

---

$$P + 12 + 59 = 147 \text{ giving } P = 76$$
$$Q + 59 + 37 = 102 \text{ giving } Q = 6$$

i) Those who did not vote  
 $= 250 - (76 + 12 + 14 + 59 + 6 + 37)$   
 $= 250 - 204 = 46$

ii)  $x = 76 + 12 + 14 = 102$   
 $y = 12 + 59 + 6 = 77$   
 $z = 37 + 14 + 6 = 57$

iii) X won the election

## **BINOMIAL THEOREM**

## **PROGRESSION**

## **CHAPTER TWO**

### **DATA COLLECTION AND PRESENTATION**

#### Specific objectives

At the end of this topic the trainee should be able to:

- ❖ Discuss the basic consideration for data collection.
- ❖ Classify collected data into various categories.
- ❖ Tabulate collected data.
- ❖ Diagrammatically and graphically present data.

#### Introduction

---

## a) Statistics

Statistics viewed as a subject is a process of collecting, tabulating and analyzing numerical data upon which significant conclusions are drawn. Statistics may also be defined as numerical data, which has been, collected from a given source and for a particular purpose e.g. population statistics from the ministry of planning, Agricultural statistics from the ministry of Agriculture

Statistics may also refer to the values, which have been obtained from statistical calculations e.g. the mean, mode, range e.t.c.

## b) Application of statistics

### 1. Quality Control

Usually there is a quality control departments in every industry which is charged with the responsibility of ensuring that the products made do meet the customers standards e.g. the Kenya bureau of standards (KeBS) is one of the national institutions which on behalf of the government inspects the various products to ensure that they do meet the customers specification. The KeBS together with other control department have developed quality control charts. They use these charts to check whether the products are up to standards or not.

### 2. Statistics may be used in making or ordering economic order quantities (EOQ).

It is important for a business manager to realize that it is an economic cost if one orders a large quantity of items which have to be stored for too long before they are sold. This is because the large stock holds a lot of capital which could otherwise be used in buying other items for sale.

It is also important to realize that the longer the items are stored in the stores the more will be the storage costs

On the other hand if one orders a few items for sale he will incur relatively low storage expenses but may not be able to satisfy all the clients. These may lose their customers if the goods are out of stock. Therefore it is advisable to work out the EOQ which will be sufficient for the clients in a certain period before delivery.

The EOQ will also ensure that minimal costs are incurred in terms of storage

### 3. Forecasting

Statistics is very important for business managers when predicting the future of a business for example if a given business situation involves a dependent and independent variables one can develop an equation which can be used to predict the output under certain given conditions.

### 4. Human resource management

---

Statistics may be used in efficient use of human resources for example we may give questionnaires to workers to find out where the management is weak

By compiling the statistics of those who were signing it may be found useful to analyze such data to establish the causes of resignation thus whether it is due to frustration or by choice.

## **BASICS FOR DATA COLLECTION**

### Introduction

A statistical investigation involves a number of stages:

- ❖ Definition of the problem or issue;
- ❖ Collection of relevant data;
- ❖ Classification and analysis of the collected data;
- ❖ Presentation of the results.

Even before the collection of data starts, then, there are some important points to consider when planning a statistical investigation.

### Preliminary considerations.

It is important to be aware of these issues as they impact on the data which is to be collected and analyzed.

- ❖ Exact definition of the problem  
This is necessary in order to ensure that nothing important is omitted from the enquiry, and effort is not wasted by collecting irrelevant data. The problem as originally put to the statistician is often of a very general type and it needs to be specified precisely before work can begin.
- ❖ Definition of the units  
The results must appear in comparable units for any analysis to be valid. If the analysis is going to involve comparisons, then the data must all be in the same units. It is no use just asking for 'output' from several factories - some may give their answers in numbers of items, some in weigh of items, some in number of inspected batches and so on.
- ❖ Scope of the enquiry.  
No investigation should be got under way without defining the field to be covered. Are we interested in all departments of our business, or only some? Are we to concern ourselves with our own business only, or with others of the same kind?
- ❖ Accuracy of the data  
To what degree of accuracy is data to be recorded? For example, are ages of individuals to be given to the nearest year or to the nearest month or as the number of completed years? If some of the data is to come from measurements, then the accuracy of the measuring

---

instrument will determine the accuracy of the results. The degree of precision required in an estimate might affect the amount of data we need to collect. In general, the more precisely we wish to estimate a value, the more readings we need to take.

### Data sources and types

Data constitute the foundation of statistical analysis and interpretation.

Data can be obtained from three sources namely;

- Primary source
- Secondary source
- Internal records

### Primary data

Primary data are measurement, observed and recorded as part of an original study. There are two basic methods of obtaining primary data namely;

- Questionnaires
- Observation
- Interviews
- Sampling

Primary data is data which is both original and has been obtained in order to solve the specific problem in hand. Primary data is, therefore, raw data and has to be classified and processed using appropriate statistical methods in order to reach a solution to the problem.

### Secondary sources

Secondary data can be obtained from journals, reports, and government publications, publications of research organization, trade and professional bodies.

Secondary data must be used with utmost care, before using secondary data the investigator should examine the following;

1. Whether the data are suitable for the purpose of investigation.
2. whether the data are adequate for the purpose of investigation
3. Whether the data are reliable.

Secondary data is any data other than primary data. Thus, it includes any data which has been subject to the processes of classification or tabulation or which has resulted from the application of statistical methods to primary data, and all published statistics.

### Internal data

Internal data refers to the measurement that are the by products of routine businesses records keeping like accounting , finance, production personnel, quality control , sales , R&D.

Since internal data originate within the business collecting the desired information does not usually offer much difficult. The particular procedure

---

depends largely upon the nature of the facts being collected and the form in which they exist.

#### Data collection methods and limitations

The methods usually available are as follows:

- ❖ Questionnaires
- ❖ Observation
- ❖ Interview
- ❖ Use of published statistics

##### a. Questionnaire

As the name suggest is distinguished by the fact that data are collected by asking questions from people who are thought to have the desired information.

A formal list of such questions is called a questionnaire

Questionnaire refers to a device for securing answers to questions by using a form which the respondent fills in it.

##### b. Observation

The investigator observes the object or action in which he is interested.

Sometime individual makes the observation on other occasions mechanical device observe and record the desired information.

Observation method does not automatically produce accurate data. Physical difficulties n the observation situation on the part of the observer may result in errors.

#### Classification of Data

Classification is grouping of related facts into different classes. Facts in one class differ from another class with respect to some characteristics called a basis of classification.

Sorting facts on one basis of classification and then another basis is called cross-classification

#### Rules of Classification

1. the number of classes should preferably be between 5 and 15
2. as far as possible one should avoid odd values of class intervals
3. the starting point i.e. the lower limit of the first class should either be zero or 5 or multiple of 5
4. to ensure continuity and to get correct interval we should adopt exclusive method of classification
5. Whenever possible all classes should be of the same size.

#### Types of classification

- ❖ Geographical
- ❖ Chronological

- 
- ❖ Qualitative
  - ❖ Quantitative

Geographical classification; in geographical classification data are classified on the basis of geographical or location difference between the various items.

Chronological classification; when data are observed over a period of time, the type of classification is known as chronological classification.

Qualitative classification; in qualitative classification data are classified on the basis of some attributes or quality such as sex, color of hair, literacy, religion.

#### QUALITATIVE ANALYSIS DATA COLLECTION LIMITATION OF STATISTICS.

Despite the usefulness of statistics in many fields, impression should not be carried that statistics are like magical devices which always provide the correct solution of problems. Unless the data are properly collected and critically interpreted there is every likelihood of drawing wrong conclusion. Therefore it is necessary to know the limitation and the possible misuse of statistics.

The following are the important limitation of the science of statistics:

- Statistics does not deal with isolated measurement.
- Statistics deals only with quantitative characteristic i.e quantitative characteristic such as honesty, efficiency, intelligence, blindness and deafness cannot be studied directly
- Statistics result are true only an average.
- Statistic is only a means not a end.
- Statistics can be misused.

#### Classification functions.

Classification of data is a function very similar to that of sorting letters in a post office. It is well known that the letters collected in a post office are sorted into different geographical basis i.e in accordance with their destination as Nairobi, Mombasa, Kampala etc. They are then put in separate bags each containing letters with a common characteristic, or having the same destination.

Classification of statistical data is comparable to the sorting operation, the process of classification gives prominence to important information gathered while dropping unnecessary details facilitates comparison and enables a statistical treatment of the material collected.

#### Formation of frequency distribution.

Here we just count the number of times a particular value is repeated which is called the frequency. In order to facilitate counting prepare a column of "tally" in another column all possible values of variables from

---

the lowest to the highest. Then put a bar( vertical line ) opposite the particular value to which it relates.

To facilitate counting blocks of five bars are prepared and some space is left in between each block.

We finally count the number of blocks and bars corresponding to each value to each value of the variable and place it in the column and frequency.

Example .

Construct a frequency distribution from the following data.

23, 30,20,26,30,30,20,23,40,40,26,20,23,40,28,26,30,40,28,28,30.

class	tally	frequency
20	111	3
23	1111	4
26	111	3
28	111	3
30	1111	5
40	1111	4
	total	22

Classification according to class interval.

The following technical terms are important when data are classified according to class intervals:

- (a) Class limit. these are the lowest and the highest values that can be included in the class e.g take the class 20-40. the lowest value of this class is 20 and the highest is 40.

The two boundaries of a class is the value below which there can be no value in that class, while the upper limit of a class is the value above which no value can belong to that class.

The class 70- 89 .70 is the lower limit and 80 is the upper limit i.e in this class there can be no value which is less than 70 or more than 89.

- (b) Class interval .it is the span of a class i.e the difference between the upper limit and the lower limit is known as class interval , for example in the class 20-40 , the class interval is 20 (i.e 40 - 20) the size of the class interval is dertemined by the number of classes and the total range in the data.

- (c) Class frequency . this is the number of observation corresponding to the particular class.it is also known s the frequency of that class or

---

the class frequency .if we add together the frequency of all individual classes , we obtain the total frequency.

(d) Class mid - point.it is the value lying half way between the lower and the upper class limit of a class interval.

Mid point of a class is ascertained as follows:

Mid point of a class =  $\frac{\text{upper limit of the class} + \text{lower limit of the class}}{2}$

2

There are two methods of classifying the data according to class interval namely

- Exclusive method.
- Inclusive method.

(a) Exclusive method.

When the class interval are so fixed that the upper limit of one class is the lower limit of the next class, it is known as the exclusive method of classification .this can be illustrated as follow.

Class interval	No of items
10-20	5
20-30	3
30-40	4
40-50	6
50-60	2
60-70	1

It is clear that the exclusive method ensures continuity of data in as much as the upper limit of one class is the lower limit of the next class. Whenever this method is used it is always assumed that the upper limit is exclusive i.e the observation exactly equals to the upper limit is not included in that class.

(b) Inclusive method.

Under this method the upper limit of one class is included in that class itself.this can be illustrated as below:

Class interval	frequency
1-10	2
11-20	5
21-30	4
31-40	10
41-50	15
51-60	30



---

61-70	12
71-80	3
81-90	2

Whenever inclusive method is used for equal class interval is obtained by taking the difference between the two upper limit or lower limit.

#### DATA TABULATION.

- a table is a systematic arrangements of statistical data in column and rows.
- Rows are horizontal arrangements whereas columns are vertical ones.
- The purpose of a table is to simplify the presentation and to facilitate comparisons.
- The simplification result from the clear cut and systematic arrangements which enables the reader to quickly locate desired information.

#### Parts of a table.

- The various parts of a table may vary from case to case depending upon the given data, but a good table must contain atleast te following parts:
  - (a) Table number
  - (b) Title of the table
  - (c) Caption
  - (d) Stub
  - (e) Body of the table
  - (f) Headnote
  - (g) Footnote

- (a) table number. each table should be numbered.
- (b) Title of the table. every table must have a suitable title.
- (c) Caption. it referstothe column heading .it represent what the column represent.
- (d) Stub .are the designation of the rows or row heading .they are at the extreme left and perform the same function for the horizontal rows or numbers l the table as the ollumn heading do for the horizontal rows or numbers in the table as the column haeding do for the vertical column or numbers.
- (e) Body of the table. It contains the numerical information .data presented in the body arranged according to description are classification of the captions and stubs.

- 
- (f) Head note. It is a brief explanatory statement applying to all or a major part of the material in the table and is placed below the title entered and closed in brackets.
- (g) Footnote .they are placed directly below the body of the table.

## DIAGRAMATIC AND GRAPHICAL PRESENTATION .

### Diagrams.

#### Rules for construction of diagrams.

- There must be the title of the diagram .
- Proper proportion between width and height should be maintained.
- The scale selected should be appropriate
- If necessary footnotes should be given at the bottom of the diagram.
- Diagram should be absolutely neat and clean
- An index illustration different types of lines or different shades , column , should be give so that the reader can easily make out the meaning of the diagram.
- Diagram should be as simple as possible.

#### Types of diagrams.

- One dimensional diagram e.g bar diagram.
- Two dimensional diagram e.g squares .
- Picto grams and cartograms.

#### One dimensional diagrams.

- Bar diagrams are the most common type of diagram used in practice , a bar is a thick line whose width is shown merely for attention.
- They are called one dimensional because it is only the length of the bar that matters and not the width.

#### Merits of bar diagrams.

- They are readily understood even by those unaccustomed to reading charts or those who are not chart minded.
- They possess the outstanding advantage that they are the simplest and the easiest to make.
- When a large number of observations are to be compared they are the only form that can be used effectively..

#### Points to mind when constructing bar diagrams.

- The width of the bars should be uniform throughout the diagram.
- The gap between one bar and another should be uniform throughout.

- 
- Bars may either horizontal or vertical .the vertical bars should be preferred because they give a better look an also facilitate comparison.
  - While constructing the bar diagram it is desirable to write the respective figure at the end of each bar so that the reader can know the precise value without looking at the scale.

Types of bar diagrams.

- (a) Simple bar charts.
- (b) Sub divided or component bar charts.
- (c) Multiple bar charts
- (d) Percentage component barcharts.

(a) Simple bar charts.

It is used to represent only one variable i.e show s only totals. However an important limitation of such diagrams is that they can prefer only one classification or one category of data e.g for example while presenting the population for the last five decade , one can only depict the total population in the simple bar diagram and not its sex- wise distribution.

(b) Component bar charts.

These diagrams are used to represent various parts of the total , while constructing such charts, th various components in each bar should be kept in the same order.

A common and helpful arrangement is that of presenting each bar in the order of magnitude from the largest component at the base of the bar to the smallest at the end.

To distinguish between the different component , it is useful to use different shade or colours.

They usually show component totals and the totals.

(c) multiple bar charts

in multiple bar chart two or more set of interrelated data are represented .the technigue of drawing such a diagram is the same as that of simple bar charts.

The only difference is that since more than one phenomenon is represented ,different shades , colours .,dots or crossing are used to distinguish between the bars.they normally show the component totals only.

(d) Percentage component bar charts.

They are particularly useful in statistical work which require the portrayal of relative changes in data.

---

When such diagrams are prepared , the length of the bars is kept equal to 100 and segment are cut in these bars to represent the component of an aggregate.

Illustration.

Using the data below , construct the following charts.

- (a) Simple bar chart.
- (b) Component bar chart.
- (c) Multiple bar charts.
- (d) Percentage component bar chart.

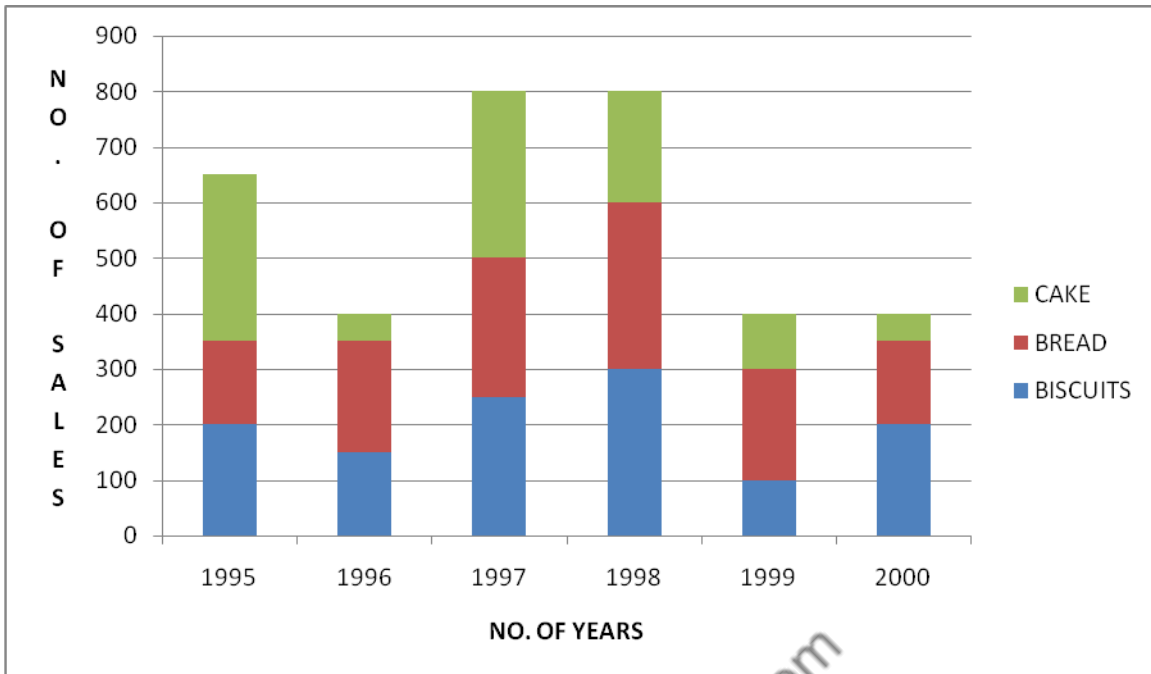
Shows the sales of Xyz Ltd.

year	biscuits	bread	cake	total
1995	200	150	300	650
1996	150	200	250	600
1997	250	250	300	700
1998	300	300	200	750
1999	100	200	100	400
2000	200	150	50	400

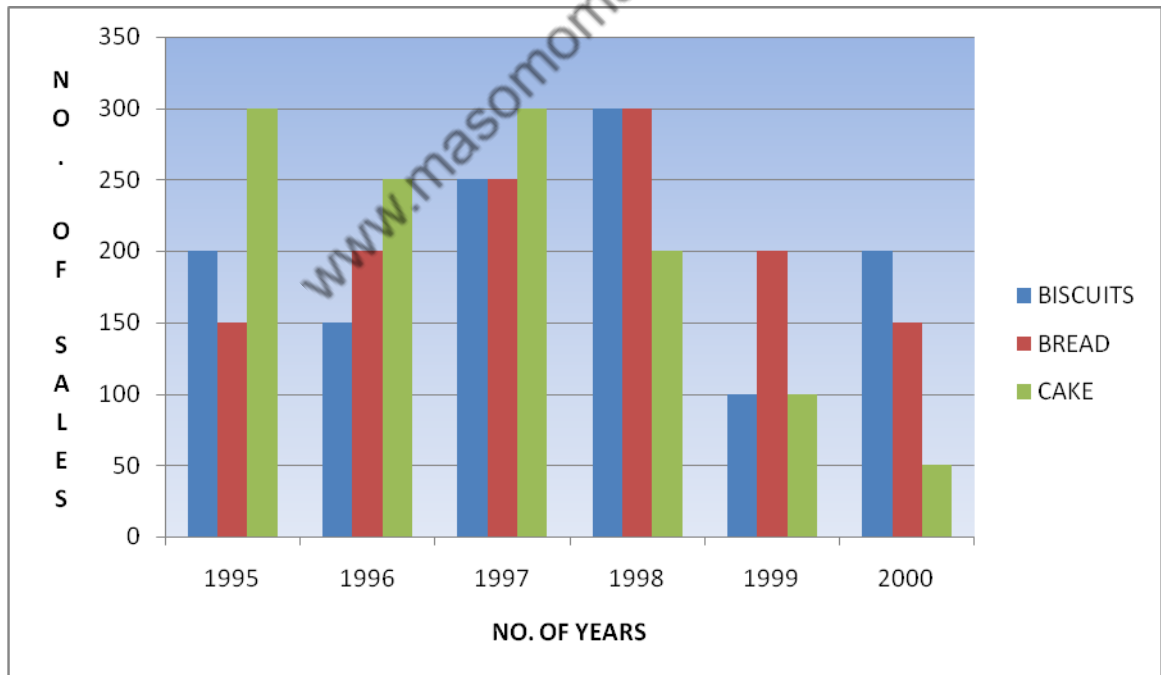
Solution

- (a) Component chart

www.masomomsingji.com



b) simple chart



---

## TWO DIMENSIONAL DIAGRAMS.

As distinguished from one dimensional diagram in which only the length of the bars is taken into account in two dimension diagram, the length as well as the width of the bars is considered.

Thus the area of the bar represents the given data. Two dimensional diagrams are also known as surface diagrams example of this are-

- Rectangles
- Squares
- Circles

### Pie chart or diagram

This type of diagram enables us to show the partitioning of a total into component parts.

In constructing a pie chart , the steps involved are:

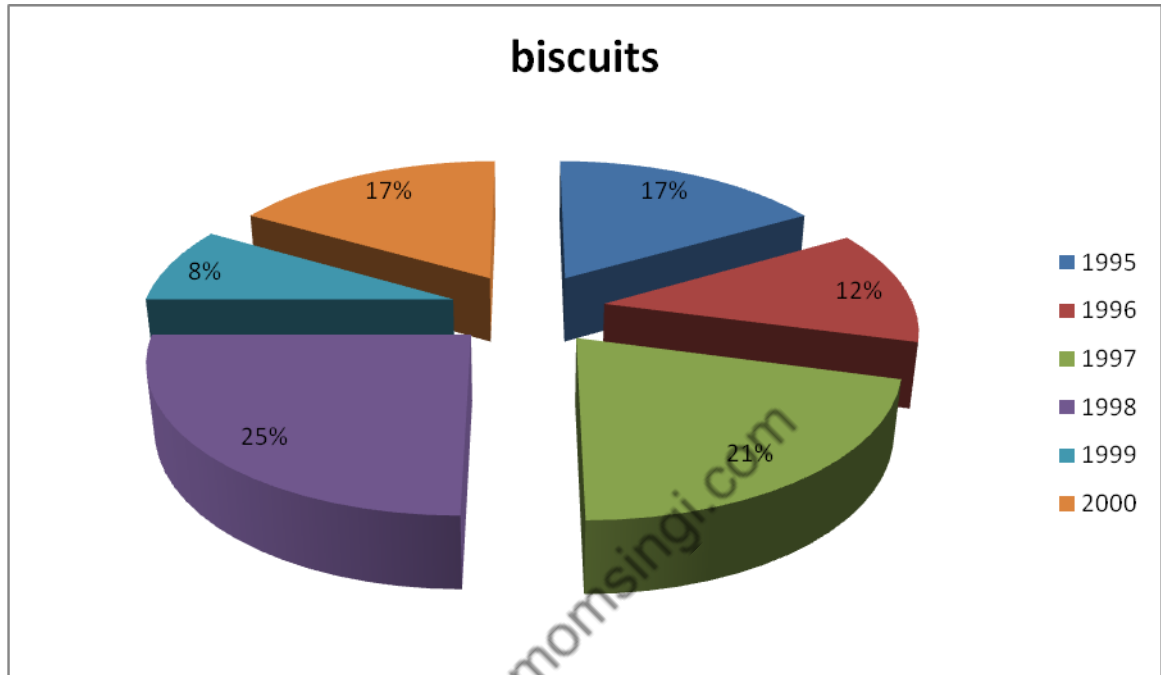
- Prepare the data so that the various components values are in percentage by applying

$$\% \text{ component} = \frac{x}{\text{Total}} \times 360$$

- Draw a circle of appropriate size with a compass.
- Measure points on the circle representing the size of each component with the help of a protractor.

---

Pie chart of the above.



Quantitative classification; it refers to the classification of data according to some characteristics that can be measured such as height, weight, income, sales.

### Tabulation of Data

One of the simplest and most revealing devices of summarizing data and presenting them in meaningful fashion is the statistical tables.

### Types of tabulation

Tabulation may be classified as:

1. simple tabulation
2. complex or matrix tabulation

**Simple tabulation** only one characteristic is shown hence this type of table is also known as one-path table. This can be illustrated as follows:

---

**Table 1: Distribution of workers by workshop**

Workshop	Number Of Employees
A	600
B	360
C	660
D	840
E	540
Total	3,000

N/B: simple tabulation does not tell us very much - although it may be enough for the question of the moment.

**Complex tabulation** has two or more characteristics are shown i.e. two or more aspects of a problem are dealt with at the same time. It is some times called two path tables. Such tables show two characteristics and are formed when either the stub or the caption is divided into two coordinate parts.

Age (in years)	Employees		total
	Males	females	
Bellow 25	32	18	50
25-35	40	27	67
35-45	25	18	43
45-55	10	5	15
55 and above	5	-	5
Total	112	68	180

### Presentation of data

After the data has been collected the next step is to present them in a suitable form.

### Charting data

One of the most convincing and appealing ways in which data may be presented is through charts

A chart can take the shape of either a diagram or a graph.

### Types of diagrams

- ❖ One dimensional diagram e.g. bar graphs
- ❖ Two dimensional diagram e.g. rectangles squares
- ❖ Pictograms and cartograms (circles)



---

### One dimensional diagram

Bar diagrams is the most common type of diagram used in practice.

#### Merits of bar diagrams

- ❖ They're readily understood
- ❖ They are simplest and easiest to make
- ❖ They are effective if the large number of observation is to be compared.

#### Example

### Two dimensional diagrams

As distinguished from one dimensional diagrams in which only the length of the bar is taken into account in two dimensional diagram the length as well as the width of the bars is considered. Thus the area of the bar represents the given data.

### Pictograms

Also known as picture grams are very popularly used in presenting situation data. They are no abstract presentation such as lines or bars

When pictograms are used data are represented through a pictorial symbol that is carefully selected.

Illustration

The following table gives the production of tea in India by a leading company.

Year	Production ( million kgs)
2004	421
2005	561
2006	587
2007	645
2008	660

Solution for representing the above data by a pictogram we will use the symbol of a star

Year	Production of tea
2004	◇ ◇ ◇ ◇
2005	◇ ◇ ◇ ◇ ◇ ◇
2006	◇ ◇ ◇ ◇ ◇ ◇
2007	◇ ◇ ◇ ◇ ◇ ◇ ◇

**Merits of diagrams**

- ❖ Pictograms have a greater attraction thus stimulate interest in the information being presented
- ❖ Facts portrayed in pictorial form are generally remembered longer.

**Limitations**

- ❖ They are difficult to construct

**Graphs**

A large variety of graphs are used in practice. Graphs can be divided under the following headings

- ❖ Time series.
- ❖ Z- Charts.
- ❖ Scatter graphs
- ❖ Semi- logarithmic graphs
- ❖ Graphs of frequency distribution

**Graphs of time series**

When we observe the values of a variable or different point of time the series forms is known as time series. The technique of graphic presentation is extremely helpful in analyzing change at different point of time.

Graphs of time series can be constructed either on a natural scale or on a ratio scale.

**Graphs of one variable**

When one variable is to be represented on x-axis measure time and on y-axis in that various points are plotted and joined by one straight line, the fluctuation of this line show the variation in the variable

**Illustration**

Represent the following data by suitable graph

Year	Imports(in million tones)
2003-04	1.5
2004-05	2.5
2005-06	2.0
2006-07	2.7

Solution

Draw a graph

Graphs of frequency distribution

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

Histogram

Frequency polygon

Smoothed frequency curve

Cumulative frequency curve

---

## Histogram

Histograms are a set of vertical bars whose areas are proportional to the frequencies represented. Histogram is most widely used for graphical presentation of a frequency distribution.

### Illustration

Represent the following data by a histogram

Marks	number of students
0-10	8
10-20	12
20-30	22
30-40	35
40-50	40

### Solution

Draw a histogram

### Frequency polygon

A frequency polygon is a graph of frequency distribution it has more than four sides

It is particularly effective in comparing two or more frequency distribution.

### Illustration

The daily profits (in thousand rupees) of 100 shops are distributed as follows

Daily profit	No of shops
0-50	12
50-100	18
100-150	21
150-200	20
200-250	17

### Solution

The frequency polygon of the above data are shown below

Draw graph

### Smoothed frequency curve

The following points should be kept in mind while smoothing a frequency graph.

Only frequency distribution based on sample should be smoothed

Only continuous series should be smoothed

The total area under the curve should be equal to the area under the original histogram or polygon.

### Illustration

Draw a histogram frequency polygon and frequency curve representing the following information

---

Length of service (in years)	Number of employees
5-10	5
10-15	12
15-20	25
20-25	48
25-30	32
30-35	6
	1

Histogram frequency polygon and curve

Draw

### Cumulative frequency curve / ogives

Sometimes one needs to know the answers to questions like how many workers of a factory earn more than Rs 1500 per month

There are two methods of constructing a cumulative frequency curve namely;

The less than method

The more than method

Less than method; in the less than method we start with the upper limit of the classes and go on adding the frequencies. When these frequencies are plotted we get a rising curve.

More than method; in the more than method, we start with the upper limits of the classes and from the total frequencies we subtract the frequency of each class. When the frequencies are plotted we get a declining curve.

given by less than method

Yearly profit(Rs)	Frequency
Less than 850	21
Less than 900	50
Less than 950	69
Less than 1000	108
Less than 1050	151
Less than 1100	245
Less than 1150	318
Less than 1200	422
Less than 1250	467
Less than 1300	494
Less than 1350	542
Less than 1400	563
Less than 1450	575
Less than 1500	580

Ogive by more than method

---

Yearly Rs	Frequency
More than 800	580
More than 850	559
More than 900	530
More than 950	511
More than 1000	472
More than 1050	429
More than 1100	335
More than 1150	262
More than 1200	194
More than 1250	158
More than 1300	113
More than 1350	86
More than 1400	38
More than 1450	17
More than 1500	5

Less than and more than ogive and median

Draw

#### Limitation of charts

- They can present only approximate values
- They can appropriately represent only limited amount of information
- They are intended mostly to explain quantitative facts to the general public.
- They can be easily misinterpreted