MODULE IV

MEANING AND DEFINITIONS OF STATISTICS

The word statistics is derived from the Latin word 'Status' or Italian word 'Statista' or German word 'Statistik' which means a Political State. It is termed as political state, since in early years, statics indicates a collection of facts about the people in the state for administration or political purpose.

Statistics has been defined either as a singular non or as a plural noun.

Definition of Statistics as Plural noun or as numerical facts:- According to Horace Secrist, 'Statistics are aggregates of facts affected to a marked extent by multiplicity of causes numerically expressed, enumerated or estimated according to a reasonable standard of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other'.

Definition of Statistics as a singular noun or as a method:- According to Seliman, " Statistics is the science which deals with the methods of collecting classifying, comparing and interpreting numerical data collected, to know some light on any sphere of enquiry".

Characteristics of Statistics

    (1) Statistics show be aggregates of facts

    (2) They should be affected to a marked extent by multiplicity of causes.

    (3) They must be numerically expressed.

    (4) They should be enumerated or estimated according to a reasonable standard of accuracy.

    (5) They should be collected in a systematic manner.

    (6) They should be collected for a predetermined purpose.

    (7) They should be placed in relation to each other.

<u>Function of Statistics</u>

The following are the important functions of statistics:

1. It simplifies complexity:- Statistical methods make facts and figures easily understandable form. For this purpose Graphs and Diagrams, classification, averages etc are used.
2. It presents facts in a proper form:- Statistics presents facts in a precise and definite form.
3. It facilitates for comparison:- When date are presented in a simplified form, it is easy to compare date.
4. It facilitates for formulating policies:- Statistics helps for formulating policies for the companies, individuals, Govt. etc. it is possible only with the help of date presented in a suitable form.
4. It tests hypothesis:- Hypothesis is an important concept in research studies. Statistics provides various methods for testing the hypothesis. The important tests are Chi – square, Z-test, T-test and F-test.
5. It helps prediction or forecasting:- Statistical methods provide helpful means of forecasting future events.
6. It enlarges individual's knowledge:-When data are presented in a form of comparison, the individuals try to find out the reasons for the variations of two or more figures. It thereby helps to enlarge the individual's knowledge.
7. It measures the trend behavior:- Statistics helps for predicting the future with the help of present and past data. Hence plans, programs, and policies are formulated in advance with the help of statistical techniques.

Scope of Statistics or importance or utility of statistics.

The Scope of Statistics in various field are:

(1) Statistics in Business:- Statistics is most commonly used in business. It helps to take decision making of the business. The statistical data regarding the demand and supply of product can be collected and analyzed to take decisions. The company can also calculate the cost of production and then the selling price. The existing firms can also make a comparative study about their performance with the performance of others through statistical analysis.

(2) Statistics in Management:- Most of the managerial decisions are taken with the help of

statistics. The important managerial activities like planning, directing and controlling are properly executed with the help of statistical data and statistical analysis. Statistical techniques can also be used for the payment of wages to the employees of the organization.

(3) Statistics in economics:- Statistical data and methods of statistical analysis render valuable assistance in the proper understanding of the economic problems and the formulation of economic policy.

(4) Statistics in banking and finance:- Banking and financial activities use statistics most commonly.

(5) Statistics in Administration:- The govt. frames polices on the basis of statistical information.

(6) Statistics in research: Research work are undertaken with the help of statistics.

Limitation of statistics

 (1) Statistics studies only numerical data

 (2) Statistics does not study individual cases

 (3) Statistics does not reveal the entire story of the problem.

 (4) Statistics in only one of the methods of study a problem.

 (5) Statistics can be misused. Statistical result are true only an average

Statistical Enquires or Investigation

Statistical Investigation is concerned with investigation of some problem with the help of statistical methods. It implies search for knowledge about some problems through statistical device.

Different stages in statistical enquiry are:

(1) Planning the enquiry

(2) Collection of data.

(3) Organization of data.

(4) Presentation of data.

(5) Analysis of data.

(6) Interpretation of data.

(1) **Planning the enquiry**:- The first step in statistical investigation is planning. The investigator should determine the objective and scope of the investigation. He should decide in advance about the type of enquiry to be conducted, source of information and the unit of measurement.

Object and scope:- The objective of the Statistical enquiry must be clearly defined. Once the objective of enquiry has been determined, the next step is to decide the scope of enquiry. It refers to the coverage of the enquiry.

Source of information:- After the purpose and scope have been defined, the next step is to decide about the sources of data. The sources of information may be either primary or secondary.

Types of enquiry:- Selection of type of enquiry depends on a number of factors like object and scope of enquiries, availability of time, money and facilities. Enquiries may be (1) census or sample (2) original or repetitive (3) direct or indirect (4) open or confidential (5) General or special purpose.

Statistical unit:- The unit of measurements which are applied in the collected data is called statistical unit. For example ton, gram, meter, hour etc.

Degree of accuracy:- The investigator has to decide about the degree of accuracy that he wants to attain. Degree of accuracy desired primarily depends up on the object of an enquiry.

Cost of plan:- An estimate of the cost of the enquiry must be prepaid before the commencement of enquiry.

(2) **Collection of data**:- Collection of data implies accounting and systematic recoding of the information gathered in a statistical investigation. Depending on the source, the collected statistical data are classified under two categories namely primary data and secondary data.

(3) **Organization of data**:- Organization of data implies the arrangement and presentation of data in such a way that it becomes easy and convenient to use them. Classification and tabulation are the

two stages of organizing data.

(4) **Presentation of data**:- They are numerous ways in which statistical data may be displayed. Graphs and diagrams are used for presenting the statistical data.

(5) **Analysis data**:- Analysis of data means critical examination of the data for studying characteristics of the object under study and for determining the pattern of relationship among the variables.

(6) **Interpretation of data**:- Interpretation refers to the technique of drawing inference from the collected facts and explaining the significance**Classification according to variables**

Data are classified on the basis of quantitative characteristics such as age, height, weight etc.

Geographical Classification:- Classified according to geographical differences.

Chronological Classification:- Classified according to period wise.

Frequency Distribution

A frequency distribution is an orderly arrangement of data classified according to the magnitude of observations. When data are grouped into classes of appropriate size indicating the number of observations in each class we get a frequency distribution.

Components of frequency Distribution

(1) Class and class interval

(2) Class limits

Methods of classification

(1) Classification according to attributes.

(2) Classification according to variables.

Classification according to attributes

Under this methods the data are classified on the basis of attributes. For example literacy, unemployment etc. are attributes.

Following are the classification under this method.

1. Simple classification
2. Manifold classification

In simple classification the data are divided on the basis of only one attributes.

In manifold classification the data are classified on the basis more than one attributes. For example

population is divided on the basis of sex and literacy

Class boundaries

Magnitude of class interval

Class frequency.

Tabulation

Tabulation is an orderly arrangement of data in rows and columns. It is a moment of presentation of data.

Objectives

1. To simplify complex data
2. To facilitate comparison
3. To facilitate statistical analysis
4. To save time
5. To economies space

Part of a table

1. Table number
2. Title of the table
3. Caption -------- i.e. column headings
4. Sub -------------- i.e. row heading
5. Body
6. Head note
7. Foot note
8. Source data.

Collection of data

On the basis of source, data can be collected from primary and secondary source.

<u>Primary data</u>

Primary data are those collected by the investigator himself. May are original in character. May are truthful and suit for the purpose. But the collection is very expensive and time consuming.

Methods of collection of primary data

1. <u>Direct personal interview</u>:- In this method investigator collection the data personally. He was to meet the people for collecting the data. This method is suitable:

   a) When the area of investigation is limited
   b) When higher degree of accuracy is leaded.

2. <u>Indirect oral investigation:-</u> :Under this method, information are collected from When the results of investigation to be kept confidential third parties who are is touch with the facts under enquiry.

3. <u>Schedules and Questionnaires methods:-</u> Under this method, a list of questions called questionnaire is prepared and information are called from various sources. It is a printed list of questions to be filled by the informations. But schedule is filled by the enumerator.

Essentials of a good questionnaire

   (1) The person conducting the survey much introduce himself.
   (2) The number of questions should be kept to the minimum.
   (3) The question should be as short as possible and simple.
   (4) The questions must be arranged in logical order.
   (5) The questions should be clear.
   (6) Personal questions should be avoided.
   (7) Questions should be in the nature of yes or no type.
   (8) Questions must be of convenient size and easy to handle.
   (9) Questions should be attractive.
   (10) Instructions should be given for filing up the form. Specimen of questionnaire.

Secondary data

Secondary data are those data which are collected by someone for this purpose. Secondary data are usually in the shape of finished product. The collection of secondary data is less expensive and less time consuming. Secondary data are collected from published and unpublished sources.

Precautions to be taken before using secondary data

    (1) Suitability
    (2) Adequacy
    (3) Reliability

Difference between Primary and Secondary data

1. Primary data are original character. But secondary data are not original, they are collected by somebody else.
2. Primary data are in the shape of raw material. But secondary data are in the shape of finished product.
3. Collection of primary data is expanse and time consuming. But collection of secondary data is less expensive and less time consuming.
4. Primary data will be usually adequate and suitable. But secondary data need not be adequate and suitable for the purpose.

Sampling

Sampling is the process obtaining information about an entire population by examining only a part of it. It is the examination of the regenerative items and conclusion of draw for all items coming in that group.

Methods of sampling or techniques of sampling

1. Probability sampling or random sampling
2. Non probability sampling

Probability sampling

Under this method, each items has an equal chance for being selected.

Following are the random sampling.

(1) Simple random sampling

A simple random sample is a sample selected from a population in such a way that every item of the population has an equal chance of being selected. The selection depends on chance. Eg. Lottery methods.

(2) Systematic sampling

This method is popularly used in those cases where complete list of the population from which sample is to be drawn is available. Under this method the items in the population are included in intervals of magnitude K. From every interval select an item by simple random sample method.

(3) Cluster sampling

Cluster sampling consists in forming suitable clusters of units. All the units is the sample of clusters selected are surveyed.

(4) Quota sampling

In this method each investigator engaged in the collection of data is assigned a quota for investigation.

(5) Multi stage sampling

This is a sampling procedure carried but in several stages. In multistage sampling, firstly units selected by suitable methods of sampling. From among the selected units, sample is drawn by some suitable methods. Further stages are added to arrive at a sample of the desired units.

Non probability sampling

1. Judgment sampling:- Under this sampling investigator exercise this discretion in the mater of selecting the items that are to be included in the sample.

2. by selecting such units of the universe which may be conveniently located.

Organization of data

Organizing data mean, the arrangement and presentation of data. Classification and tabulation are the two stages of organizing data.

Classification

The process of arranging data in groups or classes according to similarities called classification.

Objects of classification

1. To simplify the complexity of data.
2. To bring out the points of similarity of the various items.
3. To facilitate comparison.
4. To bring out relationship.
5. To provide basis for tabulation.

Graphs and Diagrams

Graphs and diagrams is one of the statistical methods which simplifies the complexity of quantitative data and make them easily understandable.

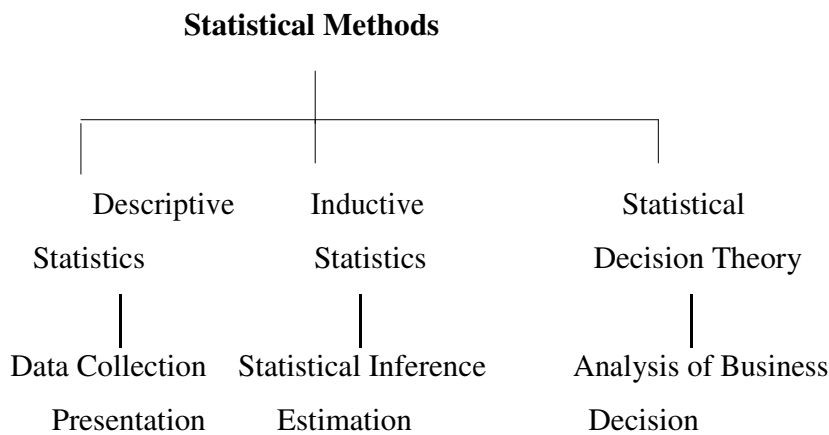**Importance of Diagrams & Graphs**

1. Attract common people
2. Presenting quantitative facts in simple.
3. They have a great memorizing effect.
4. They facilitate comparison of data.
5. Save time in understanding data.
6. Facts can be a understood without mathematical calculations.

Limitations

1. They can present only approximate values.
2. They can represent only limited amount of information.
3. They can be misused very easily.
4. They are not capable of further mathematical treatment.
5. They are generally useful for comparison purpose only

Classification of Statistical Methods

The filed of statistics provides the methods for collecting, presenting and meaningfully interpreting the given data. Statistical Methods broadly fall into three categories as shown in the following chart.

Statistical Methods

Descriptive	Inductive	Statistical
Statistics	Statistics	Decision Theory
Data Collection	Statistical Inference	Analysis of Business
Presentation	Estimation	Decision

Descriptive Statistics

There are statistical methods which are used for re-arranging, grouping and summarising sets of data to obtain better information of facts and thereby better description of the situation that can be made. For example, changes in the price- index. Yield by wheat etc. are frequently illustrated using the different types of charts and graphs. These devices summarise large quantities of numerical data for easy understanding. Various types of averages, can also reduce a large mass of data to a single descriptive number. The descriptive statistics include the methods of collection and presentation of data, measure of Central tendency and dispersion, trends, index numbers, etc.

Inductive Statistics

It is concerned with the development of some criteria which can be used to derive information about the nature of the members of entire groups (also called population or universe) from the nature of the small portion (also called sample) of the given group. The specific values of the population members are called 'parameters' and that of sample are called 'Statistics'. Thus, inductive statistics is concerned with estimating population parameters from the sample statistics and deriving a statistical inference.

Samples are drawn instead of a complete enumeration for the following reasons:

- the number of units in the population may not be known

- the population units may be too many in number and/or widely dispersed.
- Thus complete enumeration is extremely time consuming and at the end of a full enumeration so much time is lost that the data becomes obsolete by that time.
- It may be too expensive to include each population item.

Inductive statistics, includes the methods like: probability and probability distributions; sampling and sampling distribution; various methods of testing hypothesis; correlation, regression, factor analysis; time series analysis.

Statistical Decision Theory

Statistical decision theory deals with analysing complex business problems with alternative course of action (or strategies) and possible consequences. Basically,. It is to provide more concrete information concerning these consequences, so that best course of action can be identified from alternative courses of action.

Statistical decision theory relies heavily not only upon the nature of the problem on hand, but also upon the decision environment. Basically there are four different states of decision

environment as given below:

State of decision	Consequences
Certainty	Deterministic
Risk	Probabilistic
Uncertainty	Unknown
Conflict	Influenced by an opponent

Since statistical decision theory also uses probabilities (subjective or prior) in analysis, therefore it is also called a subjectivist approach. It is also known as Bayesian approach because Baye's theorem, is used to revise prior probabilities in the light of additional information.

Various Statistical Techniques

A brief comment on certain standard techniques of statistics which can be helpful to a decision- maker in solving problems is given below.

i) **Measures of Central Tendency**: Obviously for proper understanding of quantitative data, they should be classified and converted into a frequency distribution (number of times or frequency with which a particular data occurs in the given mass of data.). This type of condensation of data reduces their bulk and gives a clear picture of their structure. If you want to know any specific characteristics of the given data or if frequency distribution of one set of data is to be compared with another, then it is necessary that the frequency distribution help us to make useful inferences about the data and also provide yardstick for comparing different sets of data. Measures of average or central tendency provide one such yardstick. Different methods of measuring central tendency, provide us with different kinds of averages. The main three types of averages commonly used are:

   a) **Mean**: the mean is the common arithmetic average. It is computed by dividing the sum of the values of the observations by the number of items observed.

   b)**Median**: the median is that item which lies exactly half-way between the lowest and highest value when the data is arranged in an ascending or descending order. It is not affected by the value of the observation but by the number of observations. Suppose you

have the data on monthly income of households in a particular area. The median value would give you that monthly income which divides the number of households into two equal parts. Fifty per cent of all the households have a monthly income above the median value and fifty per cent of households have a monthly income below the median income.

c) **Mode**: the mode is the central value (or item) that occurs most frequently. When the data organised as a frequency distribution the mode is that category which has the maximum number of observations. For example, a shopkeeper ordering fresh stock of shoes for the season would make use of the mode to determine the size which is most frequently sold. The advantages of mode are that (a) it is easy to compute, (b) is not affected by extreme values in the frequency distribution, and (c) is representative if the observations are clustered at one particular value or class.

ii) **Measures of Dispersion:** the measures of central tendency measure the most typical value around which most values in the distribution tend to coverage. However, there are always extreme values in each distribution. These extreme values indicate the spread or the dispersion of the distribution. The measures of this spread are called 'measures of dispersion' or 'variation' or 'spread'. Measures of dispersion would tell you the number of values which are substantially different from the mean, median or mode. The commonly used measures of dispersion are range, mean deviation and standard deviation. The data may spread around the central tendency in a symmetrical or an asymmetrical pattern. The measures of the direction and degree of symmetry are called measures of the skewness. Another characteristic of the frequency distribution is the shape of the peak, when it is plotted on a graph paper. The measures of the peakedness are called measures of Kurtosis.

iii) **Correlation:** Correlation coefficient measures the degree to which the charge in one variable (the dependent variable) is associated with change in the other variable (independent one). For example, as a marketing manager, you would like to know if there is any relation between the amount of money you spend on advertising and the sales you achieve. Here, sales is the dependent variable and advertising budget is the independent variable. Correlation coefficient, in this case, would tell you the extent or relationship between these two variables,' whether the relationship is directly proportional (i.e. increase or decrease in advertising is associated with

decrease in sales) or it is an inverse relationship (i.e. increasing advertising is associated with decrease in sales and vice- versa) or there is no relationship between the two variables. However, it is important to note that correlation coefficient does not indicate a casual relationship, Sales is not a direct result of advertising alone, there are many other factors which affect sales. Correlation only indicates that there is some kind of association-whether it is casual or causal can be determined only after further investigation. Your may find a correlation between the height of your salesmen and the sales, but obviously it is of no significance.

iv) **Regression Analysis**: For determining causal relationship between two variables you may use regression analysis. Using this technique you can predict the dependent variables on the basis of the independent variables. In 1970, NCAER (National Council of Applied and Economic Research) predicted the annual stock of scooters using a regression model in which real personal disposable income and relative weighted price index of scooters were used as independent variable.

The correlation and regression analysis are suitable techniques to find relationship between two variables only. But in reality you would rarely find a one-to-one causal relationship, rather you would find that the dependent variables are affected by a number of independent variables. For example, sales affected by the advertising budget, the media plan, the content of the advertisements, number of salesmen, price of the product, efficiency of the distribution network and a host of other variables. For determining causal relationship involving two or more variables, multi- variable statistical techniques are applicable. The most important of these are the multiple regression analysis deiscriminant analysis and factor analysis.

v) **Time Series Analysis** : A time series consists of a set of data (arranged in some desired manner) recorded either at successive points in time or over successive periods of time. The changes in such type of data from time to time are considered as the resultant of the combined impact of a force that is constantly at work. This force has four components: (i) Editing time series data, (ii) secular trend, (iii) periodic changes, cyclical changes and seasonal variations, and (iv) irregular or random variations. With time series analysis, you can isolate and measure the separate effects of these forces on the variables. Examples of these changes can be seen, if you start measuring increase in cost of living, increase of population over a period of time, growth of

agricultural food production in India over the last fifteen years, seasonal requirement of items, impact of floods, strikes, wars and so on

vi) **Index Numbers**: Index number is a relative number that is used to represent the net result of change in a group of related variables that has some over a period of time. Index numbers are stated in the form of percentages. For example, if we say that the index of prices is 105, it means that prices have gone up by 5% as compared to a point of reference, called the base year. If the prices of the year 1985 are compared with those of 1975, the year 1985 would be called "given or current year" and the year 1975 would be termed as the "base year". Index numbers are also used in comparing production, sales price, volume employment, etc. changes over period of time, relative to a base

.

vii) **Sampling and Statistical Inference**: In many cases due to shortage of time, cost or non-availability of data, only limited part or section of the universe (or population) is examined to (i) get information about the universe as clearly and precisely as possible, and (ii) determine the reliability of the estimates. This small part or section selected from the universe is called the sample, and the process of selection such a section (or past) is called sampling.

Schemes of drawing samples from the population can be classified into two broad categories:

1   .Random sampling schemes: In these schemes drawing of elements from the population is random and selection of an element is made in such a way that every element has equal change (probability) of being selected.

2   Non-random sampling schemes: in these schemes, drawing of elements for the population is based on the choice or purpose of selector.

General rules for constructing Diagrams

1. Title

2. Proportion between width and height.

3. Selection of scale

4. Foot note

5. Index

6. Neatness and cleanliness

7. Simplicity

8. Attractiveness

Types of Diagrams

1. Dimensional Diagrams

2. Cartograms

3. Pictograms

Dimensional Diagrams

Dimensional Diagrams are those diagrams which show information in terms of length, height, area or volume. They are one dimensional two dimensional or three dimensional.

One Dimensional Diagram

In one dimensional diagram the height will represent the magnitude of observations. Must commonly used one dimensional diagrams are line diagram and Bar diagram.

Line Diagram

Line diagrams are one dimensional diagrams. They are drawn to represent values of a variable.

Ex. Draw a line diagram to the following data.

| Country:	A	B	C	D	E
Population:	10	5	15	13	12
(in million)					
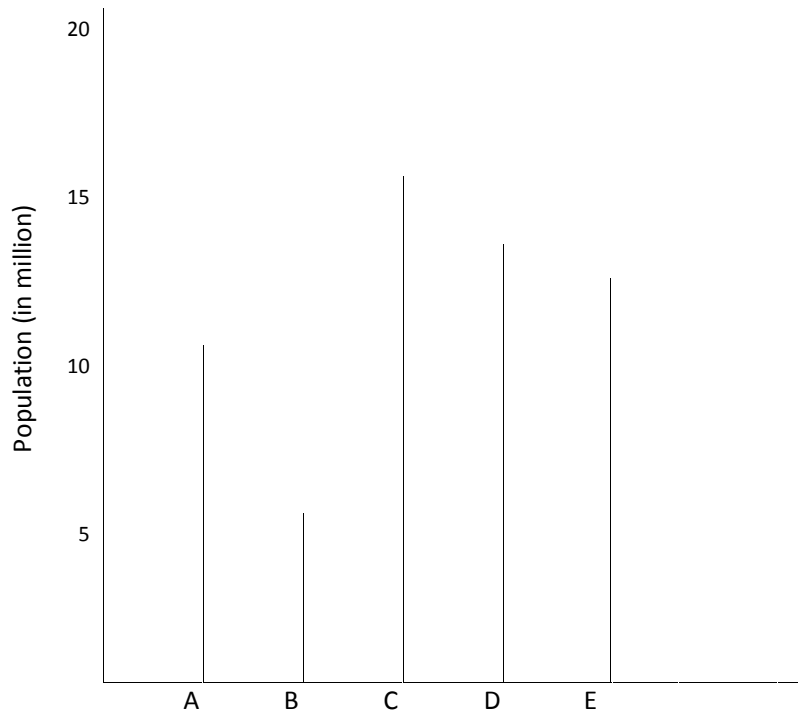
Bar Diagram

In a bar diagram only the length is considered. The width of the bar is not given any importance.
Following are the important types of bar diagrams
  (1) Simple bar diagram
      Simple bar diagram represents only one variable. For example, height, weight, etc.

| Year	2007	2008	2009	2010	2011	2012
Sales (In '0000'	45	55	65	70	50	60

School of Distance Education
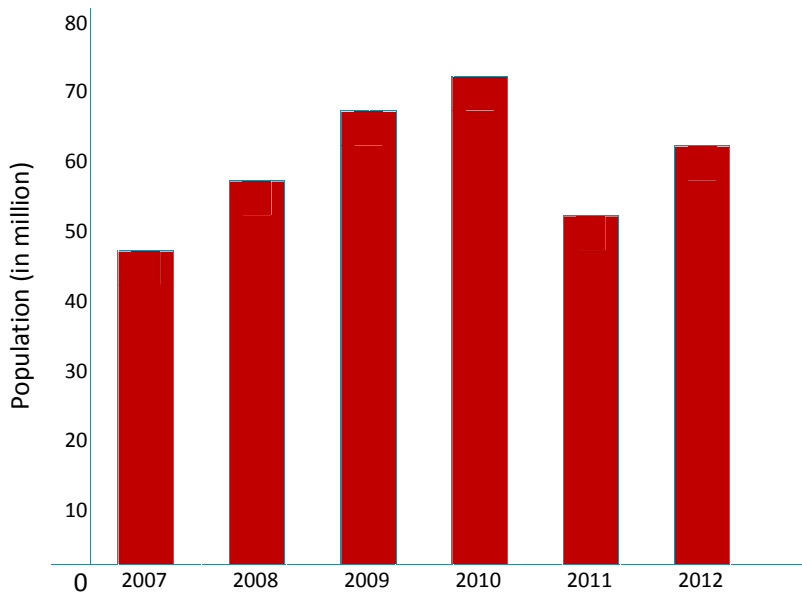
2) Multiple Bar Diagram

Two or more interrelated data are represented in a multiple bar diagram. In order to identity the data, the bars should be differentiated with colors or shades.

Eg:- From the following data draw a suitable diagram.

| Year	Production (in units)		
	A	B	C
2008	45	55	65
2009	35	60	70
2010	50	70	80
2011	55	80	60
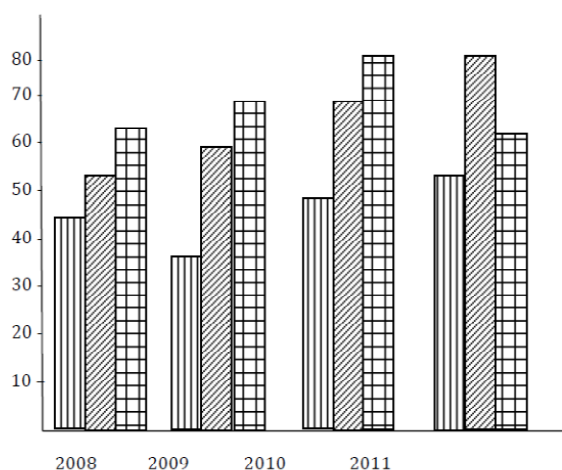
3) Sub Divided Bar Diagram

In the sub divided bar diagram each bar is subdivided into two or more parts. Each part may explain different characters.

Eg:- The number of students in Calicut University are as follows: Represent the date by suitable diagram

Year	Commerce	Arts	Science	Total
2008-09	35000	10000	9000	54000
2009-10	45000	9000	90000	64000
2010-11	55000	7000	8000	69000
2011-12	70000	5000	7000	82000
2012-13	80000	4000	6000	90000

Basic Numerical Skills
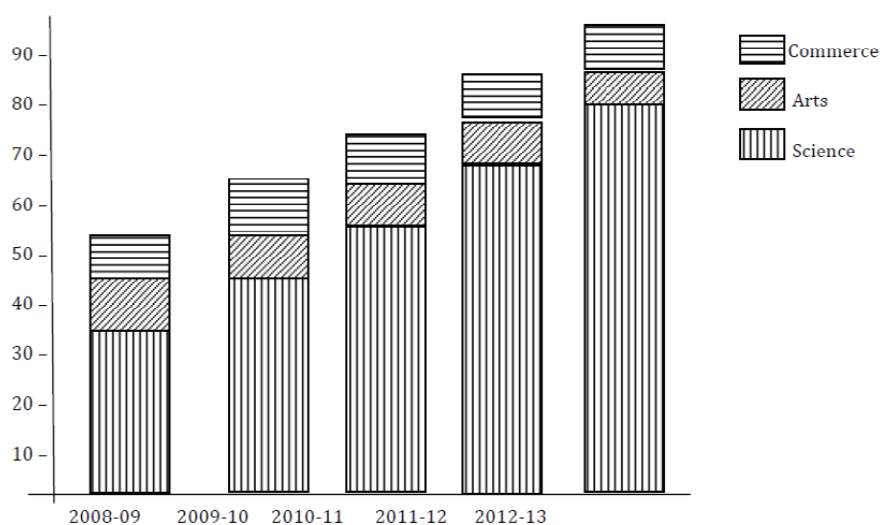

4) Percentage Bar Diagrams

In percentage bar diagram the length of all the base are equal ie each bar represent 100 percent. The component parts are expressed as percentage to the whole.

Eg:- Prepare a subdivided bar diagram on the percentage basis.

Year	Direct Cost Rs	Indirect Cost Rs	Profit Rs	Sales Rs
2009	35	15	10	60
2010	40	20	12	72
2011	32	22	8	62
2012	25	35	15	75

Answer

Year	Direct Cost in %	Indirect Cost in %	Profit in %	Sales
2009	58	25	17	100
2010	55	28	17	100
2011	52	35	13	100
2012	33	47	20	100
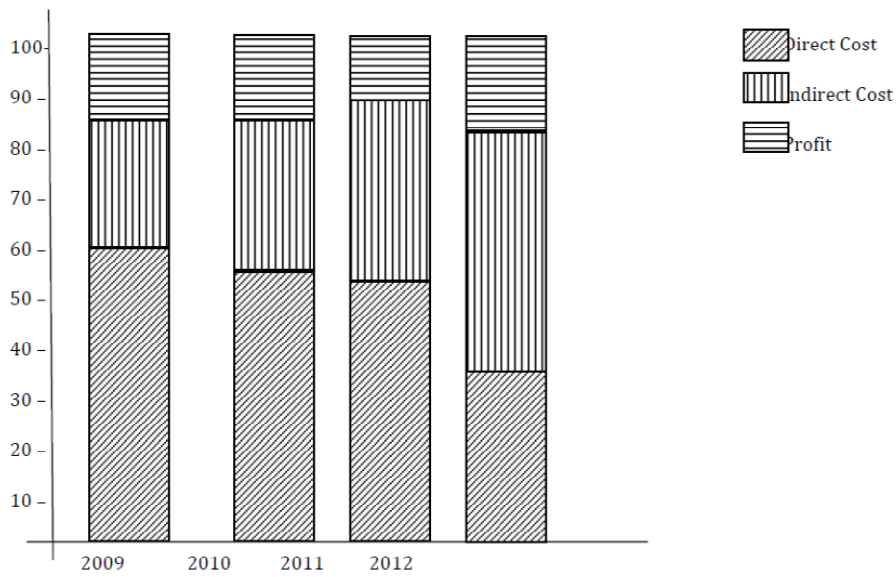
Two Dimensional Diagram

In two dimensional diagram the length as well as width have to be considered. The most commonly used two dimensional diagrams is pie diagram, Rectangles, Squares, Circles etc are also two dimensional diagrams.

Pie Diagrams

Pie diagrams are used when the aggregate and their divisions are to be shown together. The aggregate is shown by means of a circle and divisions by the sectors of the circle. For example, the selling price of a product can be divided into various segments like factory cost, administrative cost, selling cost and profit. These segments are converted into percentage in order to represent in the pie diagram.

In order to prepare the pie diagram, each percentage outlay must be multiplied by 3.6, since the pie diagram contain 360° scale.
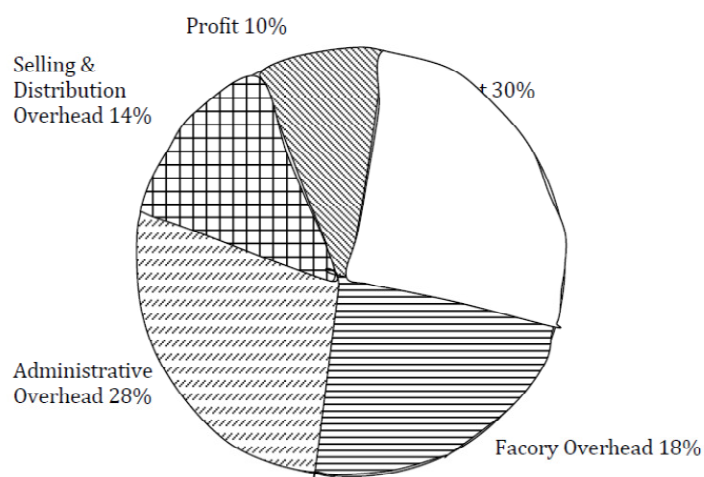
Eg:- Draw a pie diagram from the following data

Prime Cost	30%
Factory over Head	18%
Administrative overhead	28%
Selling & Distribution overhead	14%
Profit	10%

Ans:

Prime Cost	30	108°
Factory over Head	18	65°
Administrative overhead	28	101°
Selling & Distribution overhead	14	50°
Profit	10	36°
	100	360

Three Dimensional Diagrams

Three dimensional diagrams are prepared in the form of cubes, spheres, cylinders etc. In these diagrams width, length and breadth are important.

Cartograms

Cartograms means the presentation of data in a geographical basis. It is otherwise called as statistical maps. The quantities on the map may be shown through shades, dots or colours etc.

Pictograms

Under the pictograms, data are represented in the form of a appropriate pictures most suited for the data.

Basic Numerical Skills

GRAPHS

Types of Graphs

(1) Graphs of Frequency Distribution

(2) Graphs of Time Series

Graphs of Frequency Distribution

A frequency distribution can be presented graphically in any of the following ways:

(1) Histogram

(2) Frequency Polygon

(3) Frequency Curves

(4) Ogive or cumulative frequency curves.

Histogram

A histogram is a graph of frequency distributions. A histogram consists of bars erected upon the class interval columns.

While constructing histogram, the variable is always taken on the x-axis and the frequency on the y-axis. The width of the bars in the histogram will be proportional to the class interval.

Histogram for frequency Distribution having equal Class interval

1) Draw a histogram from the following information

Marks	No. of Students
0-10	7
10-20	12
20-30	15
30-40	17
40-50	20
60-70	14
70-80	10
80-90	4
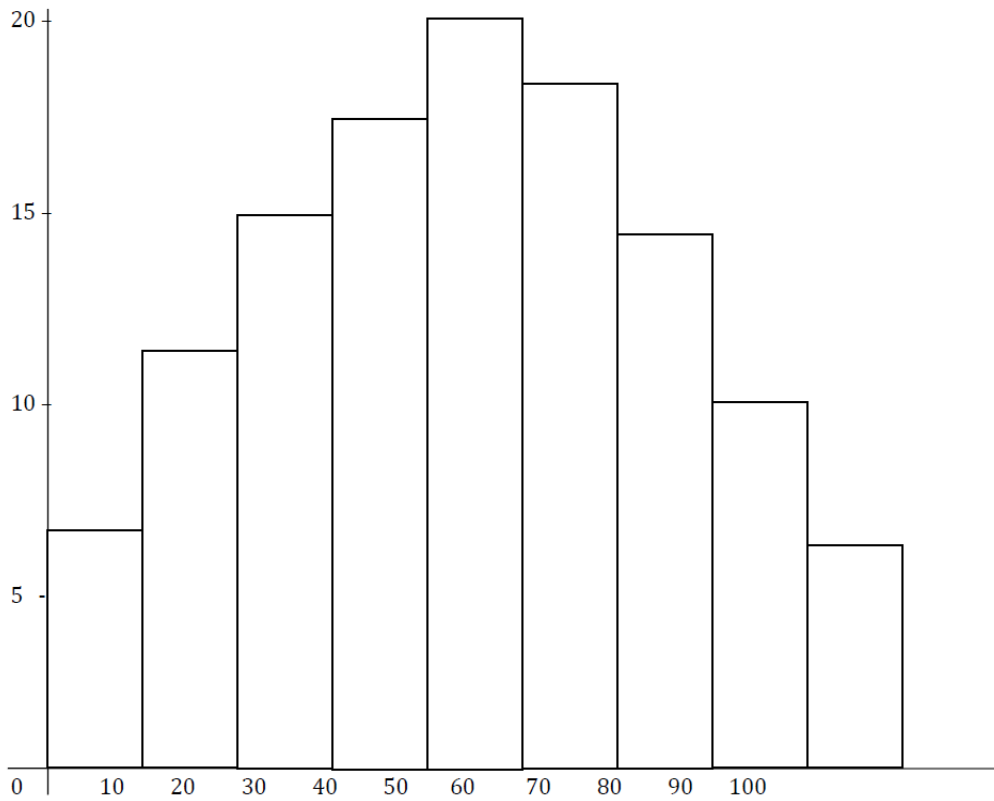
Histogram for unequal Class Interval

Unequal class intervals must be corrected.

$$\text{Unequal class intervals} = \frac{\text{Frequency unequal class intervals}}{\text{width of the unequal class intervals}} \times \text{width of the lowest class interval}$$

Draw a histogram from the following data

Daily wages	No. of workers
15-20	4
20-25	9
25-30	12
30-40	20
40-50	16
50-55	7
55-60	6
60-75	15
75-80	4
80-95	9
95-100	2

Answer:

Calculation of Frequency Density

Daily wages	No. of workers	Frequency Density
15-20	4	4
20-25	9	9
25-30	12	12
30-40	20	10
40-50	16	8
50-55	7	7
55-60	6	6
60-75	15	5
75-80	4	4
80-95	9	3
95-100	2	2
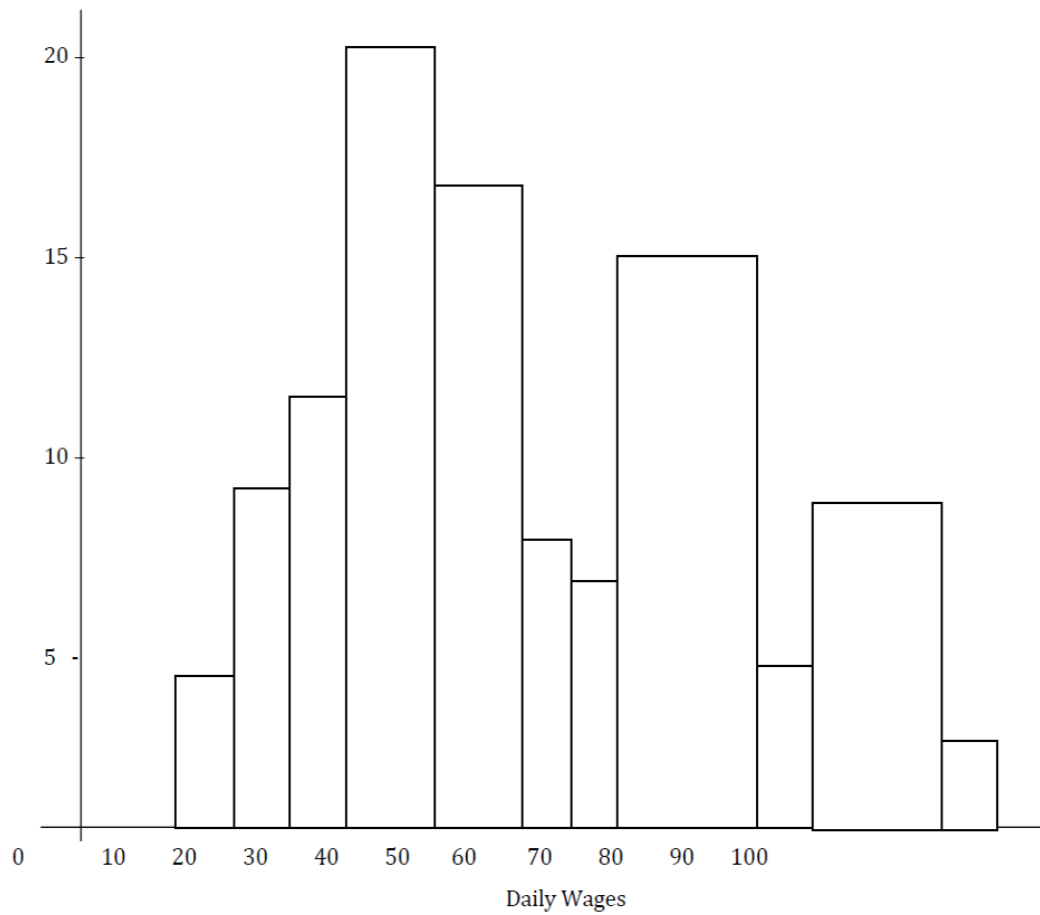
Frequency Polygon

It is a curve instead of bars. There are two methods for constructing frequency polygon. First, histogram should be drawn and mark mid point of upper side of each bar and join such joints by a curve.

In the second method, first of all plot the frequencies corresponding to midpoints of various class intervals. Then join all the plotted points to get the frequency polygon curve.

3) Ogive or Cumulative Frequency Curve

A frequency distribution when cumulated, we get cumulative frequency distribution and curve drawn is known as ogive. An ogive can either less than ogive or more than ogive. Less than ogive curve is drawn on the basis of less than cumulative frequency distribution and more than ogive is drawn on the basis of more than cumulative frequency distribution.

Example :-

From the following data drawn less than and more than ogives

Marks :	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Students	10	20	35	30	20	15	10	10

Answer :

Less than CF	F	More than CF	F
Less than 0	0	More than 0	150
Less than 10	10	More than 10	140
Less than 20	30	More than 20	120
Less than 30	65	More than 30	95
Less than 40	95	More than 40	55
Less than 50	125	More than 50	35
Less than 60	130	More than 60	20
Less than 70	140	More than 70	10
Less than 80	150	More than 80	0
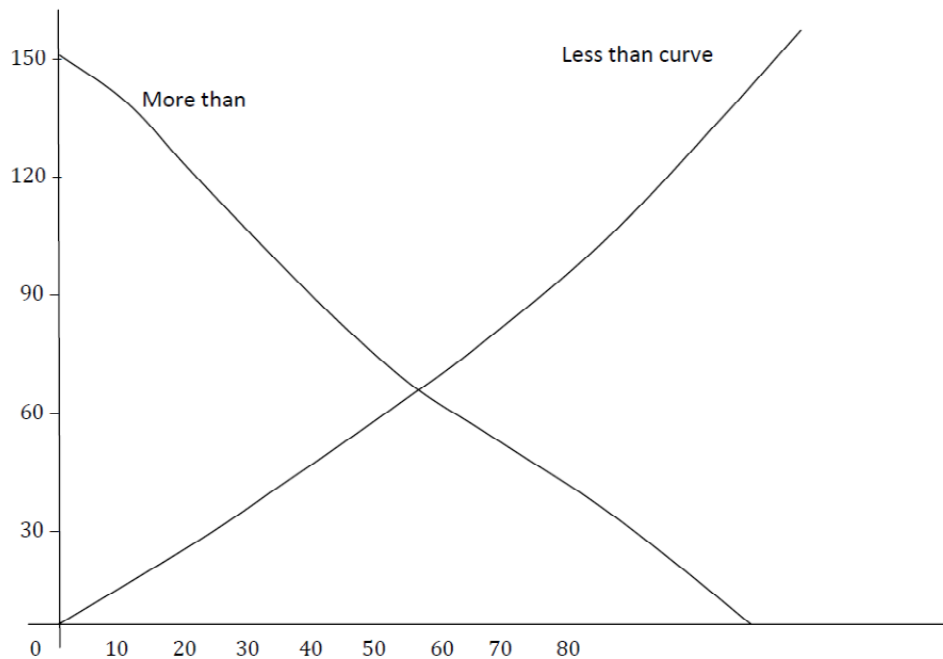
Measures of central tendency or Averages

An average is a single value that represents a group of values. It represents the whole series and conveys general idea of the whole group. Characteristics of a good average or Requisites or Essentials properties of average

(1) Clearly defined
(2) Easy to understand
(3) Simple to compute
(4) Based on all items
(5) Not be unduly affected by extreme observations.
(6) Capable of further algebraic treatment
(7) Sampling stability.

Types of averages

1) Arithmetic Mean

2) Median

3) Mode

4) Geometric mean

5) Harmonic Mean

Arithmetic Mean (AM)

It is the value obtained by adding together all the items and by dividing the total number of items.

Arithmetic mean may either be

(1) Simple arithmetic Mean   or

(2) Weighted arithmetic Mean

Simple Arithmetic Mean

It is the mean of items which give equal importance to all items.

It is denoted by $\bar{x}$

$$\bar{x} = \frac{\sum x}{N}$$

Where = Sum of given variables

N = Number of items

Calculation of Arithmetic Mean

(a) Individual Series :-

(i)   Direct Method

$$\bar{x} = \frac{\sum x}{N}$$

(ii)   Short Cut Method

$$\bar{x} = A + \frac{\sum d}{n}$$

A = Assumed mean

D = X – A

n = total number of items

(b) Discrete Series

(i)   Direct Method

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

(ii)   Short Cut method

$$\bar{x} = A + \frac{\sum fd}{N}$$

d = X – A

(iii)   Step deviation method

$$\bar{x} = A + \frac{\sum fd'}{N} \times C$$

$$d' = \frac{X-A}{C}$$

c = common factor

(c) Continuous Series

(i)   Direct method

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

m = midpoint of X

N = Total frequency

(ii) Short cut method

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

d = m – A

(iii) Step deviation method:

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

$$d' = \frac{m-A}{c}$$

C = Common factor or class interval

Practical Problems

1) Calculate A.M. of the weight of 10 students in a Class

Sl. No.	1	2	3	4	5	6	7	8	9	10
Weight in Kg	42	56	49	50	49	53	52	48	47	54

Ans: This is an individual series.

$$\bar{X} = \frac{\Sigma X}{n}$$

Σx = 42+56+49+50+49+53+52+48+47+54

= 500

n = 10

$$\bar{x} = \frac{500}{10} = 50Kg.$$

==========

2) Calculate mean from the following data.

Marks	25	30	35	40	45	50	55	60	65	70
No. of students	3	8	12	9	4	7	15	5	10	7

Ans:

Marks x	No. of students f	d (x – 55)	d'	fd'
25	3	-30	-6	-18
30	8	-25	-5	-40
35	12	-20	-4	-48
40	9	-15	-3	-27
45	4	-10	-2	-8
50	7	-5	-1	-7
55	15	0	0	0
60	5	5	1	15
65	10	10	2	20
70	7	15	3	21
	80			-120

$$\bar{X} = A + \frac{\sum fd'}{N} \times C$$

$$\bar{X} = 55 + \frac{-102}{80} \times 5$$

$$= 55 + \frac{-510}{80}$$

$$= 55 + -6.375$$

$$= 48.625$$

=========

3. Calculate Arithmatic Mean

Production in tons	No. of factories
10 – 20	5
20 – 30	4
30 – 40	7
40 - 50	12
50 – 60	10
60 – 70	8
70 – 80	4

Ans:

X	f	m	fm
10 – 20	5	15	75
20 – 30	4	25	100
30 – 40	7	35	245
40 - 50	12	45	540
50 – 60	10	55	550
60 – 70	8	65	520
70 – 80	4	75	300
	80		2330

$$\bar{X} = \frac{\sum fm}{N} = \frac{2330}{50} = 46.6$$

==========

4. Following are the data related with the production of a product during January in 100 factories

Production in tons	No. of factories
0 – 100	7
100 – 200	15
200 – 300	10
300 – 400	9
400 – 500	10
500 – 600	12
600 - 700	8
700 – 800	13
800 – 900	9
900 - 1000	7

Ans:

x	f	m	d (m – A)	d'	fd'
0 – 100	7	50	-500	-5	-35
100 – 200	15	100	-400	-4	-60
200 – 300	10	250	-300	-3	-30
300 – 400	9	350	-200	-2	-18
400 – 500	10	450	-100	-1	-10
500 – 600	12	550	0	0	0
600 - 700	8	650	100	1	8
700 – 800	13	750	200	2	26
800 – 900	9	850	300	3	27
900 - 1000	7	950	400	4	28
	100				-64

$$\bar{X} = A + \frac{\Sigma fd'}{N} \text{ X C}$$

$$\bar{X} = 550 + \frac{-64}{100} \text{ X } 100$$

$$= 486$$
========

Calculation of Arithmatic Mean for open end classes

If the lower limit of the first class and upper limit of the last class are not known, it is called open end classes.

1. Calculate A.M.

Below 10	5
10 – 20	12
20 – 30	14
30 – 40	10
Above 40	8

Ans:

X	f	m	fm
0 – 10	5	5	25
10 – 20	12	15	180
20 – 30	14	25	350
30 – 40	10	35	350
40 - 50	8	45	360
	49		1265

$$\bar{X} = \frac{\Sigma fm}{N} = \frac{1265}{49} = 25.82$$

==========