIVEN

MID POINTS	5	15	25	35	45	55
FREQUENCY	6	7	18	11	30	20

CORRECTION FACTOR= (15-5)/2=10/2=5

SOLUTION

LOWER LIMIT = 5-5=0

MID POINTS	CLASS INTERVAL	F	CF
5	0-10	6	6
15	10-20	7	13
25	20-30	18	31 Q1 LIES HERE
35	30-40	11	42
45	40-50	30	72
55	50-60	20	92
		N=∑f=92	

Q1 =LIES IN N/4TH TERM = 92/4TH TERM=23RD TERM

L = 20, H = 30-20=10,, F=18, N/4 =23, cf=13

Q1 = $L + \frac{\left(\frac{N}{4}\right) - cf}{f} xh$

Q1= 20 +(23-13)/18 x10

Q1= 20+(10x10)/18

Q1 = 20+5.55

Q1= 25.55

Me=

MISSING FEQUENCIES IN MEDIAN

MARKS	F
10-20	12
20-30	30
30-40	? =F1
40-50	65
50-60	?=F2
60-70	25
70-80	18

GIVEN MEDIAN =46, TOTAL NO OF STUDENTS = 229
FIND THE MISSING FREQUENCIES

SOLUTION
GIVEN N= 229
LET THE MISSING FREQUENCIES BE F1 , F2
N= 12+30+F1+65+F2+25+18=150+F1+F2
229-150=F1+F2
79=F1+F2
79-F1 = F2

MARKS	F	CF
10-20	12	12
20-30	30	42
30-40	F1	42+F1=cf
40-50	65 =f	107+F1 ME LIES HERE
50-60	79-F1	186
60-70	25	211
70-80	18	229

GIVEN MEDIAN = 46
HENCE MEDIAN CLASS IS 40-50 L=40, H=10, F=65, N/2=229/2=114.5, CF=(42+F1)

Me= $L + \frac{\left(\frac{N}{2}\right) - cf}{f} xh$

46= 40+ $\frac{(114.5) - (42+F1)}{65} x10$

46-40 = $\frac{(114.5-42-F1)}{65} X10$

6= $\frac{(72.5-f1)}{65} x10$ = $\frac{6x65}{10}$ = 72.5-F1

39 = 72.5-F1

F1=72.5-39

F1=33.5=34

F2= 79-F1=79-34=45

Q

expenditure	0-20	20-40	40-60	60-80	80-100				
No. of families	14	?	27	?	15				

if median is 50 and the no. of families are 100 find the missing frequencies

23, 21

median is24 find missing frequency

Marks	No. of students	
0-10	5	
10-20	25	
20-30	?	
30-40	18	
40-50	7	

f= 25

Mode is most repeated value

Individual series

12,23, 45, 65, 67, 43, 12, 70,67,42, 50,67

Arrange in order

12,12,23,42,43,45,50,65,67,67,67,70

Mode = 67 (3times)

Discrete series

X	1	2	3	4	5	6	7	8	9	10
f	0	2	2	3	1	2	3	4	1	2

Mode = Z = 8 (f= highest = 4)

This not correct way

Grouping Table

X	F=I	II	III	IV	V	VI
10	7	7+15=22	X	7+15+21=43	X	X
12	15		15+21=36		15+21+38=74	X
14	21	21+38=59				21+38+34=93
16	38		38+34=72	38+34+34=106		
18	34	34+34=68			34+34+11=79	
20	34		34+11=45			34+11+19=64
22	11	11+19=30		11+19+10=40		
24	19		19+10=29		19+10+38=67	
26	10	10+38=48				10+38+5=53
28	38		38+5=43	38+5+2=45		
30	5	7				
32	2					

Analysis Table

x	10	12	14	16	18 =z	20	22	24	26	28	30	32
I				1						1		
II					1	1						
III				1	1							
IV				1	1	1						
V					1	1	1					
VI			1	1	1							
total			1	4	5	3	1			1		

Grouping Table

X	F=I	II	III	IV	V	VI
32	7		X		X	X
33	14					X
34	30					
35	28					
36	35					
37	34					
38	16					
39	14					
40	36					
41	16					

Analysis Table

X	32	33	34	35	36	37	38	39	40	41
I										
II										
III										
IV										
V										
VI										
Toal										

Continuous series

Grouping Table

CI	F= I	II	III	IV	V	VI
0-10	8	8+6= 14	X	8+6+10= 24	X	X
10-20	6		6+10= 16		6+10+12= 28	X
20-30	10	10+12= 22				10+12+20= 42
30-40	12=f0		12+20= 32	12+20+12= 44		
40-50	20=f1	20+12= 32			20+12+5= 37	
50-60	12=f2		12+5= 17			12+5+3= 20
60-70	5	5+3=8		5+3+2= 10		
70-80	3		3+2= 5		3+2+4= 9	
80-90	2	2+4=6				
90-100	4					

Analysis Table

X	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
I					1					
II					1	1				
III				1	1					
IV				1	1	1				
V					1	1	1			
VI			1	1	1					
Toal			1	3	6	3	1			

Modal class is 40-50

$Z = L + \frac{f1-f0}{2f1-f0-f2} xh$

$Z = 40 + \frac{20-12}{2(20)-12-12} x10$

$Z = 40 + \frac{8x10}{40-24}$

$Z = 40 + \frac{80}{16}$

Z = 40 +5=45

Marks	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50
No. of students	22	45	67	73	85=f0	190=f1	64=f2	55

modal class is 35-40

CI	F (I)	II	III	IV	V	VI	
10-15	22	67	X	134	X	X	
15-20	45		112		185	X	
20-25	67	140				225	
25-30	73		158	348			
30-35	85=f0	275			339		
35-40	190 =f1		254			309	
40-45	64 =f2	119					
45-50	55						

C1	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	
I						1			
II					1	1			
III						1	1		
IV				1	1	1			
V					1	1	1		
VI						1	1	1	
TOTAL				1	3	6	3	1	

Modal class is 35-40

$Z = L + \frac{f1-f0}{2f1-f0-f2} xh$

$Z = 35 + \frac{190-85}{2(190)-85-64} x5$

$Z = 35 + \frac{105x5}{231}$

Z= 35 +2.27=37.27

question

Marks below	10	20	30	40	50	60	70	80	90
No. of students (cf)	4	6	24	46	67	86	96	99	100

exclusive series
equal heights

CI	F (I)	II	III	IV	V	VI
0-10	4	6	X	24	X	X
10-20	6-4=2		20		42	X
20-30	24-6=18	40				61
30-40	46-24=22		43	62		
40-50	67-46=21	40			50	
50-60	86-67=19		29			32
60-70	96-86=10	13		14		
70-80	99-96=3		4			
80-90	100-99=1					

CI	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
I				1					
II			1	1	1	1			
III				1	1				
IV				1	1	1			
V					1	1	1		
VI			1	1	1				
TOTAL			2	5	5	3	1		

THIS IS CALLED AS BI- MODAL SERIES

Z=3ME-2MEAN

CI	M=midvalues	d=m-A	D =D/10	F	fd	cf	
0-10	5	-30	-3	4	-12	4	
10-20	15	-20	-2	2	-4	6	
20-30	25	-10	-1	18	-18	24	
30-40	35 = A	0	0	22	0	46=cf	
40-50	45	10	1	21 =f	21	67 median lies here	
50-60	55	20	2	19	38	86	
60-70	65	30	3	10	30	96	
70-80	75	40	4	3	12	99	
80-90	85	50	5	1	5	100	
				∑f=100	∑fd =72		

$\bar{X} = A + \frac{\sum fd}{\sum f} xh$

$35 + \frac{72}{100} x10$

35 +7.2

42.2

Median lies in n/2th term =100/2th term = 50th term

L= 40, n/2=50, cf=46, f=21, h = 10

$Me = L + \frac{\frac{n}{2} - cf}{f} xh$

$me = 40 + \frac{50-46}{21} x10$

40+ 1.9

41.9

Mode = 3Median - 2Mean

3(41.9)- 2(42.2)=41.3

Size	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of items	4	6	20	32	33	17	8	2

GROUPING TABLE

CI	F (I)	II	III	IV	V	VI
0-10	4	4+6= 10	X	4+6+20= 30	X	X
10-20	6		6+20= 26		6+20+32= 58	X
20-30	20	20+32= 52				20+32+33= 85
30-40	32		32 +33= 65	32+33+17= 82		
40-50	33	33+17= 50			33+17+8= 58	
50-60	17		17+8= 25			17+8+2= 27
60-70	8	8+2= 10				
70-80	2					

CI	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
I					1			
II			1	1				
III				1	1			
IV				1	1	1		
V		1	1	1	1	1	1	
VI			1	1	1			
TOTAL		1	3	5	5	2	1	

THIS IS A BIMODAL SERIES

MODE = 3(ME)-2(MEAN)

CI	F						
0-10	4						
10-20	6						
20-30	20						
30-40	32						
40-50	33						
50-60	17						
60-70	8						
70-80	2						

MEAN = 39.5

MEDIAN=40.3

MODE = 3(ME)-2(MEAN)= 3(40.3)-2(39.5)=41.9

UNIT 3

MEASURES OF CENTRAL TENDENCY -I

MEANING AND DEFINITION

Meaning and definition

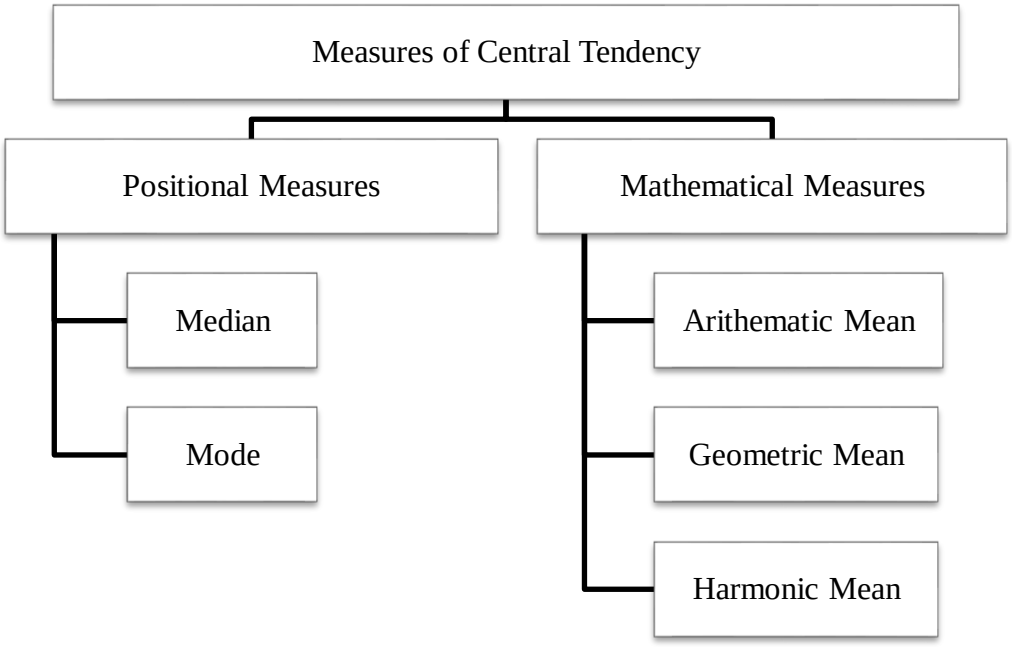
According to Simoson and Kofka - "A measure of Central tendency is the typical value around which the other figures concentrate. The average is also called as the central value

According to A.L. Bowley "Statistics maybe rightly called as science of averages. Science of averages is also called as Measures of Central Tendency which means measure of location. The average always lies between the two extreme values and represents the whole group."

Essentials of good average

1. Average should be based on all the observations as it should not be affected by the extreme values
2. Averages should be rigidly defined
3. Average should be easy to calculate and simple to understand
4. Average can be formed by the graphical methods also
5. Averages should be capable of further algebraic treatment
6. Averages should not be affected by skewness
7. Averages should be popular

DIFFERENT TYPES OF AVERAGES



ARITHEMATIC MEAN

Arithmetic mean is the most widely used measure to represent the entire data. It is obtained by adding all the items with the number of items.

NOTE: For discrete and continuous series the method is different.

Individual series

Individual series are those where only variables are given without any frequencies. Serial nos. or Roll nos. is not frequencies.

DIRECT METHOD

Here the mean can be found by three methods.

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

$\bar{X}$ = **Arithmetic mean**

$\sum X$ = Sum of all values given

N = Number of values

Question 1: For the following calculate arithmetic mean.

Roll No	1	2	3	4	5	6	7	8	9	10
Marks	10	9	12	15	20	17	19	20	10	9

Solution:

Roll No	Marks
1	10
2	9
3	12
4	15
5	20
6	17
7	19
8	20
9	10
10	9
	Total or $\sum X = 141$

$$\therefore \bar{X} = \frac{\sum X}{N} = \frac{141}{10} = 14.1$$

PROBLEMS

1. Six month income of a departmental store is given below. Find mean income of store.

Month	Jan	Feb	Mar	Apr	May	June
Income (Rs.)	10000	20000	15000	20000	30000	25000

Answer: Rs. 20000

2. For the following calculate arithmetic mean.

Roll No	1	2	3	4	5	6	7	8	9	10
Marks in Statistics	20	22	23	25	21	26	24	28	21	20

Answer: 23.

3. Following are the marks obtained by a student.

English	Hindi	Statistics	Economics	Accounts	Maths
123	156	127	160	102	124

Answer: 132.

4. Calculate Arithmetic Mean from your following.

SL no	0	1	2	3	4	5	6	7
Monthly Wages	150	155	159	142	160	153	154	140

Answer: Rs. 151.625.

SHORT CUT METHOD:

Here $\overline{X}$ is calculated using an Assumed Mean; taking deviations from it, the following formula is used.

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

Where A is assumed mean

And d = the deviation of items from assumed mean (X – A),

(Note: – Value of Assumed Mean may be taken of any magnitude; but we often take whole number which is repeated more no. of times.)

Question 1: For the following calculate arithmetic mean.

Roll No	1	2	3	4	5	6	7	8	9	10
Marks	30	70	40	20	60	40	30	80	50	90

Solution:

Here A =40 **A=50**

Roll no.	Marks	d= X-A (A= 40)	Marks	d = X-A (A= 50)
1	30	30-40=-10	30	30 - 50 = -20
2	70	70-40=30	70	70 - 50 = 20
3	40	40-40=0	40	40 - 50 = -10
4	20	20-40=-20	20	20 - 50 = -30
5	60	60-40=20	60	60 - 50 = 10
6	40	40-40=0	40	40 - 50 = -10
7	30	30-40=-10	30	30 - 50 = -20
8	80	80-40=40	80	80 - 50 = 30
9	50	50-40=10	50	50 - 50 = 0
10	90	90-40=50	90	90 - 50 = 40
	N = 10	∑d=110		∑d = 10

Case I

$\overline{X} = A + \frac{\sum d}{N}$, **N= 10 , Here A= 40 ∑d =110**

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

40+ (110/10) = 40+11=51

Case II

Here, **A= 50, ∑d = 10**

$\overline{X} = 50 + \frac{10}{10}$, $\overline{X} = 50 + 1$

$$\overline{X} = 51$$

Hence the average marks are 51.

PROBLEMS

1. For the following calculate arithmetic mean.

Roll No	1	2	3	4	5	6	7	8	9	10
Marks in Statistics	20	22	23	25	21	26	24	28	21	20

Answer: 23.

2. Following are the marks obtained by a student.

English	Hindi	Statistics	Economics	Accounts	Maths
123	156	127	160	102	124

Answer: 132.

3. Calculate Arithmetic mean for the following by using Short-cut method

X	150	210	140	190	170	400	310	230	310	420	190

Answer: 247.27.

4. Calculate Arithmetic Mean from your following.

SL no	0	1	2	3	4	5	6	7
Monthly Wages	150	155	159	142	160	153	154	140

Answer: Rs. 151.625.

STEP DEVIATION METHOD:

This method is not applicable to all the problems.

Here

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

Where A is assumed mean

And d = the deviation of items from assumed mean (X – A),

Where i or c is common factor in d, such that d/i or d/c = d '

h= height of the class

Question 1: For the following calculate arithmetic mean.

Roll No	1	2	3	4	5	6	7	8	9	10
Marks	30	70	40	20	60	40	30	80	50	90

Solution: **i=c=10**

Roll no.	Marks	d= X-A (A= 40)	d'=d/10
1	30	30-40=-10	-10/10=-1
2	70	70-40=30	30/10=3
3	40	40-40=0	0
4	20	20-40=-20	-20/10=-2
5	60	60-40=20	20/10=2
6	40	40-40=0	0
7	30	30-40=-10	-10/10=-1
8	80	80-40=40	40/10=4
9	50	50-40=10	10/10=1
10	90	90-40=50	50/10=5

	∑X=510	∑d=110	∑d'=11

Direct Method

$$\overline{\mathbf{X}} = \frac{\sum \mathbf{X}}{\mathbf{N}}$$

$$= \frac{\mathbf{510}}{\mathbf{10}}$$

$$\mathbf{=51}$$

Short cut Method

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

$$= \mathbf{40} + \frac{\mathbf{110}}{\mathbf{10}}$$

$$= \mathbf{40} + \mathbf{11}$$

$$= \mathbf{51}$$

Step Deviation Method

$$\overline{\boldsymbol{X}} = \boldsymbol{A} + \frac{\sum \boldsymbol{d'}}{\boldsymbol{N}} \times \boldsymbol{C}$$

$$= \mathbf{40} + \frac{\mathbf{11}}{\mathbf{10}} \boldsymbol{x}\mathbf{10}$$

$$= \mathbf{40} + \mathbf{11}$$

$$= \mathbf{51}$$

Hence the average marks are 51.

PROBLEMS:

1. The following are the monthly salaries in rupees of employees of a mill. Calculate the average salary per employee.

Employee id.	100	101	102	103	104	105	106	107	108
Rupees	100	187	229	245	300	305	325	954	375

Answer: Rs. 335.56

2. Following are the marks obtained by a student.

Student id	011	012	013	014	015	016	017	018
Marks	150	155	159	142	160	153	154	140

Answer: 151.625.

3. Following are the marks obtained by a student.

Roll No	1	2	3	4	5	6	7	8	9	10
Marks in Statistics	20	22	23	25	21	26	24	28	21	20

Answer: 23.

4. Calculate mean for the following

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

Answer: 16.86

5. Calculate mean for the following

Roll no.	1	2	3	4	5	6	7	8	9
Height	105	120	110	130	115	125	135	140	95

Roll no.	Height	d=X-A (A=120)	d'=d/5 (i=5)
1	105	-15	-3
2	120	0	0
3	110	-10	-2
4	130	10	2
5	115	-5	-1
6	125	5	1
7	135	15	3

8	140	20	4
9	95	-25	-5
	∑X=1075	∑d=-5	∑d'=-1

Direct method

$$\overline{X} = \frac{\sum X}{N} = \frac{1075}{9} = 119.444$$

Short cut Method

$$\overline{X} = A + \frac{\sum d}{N}$$
$$= 120 + \left(-\frac{5}{9}\right)$$
$$= 120 - 0.5555$$
$$= 119.444$$

Step Deviation Method

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

= 120+(-1/9)x5

= 120-0.55555

=119.4444

DISCRETE SERIES (frequency will be given i.,e Number of …..)

In discrete series, arithmetic mean maybe calculated applying either

a) Direct method

b) Short-cut method

c) Step Deviation method

Direct Method

$$\overline{X} = \frac{\sum fx}{\sum f}$$

∑fx = Sum of the product of the values and their corresponding frequencies
Sum of the frequencies i.e., ∑f = total number of observations.

Question 1: Calculate mean for the following data

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

Solution: Arrange in order along with their respective frequencies
Given question is in order only

Marks X	No. of students f	fX
4	6	24
8	12	96
12	18	216
16	15	240
20	9	180
	N = 60	ΣfX = 756

As $\bar{X} = \frac{\Sigma f X}{N}$

$\therefore \quad \bar{X} = \frac{756}{60} = 12.6$

PROBLEMS:

1. Calculate the arithmetic mean by direct method from the following data

Marks	No.of students
35	8
8	5
15	10
30	12
24	15
7	3
15	12

40	5

Solution Arrange in order (Lowest to height in Marks))

Marks	No.of students	order
~~35~~	8	7th
~~8~~	5	2nd
~~15~~	10	3rd
~~30~~	12	6th
~~24~~	15	5th
~~7~~	3	1st
~~18~~	12	4th
~~40~~	5	8th

After arranging from lowest to heigest marks along with their corresponding frequecies

Marks X	No.of students f	fx
7	3	21
8	5	40
15	10	150
18	12	216
24	15	360
30	12	360
35	8	280
	5	200
	∑f=70	$\sum fx = 1627$

$$\overline{X} = \frac{\sum fx}{\sum f}$$

$$\frac{1627}{70} = 23.2428 = 23.243$$

$$\overline{X} = 23.243$$

2. Calculate the arithmetic mean for the following data

Wages	20	40	30	50	10
No. of Employees	5	2	3	5	4

Answer: 29.47

3.

Items	5	10	20	30	40	50	60	70
Frequency	2	5	1	3	12	0	5	7

For discrete series, the Arithmetic Mean can be calculated using the following formula.
Answer: 41.143.

4. Calculate Arithmetic Mean for the following discrete data:

Items	14	36	45	70
Frequency	2	5	1	3

Answer: 42.09

Short Cut Method:

Here Assumed Mean is taken and taking deviations of variable from it. We obtain X by using the following formula.

$$\overline{X} = A + \frac{\sum fd}{\sum f}$$

Where A = Assumed Mean

d = (X-A);

f = frequency $\sum f$ = Total number of terms,

(Note:-This formula is often used when the variables are large in size or infractions and direct formula is not easy to use.)

Question 1: Calculate the mean for the following data

Wages (Rs.) :	10	20	30	40	50	60	70	80	90
Workers :	4	5	3	2	5	2	3	1	2

Solution:

Wages X	d= (X-A)	Workers f	fd
10	10-50=-40	4	-160
20	20-50=-30	5	-150
30	30-50=-20	3	-60
40	40-50=-10	2	-20
50	50-50=0	5	0
60	60-50=10	2	20
70	70-50=20	3	60
80	80-50=30	1	30
90	90-50=40	2	80
		∑f = 27	∑f d=-200

Here A = 50,$\sum fd = -200$ ∑f = 27.

$$\overline{X} = A + \frac{\sum fd}{\sum f}$$

$\overline{X} = 50 + (\frac{-200}{27})$ 50-7.407 = 42.593=42.6

Hence the Average wage is 42.6

PROBLEMS:

1. Calculate the arithmetic mean for the following data

Wages	10	20	30	40	50
No. of	4	5	3	2	5

employees					

Answer: 29.47.

2. Find the mean for the following Data.

X	9	10	11	12	13	14	15	16	17	18
f	1	2	3	6	10	11	7	3	2	1

Answer: 13.54.

3. Calculate Arithmetic mean from the following frequency distribution of marks in statistics.

Marks	5	10	15	20	25	30	35	40
No. of Students	5	7	9	10	8	6	5	2

Answer: 20.48.

4. Eight coins are tossed together; the number of heads resulting should theoretically be given by the following frequency distribution. Find the Mean.

X	0	1	2	3	4	5	6	7	8
f	1	9	26	59	72	52	29	7	1

Answer: 3.937.

Step Deviation Method

$$\overline{X} = A + \frac{\sum fd'}{\sum f} \times C$$

Where A = Assumed Mean
d = (X-A)
d'= d/c

c= common factor

f = frequency $\sum f$ = Total number of terms,

Question 1. Calculate arithmetic average for the following.

Wages	20	30	40	50	60	70	80
No. of Persons	5	2	3	10	3	2	5

Solution: **A= 50** **c= 10**

Wages X	**d=X-A**	**d′= d/c**	No .of Persons f	f **d′**
20	20-50=-30	-30/10=-3	5	-15
30	30-50=-20	$\frac{-20}{10} = -2$	2	-4
40	40-50=-10	$\frac{-10}{10} = -1$	3	-3
50	50-50=0	$\frac{0}{10} = 0$	10	0
60	60-50=10	$\frac{10}{10} = 1$	3	3
70	70-50=20	$\frac{20}{10} = 2$	2	4
80	80-50=30	$\frac{30}{10} = 3$	5	15
			∑f= 30	**∑fd′ = 0**

Here A = 50;$\sum fd' = 0$, N = 30, C = 10.

$$\overline{X} = A + \frac{\sum \mathbf{fd'}}{N} X\ C$$

$$\overline{X} = 50 + \left(\frac{0}{30}\right) X\ 10$$
$$= 50 + 0$$

$= 50.$

Hence average wage is 50.

PROBLEMS

1. Calculate Arithmetic mean from the following frequency distribution of marks in statistics.

Marks	5	10	15	20	25	30	35	40
No. of Students	5	7	9	10	8	6	5	2

Answer: 20.48.

2. Calculate the arithmetic mean for the following data

Wages	10	20	30	40	50
No. of employees	4	5	3	2	5

Answer: 29.47.

3. Find the mean for the following Data.

X	9	10	11	12	13	14	15	16	17	18
f	1	2	3	6	10	11	7	3	2	1

Answer: 13.54.

4. Calculate Arithmetic Mean for the following discrete data:

Items	14	36	45	70
Frequency	2	5	1	3

Answer: 42.09

CONTINUOUS SERIES

In continuous series arithmetic mean can be calculated by applying

Direct Method
Short Cut Method
Step Deviation Method

Direct Method

Steps:

Find the Mid value of each class = m

Mid value = $\frac{LL+UL}{2}$ where LL = Lower Class limit.

UL = Upper class limit.

Find $\sum \mathbf{f}$ **= sum of all the frequencies.**
Multiply each mid-value by the corresponding frequency to find out fm
Find out $\sum \mathbf{fm}$
Find the Arithmetic mean by applying the formula $\overline{X} = \frac{\sum \mathbf{fm}}{\sum \mathbf{f}}$

Question 1. Calculate mean for the following frequency distribution of marks.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No of Students	5	12	30	45	50	37	21

Solution:

Marks	Mid –value m	No of Students f	fm
0-10	5	5	25
10-20	15	12	180
20-30	25	30	750
30-40	35	45	1575
40-50	45	50	2250
50-60	55	37	2035

60-70	65	21	1365
		$\sum$ f = 200	$\sum$ fm= 8180

$\overline{X} = \frac{\sum \mathbf{fX}}{\sum f} = \frac{8180}{200} = 40.9$

Hence, the average wage is Rs.40.9 .

Short cut method

In case, Short cut method is used then we apply the following formula to calculate Arithmetic mean.

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

Where A is assumed mean, dx is deviation of mid points from assumed mean i.e.; d

$\sum$ f =Total no. of observations.

STEPS:

Find out mid-values of each class= m
Take an assumed mean A
From the mid value of each class deduct the assumed mean i.e, (m-A) and find out deviations =d
Multiply respective frequencies of each class by these deviations and obtain the total i.e, $\sum$fd
Apply the formula $\overline{X} = A + \frac{\sum \mathbf{fd}}{N}$

Question 1:.Calculate mean for the following frequency distribution of marks.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No of Students	5	12	30	45	50	37	21

Solution

Marks	Mid –value m	d=m-A (A=35)	No of Students f	fd
0-10	5	5-35=-30	5	-150
10-20	15	15-35=-20	12	-240
20-30	25	25-35=-10	30	-300
30-40	35	35-35=0	45	0
40-50	45	45-35=10	50	500
50-60	55	55-35=20	37	740
60-70	65	65-35=30	21	630
			$\sum$ f = 200	$\sum$ fd=1180

$$\overline{X} = A + \frac{\sum \mathbf{fd}}{\sum \mathrm{f}}$$

$$= 35 + \frac{1180}{200}$$
$$= 35 + 5.9$$
$$= 40.9$$

Step Deviation Method

In Step Deviation Method

we apply the following formula to calculate Arithmetic mean.

$$\overline{X} = A + \frac{\sum \mathbf{fd'}}{N} xh$$

Where A is assumed mean, d is deviation of mid points from assumed mean i.e.; d= (m-A),

$\sum$ f =Total no. of observations.**STEPS:**

Find out mid-values of each class= m
Take an assumed mean A
From the mid value of each class deduct the assumed mean i.e, (m-A) and find out deviations =d
Then find d'=d/ h, h is height of the class=UL-LL
Multiply respective frequencies of each class by these d' and obtain the total i.e, $\sum$fd'

Apply the formula $\overline{X} = A + \frac{\sum \mathrm{fd'}}{\sum \mathrm{f}} xh$

A=35, h=height of the class = UL-LL=10-0=10

Marks	Mid –value m	d=m-A (A=35)	$d' = d/10$	No of Students f	fd'
0-10	5	5-35=-30	-3	5	-15
10-20	15	15-35=-20	-2	12	-24
20-30	25	25-35=-10	-1	30	-30
30-40	35	35-35=0	0	45	0
40-50	45	45-35=10	1	50	50
50-60	55	55-35=20	2	37	74
60-70	65	65-35=30	3	21	63
				$\sum f = 200$	$\sum fd' = 118$

$$\overline{X} = A + \frac{\sum fd'}{\sum f} xh$$

$$= 35 + \frac{118}{200} x10$$

$$= 35 + 5.9$$
$$= 40.9$$

PROBLEMS

Question 1. Calculate mean for the following frequency distribution of marks.

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

Answer 43.43

2. Calculate the arithmetic mean for the following data

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No of Students	5	12	30	45	50	37	21

Answer: 40.9

3.. Calculate the arithmetic mean for the following data

Marks	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No of Students	1	2	3	5	6	12	16	10	4

Answer: 66.5

4. Calculate mean from following data

Marks	5-15	15-25	25-35	35-45	45-55	55-65
No of Students	8	12	6	14	7	3

Answer: 31.8

5. Find the value of mean from the following distribution

Class interval	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequencies	8	12	14	16	15	9	6

Answer: 33.625

6. The following table gives the information about the marks obtained by 100 students in an examination.

Class	0-10	10-20	20-30	30-40	40-50
Frequency	12	28	32	25	13

Find the mean marks of the students using assumed mean method.
Answer=24.9

Mean in Case of Unequal Class Interval

Question 1: Find the mean from the following data

Class Interval :	4–8	8–20	20–28	28–44	44–68	68–80
Frequency :	3	8	12	21	10	6

Solution

Let Assumed mean (A) be = 36 ; And Class Interval (i) = 2 (For IInd Method)

C.I.	M.P. m	f	fm	m–A =dx	fdx	$d'x = \frac{dx}{2}$	$fd'x$
4–8	6	3	18	–30	–90	–15	–45
8–20	14	8	112	–22	–176 –410	11	–88 –205
20–28	24	12	288	–12	–144	–6	–72
28–44	36	21	756	0	0	0	0
44–68	56	10	560	20	+200 +428	10	100 214
68–80	74	6	444	38	+228	19	114
		N = 60	Σfm =2178		Σfdx = 18	$\Sigma fd'x$ =9	$\Sigma fd'x$ =9

(i) By Direct Method

$$\bar{X} = \frac{\Sigma fm}{N} = \frac{2178}{60} = 36.3$$

(ii) By Short Cut Method

$$\bar{X} = A + \frac{\Sigma fdx}{N} = 36 + \frac{18}{60} = 36 + .3 = 36.3.$$

(iii) By Step-deviation Method

$$\bar{X} = A + \frac{\Sigma fd'x}{N} \times i = 36 + \frac{9}{60} \times 2 = 36 + \frac{9}{30} = 36 + .3 = 36.3.$$

A=36

CLASS INTERVAL	m (2)	f (3)	fm (2)X(3)	d=m-A (4)	fd (3)X(4)	d'=d/2 (c=2) (5)	fd' (3)X(5)
4-8	6	3	18	6-36=-30	-90	-15	-45
8-20	14	8	112	14-36=-22	-176	-11	-88
20-28	24	12	288	24-36=-12	-144	-6	-72
28-44	36 = A	21	756	36-36=0	0	0	0
44-68	56	10	560	56-36=20	200	10	100
68-80	74	6	444	74-36=38	228	19	114

		$\sum$ f = 60	$\sum$ fm=2178		$\sum$fd = 18		$\sum fd' = 9$

Direct Method

$$\overline{X} = \frac{\sum \mathbf{fm}}{\sum \mathrm{f}} = \frac{2178}{60} = 36.3$$

Short cut Method

$$\overline{X} = A + \frac{\sum \mathbf{fd}}{\sum \mathrm{f}} = 36 + \frac{18}{60} = 36 + .3 = 36.3$$

Step Deviation Method

$$\overline{X} = A + \frac{\sum \mathbf{fd'}}{\sum \mathrm{f}} xc = 36 + \frac{9}{60} \mathrm{x} 2 = 36 + .3 = 36.3$$

PROBLEMS

1. From the following data calculate the mean

Income	0-10	10-30	30-60	60-100	100-150	150-210
No of Persons	4	6	8	10	4	8

Answer: Rs.81

2. Find Mean for the following data

X :	Less than 20	20–30	30–50	50–90	90–120	above 120
f :	2	8	20	25	16	9

.

Answer= 70.---

INCLUSIVE CLASS INTERVAL

Question 1: Find the mean from the following data

Class Interval :	4–6	7–9	10–12	13–15	16–18	19–21	22–24
Frequency :	1	3	7	15	11	3	2

Solution

As the difference between upper limit of one interval and lower limit of next interval is one, we will deduct and add .5 from lower limit and to upper limit of every interval to make intervals exclusive.

Let Assumed mean (A) be = 14 ; $i = 3$

C.I.	C.I. Improved	Mid Points m	Frequency f	Step deviation $d'x$	$fd'x$
4–6	3.5–6.5	5	1	–3	–3
7–9	6.5–9.5	8	3	–2	–6 –16
10–12	9.5–12.5	11	7	–1	–7
13–15	12.5–15.5	14	15	0	0
16–18	15.5–18.5	17	11	1	11
19–21	18.5–21.5	20	3	2	6 +23
22–24	21.5–24.5	23	2	3	6
			N = 42		$\Sigma fdx = 7$

As $\bar{X} = A + \frac{\Sigma fd'x}{N} \times i$

$\therefore \bar{X} = 14 + \frac{7}{42} \times 3$

$= 14 + .5 = \mathbf{14.5}$

CI	Improved CI	m	d= m-A	d' = d/ h	f	fd'
4-6	3.5 – 6.5	5	5-14=-9	-3	1	

7-9	6.5 – 9.5	8	8-14=-6	-2	3	-6
10-12	9.5 -12.5	11	11-14=-3	-1	7	-7
13-15	12.5 – 15.5	14	14-14-=0	0	15	0
16-18	15.5 – 18.5	17	17-17=3	1	11	11
19-21	18.5 – 21.5	20	20-14=6	2	3	6
22-24	21.5 -24.5	23	23-14=9	3	2	6
					∑ f=42	∑ fd'=7

Correction factor= (LL of next class – U L of 1st class) / 2=(7-6)/2=1/2=.5

LL-CF and UL +Cf

Step Deviation Method

$$\overline{X} = A + \frac{\sum \mathbf{fd'}}{\sum \mathrm{f}} xh$$

$$= 14 + \frac{7}{42}\text{x}3$$

=14+0.5

=14.5

Problem 1.

Marks	1-7	8-14	15-21	22-28	29-35	36-42	43-49
No. of students	7	11	13	20	14	12	9

2.. Calculate the mean from the following data:

Thickness	4 -9	10 - 15	11 - 16	17 - 22	23 -28	29 -35
Frequencies	9	18	34	20	15	7

Answer:

3. Calculate average wages from the following data

Wages	10-19	20-29	30-39	40-49	50-59	60-69	70-79
No of Persons	5	2	3	10	3	2	5

Answer: 44.5

Mean in Open End Classes

Question 1: Find mean for the following distribution

Class Intervals :	Below 20	20–30	30–40	40–50	50–60	60–70	Above 70
Frequency :	6	10	24	28	14	5	3

Solution

As each class interval has length of 10 units. Hence lowest class interval will be 10–20 and highest as 70–80 and Let Assumed Mean (A) be = 45 ; i = 10

C.I.	M.P. m	f	d′x	fd′x	
10–20	15	6	–3	–18	
20–30	25	10	–2	–20	–62
30–40	35	24	–1	–24	
40–50	45	28	0	0	
50–60	55	14	1	14	
60–70	65	5	2	10	33
70–80	75	3	3	9	
		N = 90		$\Sigma fd'x = -29$	

$$\text{As } \bar{X} = A + \frac{\Sigma fdx'}{N} \times i$$

$$\therefore \bar{X} = 45 + \frac{-29}{90} \times 10$$

$$= 45 + \frac{-29}{9}$$

$$= 45 - 3.22 = \mathbf{41.78.}$$

PROBLEMS:

1. Calculate the arithmetic mean for the following

Wages	Below 10	10-20	20-30	30-40	40-50	50-60	60-70
No of Persons	5	2	3	10	3	2	5

Answer= 35

2. Calculate the arithmetic mean for the following data

Marks	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	Above 90
No of Students	1	2	3	5	6	12	16	10	4

Answer: 66.5

CUMULATIVE FREQUENCY DISTRIBUTION

When the data is given in form of more than or less than, above or below for all items, in the series, it is called as cumulative frequency.

Question 1: Find the arithmetic mean for the following data

Class Interval :	Below 50	Below 60	Below 70	Below 80	Below 90	Below 100
Frequency :	3	11	34	59	72	80

Solution

Let Assumed Mean (A) be = 75 ; $i = 10$

C.I.	Mid point m	f	$d'x$	$fd'x$
40–50	45	= 3	–3	–9
50–60	55	11–3 = 8	–2	–16 –48
60–70	65	34–11 = 23	–1	–23
70–80	75	59–34 = 25	0	0
80–90	85	72–59 = 13	1	13 +29
90–100	95	80–72 = 8	2	16
		N = 80		$\Sigma fd'x = -19$

As $\bar{X} = A + \frac{\Sigma fd'x}{N} \times i$

$\therefore \bar{X} = 75 + \frac{-19}{80} \times 10$

$= 75 - \frac{19}{8}$

$= 75 - 2.375 = 72.625.$

Class	Less than CF	Improved Class	f
Below 50	3	40-50	3
Below 60	11	50-60	11-3=8
Below 70	34	60-70	34-11=23
Below 80	59	70-80	59-34=25
Below 90	72	80-90	72-59=13
Below 100	80	90-100	80-72=8

PROBLEMS

1. Find the arithmetic mean for the following data

Intervals :	Above 40	Above 50	Above 60	Above 70	Above 80	Above 90
Frequency :	80	77	69	46	21	8

Class	More than	Improved	m	d=m- A	d'=d/h	f	fd'

	CF	Class			h=10		
above 40	80	40-50	45	-30	-3	80-77=3	-9
above 50	77	50-60	55	-20	-2	77-69=8	-16
above 60	69	60-70	65	-10	-1	69-46=23	-23
above 70	46	70-80	75=A	0	0	46-21=25	0
above 80	21	80-90	85	10	1	21-8=13	13
above 90	8	90-100	95	20	2	8	16
						$\sum$ f=80	$\sum$ fd'=-19

$$\overline{X} = A + \frac{\sum \mathbf{fd'}}{\sum \text{f}} xh$$

$$\overline{X} = 75 + \frac{(-\mathbf{19})}{\mathbf{80}} x10$$

$$\overline{X} = 75 + \frac{(-\mathbf{19})}{\mathbf{8}}$$

$$\overline{X} = 75 - 2.375$$

$$\overline{X} = 72.625$$

Answer= 72.625

2.

Wages in000' Rs.	Less than 10	Less than 20	Less than 30	Less than 40	Less than 50	Less than 60	Less than 70	Less than 80	Less than 90	Less than 100	Less than 110
No. of wrokers	1	17	55	113	173	219	241	256	271	278	290

3. From the following distribution of marks of Economics calculate Arithmetic mean

Marks (More than)	10	20	30	40	50	60
No of Persons	100	92	80	40	20	6

Answer= 39

4. The following table gives the life time in hours of 400 electrical bulbs of certain make. Find mean life time of these bulbs

Life time (Less than)	300	400	500	600	700	800	900	1000	1100	1200
No. of electric bulbs	0	20	60	116	194	265	324	374	392	400

Answer: 38.8 or 39

WHEN MID POINTS ARE GIVEN

When middle points are given, we convert it into exclusive series noting the difference between each midpoint.

Question 1: Calculate the arithmetic mean for the following

Mid points :	6	10	14	18	22	26	30
Frequency :	2	7	18	29	17	11	6

Solution

Difference between each mid point = 4.

∴ Deducting and adding $2 = \left(\frac{4}{2}\right)$ to each mid point ; We get the intervals as below.

C.I.	Mid Point m	Frequency f	Step-deviation d 'x	fd 'x	
4–8	6	2	–3	–6	
8–12	10	7	–2	–14	–38
12–16	14	18	–1	–18	
16–20	18	29	0	0	
20–24	22	17	1	17	
24–28	26	11	2	22	+57
28–32	30	6	3	18	
		N = 90		Σfd 'x = 19	

Let Assumed Mean (A) be = 18 ; $i = 4$.

$$\bar{X} = A + \frac{\Sigma fd'x}{N} \times i$$

$$= 18 + \frac{19}{90} \times 4 = 18 + \frac{76}{90}$$

= 18+.84 = **18.84.**

Note. It is not necessary to obtain class intervals, but it is necessary to do if lower limit is needed to obtain Median Decile, Percentile or Mode.

Mid Point m	d	**d'**	**f**	**fd'**

PROBLEMS

Find mean for the following data

Mid values	5	20	45	80	125	180
frequency	4	7	10	16	8	5

Answer: 75.8

Calculate mean for the following data

Mid values	349.5	449.5	549.5	649.5	749.5	849.5	949.5	1049.5	1149.5
frequency	20	40	56	75	71	59	50	18	8

Answer: 713.25

WHEN CLASS INTERVALS ARE UNEQUAL

Question 1: Find the following arithmetic mean for the given data
PEOBLEMS
Find the average marks of a student

Marks	Below 10	10-30	36-60	60-100	100-150	150 and above
No. of students	4	7	10	16	8	2

Answer: 75.8

1. A class teacher has the following absentee record of students:

No. of days	0-6	6-10	10-14	14-20	20-28	28-30	38- 40
No. of students	11	10	7	4		3	1

Answer: 12.175 or appox. 12

MEAN IN CASE OF MISSING FREQUENCY

Find Missing frequency if mean is 30.5

Marks	10	20	30	40	50
No. of students	8	10	?	15	7

Marks X	**d=** X -A	d'=d/10	f	fd'
10	-20	-2	8	-16
20	-10	-1	10	-10
30 = A	**0**	0	F	0
40	**10**	1	15	15
50	**20**	2	7	14
			$\sum f = 40 + F$	$\sum fd' = 3$

$$\overline{X} = A + \frac{\sum fd'}{\sum f} \times c$$

$$30.5 = 30 + \frac{3}{(40+F)} \times 10$$

$$30.5 - 30 = \frac{3}{(40+F)} \times 10$$

$$0.5 = \frac{3x10}{(40+F)}$$

0.5X(40+F) =30
20 + 0.5F = 30

0.5 F = 30 – 20 =10

F = 10/0.5=20

Find the missing frequency if mean is 50

Wages '000 Rs	**20**	**30**	**40**	**50**	**60**	**70**	**80**
No . of workers	**5**	**2**	**3**	**?**	**3**	**2**	**5**

Question 1: Calculate the number of students against class 30-40 of the following data where

$\overline{X}$= 28

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	12	18	27	?	17	6

Solution:

Marks	Mid values	D= m-25	d′ = d/10	No of students f	fdx′
0-10	5	-20	-2	12	-24
10-20	15	-10	-1	18	-18
20-30	25	0	0	27	0
30-40	35	10	1	F_1	F_1
40-50	45	20	2	17	34
50-60	55	30	3	6	18
				$\sum f$ = 80+F_1	$\sum fd'$ $= 10 + F1$

Here A=25; h=10

$\overline{X} = 28$, $\sum f$ = 80+F_1

$\sum fd' = 10 + F_1$

$$\overline{X} = A + \frac{\sum fd'}{\sum f} \times h$$

$$28 = 25 + \frac{(10 + F1)}{(80 + F1)} \times 10$$

$$28\text{-}25 = \frac{10+F1}{80+F1} \times 10$$

$$3 = \frac{10+F1}{80+F1} \times 10$$

$$3x(80+F_1) = (10+F_1)x10$$

$$240 + 3f_1 = 100 + 10f_1$$

$$240\text{-}100 = 10F_1 - 3F_1$$

$$140 = 7F_1$$

$$7F_1 = 140$$

$$\therefore F_1 = \frac{140}{7} = 20$$

Hence 20 students are against 30-40 class intervals.

Find missing frequency if $\overline{X} = 35$

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70		
No.of students	5	2	?	10	3	2	5		

Solution

Marks	m	d=m-25	d'=d/10	f	fd'
0-10	5	-20	-2	5	-10
10-20	15	-10	-1	2	-2
20-30	25 = A	0	0	F_1	0
30-40	35	10	1	10	10
40-50	45	20	2	3	6
50-60	55	30	3	2	6
60-70	65	40	4	5	20
				$\sum f=27+F_1$	$\sum fd'=30$

$$\overline{X} = A + \frac{\sum fd'}{\sum f} \times h$$

$$35 = 25 + \frac{30}{27 + F1} \times 10$$

$$35 - 25 = \frac{30X10}{(27 + F1)}$$

$$10x(27 + F1) = 300$$

$$270+10F_1 = 300$$

$$10F_1 = 300-270=30$$

$$F_1 = 30/10=3$$

find the missing frequencies if

mean is 1.46 and N= 200

No. of accidents	0	1	2	3	4	5

Frequency	46	?	?	25	10	5

Solution

Given N= 200

N= 46 +F_1 + F_2 +25 +10+5 = 200

86 + F_1 + F_2 = 200

F_1 + F_2 = 200 – 86=114

F_2 = 114 - F_1

X	f	fX
0	46	0
1	F1	F1
2	114 - F_1	228-2 F_1
3	25	75
4	10	40
5	5	25
	∑f = 200	∑fX $= 368 - F1$

$$\overline{X} = \frac{\sum fX}{\sum f}$$

$$1.46 = \frac{368 - F1}{200}$$

1.46x200 = 368-F_1

292 = 368-F_1

F_1 = 368-292 = 76

$F_2 = 114 - F_1 = 114 - 76 = 38$

PROBLEMS

1. From the following data calculate arithmetic mean for the following find the missing item . $\overline{X} = 219.1$

Wages	210	212	213	217	x	225	228	230
No. of workers	5	2	7	6	5	9	4	2

Answer = 220

2. Find missing frequency $\overline{X}$=28. AM=25

X	0-10	10-20	20-30	30-40	40-50	50-60
f	12	18	27	?	17	6

Answer: 20

3. Calculate missing frequency when $\overline{X}$ = 35 and N= 25

X	10-20	20-30	30-40	40-50	50-60
f	6	a	5	b	4

Answer: 3, 7

CORRECT MEAN

Question 1: The mean wage of 100 workers per day was found to be Rs. 80. But later on, it was found that the wages of two labourers Rs. 93 and Rs. 59 were misread as Rs. 39 and Rs. 95. Find correct mean wage per day.

Solution:

We are given N = 100, $\overline{X}$ = 80

Correct items = 93 and 59

Incorrect items = 39 and 95

Since $\overline{X} = \frac{\sum X}{N}$ $\sum X = N.\overline{X}$ = 100(80) = 8000

But this is Incorrect $\sum X$

Correct$\sum X$ = Incorrect $\sum X$ - Wrong Items + Correct Items

Correct$\sum X$ = 8000 – (39 + 95) + (93 + 59)

= 8000 – 134 +152 = 8000 + 18 = 8018

Correct $\overline{X} = \frac{\text{Correct}\sum X}{N} = \frac{8018}{100} = 80.18$

Hence correct mean wage per day = Rs. 80.18

PROBLEMS

1. The mean salary of 500 employees per day was found to be Rs. 150. But later on, it was found that the wages of two employees Rs. 104 and Rs. 117 were misread as Rs. 140 and Rs. 171. Find correct mean wage per day.
   Answer: Rs. 149.82
2. The mean weight of 25 boys was calculated to be Rs. 78.4. But later on, it was found that the weight of one boy was misread as 69 lbs instead of 96 lbs. Find correct average.
   Answer: 79.48 lbs
3. The arithmetic mean of 100 students is given as 45 marks. But later it was found that the marks of students getting 64 were wrongly written as 46. Find correct mean.
   Answer: 45.18 marks

COMBINED AVERAGE

$$\overline{X}_{12....k} = \frac{N_1\overline{X}_1 + N2\overline{X}_2 + N_3\overline{X}_3 + ...+ N_K\overline{X}_K}{N_1 + N2 + N3 + ...+ N_k}$$

Where $\overline{X}_{12}$ = Combined mean

$\overline{X}_1$, $\overline{X}_2$.... = A.M of First, Second….Groups

N_1, $N_{2...}$= Number of items in first, second…..groups

Question 1: The mean height of 25 male workers in a factory is 61 cm, and the mean height of 35 female workers working in the same factory is 58 cm. Find the combined mean height of 60 workers in the factory.

Solution:

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

$N_1 = 25$, $N_2 = 35$

$\overline{X}_1 = 61$, $\overline{X}_2 = 58$

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

$$\bar{X}_{12} = \frac{(25\times61)+(35\times58)}{25+35} = \frac{1525+2030}{60} = \frac{3555}{60} = 59.25$$

Thus the combined mean height of 60 workers is 59.25

PROBLEMS

1. The following table shows certain data collected for the three regions of a country:

Region	No. of inhabitants (in millions)	Percent of literates	Average annual income per person (Rs.)
A	10 =N1	52	8000= X1
B	5 =N2	68	5800 =X2
C	20= N3	41	7500 =X3

   Answer: Rs. 7400

2. The average expenditure of 100 laborers Rs.20. If average expenditure of 150 laborers is Rs.25, Calculate the average expenditure of remaining 50 workers.
   Answer: Rs.35

3. 50 Students took up a test. The result of those is given below.

Marks	4	5	6	7	8	9

No of Students	8	11	8	6	4	3

If the average of all the 50 students were 5.16, find the average marks of those who failed.

Answer: 2.2 Marks.

4. 100 Students took a test. The result of those who secured less than 60% marks is given as

Marks	0-20	20-40	40-60
Students	16	24	40

If the average mark of students were 50, find out those who scored 60 or more than 60% marks.

Answer: 106.

WEIGHTED ARITHMETIC MEAN

The formula for computing weighted arithmetic mean is given below:

$$\bar{X}w = \frac{\sum WX}{\sum W},$$

where,

$\bar{X}_w$ **=The weighted arithmetic mean**

X = The variable.

W = Weights attached to the variable X.

Ch 4_37

Steps for Calculating:

- Denote Variants as X.

- Weights as W
- Multiply X by W and get WX
- Sum up weights and get $\sum \mathbf{WX}$
- Apply the formula $\overline{X} = \frac{\sum \mathbf{WX}}{\sum \mathbf{W}}$

Question 1: A student obtained the marks 40, 50, 60, 80, and 45 in math, statistics, physics, chemistry and biology respectively. Assuming weights 5, 2, 4, 3, and 1 respectively for the above mentioned subjects, find the weighted arithmetic mean per subject.

Solution:

Subject	**Mark Obtained** x	**Weight** w	wx
Math	40	5	200
Statistics	50	2	100
Physics	60	4	240
Chemistry	80	3	240
Biology	45	1	45
Total		∑w=15	∑wx=825

Now we will find the weighted arithmetic mean as:

$\overline{X}w = \frac{\sum \mathbf{WX}}{\sum \mathbf{W}} = \frac{\mathbf{825}}{\mathbf{15}}$ = 55 marks/subject.

PROBLEMS

1. A student gets the following percentage in Bsc : English: 75, Hindi :90 , Maths: 70, Physics: 80, Chemistry : 60,Find out the weighted mean if the weights are 1,1,3,2,3.

Answer: 71.5

2. A candidate obtains the following percentage , English : 60, Hindi: 75, Maths: 63,Physics : 60, Chemistry: 55.Find out the weighted mean if the weights are 1,1,2,3,3.
   Answer: 60.6.

Geometric Mean

- Individual Series

$$G = (x_1.x_2.x_3.....x_n)^{1/n}$$

$$\log G = \frac{1}{n}(\log x_1 + \log x_2 + ... + \log x_n)$$

$$G = \text{antilog}\left(\frac{1}{n}\Sigma \log x\right)$$

Steps:

- Take log of all given items of the data
- Add the log values i.e, find $\sum logX$
- Divide $\sum logX$ by the number of items (N) is a given problem.
- Read the antilog value table. This will give us the value of g which is the required geometric mean.

Question 1: Calculate Geometric Mean for the following individual data:

Items	14	36	45	70	105

Solution:

Based on the given data, we have:

X	Logx
14	1.1461
36	1.5563
45	1.6532
70	1.8450
105	2.0211
Total	8.2217

G.M = Antilog of $\frac{\sum logX}{n}$

= Antilog of $\frac{8.2217}{5}$ = Antilog of 1.6443 = 44.09

PROBLEMS

1. Calculate geometric mean for the following data:

X	10	110	135	120	50	59	60	7

Answer: 46.56

2. Calculate geometric mean for the following data:

x	10	52	110	120	50	80	60	37

Answer: 52.85

3. From the following data calculate geometric mean

X	135	79	38	286	59	176	7

Answer: 70.23

4. Calculate GM for the following

X	15	70	85	75	500

Answer: 80.34

CALCULATION OF GEOMETRIC MEAN – DISCRETE SERIES

- Discrete Frequency Distribution

$$G = (x_1^{f1}.x_2^{f2}.....x_n^{fn})^{1/N}$$

$$\log G = \frac{1}{N}(f_1\log x_1 + f_2\log x_2 + ...+x_n\log f_n)$$

$$G = \text{antilog} \left(\frac{1}{N} \Sigma f_i \log x_i \right)$$

STEPS:

- Take log values of all items of a given series
- Multiply each log values to their respective frequencies
- Add the values and divide by the total number of frequencies i.e,$\sum \boldsymbol{X} = \boldsymbol{N}$.
- Take the values of antilog from antilog table and result would be geometric mean.

Question1: Find the geometric mean from the following data

Variable	2	3	5	6	4
Frequencies	10	15	18	12	7

Solution:

Variable	Frequencies	logX	f × $logX$
2	10	0.3019	3.0100
3	15	0.4771	7.1565
5	18	0.6990	12.5820
6	12	0.7782	9.3384
4	7	0.6021	4.2147
	N = 62		$\sum f \times logX$ = 36.3016

The formula for geometric mean is

g = Antilog $\left[\frac{\sum f\ logX}{N}\right]$

= Antilog $\left[\frac{36.3016}{62}\right]$

= Antilog [0.5855] = 3.850

Thus G.M is g = 3.850

PROBLEMS

1. Compute Geometric mean

Variable	5	15	25	35	45
Frequencies	5	7	15	25	8

Answer: 25.63

2. Compute Geometric mean

Variable	10	15	18	25

Frequencies	2	3	5	4

Answer: 18.2

3. Calculate Geometric mean for the following data

Variable	5	15	25	35	45
Frequencies	7	12	15	11	8

Answer: 23.59

CALCULATION OF GEOMETRIC MEAN – CONTINUOUS SERIES

In case of continuous series, the following steps are observed.

- Take the mid-value of each of the following class interval
- Take log values of all mid-values of a given series
- Multiply each log values to their respective frequencies
- Add the values to find $\sum f\ logX$
- Divide by the total number of frequencies i.e.,$\sum \boldsymbol{f = N.}$

Apply formula g = Antilog $\left[\frac{\sum f\ logX}{N}\right]$

Question 1: Calculate geometric mean for the following

Variable	10-20	20-30	30-40	40-50	50-60
Frequencies	5	10	15	7	4

Solution:

Variable	Frequencies	Mid- values	logX	f × $logX$
10-20	5	15	1.1761	5.8805
20-30	10	25	1.3979	13.9790
30-40	15	35	1.5441	23.1615
40-50	7	45	1.6532	11.5724
50-60	4	55	1.7404	6.9616
	N = 41			$\sum f\ logX = 61.5550$

g = Antilog $\left[\frac{\sum f\ logX}{N}\right]$

= Antilog $\left[\frac{61.5550}{41}\right]$ = Antilog [1.5013] = 31.72

The geometric mean is 31.72

PROBLEMS

1. Calculate geometric mean for the following

Variable	10-20	20-30	30-40	40-50	50-60
Frequencies	5	10	15	7	4

Answer = 31.72

2. The following distribution relates to marks in Accounts of 60 students

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No. of Students	3	8	15	20	10	4

Answer: 28.02

3. Find the geometric mean of the following grouped data for the frequency distribution of weights.

Weights of ear heads (g)	No of ear heads (f)
60-80	22
80-100	38
100-120	45
120-140	35
140-160	20
Total	160

Answer= 106.23

4. A record of dividend declared (in percentage) in the past four years by 20 companies is as follows:

Dividend declared (in %age)	**0-10**	**10-20**	**20-30**	**30-40**
No. of companies	5	8	3	4

Answer: 14.8

WEIGHTED GEOMETRIC MEAN

STEPS:

- Take log values of all the items of given series
- Multiply each of the values to its respective weights
- Add these values and divide by total weights
- Take antilog of these values and the result would be weighted geometric mean

Apply the formula

$$g = \text{Antilog}\left[\frac{\sum W\ logX}{\sum W}\right]$$

Question1: Find the weighted arithmetic mean for the following data

X	8	10	52	25	37
W	5	3	4	2	1

Solution:

X	W	log X	W × log X
8	5	0.9.31	4.5155
10	3	1.0000	3.0000
52	4	1.7160	6.8640
25	2	1.3979	2.7959
37	1	1.5682	1.5682

	$\sum W$ = 15		$\sum W$ logX = 18.7435

Geometric mean g = Antilog $\left[\frac{\sum W log X}{\sum w}\right]$ = Antilog $\left[\frac{18.7435}{15}\right]$ = 17.74.

PROBLEMS:

1. Find the weighted Geometric Index number from the following:

X:	260	120	180	230	220	200
W:	46	12	10	20	8	4

Answer: 215.6.

2. Find the weighted Geometric Index from the following table given below:

X:	5	8	44	160	500
W:	10	9	3	2	1

Answer: 12.20.

HARMONIC MEAN

Individual Series

STEPS:

- Take reciprocal of the items
- Add the reciprocal values
- Number of items (N) divideD by the SUM OF RECIPROCALS
- The result would be Harmonic mean of the given series.

The formula of Harmonic mean is

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

Question1: Find the Harmonic mean for the following data

Values	35	45	89	76	87	66	110	135

Values (X)	Reciprocals $\left(\frac{1}{X}\right)$
35	1/35=0.0286
45	1/45=0.0222
89	1/89=0.0112
76	1/76=0.0131
87	1/87= 0.0115
66	1/66=0.0151
110	1/110=0.0031
135	1/135=0.0074
N = 8	$\Sigma\left(\frac{1}{X}\right)$ = 0.1184

Formula of Harmonic mean is $\mathbf{H.M} = \frac{\mathbf{N}}{\Sigma\left(\frac{1}{x}\right)} = \frac{8}{\mathbf{0.1184}}$ = 67.57

PROBLEMS:

1. Calculate Harmonic mean for the following data

x	10	20	40	60	120

Answer: 25

2. Find Harmonic mean for the following

Values	2574	475	75	5	0.8	0.08	0.005	0.009

Answer: 0.006

3. Compute harmonic mean for the following data

Values	3834	382	63	0.8	0.4	0.03	0.009	0.0005

Answer: 0.0037

4. Find harmonic mean

x	10	68	40	60	38

Answer: 27.37

Discrete Series

STEPS:

- Find the reciprocal of the given data
- Multiply each reciprocal value by its frequencies.
- Add the products obtained
- Divide the result by the total frequencies
- Take its reciprocal which is required for H.M

The formula is

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

Question 1: Find the Harmonic mean for the following data

X	10	20	40	60	120
F	1	3	6	5	4

Solution:

X	F	f/X
10	1	1/10=0.1000
20	3	3/20=0.1500
40	6	6/40=0.1500
60	5	5/60=0.0833
120	4	4/120=0.0333
	$\Sigma f = 19$	$\Sigma(f/X) = 0.5166$

$$\mathbf{H.M} = \frac{\mathbf{N}}{\Sigma\left(\frac{f}{x}\right)} = \frac{\mathbf{19}}{\mathbf{0.5166}} = 36.78$$

PROBLEMS:

1. Calculate the harmonic mean for the following data:

x	1	3	5	7	9	11
f	2	4	6	8	10	12

Answer: 5.331

2. Calculate Harmonic Mean for the following data:

Items	14	36	45	70	105
Frequency	2	5	1	3	2

Answer: 13.67

3. Calculate Harmonic mean for the following

X	10	11	12	13	14
F	5	8	10	9	6

Answer: 11.94

4. Calculate H.M

X	10	20	25	40	50
F	20	30	50	15	5

Answer: 20.08

Continuous series

STEPS:

- Take the mid values of all class intervals
- Find the reciprocals of the mid values
- Multiply the reciprocal values with respective frequencies of the class interval
- Divide N by the sum of the product obtained. This is required H.M

Use the formula $\mathbf{H.M} = \frac{\mathbf{N}}{\Sigma\left(\frac{f}{x}\right)}$

Question1: Calculate H.M for the following

X	0-10	10-20	20-30	30-40	40-50
F	2	7	13	5	3

Solution:

Marks CI	Mid- values m	No. of students f	f/m
0-10	5	2	⅖=0.400
10-20	15	7	7/15=0.467
20-30	25	13	13/25=0.520
30-40	35	5	5/35=0.143
40-50	45	3	3/45= 0.067
		∑f= 30	∑f/m = 1.597

$$\mathbf{H.M} = \frac{\mathbf{N}}{\Sigma\left(\frac{f}{x}\right)} = \frac{\mathbf{30}}{\mathbf{1.597}} = 18.79$$

PROBLEMS

1. Calculate Harmonic Mean for the following continuous data:

Items	0-10	10-20	20-30	30-40
Frequency	2	5	1	3

Answer: 12.80

2. Find the harmonic mean of the following distribution of data

Dividend yield (percent)	2 – 6	6 – 10	10 – 14
No. of companies	10	12	18

Answer: 7.27

3. Calculate Harmonic mean for the following data

Marks X	0-10	10-20	20-30	30-40	40-50
No of Students f	4	6	10	7	3

Answer: 16.07

4. Find out the H.M from the following data

Class Intervals	10-20	20-30	30-40	40-50	50-60
Frequency f	4	6	10	7	3

Answer: 29.88

.

UNIT 4

MEASURES OF DISPERSION

Dispersion

Dispersion or spread is the degree of the scatter or variation of the variables about a central value. Dispersion is the measure of variability.

Range

R= L -S

Coefficient of Range $=\frac{L-S}{L+S}$

Range is the simplest measure of dispersion. Range is the difference between the extreme values. it is the difference between largest value and smallest value

The difference between the highest value and lowest values of a series is known as range.

$$\textbf{Range} = \text{largest value} - \text{smallest value}$$

$$\mathbf{R} = L - S$$

$$\textbf{Coefficient of Range} = \frac{L-S}{L+S}$$

only variable is important. Ignore frequency

Individual Series

Question: Find the range and coefficient of range of the following data: 25, 67, 48, 53, 18, 39, 44

Solution:

Arrange in order

18, 25, 39, 44, 48, 53, 67

largest value = 67; smallest value =18

Range R = L -S = 67 −18 = 49

Coefficient of range = $\frac{L-S}{L+S}$

Coefficient of range = $\frac{67-18}{67+18}$

$= \frac{\mathbf{49}}{\mathbf{85}}$

= 0.576

Discrete Series

Question: Find range and its coefficient for the following

X	3	4	5	6	7	8	9	10
F	35	30	20	10	6	3	2	1

Solution:
Calculation of Range and its coefficient
largest value = 10; smallest value =3 (do not consider frequnecy)
Range R = L -S = 10 −3 = 7
Coefficient of range = $\frac{L-S}{L+S}$
Coefficient of range = $\frac{10-7}{10+7}$
= $\frac{3}{17}$
= 0.54

Continuous Series

Question: Find range and its coefficient for the following

Marks:	5-10	10-15	15-20	20-25	25-30	30-35
No.of Students	6	11	7	4	3	1

Solution:
largest value = 35 (highest Upper limit);
smallest value =5 (lowest lower limit) (do not consider frequnecy)

Range R = L -S = 35 −5 = 30
Coefficient of range = $\frac{L-S}{L+S}$
Coefficient of range = $\frac{35-5}{35+5}$
= $\frac{30}{40}$

= 0.75

Individual series
120,170,240.100.105,205,300,160,150,180
Arrange in order
100,105,120,150,160,170,180,205,240, 300

L=300, S= 100

R= L -S=300-100=200

Coefficient of Range $=\frac{L-S}{L+S}=\frac{300-100}{300+100}=\frac{200}{400}=$ 0.5

DICRETE SERIES

Marks	0	1	2	3	4	5	6	7	8
No. of students	1	5	7	15	25	10	6	5	4

L= Largest mark = 8

S = Smallest mark = 0(Ignore frequency)

R= L-S=8-0=8

Coefficient of rang e $=\frac{8-0}{8+0}=1$

Continuous series

Size	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of items	4	6	20	32	33	17	8	2

L= 80, S =0

PROBLEMS

1. Find range and its coefficient for the following

232	228	230	218	240	255	275	285

   Answer: 67 and 0.133

2. Find range and its coefficient 63, 89, 98, 125, 79, 108, 117, 68
   Answer: 62 and 0.33

3. Find range and its coefficient 43.5, 13.6, 18.9, 38.4, 61.4, 29.8
   Answer: 13.6 and 0.64

4. Find range and its coefficient for the following

Wages	5	7	8	9	10	11	12	15
No. of workers	10	15	17	80	25	20	10	4

   Answer: 9, 0.43

5. Find range and its coefficient for the following

X	10	20	30	40	50	60
f	2	4	6	8	10	12

   Answer: 50, 0.714

6. Find range and its coefficient for the following

Marks	10-20	20-30	30-40	40-50	50-60
No. of students	15	18	20	9	6

   Answer: 50, 0.71

7. Find range and its coefficient for the following

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of students	3	5	16	18	12	7	4

   Answer: 70, 1

8. Find range and its coefficient for the following

Marks	20-29	30-39	40-49	50-59	60-69
No of students	8	12	20	7	3

   Answer: 50, 0.56

9. Find range and its coefficient for the following

Wages	5	10	15	20	25	30	35	55
No. of workers	10	15	17	80	25	20	10	4

Answer: 50, 0.833

10. Find range and its coefficient for the following

Income	1200	1300	1400	1500	1600	1800	2000
No. of people	4	6	15	12	7	4	2

Answer: 800, 0.25

Quartile Deviation

Interquartile deviations or Quartile range = $Q_3 - Q_1$

The Quartile deviation gives the average amount by which the two quartiles differ from medians. For a normal distribution we have: **Q3 –M = M – Q1**

Where M = Median, Q1 = Lower Quartile and Q3 = Upper Quartile.

Semi Inter range or Quartile deviation = $\frac{Q_3 - Q_1}{2}$

Coefficient of Quartile deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$.

Calculation of Quartile Deviation

5. Calculate Quartile deviation and its coefficient

Months	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**July**	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**
Sales (in Rs.)	**55**	**60**	**70**	**90**	**90**	**110**	**120**	**130**	**145**	**145**	**155**	**170**

Solution:

Sales are arranged in ascending order as follows:

X Sales (in lakhs): 55 60 70 90 90 110 120 130 145 145 155 170.

Thus after sorting the values in ascending order.

Q1 = Size of the $\frac{N+1}{4}$ th term = Size of the $\frac{12+1}{4}$ th term

= $\frac{13}{4}$ th term = 3.25^{th} term.

= 70 + $\frac{1}{4}$ (90-70) = 70+5 =75.

Q3 = Size of the $\frac{3(N+1)}{4}$ th term = Size of the $\frac{3(12+1)}{4}$th term =

$\frac{39}{4}$th term = 9.75 item.

= Size of the 9^{th} item + ¾ (10th item – 9^{th} item) =

145 + ¾(145-145) = 145 + 0 = 145.

Q.D = $\frac{Q_3 - Q_1}{2}$ = $\frac{145 - 75}{2}$ = 35.

Coefficient of Quartile deviation =$\frac{Q_3 - Q_1}{Q_3 + Q_1}$. = $\frac{145-75}{145+75}$ = $\frac{70}{220}$ = 0.318 Ans.

Calculation of Quartile Deviation – Discrete Series

7. Calculate Quartile deviation and its coefficient from the following:

Age in Years:	**50**	**51**	**52**	**53**	**54**	**55**	**56**
No. of Teachers:	**10**	**12**	**15**	**10**	**14**	**18**	**6**

Solution:

Calculate Quartile deviation and its coefficient

Age in Years(X)	No. of Teachers(f)	Cumulative Frequency(cf)
50	**10**	**10**
51	**12**	**22**
52	**15**	**37**
53	**10**	**47**
54	**14**	**61**
55	**18**	**79**
56	**6**	**85**

Q1 = Size of the $\frac{N+1}{4}$ th term = Size of the $\frac{85+1}{4}$ th term = 21.5^{th} item = 51.

Q3 = Size of the $\frac{3(N+1)}{4}$ th term = Size of the $\frac{3(85+1)}{4}$th term = $\frac{285}{4}$th term = 64.5th item = 55.

Q.D = $\frac{\mathbf{Q_3 - Q_1}}{\mathbf{2}} = \frac{\mathbf{55 - 51}}{\mathbf{2}}$ = 2.

Coefficient of Quartile deviation =$\frac{Q_3 - Q_1}{Q_3 + Q_1}$. = $\frac{55-51}{55+51}$ = $\frac{4}{106}$ = 0.38. Ans

Calculation of Quartile Deviation – Continuous Series

8. Calculate Quartile deviation and its coefficient for the following data:

Marks:	0-10	10-20	20-30	30-40	40-50
No of Students:	4	15	28	16	7

Solution:

Calculate Quartile deviation and its coefficient

Marks(X)	No. of Students(f)	Cumulative Frequency(cf)
0-10	4	4
10-20	15	19
20-30	28	47

30-40	16	63
40-50	7	70
	N = 70	

Q1 = Size of the $\frac{N}{4}$ th term = Size of the $\frac{70}{4}$ th term = 17.5th item.

Q1 class = 10-20

Q1 = L1 + $\frac{\frac{N}{4} - cf}{f}$ X C,

Where L1 = 10, $\frac{N}{4}$ = 17.5, cf = 4, f = 15, C = 10.

Q1 = 10 + $\frac{17.5-4}{15}$ X 10 = 10 + $\frac{135}{5}$ = 10 + 9 = 19.

Q3 = Size of the $\frac{3(N)}{4}$ th term = Size of the $\frac{3(70)}{4}$th term = 52.5th item.

Q3 class = 30-40

Q3 = L1 + $\frac{\frac{3N}{4} - cf}{f}$ X C

= Q3 = 30 + $\frac{52.5-47}{16}$ X 10 = 30 + $\frac{55}{16}$ = 30 + 3.44 = 33.44.

Q.D = $\frac{Q_3 - Q_1}{2}$ = $\frac{33.44 - 19}{2}$ = 7.22 Marks.

Coefficient of Quartile deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$. = $\frac{33.44-19}{33.44+19}$ = $\frac{14.44}{54.44}$ = 0.275. Ans

PROBLEMS

1. Calculate the value of Quartile deviations and coefficient of Quartile deviation from the following figures

X	42	75	85	101	145	175	210	250	300

Answer: 75, 0.484

2. Calculate the value of Quartile deviations and coefficient of Quartile deviation from the following figures

X	22	26	14	30	18	11	35	41	12	32

Answer: quartile deviation = 2.125, Coefficient of Quartile deviation = 0.26153

3. Calculate the value of Quartile deviations and coefficient of Quartile deviation from the following figures

X	30	45	75	65	50	52	28	40	39	35	42

Answer: Quartile deviation = 9, Coefficient of Quartile deviation = 0.10741

4. Calculate the value of quartile deviation and its coefficient

Marks	20	25	13	27	30	23	17	21	18	16	24
No of students	4	6	6	3	6	4	5	6	3	2	5

Answer: 4, 0.19

5. Calculate the value of quartile deviation and its coefficient

Marks	10	20	30	40	50	80
No of students	4	7	15	8	7	2

Answer: 10, 0.43

6. Calculate the value of quartile deviation and its coefficient

Marks	10	20	30	40	50	60
No of students	4	7	15	8	7	2

Answer: 10, 0.333

7. In the following table shows the wages of 60 workers. Calculate Q.D. and its Coefficient.

Wages:	20-25	25-30	30-35	35-40	40-45
No. of workers	2	10	26	16	7

Answer: 3.40, 0.1

8. Calculate Q.D and its Coefficient for the following

Class interval	0-5	5-10	10-15	15-20	20-25
Frequency	44	60	36	44	8

9. Calculate Q.D and its Coefficient

Age X	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60

No of Persons	60	80	90	150	180	130	140	60

Answer: 7.276, 0.1738

10. Calculate Q,D and its Coefficient

Wages (in thousands)	0-10	10-20	20-30	30-40
No of Persons	5	2	5	4

MEAN DEVIATION

The formula for calculating mean deviation is

$$\mathbf{M.D} = \frac{\sum |\mathbf{X} - \bar{\mathbf{X}}|}{N}$$

M.D is Mean Deviation

$|\mathbf{X} - \bar{\mathbf{X}}|$ **= with parallel lines read as mod** $(\mathbf{X} - \bar{\mathbf{X}})$ **is the modulus value or absolute value i.e, deviation from median or mean by ignoring signs.**

COEFFICIENT OF MEAN DEVIATION

If deviations are taken from arithmetic mean then:

Coefficient of $MD_{\bar{X}}$ = $\frac{MD}{\bar{X}}$ Where $\bar{X}$ **= Mean**

If deviations are taken from median then:

Coefficient of MD_m = $\frac{MD_m}{\text{Median}}$

If deviations are taken from mode then:

Coefficient of MD_z = $\frac{MD_z}{\text{Mode}}$

Calculation of Mean Deviation

Individual Series

STEPS

- Compute Mean or Median or Mode of the series
- Find the deviations of each item from mean or median or mode by ignoring plus or minus signs and obtain $\sum |X - \bar{X}|$
- Apply the formula

$$M.D = \frac{\sum |X - \bar{X}|}{N}$$

- For coefficient of MD

  - If deviations are taken from arithmetic mean then:

    Coefficient of $MD_{\bar{X}}$ = $\frac{MD}{\bar{X}}$ Where $\bar{X}$ **= Mean**

  - If deviations are taken from median then:

Coefficient of MD_m = $\frac{MD_m}{Median}$

- If deviations are taken from mode then:

  Coefficient of MD_z = $\frac{MD_z}{Mode}$

NOTE: When the question is silent then that situation mean deviation calculated from median.

Question1: The following are the room rents charged by a hotel for various types of accommodation provided to the occupants: Rs. 300, 400, 700, 200, 600, 500, 100. Calculate the mean deviation and its coefficient.

Solution:

Rent (Ascending Order) X	$\lvert X - \bar{X} \rvert$	$\lvert X - \mathbf{Me} \rvert$	$\lvert X - \mathbf{Z} \rvert$
100	$\lvert 100 - 400 \rvert$ = **300**	$\lvert 100 - 400 \rvert$ = **300**	$\lvert 100 - 400 \rvert$ = **300**
200	$\lvert 200 - 400 \rvert$ = **200**	$\lvert 200 - 400 \rvert$ = **200**	$\lvert 200 - 400 \rvert$ = **200**
300	$\lvert 300 - 400 \rvert$ = **100**	$\lvert 300 - 400 \rvert$ = **100**	$\lvert 300 - 400 \rvert$ = **100**
400	$\lvert 400 - 400 \rvert$ = **0**	$\lvert 400 - 400 \rvert$ = **0**	$\lvert 400 - 400 \rvert$ = **0**
500	$\lvert 500 - 400 \rvert$ =**100**	$\lvert 500 - 400 \rvert$ =**100**	$\lvert 500 - 400 \rvert$ =**100**
600	$\lvert 600 - 400 \rvert$ = **200**	$\lvert 600 - 400 \rvert$ = **200**	$\lvert 600 - 400 \rvert$ = **200**
700	$\lvert 700 - 400 \rvert$ = **300**	$\lvert 700 - 400 \rvert$ = **300**	$\lvert 700 - 400 \rvert$ = **300**
$\sum \mathbf{X}$ =2800	$\sum \lvert X - \bar{X} \rvert$ =**1200**	$\sum \lvert X - \mathbf{Me} \rvert$ = **1200**	$\sum \lvert X - \mathbf{Z} \rvert$ = **1200**

$\bar{X} = \frac{\sum X}{N} = \frac{2800}{7}$ **=400**

M D from Mean = $\frac{\sum \lvert X - \bar{X} \rvert}{N} = \frac{1200}{7}$ =171.43

Coeficient of MD from mean= $\frac{\text{MD from Mean}}{Mean} = \frac{171.43}{400}$ = 0.42

Median = Size of the $\left(\frac{N+1}{2}\right)^{th}$ term

= Size of the $\left(\frac{7+1}{2}\right)^{th}$ **term**

= Size of the 4 th term = 400

$$\mathbf{M.D} = \frac{\sum|X-\mathbf{Me}|}{N} = \frac{1200}{7} = \mathbf{Rs.\ 171.43}$$

$$\mathbf{Coefficient\ of\ MD_m} = \frac{\mathbf{MD_m}}{\mathbf{Median}} = \frac{171.43}{400} = \mathbf{0.429}$$

Mode = 3meadian - 2 Mean

Question2: From the following marks 20, 22, 27, 30, 31, 32, 35, 40, 45, 48 obtained by 10 students, calculate mean deviation from mean and coefficient of mean deviation.

Solution:

Marks X	Deviation from Mean = 33 $\lvert d \rvert$
20	13
22	11
27	6
30	3
31	2
32	1
24	2
40	7
45	12
48	15
	$\sum \lvert d \rvert = 72$

$$\bar{X} = \frac{\sum X}{N} = \frac{330}{10} = 33$$

$$\mathbf{M.D} = \frac{\sum|d|}{N} = \frac{72}{10} = 7.2$$

$$\mathbf{Coefficient\ of\ MD_{\bar{X}}} = \frac{MD}{\bar{X}} = \frac{7.2}{33} = 0.218$$

Discrete Series

STEPS:

- Calculate mean or median or mode

- Take deviations ignoring signs and denote it by $|d|$
- Multiply $|d|$ by respective frequencies and obtain its total i.e, $f|d|$
- Apply the formula

$$\mathbf{M.D} = \frac{\sum f|d|}{N}$$

$$\textbf{Coefficient of MD} = \frac{\textbf{MD}}{\textbf{Mean or Median or Mode}}$$

Question1: Calculate mean deviation and its coefficient from the following data

Height (in inches)	61	64	67	70	73
No. of persons	5	18	42	27	8

Solution:

Height X	No of persons F	Cf	$\|d\| = \|X - 67\|$	$f\|d\|$
61	5	5	6	30
64	18	23	3	54
67	42	65	0	0
70	27	92	3	81
73	8	100	6	48
	N= 100			$\sum f\|d\| = 213$

Median = Size of the $\left(\frac{N+1}{2}\right)^{th}$ term

= Size of the $\left(\frac{100+1}{2}\right)^{th}$ term or Size of the 50.5 **th** term = 67

$$\mathbf{M.D} = \frac{\sum f|d|}{N} = \frac{213}{100} = 2.13$$

$$\textbf{Coefficient of } \mathbf{MD_m} = \frac{\mathbf{MD_m}}{\textbf{Median}} = \frac{2.13}{67} = 0.32$$

Continuous Series

Apply the formula

$$\mathbf{M.D} = \frac{\sum f|d|}{N} \times C$$

$$\textbf{Coefficient of MD} = \frac{\textbf{MD}}{\textbf{Mean or Median}}$$

Question1: Calculate the mean deviation and its coefficient from mean for the following

Marks	0-20	20-40	40-60	60-80	80-100
No of students	10	16	30	32	12

Solution:

Marks	F	Mid Value X	$\frac{X-A}{C} = \frac{X-50}{20}$ dx′	fdx′	$\lvert X - \overline{X} \rvert = \lvert X - 54 \rvert$ $\lvert d \rvert$	$f \lvert d \rvert$
0-20	10	10	-2	-20	44	440
20-40	16	30	-1	-16	24	384
40-60	30	50	0	0	4	120
60-80	32	70	1	32	16	512
80-100	12	90	2	24	36	432
	N=100					$\sum \mathbf{f} \lvert d \rvert = \mathbf{1888}$

$$\overline{X} = A + \frac{\sum dx'}{N} \times C = 50 + \frac{20}{100} \times 20 = \mathbf{54}$$

$$\text{M.D} = \frac{\sum f \lvert d \rvert}{N} = \frac{1888}{100} = \mathbf{18.88\ or\ 19}$$

$$\text{Coefficient of MD}_{\overline{X}} = \frac{\text{MD}_{\overline{X}}}{\text{Mean}} = \frac{18.88}{54} = \mathbf{0.349}$$

Mean deviation

$\text{MD} = \frac{\sum \lvert x - \text{Mean} \rvert}{N} = \frac{\sum \lvert x - \text{Me} \rvert}{N} = \frac{\sum \lvert x - \text{Z} \rvert}{N}$

$\text{coefficient of MD} = \frac{\text{MD}}{Mean\ or\ Median\ or\ mode}$

Individual series

Question

120,170,240.100.105,205,300,160,150,180

Arrange in order

100,105,120,150,160,170,180,205,240, 300

N= 10

case I from Mean

Mean = (100+105+120+150+160+170+180+205+240+300)/10=1730/10=173

X	$\lvert X - \overline{X} \rvert$	$\lvert X - \mathbf{Me} \rvert$	$\lvert X - \mathbf{Z} \rvert$
100	\100-173\=73	$\lvert 100 - 165 \rvert$ **= 65**	$\lvert 100 - 149 \rvert$ = 49
105	\105-173/=68	$\lvert 105 - 165 \rvert$ **=60**	$\lvert 105 - 149 \rvert$ =44
120	\120-173/=53	$\lvert 120 - 165 \rvert$ =45	$\lvert 120 - 149 \rvert$ =29
150	\150-173/=23	$\lvert 150 - 165 \rvert$ **= 15**	$\lvert 150 - 149 \rvert$ = 1
160	\160-173/=13	$\lvert 160 - 165 \rvert$ **= 5**	$\lvert 160 - 149 \rvert$ = 11
170	\170-173/=3	$\lvert 170 - 165 \rvert$ **= 5**	$\lvert 170 - 149 \rvert$ = 21
180	\180-173/=7	$\lvert 180 - 165 \rvert$ **=15**	$\lvert 180 - 149 \rvert$ = 31
205	\205-173/=32	$\lvert 205 - 165 \rvert$ **= 40**	$\lvert 205 - 149 \rvert$ = 56
240	\240-173/=67	$\lvert 240 - 165 \rvert$ **= 75**	$\lvert 240 - 149 \rvert$ = 91
300	\300-173/=127	$\lvert 300 - 165 \rvert$ **= 135**	$\lvert 300 - 149 \rvert$ = 151
$\sum X = 1730$	∑\X-Mean\=466	$\sum \lvert X - Me \rvert$ =460	$\sum \lvert X - \mathbf{Z} \rvert$ **= 484**

$$\overline{\mathbf{X}} = \frac{\sum \mathbf{X}}{\mathbf{N}}$$

$$\overline{\mathbf{X}} = \frac{\mathbf{1730}}{\mathbf{10}} \mathbf{=173}$$

MD from Mean $= \frac{\sum \lvert X - \overline{X} \rvert}{N} = \frac{466}{10} = 46.6$

Coeeficient of MD from mean= $\frac{\text{MD from Mean}}{Mean} = \frac{46.6}{173} = 0.26$

Case II from Median

Me = (n+1)/2th term=(10+1)/2th =5.5th term

5th term + 0.5(6th term - 5th term)

160+ 0.5(170-160) =160 + 0.5 x 10=160+5=165

MD from Median $= \frac{\sum \lvert X - \mathbf{Me} \rvert}{N} = \frac{460}{10} = 46$

Coeeficient of MD from median= $\frac{\text{MD from Me}}{Me} = \frac{46}{165} = 0.27$

Case III from Mode

Mode is ill defined

Mode = 3(ME) - 2(MEAN) = 3(165) - 2(173) = 149

MD from Mode $=\frac{\sum|X-\mathbf{Z}|}{N}=\frac{484}{10}=48.4$

Coeficient of MD from mode= $\frac{\text{MD from Z}}{Z}=\frac{48.4}{149}=0.32$

Discrete series

Marks	15	30	50	40	20	10	25
No. of students	12	10	2	3	15	8	10

Arranging in order

X	f	fX	\X-Mean\	f\X-Mean\
10	8	80	\10-22.2\=12.2	97.6
15	12	180	\15-22.2\=7.2	86.4
20	15	300	\20-22.2\=2.2	33
25	10	250	\25-22.2\=2.8	28
30	10	300	\30-22.2\=7.8	78
40	3	120	\40-22.2\=17.8	53.4
50	2	100	\50-22.2\=27.8	55.6
	∑f=60	$\sum fX = 1330$		$\sum f\backslash X-\text{Mean}\backslash=432$

Case I FROM Mean

Mean $=\frac{\sum fX}{\sum f}=\frac{1330}{60}=22.16=22.2$

MD $=\frac{\sum f\backslash X-\text{Mean}\backslash}{\sum f}=\frac{432}{60}=7.2$

Coeficient of MD from mean= $\frac{\text{MD from Mean}}{Mean}=\frac{7.2}{22.2}=0.32$

caseII from Median

X	f	Cf	\X-Me\	f\X-Me\
10	8	8	\10-20\=10	80
15	12	20	\15-20\=5	60
20= me	15	35 me lies here	\20-20\=0	0
25	10	45	\25-20\=5	50
30	10	55	\30-20\=10	100
40	3	58	\40-20\=20	60
50	2	60	\50-20\=30	60

	∑f=60			∑ f\X-Me\=410

Median lies in (n+1)2 th term=(60+1)/2 th term=30.5th term

Median = 20

MD $=\frac{\sum f\backslash X-Me\backslash}{\sum f}=\frac{410}{60}=6.83$

Coeeficient of MD from median= $\frac{\text{MD from Me}}{Me}=\frac{6.83}{20}=0.34$

Case III from Mode

GROUPING TABLE

X	F (I)	II	III	IV	V	VI
10	8	20	X	35	X	X
15	12		27		37	X
20	15	25				35
25	10		20	23		
30	10	13			15	
40	3		5			
50	2					

ANALYSIS TABLE

X	10	15	20 =Z	25	30	40	50
I			1				
II			1	1			
III		1	1				
IV	1	1	1				
V		1	1	1			
VI			1	1	1		
TOTAL	1	2	6	3	1		

X	F	$\lvert X - \mathbf{Z} \rvert$	$\boldsymbol{F}\lvert X - \mathbf{Z} \rvert$
10	8	$\lvert 10\text{-}20 \rvert = 10$	**80**
15	12	$\lvert 15\text{ -}20 \rvert = 5$	**60**
20	15	$\lvert 20 - 20 \rvert = 0$	**0**
25	10	$\lvert 25\text{-}20 \rvert = 5$	**50**
30	10	$\lvert 30 - 20 \rvert = 10$	**100**

40	3	$\lvert 40-20 \rvert = 20$	**60**
50	2	$\lvert 50-20 \rvert = 30$	**60**
		$\lvert X-\mathbf{Z} \rvert$	$\sum \boldsymbol{F} \lvert X-\mathbf{Z} \rvert$ **=410**

MD FROM mode = MD from Mode $=\frac{\sum f \lvert X-\mathbf{Z} \rvert}{\sum f} = \frac{410}{60} = 6.83$

Coeficient of MD from mode= $\frac{\text{MD from Z}}{\sum f} = \frac{6.83}{60} = 0.34$

PROBLEMS

1. Find mean deviation and its coefficient for the following data

S. No	1	2	3	4	5	6	7	8	9	10
Marks	17	35	38	16	42	27	19	11	40	25

Answer: 9.4, 0.361

2. Find mean deviation and its coefficient for the following two series

Series A	3484	4572	4124	3682	5624	4388	3680	4008
Series B	508	487	620	382	408	266	186	318

Answer: Series A: 481.75, 0.118

Series B: 121.375, 0.307

3. Find mean deviation and its coefficient for the following data

Price	25	28	32	32	36	45	45	48	50	50	50

Answer: 8.18, 0.18

4. Find coefficient of mean deviation for the following

A	3484	4572	4124	3682	5624	4388	3680	4308
B	487	508	620	382	408	266	186218	

Answer: 9.4, 0.361

5. Find mean deviation and its coefficient for the following data

X	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5	10.5
F	2	3	5	6	6	4	6	4	14

6. Find mean deviation and its coefficient for the following data

X	10	11	12	13
F	3	12	18	12

7. Find mean deviation and its coefficient for the following data

X	10	11	12	13	14
F	3	12	18	12	3

8. Find mean deviation and its coefficient from mean and median the following data

CI	140-150	150-160	160-170	170-180	180-190	190-200
F	4	6	10	18	9	3

9. Find mean deviation and its coefficient from mean the following data

CI	0-10	10-20	20-30	30-40	40-50	50-60	60-70
F	4	8	11	15	19	6	3

10. Find mean deviation and its coefficient from mean and median the following data

CI	4-6	6-8	8-10	10-12	12-14	14-16	16-18
F	30	90	120	150	80	60	20

STANDARD DEVIATION AND VARIANCE

Calculation of Standard Deviation

Individual Series

a) Calculation of SD from Actual mean (Direct Method)

STEPS:

- Calculate actual of the series i.e, $\overline{\mathbf{X}}$
- Take deviations of the items from the mean to find $\boldsymbol{d = X - \overline{X}}$
- Square the deviations and obtain their total $\sum d^2$
- We apply the formula

$$\sigma = \sqrt{\frac{\sum d^2}{N}}$$

Where σ = Standard deviation, $\sum d^2$ = sum of the squared of thedeviations from the actual mean and N = total number of observations.

b) Calculation of SD from Assumed mean

STEPS:

- Find the total number of items i.e, N
- Take deviations of the items from the mean to find $\boldsymbol{dx = (X - A)}$
- Square the deviations and obtain their total $\sum dx^2$
- We apply the formula

$$\sigma = \sqrt{\frac{\sum d^2}{N} - \left(\frac{\sum d}{N}\right)^2}$$

Where ϭ = Standard deviation, $\sum dx^2$ = sum of the squared deviations from the assumed mean and N = total number of observations.

Coefficient of Standard Deviation

$$\textbf{Coefficient of SD} = \frac{\sigma}{\overline{X}}$$

Where ϭ = Standard Deviation and $\overline{X}$ = Mean

Question 1: Calculate standard deviation by direct and short cut method

Wages	40	44	54	60	62	64	70	80	90	96

Solution:

a) Direct MetOD

Arrange in order

Wages	d=X- $\overline{X}$	d^2
40	-26	676
44	-22	484
54	-12	144
60	-6	36
62	-4	16
64	-2	4
70	4	16
80	14	196
90	24	576
96	30	900

$\sum X = 660$	$\sum d = \sum(\boldsymbol{X} - \overline{\mathbf{X}}) = 0$	$\sum d^2 = \sum(\boldsymbol{X} - \overline{\mathbf{X}}) = 3048$

$$\overline{X} = \frac{\sum X}{N} = \frac{660}{10} = 66$$

NOTE:- SUM OF DEVIATIONS FROM ACTUAL MEAN =
$\sum d = \sum(\boldsymbol{X} - \overline{\mathbf{X}}) = 0$ AND ONLY FOR INDIVIDUAL SERIES

Now,

$$\sigma = \sqrt{\frac{\sum d^2}{N}} = \sqrt{\frac{3048}{10}} = \sqrt{304.8} = 17.46$$

$$\text{Coefficient of SD} = \frac{\sigma}{\overline{X}} = \frac{17.46}{66} = 0.2645$$

b) Short cut Method

Wages	D X- A	d^2 =(X-A)^2
40	-20	400
44	-16	256
54	-6	36
60=A	0	0
62	2	4
64	4	16
70	10	100
80	20	400
90	30	900
96	36	1296
$N = 10$	$\sum d = 60$	$\sum d^2 = 3408$

$$\overline{X} = A + \frac{\sum d}{N} = 60 + \frac{60}{10} = 60 + 6 = 66$$

$$\sigma = \sqrt{\frac{\sum d^2}{N} - \left(\frac{\sum d}{N}\right)^2} = \sqrt{\frac{3408}{10} - \left(\frac{60}{10}\right)^2} = \sqrt{340.8 - 36} = 17.46$$

$$\text{Coefficient of SD} = \frac{6}{\bar{X}} = \frac{17.46}{66} = 0.2645$$

Question2: From the following marks 20, 22, 27, 30, 31, 32, 35, 40, 45, 48 obtained by 10 students, calculate STANDARD deviation and coefficient of Standarddeviation.

Marks X	Deviation from Mean = 33 $\boldsymbol{d = x - \bar{X}}$	d^2
20	-13	169
22	-11	121
27	-6	36
30	-3	9
31	-2	4
32	-1	1
24	-2	4
40	7	49
45	12	144
48	15	225
	$\boldsymbol{\sum d = 0}$	$\boldsymbol{\sum d^2 = 762}$

$$\bar{X} = \frac{\sum X}{N} = \frac{330}{10} = 33$$

$$\sigma = \sqrt{\frac{\sum d^2}{N}} = \sqrt{\frac{762}{10}} = \sqrt{76.2} = 8.729 = 8.73$$

$$\text{Coefficient of SD} = \frac{6}{\bar{X}} = \frac{8.73}{33} = \mathbf{0.2645}$$

Discrete Series

a) Calculation of SD from Actual mean

- We apply the formula

$$\sigma = \sqrt{\frac{\sum fd^2}{N}}$$

Where σ = Standard deviation, $\sum fd^2$ = sum of the squared deviations from the actual mean and N = total number of observations and $\boldsymbol{d = (X - \overline{X})}$

b) Calculation of SD from Assumed mean

- We apply the formula

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$$

Where σ = Standard deviation, $\sum fdx^2$ = sum of the squared deviations from the assumed mean and N = total number of observations and $\boldsymbol{d = (X - A)}$ **and** $\boldsymbol{d^2 = (X - A)^2}$

STEPS:

- Find the total number of items i.e, $\sum f$
- Take deviations of the items from the mean to find $\boldsymbol{d = (X - A)}$
- Square the deviations and obtain their total $\sum d^2$
- Multiply these deviations with their respective frequencies and obtain total $\sum fd$
- Take sum of multiple frequencies and squared deviations i.e, $\sum fd^2$

Question 1: Calculate SD and it's coefficient from the following data

X	12	14	16	18	20	22	24
f	6	12	18	26	16	10	8

Solution: short cut method

I	II	III	IV=IIxIII	V= IIxIV
X	d=X-A	F	fd	$d(fd)=fd^2$
12	-8	6	-48	384
14	-6	12	-72	432
16	-4	18	-72	288
18	-2	26	-52	104
20 =A	0	16	0	0
22	2	10	20	40
24	4	8	32	128
		$\sum f$= 96	$\sum fd = -192$	$\sum fd^2 = 1376$

$$\overline{X} = A + \frac{\sum fd}{\sum f} = 20 + \frac{-192}{96} = 20 - 2 = 18$$

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2} = \sqrt{\frac{1376}{96} - \left(\frac{-192}{96}\right)^2} = \sqrt{14.33 - 4} =$$

$\sqrt{10.33} = 3.2145$

$$\text{Coefficient of SD} = \frac{\sigma}{\bar{X}} = \frac{3.2145}{18} = 0.1786$$

Actual mean method Direct Method

X	f	fX	$d=X-\overline{X}$	$d^2=(X-\overline{X})^2$	$Fd^2=f(X-\overline{X})^2$
12	6	72	-6	36	216
14	12	168	-4	16	192
16	18	288	-2	4	72
18	26	468	0	0	0
20	16	320	2	4	64
22	10	220	4	16	160
24	8	192	6	36	288
	$\sum f = 96$	$\sum fX = 1728$			$\sum fd^2 = 992$

$\overline{X} = \frac{\sum fX}{\sum f} = \frac{1728}{96} = 18$

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f}} = \sqrt{\frac{992}{96}} = \sqrt{10.33} = 3.2145$$

Continuous Series

c) Calculation of SD from Actual mean

STEPS:

- Find out the mid values of different classes

- Calculate arithmetic mean and take deviations from mid values
- Multiply frequencies of each class with these deviations
- We apply the formula

$$\sigma = \sqrt{\frac{\sum f d^2}{\sum f}}$$

Where σ = Standard deviation, $\sum f d^2$ = sum of the squared deviations from the actual mean and N = total number of observations and $\boldsymbol{d} = (\boldsymbol{X} - \overline{\mathbf{X}})$

Calculation of SD from Assumed mean

STEPS:

- Find out the mid values of different classes
- Calculate assumed mean and take deviations from mid values
- Multiply frequencies of each class with these deviations
- Multiply frequencies of each class with the square of deviations
- We apply the formula

$$\sigma = \sqrt{\frac{\sum f d^2}{\sum f} - \left(\frac{\sum f d}{\sum f}\right)^2}$$

Where σ = Standard deviation, $\sum f d^2$ = sum of theproduct of frequency with squared deviations from the assumed mean, N = total number of observations and $\boldsymbol{d} = (\boldsymbol{X} - \mathbf{A})$ **and** $\boldsymbol{d^2} = (\boldsymbol{X} - \mathbf{A})^{\mathbf{2}}$

STEP DEVIATION IN CASE OF STANDARD DEVIATION

STEPS:

- Take mid value of all items
- When assumed mean is taken take deviations from assumed mean
- Take a common factor if the class intervals are equal. Divide the deviations by the common factor
- Multiply the frequencies of each class with deviations
- Square the deviations and multiply them with respective frequencies of each class
- We apply the formula

$$\sigma = \sqrt{\frac{\sum f d'^2}{\sum f} - \left(\frac{\sum f d'}{\sum f}\right)^2} \times \mathrm{h}$$

Question1: Calculate SD and it's coefficient from the following

CI	10-20	20-30	30-40	40-50	50-60	60-70
f	2	4	8	10	12	4

Solution:

S.D from Actual Mean

CI	f	Mid Value X	FX	$(X-\overline{X}) = d$	d^2	fd^2
10-20	2	15	30	-29.5	870.25	1740.5
20-30	4	25	100	-19.5	380.25	1521.0
30-40	8	35	280	-9.5	90.25	722.0
40-50	10	45	450	0.5	0.25	2.5
50-60	12	55	660	10.5	110.25	1323.0
60-70	4	65	260	20.5	420.25	1681.9
	$\sum f$ = 40		$\sum fX = 1780$			$\sum fd^2$ = 6990.9

$$\overline{X} = \frac{\sum fX}{\sum f} = \frac{1780}{40} = 44.5$$

$$\sigma = \sqrt{\frac{\sum f d'^2}{\sum f}} =$$

$$\sigma = \sqrt{\frac{6990.9}{40}} =$$

$$\sigma = \sqrt{174.77} = 13.22$$

$$\text{Coefficient of SD} = \frac{\sigma}{\overline{X}} = \frac{13.221}{44.5} = \mathbf{0.297}$$

short cut method

CI	10-20	20-30	30-40	40-50	50-60	60-70
f	2	4	8	10	12	4

CI	Mid Value X	d= X-A	F	Fd	fdxd=fd'^{2}
10-20	15	-30	2	-60	1800
20-30	25	-20	4	-80	1600
30-40	35	-10	8	-80	800
40-50	45 = A	0	10	0	0
50-60	55	10	12	120	1200
60-70	65	20	4	80	1600
			$\sum f$ = 40	$\sum fd$ =(-20)	$\sum$fd^2 =7000

$$\overline{X} = A + \frac{\sum fd}{\sum f} = \overline{X} = 45 + \frac{(-20)}{40} = 45\text{-}0.5 = 44.5$$

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$$

$$\sigma = \sqrt{\frac{7000}{40} - \left(\frac{-20}{40}\right)^2}$$

$$\sigma = \sqrt{175 - (0.5)^2}$$

$$\sigma = \sqrt{175 - 0.25}$$

$$\sigma = \sqrt{174.75}$$

$$\sigma = 13.221$$

$$\text{Coefficient of SD} = \frac{\sigma}{\overline{X}} = \frac{13.221}{44.5} = \mathbf{0.297}$$

CI	0-10	10-20	20-30	30-40	40-50	50-60	60-70
F	4	8	11	15	19	6	3

CI	Mid Value X	d= X-A	F	Fd	fdxd=fd^2
0-10	5	-30	4	-120	3600
10-20	15	-20	8	-160	3200
20-30	25	-10	11	-110	1100
30-40	35=A	0	15	0	0
40-50	45	10	19	190	1900
50-60	55	20	6	120	2400
60-70	65	30	3	90	2700
			$\sum f$ = 66	$\sum fd$ =10	$\sum fd^2$ =14900

$$\overline{X} = A + \frac{\sum fd}{\sum f} = 35.15$$

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2} = 15.02$$

$$\text{Coefficient of SD} = \frac{\sigma}{\overline{X}} = \mathbf{0.42}$$

Calculate mean and standard deviation

Below10	Below20	Below30	Below40	Below50	Below60
15	32	51	78	97	109

CI	m	D=m-A	F	fd	Fd^2
0-10	5	-30	15	-450	13500
10-20	15	-20	32-15=17	-340	6800
20-30	25	-10	51-32=19	-190	1900
30-40	35 =A	0	78-51=27	0	0
40-50	45	10	97-78=19	190	1900
50-60	55	20	109-97=12	240	4800
			$\sum f$ **=109**	$\sum fd$=-550	$\sum fd^2$ =28900

$$\overline{X} = A + \frac{\sum fd}{\sum f} = \overline{X} = 35 + \frac{-550}{109} = 35\text{-}5.04587 = 35\text{-}5.0459 = 29.9541$$

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$$

$$\sigma = \sqrt{\frac{28900}{109} - \left(\frac{-550}{109}\right)^2}$$

$$\sigma = \sqrt{265.1376 - (5.0459)^2}$$

$$\sigma = \sqrt{265.1376 - 25.4611}$$

$$\sigma = \sqrt{239.6765}$$

$$\sigma = 15.481488 = 15.4815$$

$$\text{Coefficient of SD} = \frac{\sigma}{\overline{X}} = \frac{15.4815}{29.9541} = \mathbf{0.5168}$$

COMBINED STANDARD DEVIATION

NUM= $[N_1(\sigma_1^2 + D_1^2) + N_2 (\sigma_2^2 + D_2^2) + N_3 (\sigma_3^2 + D_3^2)]$

DEN = $[N_1 + N_2 + N_3]$

$$\sigma_{123} = \sqrt{\frac{NUM}{DEN}} = \sqrt{\frac{[N1 (\sigma 1\text{^}2 + D1\text{^}2) + N2 (\sigma 2\text{^}2 + D2\text{^}2) + N3 (\sigma 3\text{^}2 + D3\text{^}2)]}{[N1 + N2 + N3]}}$$

$D_1 = \overline{X}_{123} - \overline{X}_1$

$D_2 = \overline{X}_{123} - \overline{X}_2$

$D_3 = \overline{X}_{123} - \overline{X}_3$

where as $\overline{X}_{123} = \frac{N1\overline{X}1 + N2\overline{X}2 + N3\overline{X}3}{N1+N2+N3}$

Calculate Combined mean and standard deviation for the following

	Series A	Series B	Series C
N	200 =N_1	250=N_2	300 =N_3
$\overline{X}$	25 =$\overline{X}_1$	10 =$\overline{X}_2$	15 =$\overline{X}_3$
σ	3 = σ_1	4 = σ_2	5 = σ_3

Step 1= Combined Mean

$$\overline{X}_{123} = \frac{N1\overline{X}1 + N2\overline{X}2 = N3\overline{X}3}{N1+N2+N3}$$

$$= \frac{200x25 + 250x10 + 300x15}{200 + 250 + 300}$$

$$= \frac{5000 + 2500 + 4500}{750}$$

$$= \frac{12000}{750} = 16$$

$D_1 = \overline{X}_{123} - \overline{X}_1 = 16-25 = -9$

$D_2 = \overline{X}_{123} - \overline{X}_2 = 16 - 10 = 6$

$D_3 = \overline{X}_{123} - \overline{X}_3 = 16 - 15 = 1$

Step 2 numerator

Num= $[N_1(\sigma_1^2 + D_1^2) + N_2 (\sigma_2^2 + D_2^2) + N_3 (\sigma_3^2 + D_3^2)]$

$$200(3^2 + (-9)^2) + 250(4^2 + 6^2) + 300(5^2 + 1^2)$$

$$=200(9+81) + 250(16+36) + 300(25+1)$$

$$200x90 + 250x52 + 300x26$$

$$=18000+13000+7800=38800$$

Step 3 Denominator

Den= $[N_1 + N_2 + N_3]$ $= 200 + 250 + 300 = 750$

$$\sigma_{123} = \sqrt{\frac{[N1 (\sigma1^2 + D1^2) + N2 (\sigma2^2 + D2^2) + N3 (\sigma3^2 + D3^2)]}{[N1 + N2 + N3]}}$$

$$\sigma_{123} = \sqrt{\frac{38800}{750}} = \sqrt{51.733} = 7.1925$$

Find Combined SD

	Girls	Boys
Mean	115	113
S D	8	6
No. of students	90	60

$\overline{X}_{12} = 114.2$

$d_1 = -0.8$

$d_2 = 1.2$

Num = $[N_1(\sigma_1^2 + D_1^2) + N_2 (\sigma_2^2 + D_2^2)$ = 8064

Den = $N_1 + N_2$ = 150

$$\sigma_{12} = \sqrt{\frac{NUM}{DEN}} = \sqrt{\frac{8064}{150}} = \sqrt{53.76} = 7.332$$

LORENZ CURVE

It is graphical Method. It helps in studying Variability in different componenets of distribution.

Steps

1. Calculate Mid points
2. calculate cumulative mid points
3. Calculate% of cumulative mid points
4. On X axis and Y axis Mark from 0 to 100 taking scale of 1cm = 10 units

5. Join origin (0,0) and (100, 100) . This is called as line of equal distribution.
6. Plot the points.

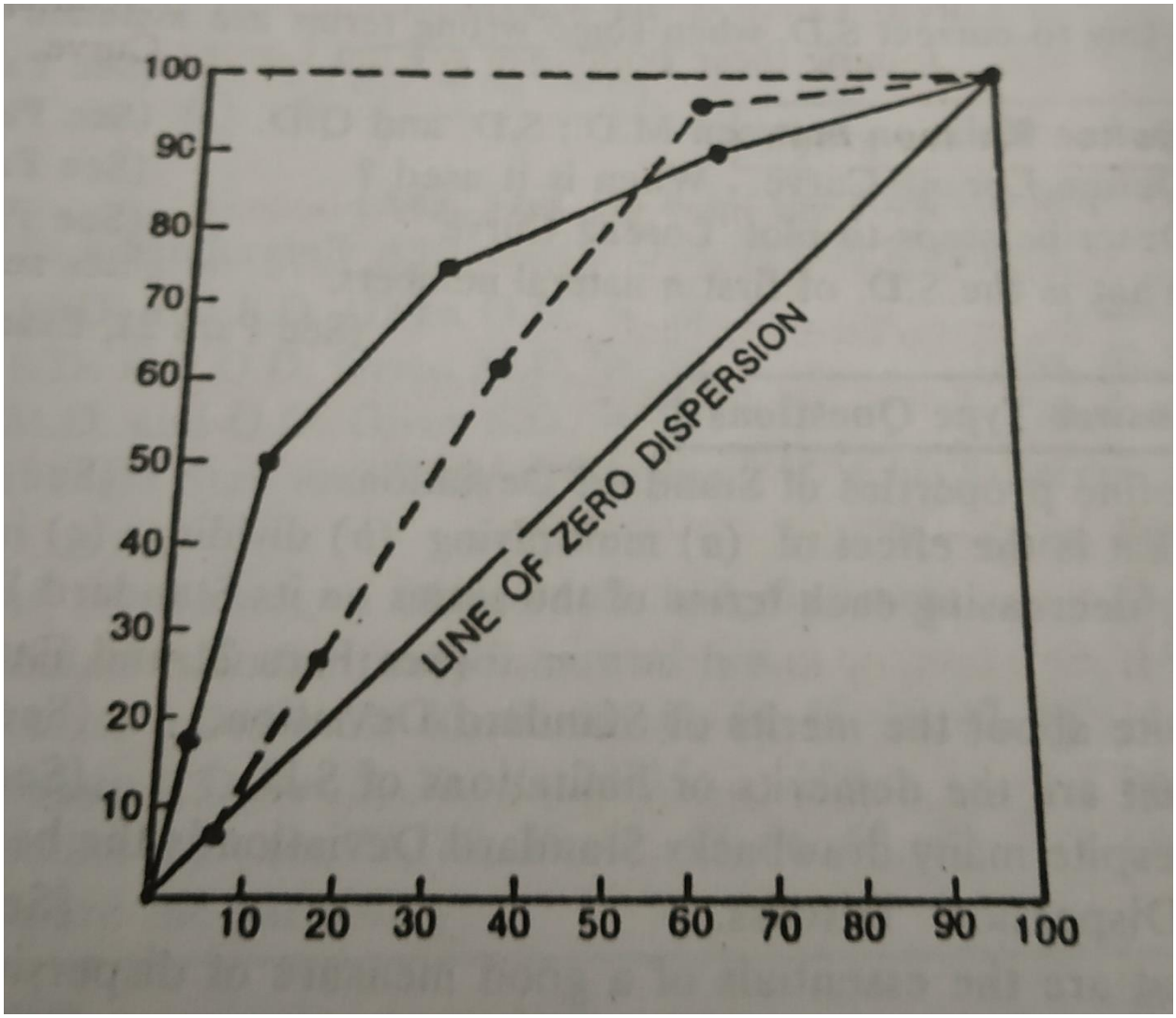
Draw Lorenz Curve

Marks	0-20	20-40	40-60	60-80	80-100	
Section H	12	27	14	13	9	75
Section IFA	4	17	43	37	5	106

Marks				Section H			Section IFA		
Marks	m	cm	%cm	No. of students h	ch	%ch	No. of students IFA	C IFA	%C IFA
0-20	10	10	$\frac{10}{250}x100 = 4$	12	12	16	4	4	4
20-40	30	40	$\frac{40}{250}x100 = 16$	27	39	52	17	21	20
40-60	50	90	$\frac{90}{250}x100 = 36$	14	53	71	43	64	60
60-80	70	160	$\frac{160}{250}x100 = 64$	13	66	88	37	101	95
80-100	90	250	$\frac{250}{250}x100 = 100$	9	75	100	5	106	100

h	ifa

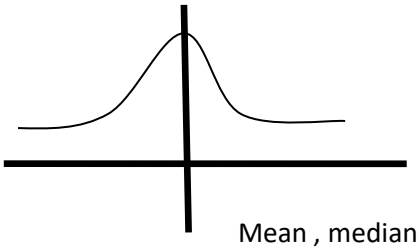
4,16	4,4
16, 52	16,20
36, 71	36,60
64, 88	64,95
100, 100	100,100



- **Meaning :** Skewness refers to lack of symmetry. Asymmetrical distribution is called as skew distribution

- **Types of skewness :**

1) **Symmetrical distribution (not a skewness) :**

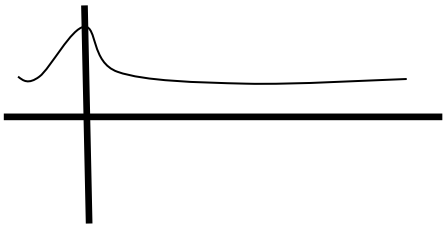


Mean , median , mode

In symmetrical distribution mean , median and mode coincide (equal) and it shows a symmetrical distribution also called as normal distribution

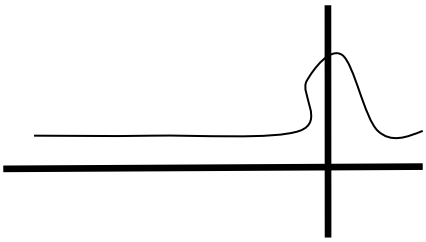
.

2) **Positively skewed distribution :**



Under the positive skewed distribution the longer tale is towards the right side and hence the left side area is not equal to the right side area . Left side area is less than right side area.

3) **Negatively skewed distribution :**



Under the negatively skew distribution the left side area has a longer tail and the left side area is greater than the right side area

Test of skewness :

In a giving distribution skewness is present if

1) The values of mean , median and mode are not equal .
2) The sum of the positive deviations from median is not equal to the sum of the negative deviations.
3) Quartiles are not equi-distant from median
4) The curve is not bell shaped .

Differences between Dispersion and skewness :

1) Dispersion indicates the amount of variation , whereas skewness indicates the direction and extent of variable.
2) Skewness are dependent on measure of dispersion , whereas dispersion is an independent one variable .
3) We cannot compute the measure of skewness .

CARL PEARSON'S COEFFICIENT OF SKEWNESS

CASE 1 : WHEN MODE IS DEFINED .

$$SK_p = \frac{\overline{X} - Z}{6}$$

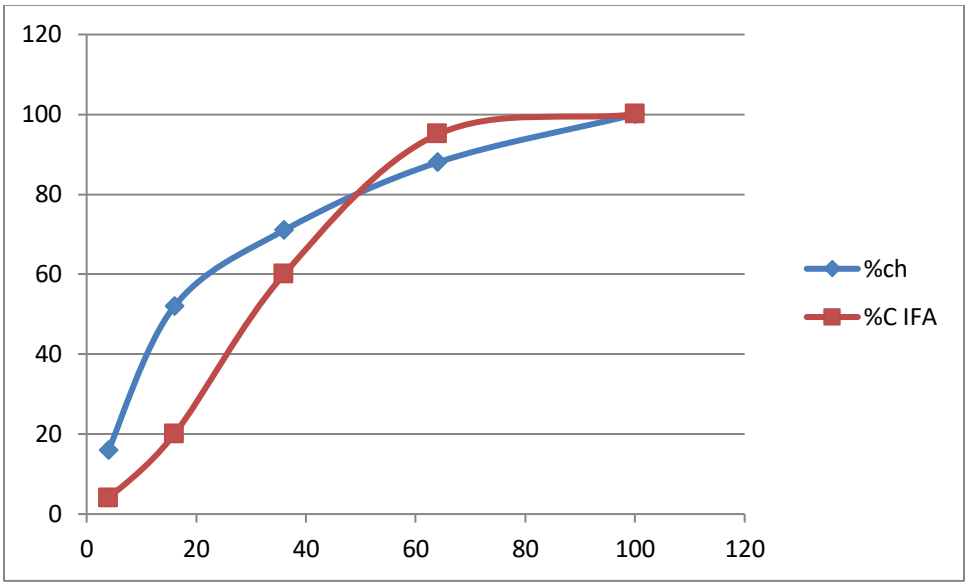
CASE 2 : WHEN MODE IS ILL DEFINED .

$$SK_p = \frac{\overline{3(X - Me)}}{6}$$

Bowleys coefficient of Skewness

$$SK_B = \frac{Q3 + Q1 - 2\,Me}{Q3 - Q1}$$

CI	0 - 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
f	4	13	21	44 44	33	22	7



step deviation

CI	Mid Value X	d= X-A	d' =d/10	f	fd'	fd'xd'=fd'^{2}
10-20	15	-30	-3	2	-6	18
20-30	25	-20	-2	4	-8	16
30-40	35	-10	-1	8	-8	8

40-50	45 = A	0	0	10	0	0
50-60	55	10	1	12	12	12
60-70	65	20	2	4	8	16
				$\sum f$ = 40	$\sum fd'$ =(-2)	$\sum fd^2$ =70

$\overline{X} = A + \frac{\sum fd'}{\sum f} xh = \overline{X} = 45 + \frac{(-2)}{40} x10 = 45 + \frac{(-2x10)}{40}$ =45-0.5=44.5

$$\sigma = \sqrt{\frac{\sum fd'^2}{\sum f} - \left(\frac{\sum fd'}{\sum f}\right)^2} \times h$$

$$\sigma = \sqrt{\frac{70}{40} - \left(\frac{-2}{40}\right)^2} \times 10$$

$$\sigma = \sqrt{1.75 - (0.05)^2} \times 10$$

$$\sigma = \sqrt{1.75 - 0.0025} \times 10$$

$$\sigma = \sqrt{1.7475} \times 10$$

$$\sigma = 1.321x10 = 13.221$$

Coefficient of SD $= \frac{\sigma}{\overline{X}} = \frac{13.221}{44.5}$ **=0.297**

BUSINESS STATISTICS

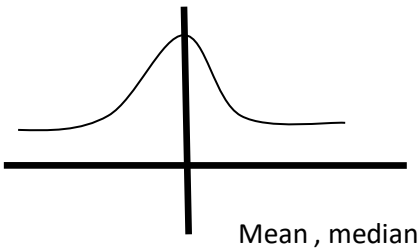
BCOM 1 H & IFA

UNIT 5 – SKEWNESS

- **Meaning :** Skewness refers to lack of symmetry. Asymmetrical distribution is called as skew distribution

- **Types of skewness :**

1) **Symmetrical distribution (not a skewness) :**

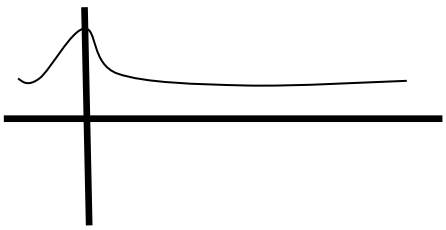


Mean , median , mode

In symmetrical distribution mean , median and mode coincide (equal) and it shows a symmetrical distribution also called as normal distribution

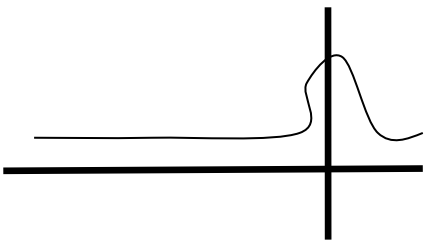
.

2) **Positively skewed distribution :**



Under the positive skewed distribution the longer tale is towards the right side and hence the left side area is not equal to the right side area . Left side area is less than right side area.

3) **Negatively skewed distribution :**



Under the negatively skew distribution the left side area has a longer tail and the left side area is greater than the right side area

Test of skewness :

In a giving distribution skewness is present if

1) The values of mean , median and mode are not equal .
2) The sum of the positive deviations from median is not equal to the sum of the negative deviations.
3) Quartiles are not equi-distant from median
4) The curve is not bell shaped .

Differences between Dispersion and skewness :

1) Dispersion indicates the amount of variation , whereas skewness indicates the direction and extent of variable.
2) Skewness are dependent on measure of dispersion , whereas dispersion is an independent one variable .
3) We cannot compute the measure of skewness .

CARL PEARSON'S COEFFICIENT OF SKEWNESS

INDIVIDUAL SERIES

CASE 1 : WHEN MODE IS MODALED / DEFINED .

1) FOLLOWING ARE THE MARKS OF 5 STUDENTS . APPLY CARL PEARSON'S SKEWNESS ?
(6 , 4 , 4 , 1 , 5)

Sol) First arrange the given data in ascending order as follows :

x	**d = x – mean**	**d^2**
1	-3	9
4	0	0
4	0	0
5	1	1
6	2	4
∑ = 20	**(0)**	**∑ = 14**

Mean = $= \overline{X} = \frac{\sum X}{N} = \frac{20}{5} = 4$

Mode = 4 **(it is the most repeated value).**

SD = $\sigma = \sqrt{\frac{\sum d^2}{N}}$

$= \sqrt{14/5}$

$= 1.673.$

SK_p = $\frac{\overline{X} - Z}{\sigma}$

$= \frac{4-4}{1.673}$

$= 0$

2) 30 , 12 , 28 , 21 , 27 , 21 , 23 . **(Ans = 0.3828)**
3) 14 , 14 , 15 , 13 , 12 , 16 , 17 .**(Ans = 0.271)**
4) 10 , 5 , 9 , 7 , 7 , 6 , 8 . **(Ans = 0.2703)**
5) 70 , 65 , 55 , 75 , 80 , 85 , 65 , 70 , 95 , 75 . **(Ans = 0.2794)**
6) 17 , 19 , 21 , 13 , 16 , 18 , 24 , 21 , 20 , 17 , 21 **(Ans = -0.7545)**

CASE 2 : WHEN MODE IS ILL DEFINED .

1) 7 , 4 , 10 , 9 , 15 , 12 , 7 , 9 , 7 , 9 .

Sol) Arrange the given data in ascending order as follows :

X	d = X - A	d2
4	-5	25
7	-2	4
7	-2	4
7	-2	4
9 (A)	0	0
9	0	0
9	0	0
10	1	1
12	3	9
15	6	36
	∑= - 1	∑ = 83

Mean = $\overline{X} = A + \frac{\sum d}{N} = 9 + \frac{-1}{10} = 9 - 0.1 = 8.9$.

Median = Q2 =(N + 1) / 2 th term .

- ➔ (10 + 1) / 2 → 5.5 th term.
- ➔ 5 th term + 0.5 (6 th term – 5 th term)
- ➔ 9 + 0 = 9

SD = $\sigma = \sqrt{\frac{\sum d^2}{N} - \left(\frac{\sum d}{N}\right)^2}$

$\sqrt{8.29}$ → 208792.

Mode is ill defined as 9 is repeated 3 times

SK_P = 3(mean – median) / SD = $\frac{\overline{3(X - Me)}}{\sigma}$

- ➔ 3(8.9 – 9) / 2.8732
- ➔ 3 (-0.1) / 2.8792
- ➔ -0.3 / 2.8792
- ➔ 0.10419.

2) 76 , 64 , 80 , 68 , 84 , 72 . **(Ans = 0)**
3) 35 , 20 , 40 , 22 , 48 , 27 , 32 , 30 , 45 , 31 . **(Ans = 0.5155)**
4) 24 , 7 , 11 , 9 , 17 , 3 , 20 , 14 , 4 , 22 , 27 . **(Ans = 0.1386)**
5) 40 , 9 , 17 , 21 , 36 , 14 , 24 , 39 , 31 , 27 . **(Ans = 0.0885)**
6) 6 , 17 , 34 , 41 , 54 , 29 , 13 . **(Ans = 0.2466)**
7) 19 , 9 , 15 , 21 , 4 , 3 , 23 , 11 , 7 , 26 , 14 . **(Ans = - 0.0736)**
8) 12 , 14 , 16 18 , 26 , 16 , 20 , 16 , 11 , 12 , 16 , 15 , 20 , 24 . **(Ans = 0.2021)**

DISCRETE SERIES :

CASE 1 : WHEN MODE IS DEFINED .

1) Find SK_P For the following :

X	**6**	**7**	**8**	**9**	**10**	**11**	**12**
F	**3**	**6**	**9**	**13**	**8**	**5**	**4**

Sol)

X	**F**	**d = X –A**	**fd**	**Fd^2**
6	**3**	**-3**	**-9**	**27**
7	**6**	**-2**	**-12**	**24**
8	**9**	**-1**	**-9**	**9**
9 = A	**13**	**0**	**0**	**0**
10	**8**	**1**	**8**	**8**
11	**5**	**2**	**10**	**20**
12	**4**	**3**	**12**	**36**
	∑f = 48	**∑d = 0**	**∑fd = 0**	**∑ fd^2= 124**

Mean = $\overline{X} = A + \frac{\sum fd}{\sum \mathbf{f}} = 9 + \frac{0}{48} = 9$

SD = $\sigma = \sqrt{\frac{\sum fd^2}{\sum \mathbf{f}} - \left(\frac{\sum fd}{\sum \mathbf{f}}\right)^2} = \sqrt{\frac{128}{\mathbf{48}} - \left(\frac{0}{\mathbf{48}}\right)^2}$

➔ $\sqrt{\frac{124}{48}} = \sqrt{2.5833}$

➔ 1.6072

Grouping table

X	f	II	III	IV	V	VI
6	3	9		18		
7	6		15		28	
8	9	22				30
9	13		21	26		
10	8	13			17	
11	5		9			
12	4	-				

Analysis table

X	6	7	8	9	10	11	12
I				✓			
II			✓	✓			
III				✓	✓		
IV				✓	✓	✓	
V		✓	✓	✓			
VI			✓	✓	✓		
TOTAL	0	1	3	6	3	1	0

Therefore mode = 9

$SK_p = \frac{\overline{X} - Z}{\sigma} = \frac{9-9}{1.6072}$

$= 0$

2)

X	15	25	35	45	55	65
f	2	4	8	20	12	4

(Ans = -0.0338)

3)

X	5	10	15	20	25	30
f	2	4	8	6	5	4

(Ans = 0.478)

4)

X	4.5	5.5	6.5	7.5	8.5	9.5	10.5	11.5
f	35	40	48	100	125	87	43	22

(Ans = - 0.245)

CASE 2 : WHEN MODE IS ill DEFINED .

1)

X	10	20	30	40	50	60	70	80
f	4	9	30	48	51	24	12	2

Sol)

X	f	cf	d = X – A	d/c= d'	Fd'c	fd(bar) 2
10	4	4	-30	-3	-12	36
20	9	13	-20	-2	-18	36
30	30	43	-10	-1	-30	30
40 (A)	48	91	0	0	0	0
50	51	142	10	1	51	51
60	24	166	20	2	48	96
70	12	178	30	3	36	108
80	2	180	40	4	8	32
	∑f = 180			∑d'= 4	∑fd' = 83	$\sum fd'^2 = 389$

Mean = $\overline{X} = A + \frac{\sum f d'}{\sum \mathbf{f}} x = 40 + \frac{83}{180} x10 = 40 + 4.611 = 44.611$

SD = $= \sigma = \sqrt{\frac{\sum f d'^2}{\sum \mathbf{f}} - \left(\frac{\sum f d'}{\sum \mathbf{f}}\right)^2} = \sqrt{\frac{389}{\mathbf{180}} - \left(\frac{83}{\mathbf{180}}\right)^2}$ $\boldsymbol{x\ 10}$

$\sqrt{1.9495}$ x 10

➔ 1.3962 x 10

➔ 13.962

Grouping table

X	F	II	III	IV	V	VI
10	4	13		43		
20	9		39		87	
30	30	78				129
40	45		99	123		
50	48	75			87	
60	51		36			38
70	12	14				
80	2					

Analysis table

X	10	20	30	40	50	60	70	80
I					✓			
II			✓	✓				
III				✓	✓			
IV				✓	✓	✓		
V		✓	✓	✓	✓	✓	✓	
VI			✓	✓	✓			
TOTAL	0	1	3	5	5	2	1	0

(mode is Bimodal)

Median = Q2 =(N + 1) / 2 th term

➔ 181 / 2

➔ 90.5 th term

➔ Median = 40

SK_P = 3(mean – median) / SD

➔ **SK_p =** $\frac{\overline{3(X-Me)}}{\sigma} = \frac{3(44.611-40)}{13.962} = \frac{3x4.611}{13.962} = \frac{13.833}{13.962}$

➔ 0.9907 .

CONTINOUS SERIES :

1)

CI	0 - 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
f	4	13	21	44 44	33	22	7

CI	M	d = m – A	f	D'	Fd'	Fd'^2
0 – 10	5	-30	4	-3	-12	36
10 – 20	15	-20	13	-2	-26	52
20 - 30	25	-10	21	-1	-21	21
30 – 40	35 (A)	0	44	0	0	0
40 – 50	45	10	33	1	33	33
50 – 60	55	20	22	2	44	88
60 – 70	65	30	7	3	21	63
			∑f = 144		∑fd' = 39	$\sum fd'^2 = 293$

$$\overline{X} = A + \frac{\sum fd}{\sum f} x\ h = \overline{X} = 35 + \frac{39}{144} x10 = 35+2.7 = 37.7$$

$$\sigma = \sqrt{\frac{\sum fd'^2}{\sum f} - \left(\frac{\sum fd'}{\sum f}\right)^2}\ \text{xh}$$

$$\sigma = \sqrt{\frac{293}{144} - \left(\frac{39}{144}\right)^2}\ \text{x10}$$

$$\sigma = \sqrt{2.03 - (0.27)^2}\ \text{x10}$$

$$\sigma = \sqrt{2.03 - .0729}\ \text{x10}$$

$$\sigma = \sqrt{1.9571}\ \text{x10}$$

$$\sigma = 1.398x10 = 13.98$$

Cl	f	II	III	IV	V	VI
0 – 10	4	17		38		
10 – 20	13		34		78	
20 - 30	21	65				98
30 – 40	44		77	99		
40 – 50	33	55			62	
50 – 60	22		29			
60 – 70	7					

Analysis table

Cl	0 - 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
f				✓			
II			✓	✓			
III				✓	✓		
IV				✓	✓	✓	
V		✓	✓	✓			
VI			✓	✓	✓		
TOTAL	0	1	3	6	3	1	0

(modal class = 30 – 40)

➔ L = 30 , fo = 21 , f1 = 44 , f2 = 33 , h = 10 .

$$Z = L + \frac{F1 - F0}{2F1 - F0 - F2} X\ h$$

$$Z = 30 + \frac{44 - 21}{2x44 - 21 - 33} x\ 10$$

$$Z = 30 + \frac{23}{34}\ x\ 10$$

$$Z = 30 + 6.7647 = 36.7647$$

$$SK_p = \frac{\overline{X} - Z}{6} \quad \frac{37.7 - 36.7647}{13.98} = 0.066$$