**Chapter- 1**

# **Business Statistics: An Introduction**

A set of numbers collected to study particular situations is known as **data**. These data are presented in systematic form in order to draw some direct inferences from the same. Also some other terms and quantities are calculated from the data to make better interpretations.

The study associated with all of the above is called **statistics**. Therefore, statistics contains collection and presentation of data, analyzing the data on the basis of the measures of central value, dimension etc.

The purpose to study business statistics in this course is to understand the basic statistical methods that are useful in decision making.

**Basic Definitions**

* **Statistics:** The collection of methods for planning experiments, obtaining data, and then organizing, summarizing, presenting, analyzing, interpreting, and drawing conclusions.
* **Data:** A set of numbers collected to study particular situations is known as data. It refers to any group of measurements that happen to interest us. These measurements provide information the decision maker uses.
* **Primary Data:** Primary data are measurements observed and recorded as part of original study. These are data not available elsewhere.
* **Secondary Data:** Data which are not originally collected but rather obtained from published or unpublished sources are called secondary data.
* **Variable:** Characteristic or attribute that can assume different values at different times, places or situations.
* **Random Variable:** A variable whose values are determined by chance.
* **Population:** All subjects possessing a common characteristic that is being studied.
* **Sample:** A sub- group or sub- set of the population.
* **Parameter:** Characteristic or measure obtained from a population.
* **Statistic** (*not to be confused with Statistics*): Characteristic or measure obtained from a sample.
* **Descriptive Statistics:** Collection, organization, summarization, and presentation of data.
* **Inferential Statistics**: Generalizing from samples to populations using probabilities. Performing hypothesis testing, determining relationships between variables, and making predictions.
* **Qualitative Variables:** Variables which assume non-numerical values.
* **Quantitative Variables:** Variables which assume numerical values.
* **Discrete Variables:** Variables which assume a finite or countable number of possible values. Usually obtained by counting.
* **Continuous Variables:** Variables which assume an infinite number of possible values. Usually obtained by measurement.
* **Nominal Level:** Level of measurement which classifies data into mutually exclusive, all-inclusive categories in which no order or ranking can be imposed on the data.
* **Ordinal Level:** Level of measurement which classifies data into categories that can be ranked. Differences between the ranks do not exist.
* **Interval Level:** Level of measurement which classifies data that can be ranked and differences are meaningful. However, there is no meaningful zero, so ratios are meaningless.
* **Ratio Level:** Level of measurement which classifies data that can be ranked, differences are meaningful, and there is a true zero. True ratios exist between the different units of measure.

**Collection of Data**

Data may be obtained either from the primary source or the secondary source. A primary source is one that itself collects the data whereas a secondary source is one that makes available data which were collected by some other agency.

**Choice between Primary and Secondary Data:** the investigator must decide at the outset whether he will use primary data or secondary data in an investigation. The choice between the two depends mainly on the following considerations:

* Nature and scope of the enquiry;
* Availability of time;
* Degree of accuracy desired; and
* The collecting agency, i.e., whether an individual, an institute or a Government body.

It may be pointed out that most statistical analysis rests upon secondary data. Primary data are generally used in those cases where the secondary data do not provide an adequate basis for analysis.

**Methods of Collecting Primary Data:**

* Direct personal interviews;
* Indirect oral interviews;
* Information from correspondents;
* Mailed questionnaire method; and
* Schedules sent through enumerators.

**Sources of Secondary Data:**

* Published sources; and
* Unpublished sources

**Editing Primary and Secondary Data:**

Once the data have been obtained either from primary or secondary source, the next step in a statistical investigation is to edit the data, i.e., to scrutinize the data. While editing primary data the following considerations need attention:

* The data should be complete;
* The data should be consistent;
* The data should be accurate; and
* The data should be homogeneous.

**Precautions in the Use of Secondary Data:**

* Whether the data are suitable for the purpose of investigation;
* Whether the data are adequate for investigation; and
* Whether the data are reliable or not.

**Sampling and Sample Designs**

When secondary data are not available for the problem under study, a decision may be taken to collect primary data. The required information may be obtained by following either the census method or the sample method.

**Census Method:**

Information on population can be collected in two ways – census method and sample method. In census method every element of the population is included in the investigation. For example, if we study the average annual income of the families of a particular village or area, and if there are 1000 families in that area, we must study the income of all 1000 families. In this method no family is left out, as each family is a unit.

**Merits and limitations of Census method:**

**Mertis:**

* + - 1. The data are collected from each and every item of the population
			2. The results are more accurate and reliable, because every item of the universe is required.
			3. Intensive study is possible.
			4. The data collected may be used for various surveys, analyses etc.

**Limitations:**

1. It requires a large number of enumerators and it is a costly method
2. It requires more money, labour, time energy etc.
3. It is not possible in some circumstances where the universe is infinite.

## **Sample:**

Statisticians use the word **sample** to describe a portion chosen from the population. A finite subset of statistical individuals defined in a population is called a sample. The number of units in a sample is called the **sample size.**

## **Sampling frame:**

For adopting any sampling procedure it is essential to have a list identifying each sampling unit by a number. Such a list or map is called sampling frame. A list of voters, a list of house holders, a list of villages in a district, a list of farmers etc. are a few examples of sampling frame.

## **Principles of Sampling:**

Samples have to provide good estimates. The following principle tell us that the sample methods provide such good estimates

## **Principle of statistical regularity:**

A moderately large number of units chosen at random from a large group are almost sure on the average to possess the characteristics of the large group.

## **Principle of Inertia of large numbers:**

Other things being equal, as the sample size increases, the results tend to be more accurate and reliable.

## **Principle of Validity:**

This states that the sampling methods provide valid estimates about the population units (parameters).

## **Principle of Optimization:**

This principle takes into account the desirability of obtaining a sampling design which gives optimum results. This minimizes the risk or loss of the sampling design.

The foremost purpose of sampling is to gather maximum information about the population under consideration at minimum cost, time and human power.

## **Types of Sampling:**

The technique of selecting a sample is of fundamental importance in sampling theory and it depends upon the nature of investigation. The sampling procedures which are commonly used may be classified as

1. Probability sampling.
2. Non-probability sampling.
3. Mixed sampling.

## **Probability sampling (Random sampling):**

A probability sample is one where the selection of units from the population is made according to known probabilities. (eg.) Simple random sample, probability proportional to sample size etc.

## **Non-Probability sampling:**

It is the one where discretion is used to select ‘representative’ units from the population (or) to infer that a sample is ‘representative’ of the population. This method is called **judgement or purposive** sampling. This method is mainly used for opinion surveys; A common type of judgement sample used in surveys is quota sample. This method is not used in general because of prejudice and bias of the enumerator. However if the enumerator is experienced and expert, this method may yield valuable results. For example, in the market research survey of the performance of their new car, the sample was all new car purchasers.

## **Mixed Sampling:**

Here samples are selected partly according to some probability and partly according to a fixed sampling rule; they are termed as mixed samples and the technique of selecting such samples is known as **mixed sampling**.

## **Methods of selection of samples:**

Here we shall consider the following three methods:

1. Simple random sampling.
2. Stratified random sampling.
3. Systematic random sampling.

## **Simple random sampling:**

A simple random sample from finite population is a sample selected such that each possible sample combination has equal probability of being chosen. It is also called unrestricted random sampling.

## **Simple random sampling without replacement:**

In this method the population elements can enter the sample only once (ie) the units once selected is not returned to the population before the next draw.

## **Simple random sampling with replacement:**

In this method the population units may enter the sample more than once. Simple random sampling may be with or without replacement.

**Frequency Distribution**

# **Introduction:**

Frequency distribution is a series when a number of observations with similar or closely related values are put in separate bunches or groups, each group being in order of magnitude in a series. It is simply a table in which the data are grouped into classes and the number of cases which fall in each class are recorded. It shows the frequency of occurrence of different values of a single Phenomenon.

# **A frequency distribution is constructed for three main reasons:**

* + 1. To facilitate the analysis of data.
		2. To estimate frequencies of the unknown population distribution from the distribution of sample data and
		3. To facilitate the computation of various statistical measures

# **Raw data:**

The statistical data collected are generally raw data or ungrouped data. Let us consider the daily wages (in SR) of 30 laborers in a factory.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 80 | 70 | 55 | 50 | 60 | 65 | 40 | 30 | 80 | 90 |
| 75 | 45 | 35 | 65 | 70 | 80 | 82 | 55 | 65 | 80 |
| 60 | 55 | 38 | 65 | 75 | 85 | 90 | 65 | 45 | 75 |

The above figures are nothing but raw or ungrouped data and they are recorded as they occur without any pre consideration. This representation of data does not furnish any useful information and is rather confusing to mind. A better way to express the figures in an ascending or descending order of magnitude and is commonly known as array. But this does not reduce the bulk of the data. The above data when formed into an array is in the following form:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 30 | 35 | 38 | 40 | 45 | 45 | 50 | 55 | 55 | 55 |
| 60 | 60 | 65 | 65 | 65 | 65 | 65 | 65 | 70 | 70 |
| 75 | 75 | 75 | 80 | 80 | 80 | 80 | 85 | 90 | 90 |

The array helps us to see at once the maximum and minimum values. It also gives a rough idea of the distribution of the items over the range. When we have a large number of items, the formation of an array is very difficult, tedious and cumbersome. The Condensation should be directed for better understanding and may be done in two ways, depending on the nature of the data.

**Example:**

***In a survey of 40 families in a village, the number of children per family was recorded and the following data obtained.***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 0 | 3 | 2 | 1 | 5 | 6 | 2 |
| 2 | 1 | 0 | 3 | 4 | 2 | 1 | 6 |
| 3 | 2 | 1 | 5 | 3 | 3 | 2 | 4 |
| 2 | 2 | 3 | 0 | 2 | 1 | 4 | 5 |
| 3 | 3 | 4 | 4 | 1 | 2 | 4 | 5 |

***Represent the data in the form of a discrete frequency distribution.***

**Solution:** Frequency distribution of the number of children

|  |  |  |
| --- | --- | --- |
| **Number of** | **Tally** | **Frequency** |
| **Children** | **Marks** |  |
| 0 |  | 3 |
| 1 |  | 7 |
| 2 |  | 10 |
| 3 |  | 8 |
| 4 |  | 6 |
| 5 |  | 4 |
| 6 |  | 2 |
|  | Total | 40 |

**b) Continuous frequency distribution:**

In this form of distribution refers to groups of values. This becomes necessary in the case of some variables which can take any fractional value and in which case an exact measurement is not possible. Hence a discrete variable can be presented in the form of a continuous frequency distribution.

Wage distribution of 100 employees

|  |  |
| --- | --- |
| **Weekly wages** | **Number of** |
| **(SR)** | **employees** |
| 50-100 | 4 |
| 100-150 | 12 |
| 150-200 | 22 |
| 200-250 | 33 |
| 250-300 | 16 |
| 300-350 | 8 |
| 350-400 | 5 |
| Total | 100 |

**Nature of class:**

The following are some basic technical terms when a continuous frequency distribution is formed or data are classified according to class intervals.

**a)** **Class limits:**

The class limits are the lowest and the highest values that can be included in the class. For example, take the class 30-40. The lowest value of the class is 30 and highest class is 40. In statistical calculations, lower class limit is denoted by L and upper class limit by U.

**b) Class Interval:**

The class interval may be defined as the size of each grouping of data. For example, 50-75, 75-100, 100-125…are class intervals. Each grouping begins with the lower limit of a class interval and ends at the lower limit of the next succeeding class interval

**c) Width or size of the class interval:**

The difference between the lower and upper class limits is called Width or size of class interval and is denoted by ‘C’.

**d) Range:**

The difference between largest and smallest value of the observation is called The Range and is denoted by ‘R’ ie

R = Largest value – Smallest value

* 1. =L - S
1. **Mid-value or mid-point:**

The central point of a class interval is called the mid value or mid-point. It is found out by adding the upper and lower limits of a class and dividing the sum by 2.

i.e., Mid- Value = $\frac{L+U}{2}$

For example, if the class interval is 20-30 then the mid-value is $\frac{20+30}{2}$ = $\frac{50}{2}$ = 25

**f) Frequency:**

Number of observations falling within a particular class interval is called frequency of that class.

Let us consider the frequency distribution of weights if persons working in a company.

|  |  |
| --- | --- |
| **Weight** | **Number of** |
| **(in kgs)** | **persons** |
| 30-40 | 25 |
| 40-50 | 53 |
| 50-60 | 77 |
| 60-70 | 95 |
| 70-80 | 80 |
| 80-90 | 60 |
| 90-100 |  30 |
|  Total |  420 |

In the above example, the class frequencies are 25,53,77,95,80,60,30. The total frequency is equal to 420. The total frequencies indicate the total number of observations considered in a frequency distribution.

In the above example, the class frequencies are 25,53,77,95,80,60,30. The total frequency is equal to 420. The total frequencies indicate the total number of observations considered in a frequency distribution.

**g) Number of class intervals:**

The number of class interval in a frequency is matter of importance. The number of class interval should not be too many. For an ideal frequency distribution, the number of class intervals can vary from 5 to 15. To decide the number of class intervals for the frequency distribution in the whole data, we choose the lowest and the highest of the values. The difference between them will enable us to decide the class intervals.

Thus the number of class intervals can be fixed arbitrarily keeping in view the nature of problem under study or it can be decided with the help of Sturges’ Rule. According to him, the number of classes can be determined by the formula

K = 1 + 3. 322 log10 N

Where N = Total number of observations; log = logarithm of the number

K = Number of class intervals.

Thus if the number of observation is 10, then the number of class intervals is

K = 1 + 3. 322 log 10 = 4.322 @ 4

If 100 observations are being studied, the number of class interval is

K = 1 + 3. 322 log 100 = 7.644 @ 8 and so on.

**h) Size of the class interval:**

Since the size of the class interval is inversely proportional to the number of class interval in a given distribution. The approximate value of the size (or width or magnitude) of the class interval ‘C’ is obtained by using Sturges’ rule as

Size of Class- Interval, C = $\frac{Range}{No of Class Interval}$

Range

=

1+3.322 log10 N

Where Range = Largest Value – Smallest Value in the distribution.

**Types of class intervals:**

There are three methods of classifying the data according to class intervals namely

* 1. Exclusive method
	2. Inclusive method
	3. Open-end classes
1. **Exclusive method:**

When the class intervals 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.

**Example:**

|  |  |
| --- | --- |
| **Expenditure** | **No. of families** |
| **(SR)** |  |
| 0 -5000 | 60 |
| 5000-10000 | 95 |
| 10000-15000 | 122 |
| 15000-20000 | 83 |
| 20000-25000 | 40 |
| Total | 400 |

It is clear that the exclusive method ensures continuity of data as much as the upper limit of one class is the lower limit of the next class. In the above example, there are so families whose expenditure is between SR.0 and SR.4999.99. A family whose expenditure is SR.5000 would be included in the class interval 5000-10000. This method is widely used in practice.

**b) Inclusive method:**

In this method, the overlapping of the class intervals is avoided. Both the lower and upper limits are included in the class interval.

**Example:**

|  |  |
| --- | --- |
| **Class interval** | **Frequency** |
| 5-9 | 7 |
| 10-14 | 12 |
| 15-19 | 15 |
| 20-29 | 21 |
| 30-34 | 10 |
| 35-39 | 5 |
| Total | 70 |

Thus to decide whether to use the inclusive method or the exclusive method, it is important to determine whether the variable under observation in a continuous or discrete one.

In case of continuous variables, the exclusive method must be used. The inclusive method should be used in case of discrete variable.

**c) Open end classes:**

A class limit is missing either at the lower end of the first class interval or at the upper end of the last class interval or both are not specified. The necessity of open end classes arises in a number of practical situations, particularly relating to economic and medical data when there are few very high values or few very low values which are far apart from the majority of observations.

**Example:**

|  |  |
| --- | --- |
| **Salary Range** | **No of** |
|  | **workers** |
| Below 2000 | 7 |
| 2000 – 4000 | 5 |
| 4000 – 6000 | 6 |
| 6000 – 8000 | 4 |
| 8000 and | 3 |
| above |  |

**Construction of frequency table:**

Constructing a frequency distribution depends on the nature of the given data. Hence, the following general consideration may be borne in mind for ensuring meaningful classification of data.

1. The number of classes should preferably be between 5 and 20. However there is no rigidity about it.
2. As far as possible one should avoid values of class intervals as 3,7,11, 26….etc. preferably one should have class-intervals of either five or multiples of 5 like 10, 20, 25, 100 etc.
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 class interval we should adopt “exclusive” method.
5. Wherever possible, it is desirable to use class interval of equal sizes.

**Preparation of frequency table:**

**Example 1:**

Let us consider the weights in kg of 50 college students.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 42 | 62 | 46 | 54 | 41 | 37 | 54 | 44 | 32 | 45 |
| 47 | 50 | 58 | 49 | 51 | 42 | 46 | 37 | 42 | 39 |
| 54 | 39 | 51 | 58 | 47 | 64 | 43 | 48 | 49 | 48 |
| 49 | 61 | 41 | 40 | 58 | 49 | 59 | 57 | 57 | 34 |
| 56 | 38 | 45 | 52 | 46 | 40 | 63 | 41 | 51 | 41 |

Here the size of the class interval as per Sturges’ rule is obtained as follows

Size of Class Interval, C = $\frac{Range}{1+3.322 Log N}$

 = $\frac{64-32}{1+3.322 Log 50}$

 =$ \frac{ 32}{6.64}$ = 5

Thus the number of class interval is 7 and size of each class is 5. The required size of each class is 5. The required frequency distribution is prepared using tally marks as given below:

|  |  |  |
| --- | --- | --- |
| **Class Interval** | **Tally marks** | **Frequency** |
| 30-35 |  | 2 |
| 35-40 |  | 6 |
| 40-45 |  | 12 |
| 45-50 |  | 14 |
| 50-55 |  | 6 |
| 55-60 |  | 6 |
| 60-65 |  | 4 |
| Total |  | 50 |

**Example 2:**

***Given below are the numbers of tools produced by workers in a factory.***

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 43 | 18 | 25 | 18 | 39 | 44 | 19 | 20 | 20 | 26 |
| 40 | 45 | 38 | 25 | 13 | 14 | 27 | 41 | 42 | 17 |
| 34 | 31 | 32 | 27 | 33 | 37 | 25 | 26 | 32 | 25 |
| 33 | 34 | 35 | 46 | 29 | 34 | 31 | 34 | 35 | 24 |
| 28 | 30 | 41 | 32 | 29 | 28 | 30 | 31 | 30 | 34 |
| 31 | 35 | 36 | 29 | 26 | 32 | 36 | 35 | 36 | 37 |
| 32 | 23 | 22 | 29 | 33 | 37 | 33 | 27 | 24 | 36 |
| 23 | 42 | 29 | 37 | 29 | 23 | 44 | 41 | 45 | 39 |
| 21 | 21 | 42 | 22 | 28 | 22 | 15 | 16 | 17 | 28 |
| 22 | 29 | 35 | 31 | 27 | 40 | 23 | 32 | 40 | 37 |

 ***Construct frequency distribution with inclusive type of class interval. Also find.***

1. ***How many workers produced more than 38 tools?***
2. ***How many workers produced less than 23 tools?***

**Solution:**

Using Sturges’ formula for determining the number of class intervals, we have

Number of class intervals = 1+ 3.322 log10N

* 1+ 3.322 log10100
* 7.6

Sizes of class interval = $\frac{Range}{No of Class Interval}$

 = $\frac{46-13}{7.6}$ = 5

Hence taking the magnitude of class intervals as 5, we have 7 classes 13-17, 18-22… 43-47 are the classes by inclusive type. Using tally marks, the required frequency distribution is obtain in the following table-

|  |  |  |
| --- | --- | --- |
| **Class** | **Tally Marks** | **Number of** |
| **Interval** |  | **tools produced** |
|  |  | **(Frequency)** |
| 13-17 |  | 6 |
| 18-22 |  | 11 |
| 23-27 |  | 18 |
| 28-32 |  | 25 |
| 33-37 |  | 22 |
| 38-42 |  | 11 |
| 43-47 |  | 7 |
| **Total** |  |  **100** |

**Cumulative frequency table:**

**Example:**

|  |  |  |  |
| --- | --- | --- | --- |
| Age group (in yrs) | No of Women | Less than Cumulative frequency | More than Cumulative frequency |
| 15-20 | 3 | 3 | 64 |
| 20-25 | 7 | 10 | 61 |
| 25-30 | 15 | 25 | 54 |
| 30-35 | 21 | 46 | 39 |
| 35-40 | 12 | 58 | 18 |
| 40-45 | 6 | 64 | 6 |

**Less than cumulative frequency distribution table**

|  |  |
| --- | --- |
| End values upper | Less than Cumulative |
| limit | frequency |
| Less than 20 | 3 |
| Less than 25 | 10 |
| Less than 30 | 25 |
| Less than 35 | 46 |
| Less than 40 | 58 |
| Less than 45 | 64 |

**(b) More than** **cumulative frequency distribution table**

|  |  |
| --- | --- |
| End values lower | Cumulative frequency |
| limit | more than |
| 15 and above | 64 |
| 20 and above | 61 |
| 25 and above | 54 |
| 30 and above | 39 |
| 35 and above | 18 |
| 40 and above | 6 |

**Conversion of cumulative frequency to simple Frequency:**

If we have only cumulative frequency ‘either less than or more than’, we can convert it into simple frequencies. For example if we have ‘less than Cumulative frequency, we can convert this to simple frequency by the method given below:

|  |  |  |
| --- | --- | --- |
| **Class interval** | **‘ less than’** | **Simple frequency** |
|  | **Cumulative frequency** |  |
| 15-20 | 3 | 3 |
| 20-25 | 10 | 10 - 3=7 |
| 25-30 | 25 | 25 - 10 = 15 |
| 30-35 | 46 | 46 - 25 = 21 |
| 35-40 | 58 | 58 - 46 = 12 |
| 40-45 | 64 | 64 - 58 =6 |

Method of converting ‘more than’ cumulative frequency to simple frequency is given below.

|  |  |  |  |
| --- | --- | --- | --- |
| **Class interval** | **‘ more than’** | **Simple frequency** |  |
|  | **Cumulative frequency** |  |  |  |  |
| 15-20 | 64 | 64 | - 61 = | 3 |
| 20-25 | 61 | 61 | - 54 = | 7 |
| 25-30 | 54 | 54 | -39 = 15 |
| 30-35 | 39 | 39 | - 18 = | 21 |
| 35-40 | 18 | 18 | - 6 | = 12 |
| 40-45 | 6 |  6 - 0 | = | 6 |

**Diagrammatic and Graphical Representation**

# **Introduction:**

In the previous chapter, we have discussed the techniques of classification and tabulation that help in summarizing the collected data and presenting them in a systematic manner.

# **Diagrams:**

A diagram is a visual form for presentation of statistical data, highlighting their basic facts and relationship. If we draw diagrams on the basis of the data collected they will easily be understood.

# **Significance of Diagrams and Graphs:**

Diagrams and graphs are extremely useful because of the following reasons.

* + 1. They are attractive and impressive.
		2. They make data simple and intelligible.
		3. They make comparison possible
		4. They save time and labour.
		5. They have universal utility.
		6. They give more information.
		7. They have a great memorizing effect.

# **Types of diagrams:**

In practice, a very large variety of diagrams are in use and new ones are constantly being added. For the sake of convenience and simplicity, they may be divided under the following heads:

* + 1. One-dimensional diagrams
		2. Two-dimensional diagrams
		3. Three-dimensional diagrams
		4. Pictograms and Cartograms

# **One-dimensional diagrams:**

In such diagrams, only one-dimensional measurement, i.e height is used and the width is not considered. These diagrams are in the form of bar or line charts and can be classified as

1. Line Diagram
2. Simple Diagram
3. Multiple Bar Diagram
4. Sub-divided Bar Diagram
5. Percentage Bar Diagram

# **Line Diagram:**

# **Example:**

***Show the following data by a line chart:***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| No. of children | 0 | 1 | 2 | 3 | 4 | 5 |
| Frequency | 10 | 14 | 9 | 6 | 4 | 2 |

**Line Diagram**

**No. of Children**

# **Simple Bar Diagram:**

# **Example:**

***Represent the following data by a bar diagram.***

|  |  |
| --- | --- |
| Year | Production (in tones) |
| 1991 | 45 |
| 1992 | 40 |
| 1993 | 42 |
| 1994 | 55 |
| 1995 | 50 |

# **Solution:**

 **Simple Bar Diagram**

1991 1992 1993 1994 1995

# **Multiple Bar Diagram:**

# **Example:**

**Draw a multiple bar diagram for the following data.**

|  |  |  |
| --- | --- | --- |
| Year | Profit before tax( in lakhs of rupees ) | Profit after tax( in lakhs of rupees ) |
| 1998 | 195 | 80 |
| 1999 | 200 | 87 |
| 2000 | 165 | 45 |
| 2001 | 140 | 32 |

# **Solution:**

**Multiple bar Diagram**

 **1998 1999 2000 2001**

**Year**

**Pie Diagram or Circular Diagram:**

# **Example:**

**Draw a Pie diagram for the following data of production of sugar in quintals of various countries.**

|  |  |
| --- | --- |
| **Country** | **Production of Sugar (in quintals)** |
| Cuba | 62 |
| Australia | 47 |
| India | 35 |
| Japan | 16 |
| Egypt | 6 |

# **Solution:**

The values are expressed in terms of degree as follows.

|  |  |
| --- | --- |
| **Country** | **Production of Sugar** |
| **In Quintals** | **In Degrees** |
| Cuba | 62 | 134 |
| Australia | 47 | 102 |
| India | 35 | 76 |
| Japan | 16 | 35 |
| Egypt | 6 | 13 |
| Total | 166 | 360 |

# **Graphs:**

A graph is a visual form of presentation of statistical data. A graph is more attractive than a table of figure. Even a common man can understand the message of data from the graph. Comparisons can be made between two or more phenomena very easily with the help of a graph.

However here we shall discuss only some important types of graphs which are more popular and they are-

* + 1. Histogram
		2. Frequency Polygon
		3. Frequency Curve
		4. Ogive
		5. Lorenz Curve

# **Histogram:**

# **Example: *Draw a histogram for the following data.***

|  |  |
| --- | --- |
| Daily Wages | Number of Workers |
| 0-50 | 8 |
| 50-100 | 16 |
| 100-150 | 27 |
| 150-200 | 19 |
| 200-250 | 10 |
| 250-300 | 6 |

**Example:**

***For the following data, draw a histogram.***

|  |  |
| --- | --- |
| Marks | Number of Students |
| 21-30 | 6 |
| 31-40 | 15 |
| 41-50 | 22 |
| 51-60 | 31 |
| 61-70 | 17 |
| 71-80 | 9 |

# **Solution:**

For drawing a histogram, the frequency distribution should be continuous. If it is not continuous, then first make it continuous as follows.

|  |  |
| --- | --- |
| Marks | Number of Students |
| 20.5-30.5 | 6 |
| 30.5-40.5 | 15 |
| 40.5-50.5 | 22 |
| 50.5-60.5 | 31 |
| 60.5-70.5 | 17 |
| 70.5-80.5 | 9 |

# **Frequency Polygon:**

# **Example:**

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

|  |  |
| --- | --- |
| Weight (in kg) | Number of Students |
| 30-35 | 4 |
| 35-40 | 7 |
| 40-45 | 10 |
| 45-50 | 18 |
| 50-55 | 14 |
| 55-60 | 8 |
| 60-65 | 3 |

# **Frequency Curve:**

# **Example:**

**Draw a frequency curve for the following data.**

|  |  |
| --- | --- |
| Monthly Income (in SR) | No. of family |
| 0-1000 | 21 |
| 1000-2000 | 35 |
| 2000-3000 | 56 |
| 3000-4000 | 74 |
| 4000-5000 | 63 |
| 5000-6000 | 40 |
| 6000-7000 | 29 |
| 7000-8000 | 14 |

# **Ogives:**

There are two methods of constructing ogive namely:

* + - 1. The ‘ less than ogive’ method
			2. The ‘more than ogive’ method.

In less than ogive method we start with the upper limits of the classes and go adding the frequencies. When these frequencies are plotted, we get a rising curve. In more than ogive method, we start with the lower limits of the classes and from the total frequencies we subtract the frequency of each class. When these frequencies are plotted we get a declining curve.

# **Example: Draw the Ogives for the following data.**

|  |  |
| --- | --- |
| Class interval | Frequency |
| 20-30 | 4 |
| 30-40 | 6 |
| 40-50 | 13 |
| 50-60 | 25 |
| 60-70 | 32 |
| 70-80 | 19 |
| 80-90 | 8 |
| 90-100 | 3 |

# **Solution:**

|  |  |  |
| --- | --- | --- |
| Class limit | Less than ogive | More than ogive |
| 20 | 0 | 110 |
| 30 | 4 | 106 |
| 40 | 10 | 100 |
| 50 | 23 | 87 |
| 60 | 48 | 62 |
| 70 | 80 | 30 |
| 80 | 99 | 11 |
| 90 | 107 | 3 |
| 100 | 110 | 0 |

**Cumulative frequency**