1. Introduction to Biostatistics

***Biostatistics** is the science of collection, analysis and interpretation of facts and numbers connected with Biology.*

Biostatistics is also called ***Biometrics***. Biometrics refers to ***biological measurements***.

A simple example to explain Biostatistics is the estimation of oxygen in a few water samples. The water samples form a ***population.*** The estimation of O_2 in each water sample is the ***collection of data***.

The amount of oxygen in the water samples is the ***data***.

Arranging values in columns is called ***tabulation***.

In one water sample, the amount of oxygen will be higher and in another it is lower. It is ***interpretation.***

Table. 1.1: A simple table to explain the science of Biostatistics.

Water Samples	Amount of O_2 in ml
1	4.5
2	6.9
3	6.2
4	5.3

Basic Concepts of Biostatistics

Population

Population is a group of individuals or study elements or observations.

In the estimation of oxygen, all the water samples form a ***population***. The value of each water sample is a ***variable.***

The population, containing limited number of individuals, is called a ***finite population***.

Eg. *Number of students in a class.*

Number of coconut trees in a grove.

Number of Tilapia fishes in a pond.

The population, containing unlimited number of individuals, is called an ***infinite population.***

Eg. *Stars in the sky. Fishes in the sea.*

Data

The values recorded in an experiment or observation are called ***data.*** The word '*data*' can be used both as singular and pleural. There is no word like '***datas***'.

The data is of two types, namely ***primary data*** and ***secondary data.***

The data collected by an investigator is called ***primary data.*** It is the first hand information. *The person collecting the data* is called ***investigator***.

The data collected from another source is called ***secondary data.*** Eg. *Data collected from News papers, journals,* etc.

Sample

*A small representative fraction of a population is called a **sample.***

Getting a sample from a population is called ***sampling.***

Eg. *Only a few rice is examined from a boiling pot to arrive at a conclusion.*

One or two grapes are tasted before buying a bunch.

Variable

The value of an item or individual is called ***variable.*** As the values vary, it is called ***variable.*** It is the ***characteristic*** of an individual.

Eg. *The mark scored by a student is a **variable.***

*The marks of many students are **variables.***

The number of mango trees in a garden.

Variables are of two types, namely

Quantitative variable and

Qualitative variable

When a variable is measurable, it is called a ***quantitative variable.*** Eg. *Number of students in a class, height, weight,* etc.

The quantitative variable is of two types, namely ***discontinuous variable*** and ***continuous variable.***

Discontinuous variables are measurable in ***whole numbers.*** They are also called ***discrete variables.***

Eg. *Number of mango trees in a garden, number of books in a Library, number of students in a class.*

Continuous variables are measured in ***decimal numbers***. The values are continuous.

Eg. *Height, weight, volume of O2 consumed, percentage of haemoglobin, length of leaves,* etc.

Variables

- **Quantitative**
  - **Discontinuous** (Discrete) Eg. *Number of students in a class*
  - **Continuous** Eg. *Length of leaves*
- **Qualitative** (Attribute) Eg. *Colour of flower*

Fig.1.1: Variables - types.

Qualitative variable is an ***unmeasurable*** variable. It is a descriptive variable. It is also called ***attribute***. Eg. *Colour of flower, colour of skin, wrinkled seed coat,* etc.

Notations Used in Biostatistics

S = Summation (pronounced as ***sigma***)

$\overline{X}$ = Mean

f = Frequency

σ = Standard deviation

χ^2 = Chi - square.

2. Collection of Data

The process of getting values and facts from an observation or experiment is called ***collection of data.***

The person collecting the data is called an ***investigator.***

The data collected by an investigator by his investigation is called ***primary data.*** If the data is obtained from already collected data, the data is called ***secondary data.***

Data is collected by the following methods:

1. Direct personal interview	*5.Experiments*
2. Indirect oral interview	*6.Census*
3. Through correspondents	*7.Sampling*
4. Questionnaire	

1. Direct Personal Interview

In direct personal interview, the investigator directly interview the individuals for collecting the data.

Eg. *The blood groups of a class of students can be obtained by personal interview.*

The data obtained by this method is ***original, reliable, authentic*** and ***accurate.***

It is more expensive and more time consuming.

2. Indirect Oral Interview

In indirect oral interview, the data is obtained from 'witnesses'.

Eg. *The number of smokers among college students can be obtained by interviewing the shop-keeper or friends.*

This method is ***economical.*** But the data may not be reliable.

3. Through Correspondents

In this method, data is collected through appointed persons called ***correspondents.*** News agents of news papers are typical correspondents. The informations, for news papers are collected by this method.

4. Questionnaire

Questionnaire is a *paper containing a set of questions to be answered by the individuals for collecting data.*

The questionnaire may be sent through post or through correspondents.

A questionnaire should possess the following characteristics:

1. The ***purpose*** of the questionnaire should be given on a covering letter.
2. The questionnaire should be ***small.***
3. The questions should be ***simple.***
4. '***Yes***' or '***No***' -type questions should be asked.
5. ***Multiple choice*** questions can be given.
6. The questions should be ***direct.***
7. Unambiguous questions should not be asked.
8. The questions should not hurt the sentiments.
9. Questions, related to age, politics, private life, etc. should be carefully worded.
10. Cross check questions may be included to bring out information about age, date of marriage, etc. For example, the year of marriage can be assessed by asking the class in which the first child is studying.

Model Questionnaire regarding Food habits of a Village

1.Name	:	
2.Address	:	
3.Father	:	
4.Sex	:	[M] [F]
5.Married	:	[Yes] [No]
6.Weight	:	
7.Height	:	
8.Food habit	:	[Veg.] [Non-Veg.]

5. Experiments

Data can be collected by doing experiments.

The amount of oxygen present in water samples can be collected by titration experiments.

Table 2.1: Data collected from an experiment.

Water Samples	Amount of O_2 in ml/l
Pond water	5.1
River water	4.8
Well water	3.7
Sea water	4.3

6. Census

Census is a method of ***collection of data.*** Census is the ***counting*** of all the members of a population one by one.

Eg.*The trees in a coconut grove are counted by census method.*

Human population is assessed by census.

The results obtained in census are reliable and accurate. Data is obtained from each and every individual.

It is very expensive. It consumes time and labour.

7. Sampling

Sampling is a method of ***collection of data.***

Sampling is the process of getting a representative fraction of a population.

Sample is the representative fraction of a population.

In sampling method a small group is taken from a large population. This small group is the ***sample.*** Analysis of the sample gives an ***idea*** of the population.

When the population is very large or infinite, sampling is the suitable method of data collection.

One rice is tested from a pot of boiling rice to arrive at a conclusion.

In an electric bulb factory, the bulbs are tested at intervals how long they will burn. If all are tested there is nothing left for selling.

One grape is tasted before buying a bunch of grapes.

The oxygen content of pond water can be found out by titrating just 100 ml of water.

The length of leaves of a neem tree can be calculated by measuring just 10 leaves.

The food habits of all students of a college can be understood by observing only about 100 students.

There are two types of sampling, namely

1. *Random sampling.*
2. *Non-random sampling.*

1. Random Sampling

Random sampling is a method of ***collection of data.***

In random sampling, a small group is selected from a large population ***without any aim or predetermination.*** The small group selected is called a ***sample.***

In this method, each item of the population has an equal and independent chance of being included in the sample.

The random sample is selected by ***lottery method.***

Each individual is given a number. The numbers are written on pieces of papers. The papers are put in a box. About 100 papers are picked out. These 100 individuals form a random sample. The analysis of 100 individuals gives an idea of the entire population.

Random sampling is of 3 types, namely

1. *Simple random sampling*
2. *Stratified random sampling*
3. *Cluster random sampling*

In ***simple random sampling,*** each individual of the population has an equal chance of being included in the sample. In this method the sample is selected by ***lottery method.***

In ***stratified random*** sampling, the population is divided into groups or strata on the basis of certain characteristics.

Then the samples are selected by simple random sampling.

For example we want to select a sample of 100 students from a college population of 1000 students consisting of 700 girls and 300 boys. The whole college population should be divided into two strata. One with 700 girls and other with 300 boys. Now by simple random sampling method select 70 girls from total of 700 girls and 30 boys from the total of 300 boys.

In ***cluster sampling,*** the whole population is divided into a number of relatively ***small clusters*** or groups. Then some of the clusters are randomly selected. For example, if we want to survey the general health of the college student in a state consisting of 5000 colleges. Here we consider each college as a cluster. Now we can randomly select several college and conduct the survey.

2. Non - Random Sampling

Non random - sampling is a method of ***collection of data.*** In this method, a sample *is collected from a large population based on the convenience, judgement and consideration of the investigator.*

In non-random sampling, each individual does not get a chance of being included in the sample.

Eg. *If 20 students are selected from a college of 1000 students, the investigator selects 20 representatives.*

Advantages of Sampling

1. Sampling is an economical method of data collection.
2. It saves time, expenditure and energy.
3. It is reliable.

Disadvantages

1. Sampling needs skill.
2. It needs experts.
3. All the individuals are not represented.

Primary Data

The data collected for the first time is called ***primary data***

The person collecting the data is called an ***investigator.***

Primary data gives the ***first hand*** information. It is ***original*** in nature. It is ***accurate*** and ***reliable.***

It consumes more money and time.

The amount of O_2 estimated by a student by titration is a primary data. The rainfall recorded by PWD department is a primary data.

Merits of Primary Data

1. It is the first hand information.
2. It is original.
3. It is accurate.
4. It is reliable.

Demerits

1. It consumes ***more money.***
2. It requires ***more time.***
3. It requires ***more labour.***
4. The investigator must be well trained.
5. Personal bias may creep in.

Secondary Data

The data obtained from already collected data is called ***secondary data.***

They are ***secondary*** in nature. They are ***economical.*** It is not ***reliable.***

The amount of O_2 estimated by a student by titration is a primary data. When this data is used by a teacher, it becomes a secondary data.

The rain fall data obtained from a PWD department is a secondary data.

Merits of Secondary data

1. They are obtained at minimum cost.
2. They require less labour.
3. They are quickly obtained.

Demerits

1. They are not reliable.
2. Errors may be there.
3. The method of collection is not known.
4. They are secondary in character.

Table.2.2: Comparison of primary and secondary data.

Sl. No	Primary Data	Secondary Data
1.	Primary data is the first hand information.	Secondary data is the second hand informtion.
2.	It is original.	It is not original.
3.	It is reliable.	It is not reliable.
4.	It consumes more money.	It is economical.
5.	It consumes more labour.	It consumes less labour.
6.	It consumes time.	It is quickly obtained.

ജൗജ

3. Classification of Data

Arranging data into groups according to their characteristics is called ***classification***. It is the ***processing*** of data.

Objectives of Classification

1. To simplify the data.
2. To condense the data.
3. To bring out the points of ***similarity*** and ***dissimilarity***.
4. To compare the data.
5. To bring out the ***cause*** and ***effect relationship***.
6. To prepare the data for tabulation.

Types of Classification

The data are classified in the following ways:

1.Geographical classification *4.Quantitative classification*
2.Chronological classification *5.Simple classification*
3.Qualitative classification *6.Manifold classification*

1. Geographical Classification

Arranging data according to the geographical ***location*** is called ***geographical classification.***

Eg. *Data of countries, states, districts*, etc.

The number of cows reared in different villages can be classified as follows:

Table 3.1: Geographical classification: Number of cows reared in three villages.

Name of Villages	No. of Cows
Village 1	7
Village 2	6
Village 3	9

2. Chronological Classification

Arranging data based on ***time*** is called ***chronological classification.***

Eg. *Data on years, months, weeks, days*, etc.

The coconut yield in four years in a garden is an example for chronological classification.

Table 3.2: Chronological classification: Coconut yield in 4 years.

Years	No. of Coconuts
1995	10000
1996	12000
1997	11000
1998	15000

3. Qualitative Classification

Arranging data based on ***qualities*** *or* ***attributes*** *is called* ***qualitative classification.***

Eg. *Arranging data based on colour, sex, literacy*, etc.

The fishes captured from a pond can be classified into males and females.

Table 3.3: Qualitative classification of fishes based on sex.

Sex	No. of Fishes
Males	250
Females	270

4. Quantitative Classification

Arranging data based on quantity is called ***quantitative classification.***

Eg. *Arranging data based on age, weight, height, price*, etc.

Table 3.4: Quantitative classification: Weight of five fishes captured from a pond.

Sl. No.	Weight in gms
1	8
2	6
3	7
4	9
5	8

5. Simple Classification

Arranging data into only two classes is called ***simple classification.***

Eg.*The students in a class are classified into two groups namely males and females or vegetarians and non-vegetarians or day scholars and hostellers,* etc.

Table 3.5: Simple classification: Students of a class are classified into groups.

Sex	No. of Students
Males	10
Females	20

6. Manifold Classification

Arranging data into many classes is called ***manifold classification.***

Eg.*The students in a class are classified into males and females. They are further classified into day scholars and hostellers. They are further subdivided into vegetarians and non-vegetarians.*(Table 3.6)

Methods of Classification

The unprocessed freshly collected data is called ***raw-data.*** *The data arranged in an order* is called ***array data.***

The raw data is classified in three methods:

1. *Individual series*
2. *Discrete series*
3. *Continuous series*

Table 3.6: Manifold classification of the students of a class based on sex, stay and food habits.

Food Habit	Males		Females	
	Day scholars	Hostellers	Day scholars	Hostellers
Vegetarians	4	2	6	4
Non-vegetarians	3	1	7	3

1. Individual Series

In individual series, the values are written individually.

For example, the weight of 10 fishes is written according to their ***serial number*** or in an ***ascending order*** or in ***descending order.***

Table 3.7: Classification of individual series. Weight of 10 fishes in gms arranged according to serial number.

Serial No.	1	2	3	4	5	6	7	8	9	10
Weight in gms	4	3	6	7	9	5	6	4	8	4

Table.3.8: Classification of individual series. Weight of 10 fishes in gms arranged in ascending order.

Weight of fishes in gms	3	4	4	4	5	6	6	7	8	9

Table.3.9: Classification of individual series.Weight of 10 fishes in gms arranged in descending order.

Weight of fishes in gms	9	8	7	6	6	5	4	4	4	3

2. Discrete Series

In discrete series, the data are given in groups. It is also called ***discontinuous series.***

In discrete series, the items having the same values are grouped together. For example in the above table, three fishes are having the same weight 4 gms. Similarly, 2 fishes are having the same weight 6 gms.So the data are arranged as follows:

Table.3.10: Classification of discrete series.

Weight in gms	3	4	5	6	7	8	9
No. of fishes	1	3	1	2	1	1	1

3. Continuous Series

In continuous series, the data are presented with ***class intervals.***

In this method, the data is divided into classes. The first class and the last class are fixed by seeing the lowest and highest values. The lowest and highest numbers of each class are called ***class limits.*** The lowest number is called ***lower limit*** and the highest number is called ***upper limit.***

The class limits are made in three methods, namely

1. *Exclusive method*
2. *Inclusive method*
3. *Open end class method*

In ***exclusive method,*** the class intervals are formed in the following method:

0	-	10		0	-	5
10	-	20	or	5	-	10
20	-	30		10	-	15

In this method, the value 10 is not included in the class 0-10, but included in the class 10-20. As *the upper limit number is excluded from the class,* it is called ***exclusive method.***

In the ***inclusive method,*** the class limits are formed in the following way:

0	-	9		0	-	4
10	-	19	or	5	-	9
20	-	30		10	-	14

Here, the upper limit value is included in the same class. Hence it is called ***inclusive method.***

The open-end class is formed in the following way:

Below	-	10	} 11
10	-	20	
Above	-	20	

In open-end class, the value does not give the upper limit and lower limit.

ജയജ

4. Tabulation

Arranging values in columns is called ***tabulation***. *A column of values* is called a ***table***. *Tabulation is a **presentation** of data.*

A table contains boxes called ***cells***. The cells are arranged in horizontal ***rows*** and ***vertical columns***.

A typical table has the following parts:

1. Table number — *5. Stub*
2. Title — *6. Body*
3. Head note — *7. Foot note*
4. Caption — *8. Source*

Table. 4.1: Food habits of III B.Sc .Zoology students.

Eggterians : Egg - eaters

(Table number) (Title) (Caption) (Head note)

Food habits	Number of students	
	Boys	Girls
(Stub) → Vegetarians	2	7
Eggterians (Egg-eaters)	1	2
Non-vegetarians	7	11

(Body)

Foot Note → Census was made during 2000-2001

Source → Source: Data collected by the class - teacher

Fig.4.1: Parts of a table.

The ***table*** has a number and it is given at the top.

The name of the table is called ***title***. It is given at the top.

Head note refers to the units of values given below the title.

The headings of vertical columns are called ***caption***.

The headings of horizontal columns are called ***stub***.

The values given in the horizontal and vertical columns are called ***body***.

Foot note is given below the table. It gives explanations on the values.

Source refers to source of information. It is given at the bottom of the table.

Tables are broadly classified into two types.

1. Simple tables
2. Complex tables

1. Simple Tables

In a simple table, only one characteristic is shown. Hence this type of table is also known as ***one-way table***. It has two factors placed in relation to each other. The following table shows the marks secured by students in a class test.

Table.4.2: Simple table.

Marks	No. of students
0 - 5	2
5 - 10	5
10 - 15	10
15 - 20	11
20 - 25	9
25 - 30	20

2. Complex Tables

In a complex table, more than two characteristics are shown. If there are two co-ordinate factors, the table is called a ***double table;*** if the number of co-ordinate group is three, it is called as ***treble table***. If it contains more than three co-ordinate factors, then it is called as ***multiple table***.

The marks of students can be classified according to the sex to get a ***double table***.

Table 4.3: Double table.

Marks	No. of Students	
	Male	Female

The males and females are further classified according to their residence into hostellers or day scholars. It is a case of ***treble table.***

Table 4.4: Treble table.

Marks	No. of Students			
	Male		Female	
	Hostellers	Day scholars	Hostellers	Day scholars

If they are again classified as belonging to different religions, nationalities, states, etc. It will be an example for ***multiple table.***

Table 4.5: Multiple Table.

Marks	Number of Students											
	Male						Female					
	Hostellers			Day Scholars			Hostellers			Day Scholars		
	Hindu	*Christian*	*Muslim*	*Hindu*	*Christian*	*Muslim*	*Hindu*	*Christian*	*Muslim*	*Hindu*	*Christian*	*Muslim*

5. Frequency Distribution

Frequency distribution is a *statistical table containing groups of values according to the number of times a value occurs.*

Frequency means the number of times a particular value occurs. It is the repetition of values.

In frequency distribution, the data are arranged in an order in a table.

In a ***frequency table*** the data are grouped into ***classes.*** The number of values in each class is called a ***frequency.***

There are 3 types of frequency distribution, namely.

1.Frequency distribution without class intervals.

2.Frequency distribution with class intervals.

3.Cumulative frequency distribution.

1. Frequency Distribution Without Class Intervals

Frequency distribution without class intervals is a table of values without classes. It is formed in the following way :

1. The data collected by an investigator is called raw data. Raw data is the ungrouped data. It is not in order. It is in a haphazard manner. The raw data does not give a clear picture.

2. *The raw data arranged in an order* is called ***array data.*** The data is arranged in ***ascending order*** or ***descending order.*** The array data gives an idea about the population.

Table.5.1: Raw data; Number of children in 30 families.

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

3.A table is formed with three vertical columns.

4. In the first column, the values of all the items are written in the ascending order.

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

Fig.5.1: Array data; number of children in 30 families. Ascending order.

No. of children	Tallies	Frequency
0	1111	4
1	~~1111~~	5
2	1111	4
3	1111	4
4	111	3
5	11	2
6	11	2
7	1	1
8	11	2
9	11	2
10	1	1
	Total	**30**

5. In the second column repetition of items are given in the form of ***tally marks*** (bars). For each variable one tally mark is given. They are written in groups of five tally marks each. The first four tally marks are written as straight lines and the fifth tally mark is crossed. This is the ***four*** and ***cross method.***

6. The tally marks for each item is counted and is given in the third column. It is the ***frequency.*** The frequency for each item is equal to the number of tally marks against each item.

7. At the bottom of the third column, the total frequency is given by adding all the frequencies.

2. Frequency Distribution With Class Intervals

Frequency distribution with class interval is a table of values with classes. It is constructed with ***class intervals.*** It is a frequency distribution of ***continuous series.***

It involves the following steps:

1. The data collected by the investigator is called ***raw data.***

2. The data are arranged in an ascending order. The arranged data is called ***array data.***

3. The data is divided into 5 to 15 groups called ***classes.***

4. The first class and the last class are fixed by seeing the lowest and highest values.

5. The lowest and highest numbers of each class are called ***class limits.*** The lowest number of a class is called ***lower limit.*** The highest number of a class is called ***upper limit.***

6. The class limit may be made in two methods, namely ***exclusive method*** and ***inclusive method.*** In the exclusive method, the class limits are formed in the following way:

0 - 10		0 - 5
10 - 20	or	5 - 10
20 - 30		10 - 15

In the exclusive method, the upper limit of one class will be the lower limit of the succeeding class. Here the values 10 and above and the values below 20 are included in the class 10 - 20. *As the upper limit value is excluded,* it is called ***exclusive method.***

In the inclusive method the class limits are formed in the following way:

0 - 9		0 - 4
10 - 19	or	5 - 9
20 - 30		10 - 14

In the inclusive method, values between 10 and 19 are included in the class 10 - 19. *As the upper limit value is included,* it is called ***inclusive method.***

7.The difference between the lower limit and upper limit of a class is called ***class interval.*** It is obtained by subtracting the lower limit from the upper limit. It is the width or size of a class. The size of the class interval should be an easy number like 5,10,20,100, etc.

8. The number of values falling in a class is called ***class frequencies.***

9. The middle value of the class is called the ***mid-value*** or class interval. The ***mid-value*** is calculated by dividing the sum of lower limit and upper limit by 2. For example the ***mid - value*** of the class 10 - 20 is 15.

$$\textbf{Mid - value} = \frac{\text{Lower limit + Upper limit}}{2} = \frac{10 + 20}{2} = 15$$

10.A table is drawn with three vertical columns.

11.The classes are entered in the first column.

12.In the second column, the values are marked in the form of ***tally marks.*** Tally marks are given in groups of five each. Four tally marks are marked straight and one is crossed. This is the ***four and cross method.***

Continuous raw data: Weight of 30 Tilapia in gms collected from a pond (Fractions are made into whole numbers).

21	8	17	16	12	21
7	12	35	31	38	23
37	20	29	23	13	13
18	31	26	30	31	26
24	24	9	10	26	14

Table 5.2: Continuous frequency distribution.

Class	Tally marks of Variables	Frequency
0 - 9	*111*	3
10 - 19	*1111 1111*	9
20 - 29	*1111 1111 1*	11
30 - 39	*1111 11*	7
	Total frequency	**30**

13.The number of tally marks in each class is counted and it is the ***frequency*** of that class. The frequency is entered in the third column.

14.The ***total frequency*** is obtained by adding all the frequencies and entered at the bottom of the table.

3. Cumulative Frequency Distribution

The ***cumulative frequency distribution*** is a statistical table where the frequencies of preceeding classes are added.

A cumulative frequency distribution for the weight of 30 fishes is constructed as follows:

1. A table of 3 columns is prepared.
2. Classes are marked in the first column.
3. Frequency is noted in the second column.
4. In the third column the frequency for the first class is entered as such .
5. For the second class, the sum of the frequencies of first class and second class are added and entered.

Table 5.3: Cumulative frequency distribution of weight of 30 fishes (less than method).

Weight	Frequency	Cumulative frequency
10 - 9	3	3
10 - 19	9	12
20 - 29	11	23
30 - 39	7	30

6. For other classes the frequencies are entered after adding the preceeding class frequency.

7. The cumulative frequency distribution containing increasing cumulative frequency is called ***less than cumulative distribution.***

8. The cumulative frequency distribution helps to find out the number of items below and above a particular weight. For example, it is clear from the table that there are 23 fishes having less than 29 gms.

6. Graphic Presentation of Data

Presenting data in the form of graphs is called ***graphic presentation of data.***

Graph

A graph is the geometrical image of a data.

A graph is a diagram consisting of lines of statistical data.

The graph is drawn on a ***graph paper.***

The graph has two intersecting lines called ***axes.***

The horizontal line is called ***X - axis***. The vertical line is called ***Y - axis.***

The point of intersection is called ***'O'.***

The ***'O'*** point is common to both X and Y axis. Hence the ***X*** axis is also called ***OX line*** and the Y axis is also called ***OY line.***

A suitable scale is given for each axis.

Usually independent variables are marked on the X- axis and dependent variables are marked on the Y - axis.

A ***title*** is given to a graph.

The values corresponding to X and Y axis are plotted on the paper.

The points are joined with straight or curved lines.

Table 6.1: Monthly fish landing in a pond.

Month	Jan	Feb	Mar	Apr	May	June
Catch in tonnes	20	25	30	10	35	40

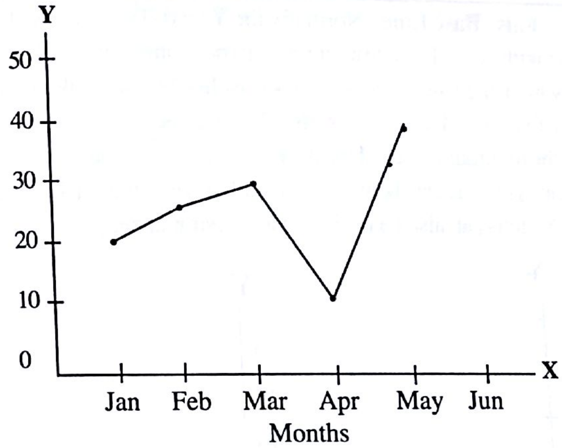


Fig.6.1: Graph showing fish landing in a pond.

The graphs are classified into two types, namely

1.Time series graphs

2.Frequency distribution graphs

1. Time Series Graphs

The graphs containing data that vary with time are called ***time series data.***

The time units such as year, month, week , days, hours, minutes, etc. are marked on the X - axis. The other units are marked on the Y- axis.

The units corresponding to X and Y axis are plotted on the graph paper. The points are joined by straight or curved lines.

The time series graphs are also called ***line graphs.***

Eg. *Monthly fish landing in a pond, yield of paddy, in different years.*

The time series graphs are of four types:

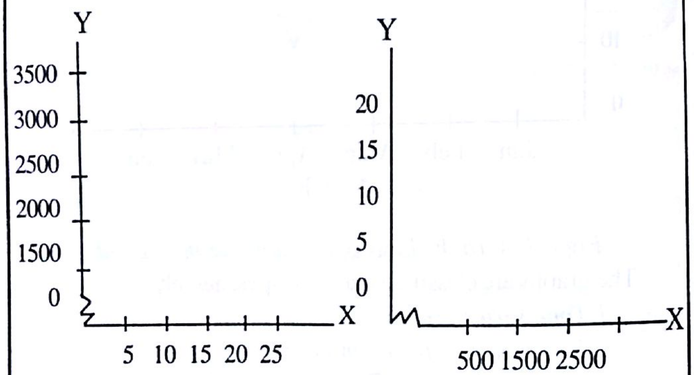
1.Graph of one variable

2.Graph of two or more variables

3. Range chart

4. Band graph

False Base Line : Normally the Y - axis (vertical) scale must start with zero. If the lowest value to be plotted on the Y - scale is very high, then the vertical scale is stretched by using false base line. Here the vertical scale is broken. The space between the origin zero and the minimum value of the dependent variable is omitted by drawing two zig - zag line between zero and the first unit on the Y- scale. The X - axis can also be broken in a similar manner.



Y - axis with false base. The space between 0 and 1500 is indicated by zig-zag line.

X - axis with false base. The space between 0 and 1500 is indicated by zig-zag line.

Fig.6.2: Graphs showing false base line.

1. Graphs of One Variable

When only one variable is to be represented, on the X- axis time is marked and on the Y - axis the value of variables are marked. The various points are joined by straight lines. (Fig.6.1)

2. Graphs of Two or More Variables

If the unit of measurement is the same, we can represent two or more variables on the same graph. For example, the data about the agricultural production over the past six years are represented by a single graph.

Table 6.2: Yield of grains in tonnes for six years.

Items	Yield (in tonnes)					
	1986	**1987**	**1988**	**1989**	**1990**	**1991**
Wheat	100	120	110	140	160	120
Pulses	130	110	120	110	140	160
Paddy	160	140	130	150	160	170

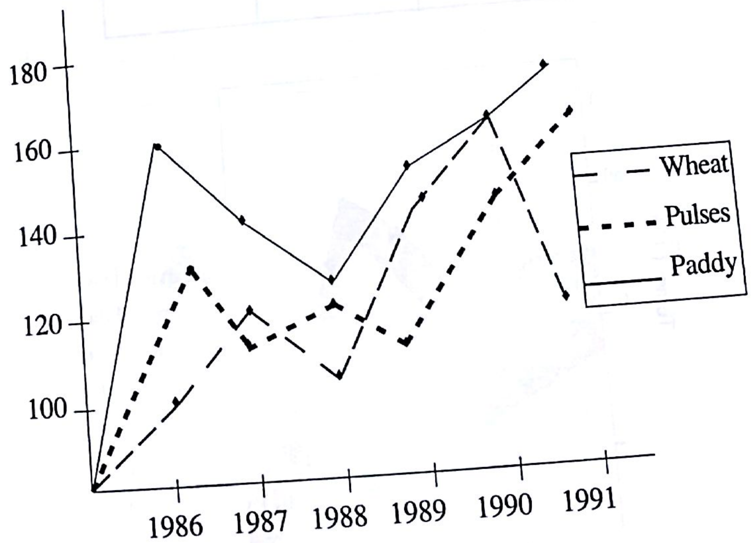


Fig.6.3: A graph of three variables.

3. Range Chart

It is used to exhibit, the minimum and maximum values of a variable. For example, to highlight the range of variation of the temperature on different days; the blood pressure readings of an individual in different days, etc.

Illustration 1 : *The following table shows the temperature for five days in a particular week.*

Table 6.3: High and low temperature for the consecutive five days.

Days	Temperature °C High	Low
Monday	50	30
Tuesday	40	25
Wednesday	55	45
Thursday	60	55
Friday	30	20

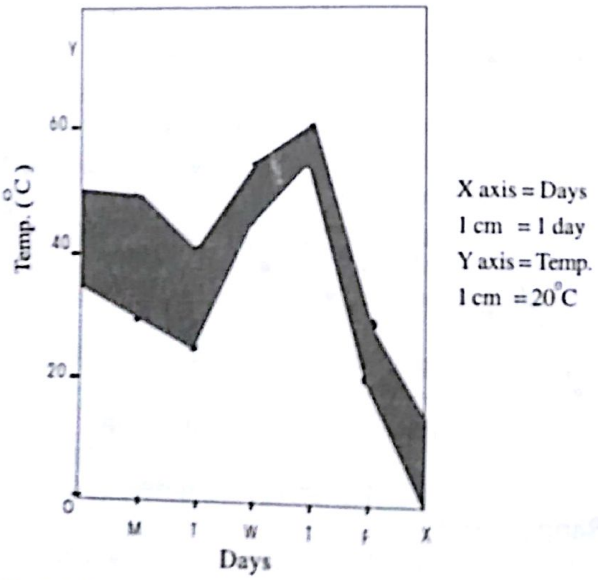


Fig.6.4:Range chart showing the range of temperature for five days.

Illustration 2 : *The following table shows the blood pressure of a hypertension patient for the consecutive seven days. Draw a range chart.*

Table 6.4:Blood pressure of a hypertension patient for the consecutive seven days.

Days	**Blood pressure (mm. Hg)**	
	Systolic	**Diastolic**
Sunday	99	71
Monday	126	74
Tuesday	108	72
Wednesday	122	68
Thursday	104	64
Friday	108	60
Saturday	116	70

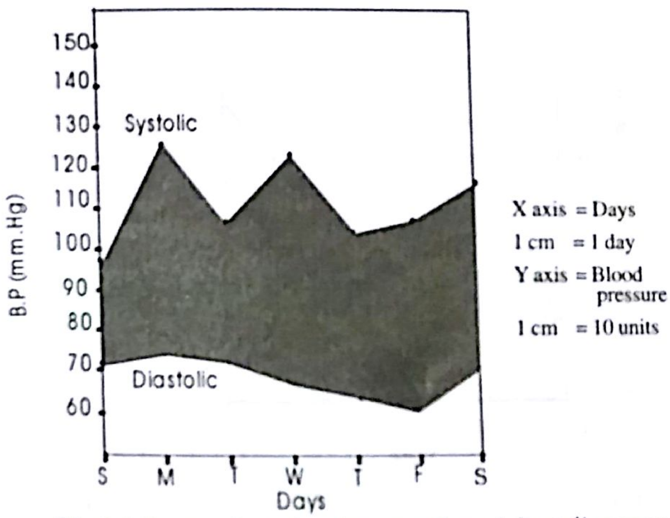


Fig.6.5:Range chart showing systolic and diastolic pressures of a hypertension patient for the consecutive seven days.

4. Band Graph / Layer Chart

In band graph, the various component parts are represented in the form of bands one above the other. It is also known as ***component part line chart*** or simply called ***layer chart.*** The various component parts are plotted one over the other and the gaps between the successive lines are filled by different shades or colours, etc., so that the chart has the appearance of a series of bands.

Illustration: *Draw band graph for the following data*

Table 6.5: Production of paddy, wheat and chillies in million tonnes.

Year	Production in million tonnes		
	Paddy	Wheat	Chillies
1987	20	10	5
1988	25	12	8
1989	10	15	10
1990	30	10	8
1991	40	20	15

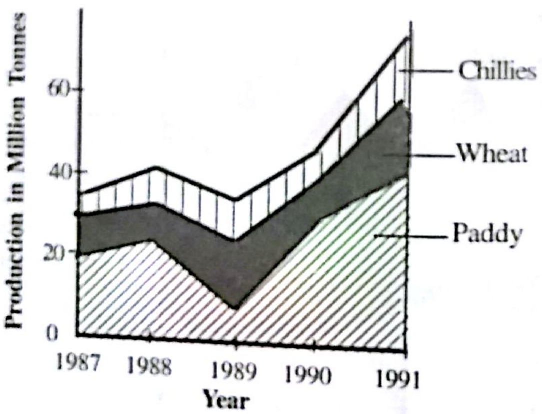


Fig.6.6: Band graph showing the production of paddy, wheat and chillies.

2. Frequency Distribution Graphs

Graphs obtained by plotting grouped data are called ***frequency distribution graphs.***

Frequency distribution is the ***grouped data.***

Frequency distribution can be represented in four types of graphs, namely:

1. *Histogram*
2. *Frequency polygon*
3. *Frequency curve*
4. *Ogives*

1. Histogram

Histogram is a ***graph*** *containing frequencies in the form of vertical rectangles.* It is an ***area diagram.***

It is a graphical presentation of ***frequency distribution.***

The X - axis is marked with ***class intervals.***

The Y - axis is marked with ***frequencies.***

Vertical rectangles are drawn as per the height of the frequency of each class. The rectangles are drawn without any gap in between.

Table 6.6. Marks of students.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No. of students	7	14	20	22	18	12	9	5	3	1

The width of the rectangle is equal to the range of the class.

The height of each rectangle is equal to the frequency of each class.

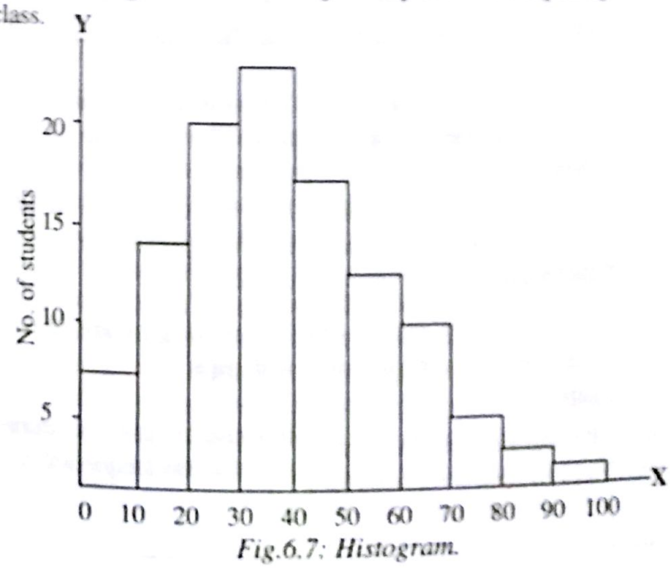


Fig.6.7: Histogram.

The histogram is a ***two dimensional*** diagram because the height and width of each rectangle are as per data.

The histogram is different from a bar diagram. The bar diagram is ***one dimensional*** because in a bar diagram the height alone is as per data and the width is not as per data.

Uses of Histogram

1. It gives a clear picture of the entire data.
2. It simplifies a complex data.
3. It is attractive and impressive.
4. It is easily memorized.
5. Median and mode can be located.
6. It facilitates comparison of two or more frequency distributions on the same graph.
7. It gives an idea of the pattern of distribution of variables in the population.

Table 6.7: Differences between histogram and bar chart.

Sl.No	Histogram	Bar Chart
1.	Histogram is a ***graph.***	Bar chart is a ***diagram.***
2.	Histogram is two dimensional.	Bar chart is one dimensional.
3.	The height and width of each rectangles are as per data.	The height alone is as per data. The width has no significance.
4.	In a histogram the total area of a rectangles represents the frequency.	In a bar chart, the ***height alone*** represents the frequency.
5.	The rectangles are drawn side by side without any gap.	The rectangles are drawn with gaps.
6.	Histogram is used to represent grouped continuous frequency distribution.	Bar chart is used to represent discrete frequency distribution.

Illustration: *Draw a Histogram for the following data.*

Table 6.8: Marks of the students.

Marks	0 -10	10-20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
No. of Students	5	7	10	15	13	10	6

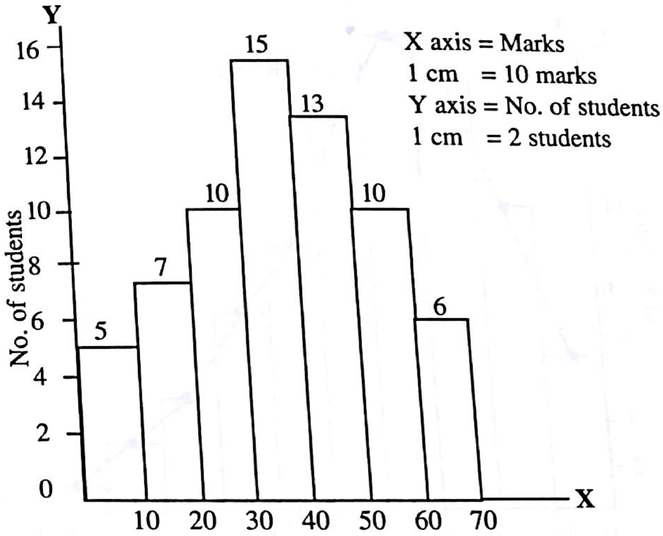


Fig.6.8: Histogram showing marks of the students.

2. Frequency Polygon

Polygon is a *histogram with straight lines joining the mid-points of the top of the rectangles.*

Polygon means a figure with ***many angles.***

It is an ***area diagram.*** Polygon is a ***graph.*** It is the ***graphical representation of frequency distribution.***

The X - axis is marked with ***class intervals.***

The Y - axis is marked with ***frequencies.***

Vertical rectangles are drawn as per height of the frequency of each class. The rectangles are drawn without any gap inbetween.

The width of each rectangle is equal to the range of the class.

The height of each rectangle is equal to the frequency of each class.

Table 6.9: Marks of students.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No.of Students	7	14	20	22	18	12	9	5	3	1

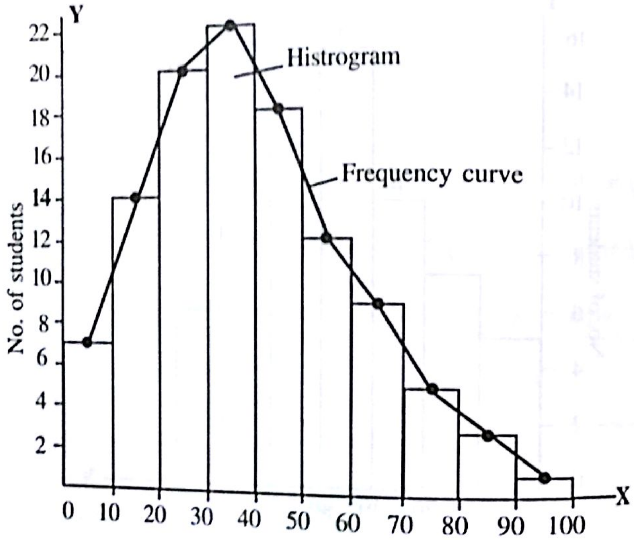


Fig.6.9: Frequency polygon.

The midpoints of the top of the rectangles are joined by ***straight lines.***

The area of the polygon is equal to the area of histogram.

Uses of Frequency Polygon

1. It gives a clear picture of the entire data.
2. It simplifies a complex data.
3. It is attractive and impressive.
4. It is easily memorized.
5. Median and mode can be located.
6. It facilitates comparison of two or more frequency distributions on the same graph.
7. It gives an idea of the pattern of distribution of variables in the population.

3. Frequency Curve

Frequency curve is a graph of frequency distribution where the line is smooth.

It is just like a frequency polygon. In the polygon the line is straight, but in the curve the line is smooth.

It is an ***area diagram*** .

It is the ***graphical representation of frequency distribution.***

The X- axis is marked with ***class intervals.***

The Y-axis is marked with ***frequencies.***

A ***histogram*** is drawn. The midpoints of the top of the rectangles are joined by a ***smooth line.***

The beginning and end of the curve should touch the X- axis at the midpoints of first and last class intervals.

Table 6.10: Marks of students.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No.of Students	7	14	20	22	18	12	9	5	3	1

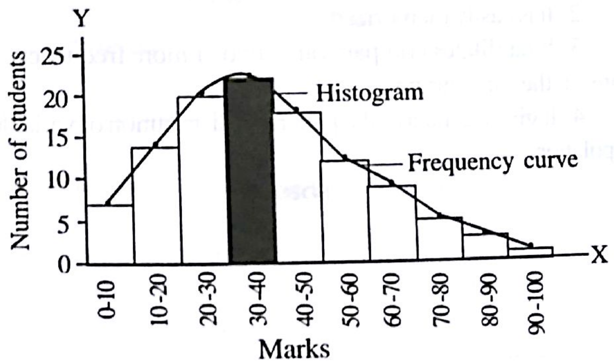


Fig.6.10: Frequency curve.

The area of the curve is equal to that of a histogram.

The frequency curve is divided into 3 types based on the shape of the curve. They are:

1. *Normal distribution curve.*
2. *Positively skewed distribution curve.*
3. *Negatively skewed distribution curve.*

The normal distribution curve is symmetrical and has an inverted ***bell - shape.***

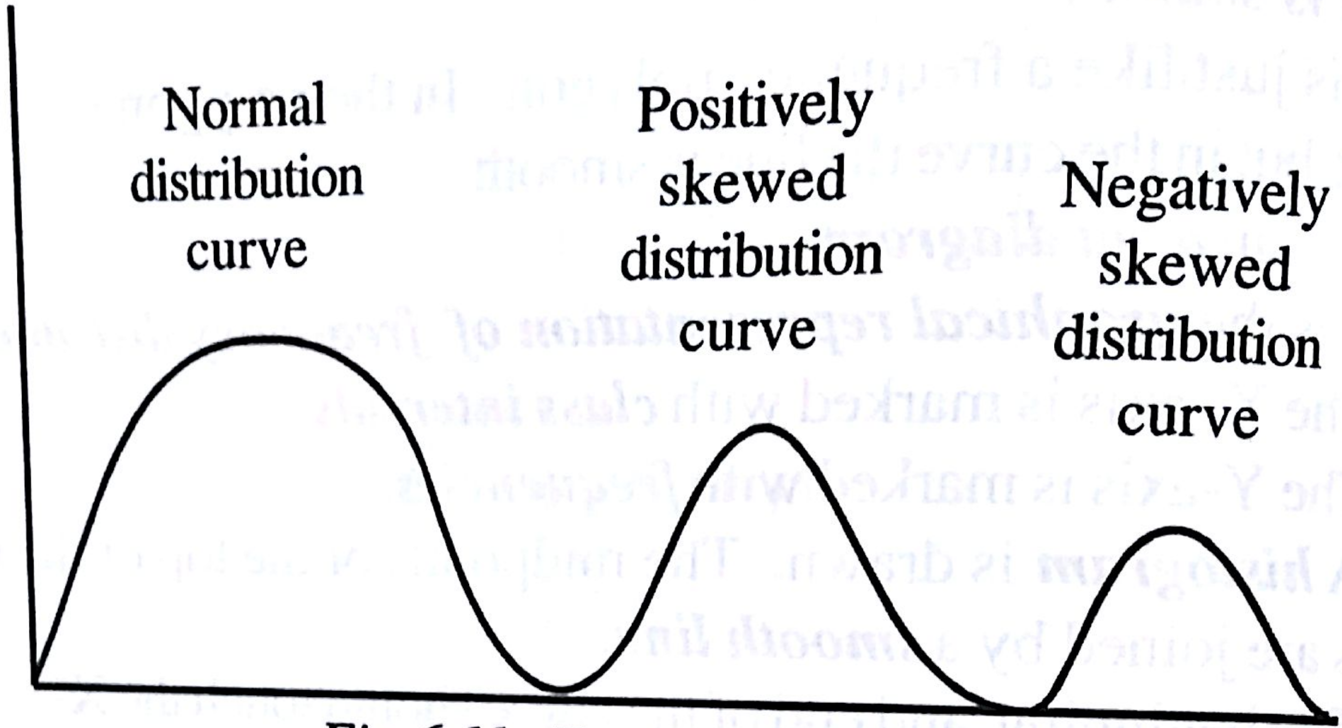


Fig.6.11: Frequency curves.

The positively skewed distribution curve is ***asymmetrical.*** The low values of the variables have the highest frequencies.

The negatively skewed distribution curve is also ***asymmetrical.*** High values of the variables have the highest frequencies.

Uses of Frequency Curve

1. It gives a clear picture of the entire data.
2. It is easily memorized.
3. It facilitates comparison of two or more frequency distributions on the same graph.
4. It gives an idea of the pattern of distribution of variables in the population.

8. Measures of Central Tendency (Average)

The average of a data is called a ***measure of central tendency.*** As the average lies in the middle of the lowest and highest values of a data, it is called a ***measure of central tendency***. All the values of a population cluster around average.

There are three types of averages, namely

1. Mean

2. Median

3. Mode

1. Mean

Mean is an ***average.***

It is a ***measure of central tendency .***

Mean is obtained by adding all the values and by dividing the total by the number of items.

Mean is represented by the symbol $\overline{X}$ (X bar).

Mean is of two types, namely ***arithmetic mean*** and ***geometric mean.***

The average obtained arithmetically is called ***arithmetic mean.*** Normally arithmetic mean is simply called ***mean.***

Geometric mean is obtained by geometric progression.

Mean can be obtained for ***ungrouped*** and ***grouped data.***

The mean can be calculated by two methods, namely-

1. Direct method

2. Indirect method or Assumed mean method.

In ***direct method***, mean is calculated for ungrouped data by the following formula:

$$\overline{X} = \frac{\Sigma x}{N}$$

$\overline{X}$ = Mean
Σ = Sum of
X = Value of an items
N = Total number of items

In ***indirect method***, mean is calculated for ungrouped data by the following formula:

$$\overline{X} = A \pm \frac{\Sigma d}{N}$$

$\overline{X}$ = Mean
A = Assumed mean
d = Deviation from assumed mean
Σd = Sum of d
N = Number of items

For grouped data, mean is calculated by the following formula:

$$\overline{X} = \frac{\Sigma fx}{\Sigma f} \quad \text{or} \quad \overline{X} = \frac{\Sigma fm}{\Sigma f}$$

$\overline{X}$ = Mean
X = Value of an item
Σ = Sum of
f = Frequency
m = Class mark
N = Number of items

Merits of Mean

1. Mean is well defined.
2. Calculation is easy.
3. All the items are considered for calculation.
4. It is based on each and every observation.
5. It is used for other statistical calculations.

Demerits of Mean

1. Mean will not be correct if certain values are very big or very small.
2. It may give false conclusion.
3. It gives absurd values. For example, the mean children in families is given sometimes as 2½ children.

Problem 1

The weight of 10 fishes in gms is given below. Find the mean by direct method and assumed mean method.

Sl. No	1	2	3	4	5	6	7	8	9	10
Weight	12	15	11	19	16	20	14	20	12	11

Direct Method

1. It is an ***ungrouped data.***
2. Add all the values to find Σx.
3. Divide the Σx by the total number of items.

$$\text{Mean } \overline{X} = \frac{\Sigma x}{N} \qquad \Sigma x = 150, \; N = 10$$

$$= \frac{150}{10}$$

Mean = 15 gms

2. Assumed Mean Method

(If calculator is not available, follow this method).

1. It is an ***ungrouped data.***
2. Draw a table with two vertical columns.
3. Write the values in the first column.
4. Consider the value of an item, as assumed mean(A).
5. Find out the deviation (d) difference of each value from the assumed mean (X - A).
6. Add all the deviations to get Σd.
7. Apply the formula:

Weight of fishes (x)	d = X - A = X - 16		
12	12 - 16 =	−4	
15	15 - 16 =	−1	
11	11 - 16 =	−5	
19	19 - 16 =		3
(16) A	16 - 16 =		0
20	20 - 16 =		4
14	14 - 16 =	−2	
20	20 - 16 =		4
12	12 - 16 =	−4	
11	11 - 16 =	−5	
		−21	+ 11
		$\Sigma d = -10$	

$$\overline{X} = A \pm \left(\frac{\Sigma d}{N}\right)$$

$$= 16 \pm \left(\frac{-10}{10}\right)$$

$$= 16 \pm (-1)$$

$$= 16 - 1 = 15 \text{gms}$$

Answer: Mean = 15 gms

Problem 2

Calculate mean for the following data:

Weight of fishes	6	7	8	9	10	11
Number of fishes	15	17	20	16	19	13

1. Direct Method

1. It is a ***grouped data*** without class interval.
2. Draw a table with three vertical columns.
3. Weight of fishes is given in the first column.

Weight (x)	Number of fishes (f)	fx
6	15	15 x 6 = 90
7	17	17 x 7 = 119
8	20	20 x 8 = 160
9	16	16 x 9 = 144
10	19	19 x 10 = 190
11	13	13 x 11 = 143
	$\Sigma f = 100$	$\Sigma fx = 846$

4. The frequency (number of fishes) is given in the second column.
5. Multiply the weight by the frequency to get fx and enter in the third column.
6. Add all the fx to get Σfx.
7. Divide the Σfx by the total number of fishes to find mean.

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

$$= \frac{846}{100}$$

$\overline{X}$ = Mean
Σ = Sum
f = Frequency
x = Weight

Answer: Mean = 8.46 gms

Problem 3

Find the mean for the following data:

Weight of fishes (x)	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60
Number of fishes (f)	15	17	16	19	13

1. Direct Method

1. It is a ***grouped data with class interval.***
2. A table with four vertical columns is drawn.
3. Weight of fishes (x, classes) is entered in the first column.
4. Class mark or mid - value (m) of each class is calculated and entered in the second column. It is calculated by adding the lower limit and upper limit of each class and dividing it by 2.
5. Enter the number of fishes (f) in the third column. Add the values to get Σf and enter at the bottom.
6. Find the *mf* by multiplying the mid value and frequency and enter in the last column.
7. Add the values of mf to get ***Σmf*** and enter at the bottom.

Weight of fishes (classes)	Mid - value (class mark) (m)	Number of fishes (f)	fm
10 - 20	$10+20 = \frac{30}{2} = 15$	15	15x15 = 225
20 - 30	$20+30 = \frac{50}{2} = 25$	17	25x17 = 425
30 - 40	$30+40 = \frac{70}{2} = 35$	16	35x16 = 560
40 - 50	$40+50 = \frac{90}{2} = 45$	19	45x19 = 855
50 - 60	$50+60 = \frac{110}{2} = 55$	13	55x13 = 715
		$\Sigma f = 80$	$\Sigma mf = 2780$

$$\bar{X} = \frac{\Sigma mf}{N}$$

$$= \frac{2780}{80}$$

Mean = 34.75 gms

$\bar{X}$ = mean
Σ = Sum of
m = mid value or class mark
f = Frequency
N = Number of fishes, Σf

2. Assumed Mean method

1. It is a grouped data with class interval.
2. A table with six vertical columns is drawn.
3. Enter the weight of fishes (x, classes) in the first column.
4. Enter mid - value of each class in the second column.

Class	Class mark (m)	Frequency (f)	Deviation d = m - A A = 45	fd (f x d)
10- 20	$10+20 = \frac{30}{2} = 15$	15	15-45=-30	- 450
20- 30	$20+30 = \frac{50}{2} = 25$	17	25-45=-20	- 340
30- 40	$30+40 = \frac{70}{2} = 35$	16	35-45=-10	- 160
40- 50	$40+50 = \frac{90}{2} = (45)$	19	45 - 45= 0	0
50- 60	$50+60 = \frac{110}{2} = 55$	13	55-45= 10	130
		$\Sigma f = 80$		130-950 $\Sigma fd = -820$

$$\bar{X} = A \pm \frac{\Sigma fd}{N}$$

$$= 45 - \frac{-820}{80}$$

$$= 45 - 10.25$$

$$= 34.75 \text{gms}$$

$\bar{X}$ = Mean
A = Assumed mean = 45
Σ = Sum of
f = Frequency
d = Deviation
N = Number of observations Σf

Answer: Mean = 34.75 gms.

5. Select an assumed mean (A) from the mid-value. Usually mid-value of the class having highest frequency is selected. Here it is 45.

6. Find the deviations(d) for each value from the assumed mean (X-A) and enter in the third column.

7. Enter frequency (*f*) in the fourth column.

8. Multiply the deviation with frequency (fxd = *fd*) to get *fd* and enter in the sixth column.

9. Add all *fd* to get Σfd and enter at the bottom.

10. Apply the formula.

Types of Mean

Mean** is the average.* It is a ***value of central tendency.

Mean is of three types, namely

1. Arithmetic mean

2. Geometric mean

3. Harmonic mean

1. Arithmetic Mean

***Arithmetic mean** is an average obtained arithmetically.*

It is the common average used in our day today life.It is commonly called 'mean'. It is represented as $\overline{x}$ (x bar).

It is calculated by adding all the values and dividing the sum by the total number of items.

The simple formula for the calculation of mean is

$$\overline{x} = \frac{\Sigma x}{N}$$

$\overline{x}$ = Mean (x bar)
Σ = Sum
x = Value
N = Number of items

Merits of Mean

1. Mean is well defined.
2. Calculation is easy.
3. All the items are considered for calculation.
4. It is based on each and every observation.
5. It is used for other statistical calculations.

Demerits of Mean

1. Mean will not be correct if certain values are very big or very small.
2. It may give false conclusion.
3. It gives absurd values. For example, the mean children in families is given sometimes as 2½ children.

2. Geometric Mean

Geometric mean is an ***average.*** *It is the antilog of the arithmetic average of log of different items in a series.* It is represented as *GM.*

It is used when dealing with ***ratios, indices*** or any such relative numbers.

GM is calculated by the following formula:

$$\text{Log GM} = \frac{\Sigma \log x}{N}$$

Merits of Geometric Mean

1. It is highly defined
2. It is based on all the items of the series.
3. It is useful in averages, percentages and ratios.
4. It is useful in economic and business statistics in index number construction.
5. It is used for further mathematical calculation.
6. It is not much affected by fluctuation of sampling.

Demerits

1. It is not easy for an ordinary man.
2. It is difficult to calculate.
3. If there are zeros or negative values in the data, GM cannot be calculated.

Solved Problems

The harvest of coconuts in 5 coconut trees is given below. Find the geometric mean.

2, 8, 5, 6, 4

$$GM = (x_1 \times x_2 \times x_3 \times x_4 \times x_5)^{1/n}$$

$$= (2 \times 8 \times 5 \times 6 \times 4)^{1/5}$$

$$GM = (1920)^{1/5}$$

$$\text{Log GM} = \frac{1}{5} \times \log 1920$$

$$= \frac{1}{5} \times 3.281$$

$$= 0.2 \times 3.281$$

$= 0.6562.$

GM = Antilog of 0.6562.

$= 4.531$

Answer Gm = ***4.531***

3. Harmonic Mean

Harmonic mean is the reciprocal of the arithmetic mean of the reciprocals of the individual observations.

It is represented by ***HM***.

It is an ***average.*** It is ***a measure of central tendency.***

It is used when dealing with rates and speeds.

It is calculated by the following formula

$$HM = \frac{N}{\Sigma 1/x}$$

N = Number of items

x = Value

Merits of Harmonic Mean

1. All the values are included in calculation.
2. It can be used for further mathematical calculations.
3. It gives better results in problems relating to rates and speeds.
4. It is not affected by sampling fluctuations.

Demerits

1. It is difficult to calculate.
2. It is difficult to understand.
3. It cannot be calculated when some values are in negatives or one value is 0.
4. It gives more weightage to smaller items.

2. Median

Median is the ***middle value*** *of a data when the values are arranged in the ascending or descending order.*

Median is an ***average.*** It is a ***measure of central value.***

Median divides a distribution into two equal halves. There will be equal number of items above and below the items.

Median is represented by the symbol ***md.***

Median can be calculated for ungrouped data and grouped data. The formula for the calculation of median for ungrouped data is

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

md = median

N = Number of items

If there are ***odd number*** of items, then median is calculated as follows:

$$Md = \frac{11+1}{2} = \frac{12}{2} = 6$$

Median = Value of the 6^{th} item, when the items are arranged in an ascending order.

When there are ***even number*** of items, the median falls between two items. The values of these two items are added and divided by 2 to get median.

If there are 12 items, the median is calculated as follows:

$$Md = \frac{12+1}{2} = \frac{13}{2} = 6.5$$

Median = Value of the 6.5^{th} item. It is obtained by adding the values of 6^{th} item and 7^{th} item and dividing the sum by 2.

Median of a grouped data with class interval can be calculated by the following formula:

$$Median = L + \left(\frac{\frac{N}{2} - cf}{f}\right) \times C$$

L = Lower limit of the median class.

N = Total frequency.

cf = Cumulative frequency prior to the median class.

C = Class interval of the median class.

f = Frequency of the median class.

Merits of Median

1. Simple to calculate.
2. It can be calculated without knowing the values of all the items.
3. It is unaffected by extreme values.
4. It can be calculated graphically.

Demerits

1. It is not based on all the items.
2. It is not used as a common average.
3. It is not used for further statistical calculation.

Table 8.1: Differences between Mean and Median.

	Mean	Median
1.	Mean is the average calculated by dividing the sum of all values by the total number of items.	Median is the middle value of the data when the data are arranged in an ascending order.
2.	Mean is represented by $\overline{x}$ (x-bar).	Median is represented as *Md.*
3.	All the items are considered for the calculation of mean.	All the items are not considered.
4.	Mean can be calculated by the simple formula. $\frac{\Sigma x}{N}$	Median is calculated by $\left(\frac{N+1}{2}\right)^{th}$ item.
5.	It is affected by extreme values.	Median is unaffected by extreme values.
6.	Mean is a mathematic average.	Median is a positional average.

Problem 1

Find the median weight of fishes from the following data:

Serial No.	1	2	3	4	5
Weight in gms	12	15	11	19	16

1. It is an ungrouped data with odd number of items.
2. The values are arranged in the ascending order.

Sl.No	1	2	3	4	5
Weight	11	12	15	16	19

3. Apply the formula:

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

$$= \frac{5+1}{2}$$

$$= \frac{6}{2}$$

= Value of 3rd item

Answer: Md = Value of 3rd item = 15 gms.

Problem 2

Find the median of the following data:

Sl.No	1	2	3	4	5	6
Weight of fish in gms	11	12	15	16	19	17

1. It is an ungrouped data with even numbers.
2. The data are arranged in an ascending order.

S.l No	1	2	3	4	5	6
Weight	11	12	15	16	17	19

3. Apply the formula:

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

$$= \frac{6+1}{2}$$

$$= \frac{7}{2}$$

= Value of 3.5^{th} item

So, median is in between 3^{rd} and 4^{th} items.

Value of 3^{rd} item = 15

Value of 4^{th} item = 16

$$\therefore \text{Median} = \frac{15+16}{2}$$

$$= \frac{31}{2}$$

Median = 15.5 gms.

Answer: Median = 15.5 gms.

Problem 3

Find the median weight of fishes from the following data:

Weight of fish in gms	6	7	8	9	10	11
Number of fishes	15	17	20	16	19	13

1. It is a grouped data without class interval.
2. Draw a table with three vertical columns.
3. Enter the weight of fishes in the first column.
4. Enter the number of fishes (N) in the second column.
5. Find out cumulative frequency (cf) and enter in the third column.
6. Apply the formula.

Weight of fishes in gms	Number of fishes (N)	Cumulative frequency (cf)
6	15	15
7	17	32
8	20	52
9	16	68
10	19	87
11	13	100
	N = 100	cf = 100

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

$$= \frac{100+1}{2}$$

$$= \frac{101}{2}$$

= 50.5^{th} item

In the data, 50.5^{th} item lies in between 32 and 52 of cumulative frequency.

So, take the highest cumulative frequency 52 and its corresponding weight 8 gms.

∴ Median = 8 gms.

So, median weight of fishes is 8 gms.

Problem 4

Find out the median weight of fishes from the following data:

Weight in gms	10-20	20-30	30-40	40-50	50-60
No.of Fishes	15	17	16	19	13

1. It is a grouped data with class interval.
2. Draw a table with three vertical columns.
3. Find the *cumulative frequency.*
4. Find the median class using the formula N/2 = 80/2 = 40.
5. The cumulative frequency 40 falls in between 32 and 48.
6. So, the class 30-40 have the highest frequency 48 is selected as the median class.
7. The lower limit of the median class is 30.
8. The actual frequency of the median class is 16.

Class	Frequency	Cumulative Frequency
10 - 20	15	15
20 - 30	17	32
30 - 40	←16→	48
40 - 50	19	67
50 - 60	13	80

9. Cumulative frequency of the class prior to median class is 32.
10. Apply the formula:

$$\text{Median} = L + \left(\frac{N/2 - Cf}{f}\right) \times C$$

L = Lower limit of median class = 30
N = Total frequency = 80
Cf = Cumulative frequency prior to median class = 32
C = Class interval of the median class = 10
f = Frequency of the median class = 16

$$\text{Median} = 30 + \left(\frac{\frac{80}{2} - 32}{16}\right) \times 10$$

$$= 30\left(\frac{40-32}{16}\right) \times 10$$

$$= 30 + \left(\frac{8}{16}\right) \times 10$$

$$= 30 + (0.5 \times 10)$$

$$= 30 + 5$$

Answer: Median weight of fishes = 35 gms.

3. Mode

***Mode** is the value of the variable which occurs most frequently in a distribution.*

The value which occurs many times in the table is the mode.

It is represented by the letter ***Mo.***

Mode is an ***average.*** It is a ***positional average.*** It is a ***measure of central value.***

When a data has one concentration of frequency, it is called ***unimodal.*** When it has two concentrations, it is called ***bimodal.*** When it has 3 concentrations of frequency, it is called ***trimodal.***

Mode can be calculated for ***ungrouped data*** and ***grouped data.***

To find out mode of an ungrouped data, the values are arranged in an ascending order. The value which occurs maximum number of times is the mode.

18, 21, 23, 23, ***25,25,25,*** 27, 29, 29.

In the above data, 25 occurs maximum number of times. So ***25*** is the ***mode.***

The mode of a ***discrete distribution*** is the value of the variable which shows maximum frequency.

No.of Count trees	10	11	12	13	14	15	16
No. of coconuts (Frequency)	8	4	12	24	26	7	11

In the above table, ***14*** is the mode because the values are maximum here.

Merits of Mode

1. Mode can be easily found out.
2. No calculation is needed.
3. It is not affected by extreme values.
4. It can be calculated graphically.

Demerits of Mode

1. It is not clearly defined.
2. It is not based on all observations.
3. It is not reliable.
4. It is not used for further statistical calculation.

Problem 1

Find the mode from the following data showing weight of fishes:

Number of fishes	10	14	11	12	13	15	14

1. It is an ungrouped data.
2. Arrange in an ascending order.

10	11	12	13	14	14	14	15

3. The value 14 is repeated 3 times So, 14 is the mode.

Problem 2

Find the mode from the following data showing weight of fishes.

Weight of fishes in gms	18	19	20	21	22	23
Number of fishes	30	35	28	26	24	25

1. It is a grouped data without class interval.
2. The mode is fixed by inspection.
3. There is a maximum number of 35 fishes with the weight 19 gms. Therefore 19 gms is the mode.

Problem 3

Calculate mode for the following data:

Weight of fishes	0 -10	10-20	20-30	30-40	40 -50	50-60
No. of fishes	2	3	6	4	1	2

1. It is a grouped data with class interval.
2. The class containing the highest frequency is the ***modal class***.

Here the modal class is 20 - 30.

3. Apply the formula.

$$\text{Mode} = L + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) \times C$$

L = Lower limit of the modal class = 20

Δ_1 = The difference between the frequency of the modal class (f_1) and the frequency of preceding modal class (f_0); $\Delta_1 = f_1 - f_0$.

Δ_2 = The difference between the frequency of the modal class (f_1) and the frequency of the succeeding modal class (f_2); $\Delta_2 = f_1 - f_2$.

C = Class interval of the modal class = 10

f_1 = Frequency of modal class = 6

f_0 = Frequency of proceding modal class = 3

f_2 = Frequency of succeeding modal class = 4

Weight		Number	
	0 - 10	2	
Preceding class	→ 10 - 20	3	→ fo
Modal class	→ 20 - 30	(6)	→ f1
Succeeding class	→ 30 - 40	4	→ f2
	40 - 50	1	
	50 - 60	2	

$$\text{Mode} = 20 + \left(\frac{6-3}{6-3 + 6-4}\right) \times 10$$

$$= 20 + \left(\frac{3}{3+2}\right) \times 10$$

$$= 20 + \left(\frac{3}{5}\right) \times 10$$

$= 20 + (0.6 \times 10)$

$= 20+6$

Answer:Mode = 26 gms.

Problems

1. Find the harmonic mean for the following data: 10, 20, 30, 40.

x	10	20	30	40	
$\frac{1}{x}$	$\frac{1}{10}$	$\frac{1}{20}$	$\frac{1}{30}$	$\frac{1}{40}$	
$\frac{1}{x}$	0.1	0.05	0.033	0.025	$\Sigma \frac{1}{x}=0.208$

$$HM = \frac{N}{\Sigma 1/x}$$

$$= \frac{4}{0.208}$$

$$= 19.27$$

Answer: ***HM = 19.27***

2. The following is the weight of 10 fishes. Find out harmonic mean.

Weight in gms	3	4	5	6
Frequency	2	4	3	1

$$HM = \frac{\Sigma f}{\Sigma fx1/x}$$

Wt. in gms (x)	Frequency (f)	1/x	F x1/x
3	2	0.33	0.67
4	4	0.25	1.00
5	3	0.20	0.60
6	1	0.17	0.17
	Σf = 10		Σfx1/x =2.44

$$HM = \frac{10}{2.44}$$

$= 4.1$ cm

Answer HM = ***4.1 cm.***

3. The following table gives the length of fishes. Calculate harmonic mean.

Length in cm	6-8	9-11	12-14	15-17
No. of fishes	2	4	3	1

$$HM = \frac{\Sigma f}{\Sigma \left(fx\frac{1}{m}\right)}$$

Length in cm	**Frequency (f)**	**Mid value (m)**	$\frac{1}{m}$	$fx\frac{1}{m}$
6-8	2	7	0.143	0.286
9-11	4	10	0.1	0.4
12-14	3	13	0.077	0.231
15-17	1	16	0.063	0.063
	Σf = 10			$\Sigma fx\frac{1}{m} = .950$

$$HM = \frac{10}{0.950}$$

$= 10.53$

Answer HM = ***10.53 cm.***

9. Measures of Dispersion

Measure of dispersion *is the deviation of the individual values around the central value of a data.*

It is also called ***measure of variation.*** It is the measure of variation of the items.

The measure of variation is easily understood by seeing the yield of coconuts in 3 coconut trees for 4 seasons.

Table.9.1: Field of coconut in all the trees.

	Yield of coconut in four seasons				**Total**	**Average**
Tree 1	10	10	10	10	40	10
Tree 2	12	8	9	11	40	10
Tree 3	22	2	13	3	40	10

The average yield is 10 in all the trees.

For the first tree, the value of individual tree is similar to average.

For the second tree, the individual values slightly differ or scatter from the average.

For the third tree, the individual values differ much from the average. This difference is called ***variation*** or ***dispersion.***

The measure of variation has the following significances:

1. It helps to determine the reliability of an average.
2. It serves as a basis for the control of variability.
3. It helps to compare two or more data regarding variability.

The measure of dispersion is of the following types:

1.Range
2.Quartile deviation
3.Mean deviation
4.Standard deviation

1. Range

Range is *the difference between the lowest value and highest value of a set of data.*

Range = Largest value - Smallest value

$$R = L - S$$

Range is a ***measure of dispersion.*** It gives an idea of the degree of variability in a set of values.

The yield of a coconut tree on 5 occasions is 22, 8, 12, 5 and 13. The range is calculated as follows:

Largest value = 22
Smallest value = 5
Range = L - S
= 22 - 5
Range = 17

When the range is divided by the sum of the extreme values, the resulting figure is called ***co-efficient of range.***

$$\text{Coefficient of Range} = \frac{\text{Largest value - Smallest value}}{\text{Largest value + Smallest value}}$$

$$\text{Coefficient of Range} = \frac{L - S}{L + S}$$

Merits of Range

1. It is easy to calculate.
2. It helps in studying variation in a set of values.
3. It helps to study the variability between data.
4. It helps to study increase in price of commodities.

Answer: Q = [illegible]

3. Mean Deviation

Mean deviation** is the average of the deviations of the individual values from the mean.* It is also called ***average deviation. *It is represented by **MD**. The negative sign is ignored*

Each value in a data deviates from the mean. The difference is called ***deviation.*** The difference above the mean is called ***negative deviation.*** The difference below the mean is called ***positive deviation.***

The deviation is zero, when the value coincides with the mean.

Mean deviation is calculated by the following formula:

$$MD = \frac{\Sigma D}{N} \text{ for ungrouped data}$$

$$MD = \frac{\Sigma fD}{N} \text{ for grouped data}$$

D = Deviation

f = Frequency

N = Total number of items

Merits and Demerits

1. It is easy to calculate.
2. It is based on all observations.
3. It is reliable.
4. It is not significant because the negative sign is ignored.

Coefficient of Mean Deviation

Coefficient of mean deviation is the ratio of mean deviation and mean.

$$\text{Coefficient} = \frac{MD}{Mean}$$

Problem : *Find out the mean deviation and coefficient of mean deviation for the following data obtained by the weight of fishes in gms.*

8, 10, 12, 13, 14, 16, 18

1. The mean is calculated using the formula.

$$\overline{X} = \frac{\Sigma x}{N}$$

$$= \frac{91}{7}$$

$$\overline{X} = 13 \text{ gms.}$$

2. The deviation (difference) between each value and the mean is calculated.

Value	8	10	12	13	14	16	18	
Mean $\overline{x}$	13	13	13	13	13	13	13	
Deviation	-5	-3	-1	0	1	3	5	18
							$\Sigma D = 18$	

3. Mean deviation is calculated using the formula:

$$MD = \frac{\Sigma D}{N}$$

$$= \frac{18}{7}$$

MD = 2.57 gms.

4. Coefficient of mean deviation is calculated by the formula:

$$\text{Coefficient of MD} = \frac{MD}{Mean}$$

$$= \frac{2.57}{18}$$

$$= 0.14 \text{ gms.}$$

Answer: MD = 2.57 gms.

Co. MD = 0.14 gms.

Problem 2 : *The following table gives the weight of 25 fishes. Calculate mean deviation and coefficient of mean deviation.*

Weight in gms	6	7	8	9	10	11
Number of fishes	2	4	5	4	3	1

1. A table with 5 vertical columns is drawn.
2. The weight (x) and number of fishes (f) marked in the first two columns.

3. The weight is multiplied by the number of fishes to get fx and it is entered in the third column.

4. The Σfx is obtained by adding all fx values and entered at the bottom of the third column.

5. Mean is calculated using the following formula.

$$\overline{X} = \frac{\Sigma fx}{N} \qquad \Sigma fx = 157 \quad N = 19$$

$$\overline{X} = \frac{157}{19}$$

$\overline{X}$ = 8.26=8.3gms

Weight X	No. of fishes f	fx	$\overline{X}$-X=D	fD
6	2	12	6-8.3 = 2.3	2x2.3 = 4.6
7	4	28	7-8.3 = 1.3	4x1.3 = 5.2
8	5	40	8-8.3 = 0.3	5x0.3 = 1.5
9	4	36	9-8.3 = 0.7	4x0.7 = 2.8
10	3	30	10-8.3 = 1.7	3x1.7 = 5.1
11	1	11	11-8.3 = 2.7	1x2.7 = 2.7
	N = 19	Σfx =157	ΣD = 9.0	ΣfD = 21.9

6. Deviation (D) is calculated by subtracting X and X entered in the fourth column.

7. ΣD is obtained by adding all deviations and entered at the bottom of 4[th] column.

8. Deviation (D) is multiplied with (f) to get (fD) and entered in the 5[th] column. ΣfD is obtained by addition and entered at the bottom of the 5[th] column.

9. Mean deviation is calculated by the following formula.

$$MD = \frac{\Sigma fD}{N} \qquad \Sigma fD = 21.9 \quad N = 19$$

$$= \frac{21.9}{19}$$

MD = 1.15 gms

$$\text{Co.of MD} = \frac{MD}{Mean} \qquad MD = 1.15 \quad Mean = 8.3$$

$$= \frac{1.15}{8.3}$$

Co. of MD = 0.14gms.

Answer: MD = 1.15 gms.

Co.of MD = 0.14 gms.

Problem 3 : *Calculate the mean deviation and coefficient of mean deviation for the following data which shows the weight of fishes in grams.*

Weight of fishes in gms	0-10	10-20	20-30	30-40	40-50
No. of fishes	5	7	8	4	3

1. Draw a table with 7 vertical columns.
2. Enter the weight in the first column.
3. Enter the mid point (x) in the second column. Σx is calculated by addition and entered at the bottom.
4. Number of fishes (f) is entered in the third column. Σf is obtained by addition and entered at the bottom.
5. The weight (x) is multiplied by f to get fx and entered in the 4[th] column fx values are added to get Σfx and entered at the bottom.
6. Mean is calculated by the following formula:

$$\overline{X} = \frac{\Sigma fx}{\Sigma f} \qquad \Sigma fx = 605 \quad \Sigma f = 27$$

$$= \frac{605}{27}$$

Mean = 22.4 gms.

7. Now deviation (D) is calculated by subtracting the X value from the mean and entered in the 5th column.

Weight	Midpoint (X)	Number of fishes (f)	fx	$D=X-\overline{X}$ $D=x-22.4$	fD
0-10	5	5	25	17.4	87
10-20	15	7	105	7.4	51.8
20-30	25	8	200	2.6	20.8
30-40	35	4	140	12.6	50.4
40-50	45	3	135	22.6	67.8
		$\Sigma f = 27$	$\Sigma fx = 605$	62.6	277.8

8. The deviation is multiplied by f to get fD and entered in the 6th column. fD values are added to get ΣfD and entered at the bottom.

9. Mean deviation is calculated by applying the following formula:

$$MD = \frac{\Sigma fD}{N} \qquad \Sigma fD = 277.8 \qquad N = 27$$

$$= \frac{277.8}{27}$$

MD = 10.29 gms

$$\text{Coefficient of Mean Deviation} = \frac{MD}{Mean} \qquad MD = 10.29 \qquad Mean = 22.4$$

$$= \frac{10.29}{22.4}$$

= 0.46 gms.

Answer: Mean Deviation = 10.29 gms.

Coefficient of Mean Deviation = 0.46 gms.

Standard Deviation

Standard deviation is a *measure of deviation. It is defined as the square root of mean of the squares of deviations from the mean.*

Standard deviation is represented by **SD** or **δ** (Sigma)

Standard deviation is calculated by the following formula.

$$SD = \sqrt{\frac{\Sigma(x-\overline{x})^2}{N}}$$

ΣD = Standard deviation

x = Value of an observation

$\overline{x}$ = Mean

N = Number of items

S = Sum of

Merits

1. Standard deviation is rigidly defined.
2. All the values are considered for calculation.
3. Squaring makes the negative signs into plus signs.
4. It is less affected by sampling.
5. It is used for further calculations.

Demerits

1. Calculation is complex.
2. It is affected by the value of each item.

Problem 1: *Calculate standard deviation for the following data:*

Weight of fishes in gms	8, 6, 7, 5, 6, 10, 8, 6, 7, 7

1. Draw a table with three vertical columns.
2. Enter the values in the first column. Add the total to get ΣX and enter at the bottom.
3. Find out mean using the following formula:

$$\overline{x} = \frac{\Sigma X}{N} \qquad X = \text{Values of items} \qquad N = \text{Number of items}$$

$$= \frac{70}{10}$$

$\overline{x}$ = 7 gms.

4. Find out the deviation of each value from mean $(x-\overline{x})$ and enter in the second column.

5. Square the deviations $(x-\bar{x})^2$ and enter in the last column. Add the squared deviations to get $\Sigma(x-\bar{x})^2$ and enter at the bottom.

Weight in gms (x)	$(x-\bar{x})$ $\bar{x} = 7$	$(x-\bar{x})^2$
8	1	1
6	-1	1
7	0	0
5	-2	4
6	-1	1
10	3	9
8	1	1
6	-1	1
7	0	0
7	0	0
$\Sigma x = 70$		$\Sigma(x-\bar{x})^2 = 18$

6. Calculate standard deviation using the formula

$$SD = \sqrt{\frac{\Sigma (x-\bar{x})^2}{N}}$$

$$= \sqrt{\frac{18}{10}}$$

$$= \sqrt{1.8}$$

$$= 1.3gms$$

Answer: SD = 1.3gms

Problem 2: *Calculate standard deviation for the following data which shows the length of fishes.*

Length in cm.	5	6	7	8	9	10	11
No.of fishes	1	2	5	5	3	3	1

1. Draw a table with six vertical columns.
2. Write the length of fishes (x) in the first column.
3. Write the number of fishes, frequency (f) in the second column. Add the values to get Σf and enter at the bottom.
4. Multiply f with x to get fx and enter in the third column. Add the products to get Σfx and enter at the bottom.
5. Calculate mean $(\bar{x})$ using the following formula.

$$\bar{x} = \frac{\Sigma fx}{N}$$

$$= \frac{160}{20}$$

$$\bar{x} = 8 \text{ cm}$$

6. Subtract the values from $\bar{x}$ for to get the deviations $(x-\bar{x})$ and enter in column number 4.

Length in cm(x)	Frequency (f)	fx	$(x-\bar{x})$ $\bar{x}= 8$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
5	1	5	-3	9	9
6	2	12	-2	4	8
7	5	35	-1	1	5
8	5	40	0	0	0
9	3	27	1	1	3
10	3	30	2	4	12
11	1	11	3	9	9
	$\Sigma f=20$	$\Sigma fx = 160$			$\Sigma f(x-\bar{x})^2 = 46$

7. Square the deviations to get $(x-\bar{x})^2$ and enter in the 5th column.
8. Multiply $(x-\bar{x})^2$ with f to get $f(x-\bar{x})^2$ and enter in the 6th column. Add all the values to get $\Sigma f(x-\bar{x})^2$ and enter at the bottom.
9. Calculate standard deviation using the following formula:

$$SD = \sqrt{\frac{\Sigma f (x-\bar{x})^2}{\Sigma f}}$$

$= \sqrt{\frac{46}{20}}$

$= \sqrt{2.3}$

Answer: SD = 1.5 cm.

Problem:3: *Calculate standard deviation for the following data which shows the weight of fishes in gms.*

Weight in gms	0-10	10-20	20-30	30-40	40-50
Number of fishes	2	4	6	5	3

1. Draw a table with seven vertical columns.
2. Write the classes (x) in the first column.
3. Find the mid value (m) of each class and enter in the second column.
4. Write the frequency (f) in the third column. Add the values to get Σf and enter at the bottom.
5. Multiply the frequency with mid value to get fm and enter in the 4th column. Add the products to get Σfm and write at the bottom.
6. Find out mean ($\bar{x}$) using the following formula.

$$\bar{x} = \frac{\Sigma fm}{\Sigma f} = \frac{530}{20}$$

$\bar{x} = 26.5$gms.

Classes (x)	Midvalue (m)	Frequency (f)	fm	(m-$\bar{x}$) $\bar{x} = 26.5$	(m-$\bar{x}$)^2	f(m-$\bar{x}$)^2
0-10	5	2	10	21.5	462.25	924.5
10-20	15	4	60	11.5	132.25	529.0
20-30	25	6	150	1.5	2.25	13.5
30-40	35	5	175	-8.5	72.25	361.25
40-50	45	3	135	-18.5	342.25	1026.75
		Σf = 20	Σfm=530			Σf(m-$\bar{x}$)^2 2855.00

7. Subtract mid values (m) from $\bar{x}$ to get (m-$\bar{x}$)
8. Square (m-$\bar{x}$) values to get (m-$\bar{x}$)^2
9. Multiply (m-$\bar{x}$)^2 values with f to get f(m-$\bar{x}$)^2. Add all the values to get Sf(m-$\bar{x}$)^2 and enter at the bottom of the last column.
10. Calculate ΣD by applying the formula

$$SD = \sqrt{\frac{\Sigma f(m-\bar{x})^2}{\Sigma f}}$$

$$= \sqrt{\frac{2855}{20}}$$

$$= \sqrt{142.75}$$

Answer:SD = 11.95 gms.

ഃഃഃ