17

COLLECTION AND PRESENTATION OF DATA

Getting information on various things around us has become a way of life. Information itself is a major source of knowledge. Without information it is difficult to take decisions. With development of science and technology the sources of information have increased and become accessible as well. Books, News papers, magazines, telephone, television, internet and mobile phones etc. are all medium of providing information of various kinds.

Information is both qualitative and quantitative in nature. Good, bad, ugly, beautiful, responsible, noble, handsome, educated etc are terms used to describe persons, can be said to be qualitative in nature. Information on income, expenditure, savings, rate of growth, height, weight, marks secured, population, food production, etc are given in quantitative or numerical terms. In the study of economics quantitative informations are mostly used for analysis.

OBJECTIVES

After completing this lesson, you will be able to:

- *understand the meaning of the term data;*
- *distinguish between various types of data;*
- *distinguish between variables and attributes;*
- *identify the areas of an economy where we cannot do without the data;*
- *classify and tabulate data;*
- *understand various forms of presentation of data.*

17.1 MEANING AND FEATURES OF DATA

Data means quantitative information providing facts in an aggregate manner. The information could be on any thing that can be given numerically and useful for decision

Notes

making. It is also called statistical data or simply statistics. Data is a plural term. The singular of data is datum.

From the meaning we can give some features of the term statistics or data below with example.

(i) Statistics are the aggregate of facts

A single fact cannot be considered as statistics or data. For example, the marks secured by a student of class X in mathematics are 95. This is given as single information which is simply a fact and not the data. However, the marks secured by all the students of class X of a school, either section wise or in total can be considered data, because it becomes an aggregate of facts. By just telling the marks of one student, we cannot know the performance of others and accordingly we cannot carry out any analysis to recommend for their betterment. This means that by giving aggregate of facts, data become meaningful as it provides scope for carrying out analysis.

See the table below. It gives the marks secured by all the 18 students of a class in mathematics. By looking at this we can compare the performance of the whole class. So this is an example of data.

Table 17.1

Students	Marks	Students	Marks
A	95	J	35
B	90	K	30
C	75	L	85
D	65	M	20
E	90	N	90
F	100	O	80
G	80	P	70
H	45	Q	100
I	40	R	50

From the above data we can know the following

(i) How many students have secured more than 90? (ii) How many students have failed? (iii) How many students secured less than 50? On the basis of the answers to these questions, the teacher can take necessary steps to improve the performance of students wherever needed. So in this way as aggregate of facts data are more meaningful than any single information.

(ii) Numerically expressed

Statistics or data are always quantitative in nature. Qualitative information such as good, bad, average, handsome, ugly are examples of some attributes, the magnitude of which can not be quantified and as such these can not be called statistics or data. When facts are put into a framework of numbers either through counting and calculation or estimation, these may be called data. In the above table marks of students are given numerically. We can give another example as in table 17.2 below which shows number of students admitted in the 1st year in different colleges in an imaginary city.

Table 17.2

College	Number of Students
Govt. College	409
Savitri College	308
J.P. College	401
N.D. College	510

(iii) Data are affected to a marked extent by multiplicity of causes

Data are not influenced by a single factor but are influenced by many factors. For Example, rise in prices of commodities may have been due to several causes like, reduction in supply, increase in demand, rise in taxes, rise in wages etc.

(iv) Reasonable standard of accuracy

100% accuracy in statistics is neither possible nor desirable. What is needed and expected is only a reasonable standard of accuracy. If a doctor has invented a new medicine to control cholesterol and statistically he ascertained that 90% of patients have responded well and statistically if 95% persons responded to the treatment, it may be considered that the new medicine is good and it has reasonable standard of accuracy as the results show that only 90% of patients have responded well and not 100%. It reflects reasonable standard of accuracy.

(v) Predetermined purpose

Data are collected for a predetermined purpose. Both the above tables serve some important purposes. The data in table 17.1 can be used to evaluate the performance of students in mathematics. Data in table 17.2 can be used to know the situation of higher education in the city to some extent on the basis of knowing number of young people entering college.

17.2 IMPORTANCE OF DATA IN ECONOMICS

Some specific areas of economics where the use of data is very important are as follows:

1. **In economic planning:** The data of the previous years are generally used to prepare future plan. For example, if we have to plan expenditure to be incurred on primary education for a year, data regarding number of students who were enrolled up to class fifth in previous years and the expenditure incurred during those years is important to look at. Forecasting is done on the basis of economic planning. For example, if we want to predict the growth of per capita income of a country, the data on the growth rate of population and the national income are also to be collected and considered.

2. **To determine national income:** In order to know the state of our economy it is important to know the national income besides various other things. But national income can be determined by using certain methods which require quantitative information on various things such as wages and salaries received by workers, rent received for use of land and building, interest received for use of funds and profit earned by the entrepreneurs in the economy in the given year.

3. **Basis of government policies**: Statistical data are widely used by government to frame policies for economic development of the country. On the basis of data on the vast number of poor and unemployed people in India, the government had to make policy to remove poverty and unemployment by enacting National Rural Employment Guarantee Act. This policy of the government guarantees an unemployed person at least 100 days of wage employment in a year. In India Census which is carried out once in every 10 years provide data on male and female population, number of literates, number of workers etc. On the basis of the data on male and female population it was found that India has 938 females per 1000 males. In some states like Haryana there are only 848 females per 1000 males. This is a very alarming situation because one of the reasons for low female population is killing of girl child before its taking birth. On the basis of this data now the government is making policy to save the girl child.

INTEXT QUESTIONS 17.1

1. Identify whether following are data or not. Write yes/no in the bracket

   (i)   Miss Monika secured 75% marks in economics   ( )

   (ii) Krish is a better player than Hari   ( )

   (iii) Lalita secured good marks   ( )

   (iv) Number of students in the records of schools are as follows; would you call these records as data?

Table 17.3

Faculty	School A	School B
Arts	400	700
Science	600	400
Commerce	300	300

Notes

17.3 TYPES OF DATA

On the basis of the source of collection data may be classified as:

(a) Primary data and

(b) Secondary data

(a) Primary data

The data which are originally collected for the first time for the purpose of the survey are called primary data. For example facts or data collected regarding the habit of taking tea or coffee in a village by an investigator.

Methods of collecting primary data

1. **Direct personal investigation:** Under this method the investigator collects the data personally from the respondent. The person who collect the information is called the investigator and the person who gives the responses/answers the questions asked by the investigator is called a respondent. The data collected in this manner are therefore most reliable. However, there is a chance that the results are influenced by the personal bias and prejudice of the investigator.

2. **Indirect investigations:** Under this method the investigator obtains information indirectly from a third person who is expected to know facts about the person about whom the enquiry is done. It is generally used by the commission appointed by the government.

3. **Through correspondent:** Under this method correspondents or agents are appointed by the investigator to obtain data from various places. These correspondents are required to collect and pass the transmit information to the investigator or the central office. This method is widely used by newspaper offices.

4. **By mailed questionnaire:** Under this method a well structured questionnaire is prepared and mailed to the respondent by post. The respondent after filling up the questionnaire send it back within the given time. However, this method can only be used when respondents are literate and can fill in the questionnaire.

5. **Through schedules:** Under this method the field workers are asked to go to the respondent with questions contained in the schedule. They collect the answers in their own hand writing and provide data to the investigator. This method is useful when the respondent is illiterate.

(b) Secondary data

When we use the data, which have already been collected by others, the data are called secondary data. This data is said to be primary for the agency which collects it first, and it becomes secondary for all the other users.

Sources of secondary data

Secondary data may exist in the form of published or unpublished form. In its published form secondary data may be obtained from

(a)   Published reports of newspapers, RBI and periodicals.

(b)   Publication from trade associations

(c)   Financial data reported in annual reports

(d)   Data available in SEBI publication

(e)   Information from official publications

(f)   Publication of international bodies such as UNO, World Bank etc.

(g)   Others

In its unpublished form secondary data may exist as

(a)   Internal reports of the government departments

(b)   Records maintained by the institutions

(c)   Research reports prepared by students in the universities

17.4 PRESENTATION OF DATA

Data collected in the form of schedules and questionnaires are not self-explanatory. These are in the form of raw data. In order to make them meaningful, these are to be made presentable Classification and tabulation are the basic tools of presenting raw data in systematic way.

17.4.1 Classification

Classification is a process of arranging data into classes or groups according to their resemblances and affinities. Mass data in its original form is called raw data.

Variable and attributes

Variable: When data is capable of being classified in the magnitude of time or size it is called as variable. Height, weight, length, distance are example of variables. Variables may be either discrete or continuous. Discrete variable usually have a specific value or measurement. Number of children per family, say for example, is a discrete variable because it cannot be broken into factors

Table 17.4

No. of children per family	0	1	2	3	4
No of families	4	8	20	38	10

This table reveals that these are four families without children, 8 families having one child and so on. Since the no. of children varies from family to family we call it the variable and denote it with symbol x. A variable can have different values. How frequently a value occurs is its frequency. Variable (x) 0 to 3 are values and their frequencies are 4, 8, 20 and 38.

Here value '0' occurs 4 times value '1 'occurs 8 times and so on.

A continuous variable on the other hand has continuity in its scale and measurement, such as scale of height, weight, length, distance etc. continuous variables are usually placed in continuous series as given below:

Height (x)	60'-62"	62"-64"	64"-66"	66"-68
Number of soldiers (frequency)	100	200	110	80

Table shows the range of heights (x) with the corresponding frequencies. It can be read as 100 soldiers having their height between 60"-62", 200 having height between 62"-64" and so on.

Attributes: When data cannot be classified in the magnitude of time or size it is known as an attribute such as beauty, bravery, intelligence, laziness etc. Attributes are difficult to be investigated in depth. These can only be numbered for a study of a limited purpose.

Statistical series: In statistics there are three types of series into which data can be organised.

Individual series: In this kind of series items are shown individually with their corresponding value. Each item has its separate and individual existence. Mass data in its original form are called raw data or unorganised data. But when they are arranged in ascending or descending order of magnitude, is called an array.

Suppose an investigator has got the following information about the marks obtained in economics out of 100 scored by 20 students in a school.

Notes

Table 17.5

Marks obtained by 20 students in Economics out of 100

40	50	35	40	48
50	80	70	75	47
45	75	90	60	57
60	50	80	55	73

The above raw data can be arranged in ascending order which starts from lowest number and goes towards highest number as shown in the following table:

Table 17.6 Arranged in ascending order (Marks out of 100)

35	47	50	60	75
40	48	55	70	80
40	50	57	73	80
45	50	60	75	90

The above data can also be arranged in descending order i.e. from highest number to lowest number as shown in the following table:

Table 17.7 Arranged in descending order (Marks out of 100)

90	75	60	50	45
80	73	57	50	40
80	70	55	48	40
75	60	50	47	35

Discrete Series: This type of series is designed to show variables with definite break with their respective frequencies. Frequency refers to the repetitiveness of a value or item. If a particular value (X) appear 4 times in a set of data X will have a frequency of 4. Theoretically this kind of series is prepared only for a discrete variable, however, in practice continuous and discrete variables are used interchangeably. Following is an example of discrete series.

Table 17.8

Marks	30	40	50	60	70	80	90	Total
Number of students (f)	4	6	10	20	10	6	4	60

Continuous Series: This kind of series is framed for placing frequency with corresponding group of variables which are classified in groups as shown below.

Table 17.9

x	0-10	10-20	20-30	30-40	40-50
f	7	13	20	13	7

This kind of series may be constructed using inclusive method or exclusive method. Above example is that of an exclusive series. In case of inclusive series frequency corresponding to the upper limit of group is included in the same group, while it is included in subsequent group in case of exclusive series.

INTEXT QUESTIONS 17.2

1. Identify whether the following items are variable or attributes?
   (i) Height of a student
   (ii) Beauty of a girl
   (iii) Intelligence level of a boy
   (iv) Mileage of a car
   (v) Weight of Mr X

17.4.2 Tabulation

After the data is collected and classified, it is always useful to put them into rows and columns in a table.

A statistical table may be a simple one or it may be a complex one, depending upon number of variable incorporated into it. Given below is a format of simple statistical table.

Table 17.10

Part of a table

Sub Heading		Caption	
	Column I II		Column I II
Rows Rows Rows Source:			

Footnote

This table may be one way or two ways or manifold. Following illustration are simple example of tabulation.

Illustration 1

During 2010-11, there were three faculties with 840 students in commerce, 660 in science and only 50 students in management.

The percentage of males is 40%, 25% and 20% respectively in each subject stream.

This data can be tabulated as follows

<div align="center">

Table 17.11

Faculty	Number of students		Total
	Male	Female	
Commerce	336	504	840
Science	165	495	660
Management	100	400	500
Total	601	1399	2000

</div>

17.4.3 Diagrammatic and Graphic Presentation of Data in Economics

Data relating to two variables may be shown with the help of a simple graph. It is usually in the form of line or curve. Data relating to a time series or a frequency distribution can be easily presented in a graph.

Diagrammatic presentation is a geometrical version of the data. Diagrams present the facts in such a manner that just by glancing at them one can understand the most complex data. Diagrams may be one-dimensional or two dimensional and even three-dimensional. Bar diagrams are usually one dimensional diagram, only height of the diagram is relevant and not the width.

Here we will discuss only about one dimensional diagram.

One dimensional diagrams

One dimensional diagrams are also called bar diagram which are most commonly used in practice. There are various types of bar diagrams but here we will study about simple bar diagrams only.

Simple Bar Diagram: They are very simple to present but only one type of variable can be presented. A simple bar can be drawn either on horizontal or vertical base. But vertical bars on horizontal base are more commonly used in practice. Bars must be of uniform thickness and they should be placed at equal distance.
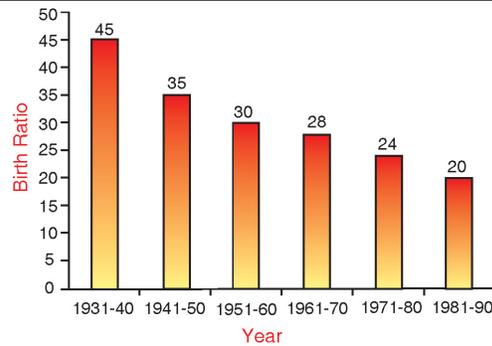
Let us now explain how a simple bar diagram can be presented from the given data. The following table gives data on birth rate in India, according to census survey of different years. This information is presented in simple bar diagram as given below.

Table 17.12

Year	1931-40	1941-50	1951-60	1961-70	1971-80	1981-90
Birth rate	45	35	30	28	24	20

Fig. 17.1 Birth rate in India

Data may also be presented graphically. In economics and statistics the values may be of time, relationship, frequencies etc. In case of time series graph, x-axis represents time and y-axis the variable. It is necessary to decide a convenient scale for each axis to accommodate the complete data given. The scale of two axis can be different.
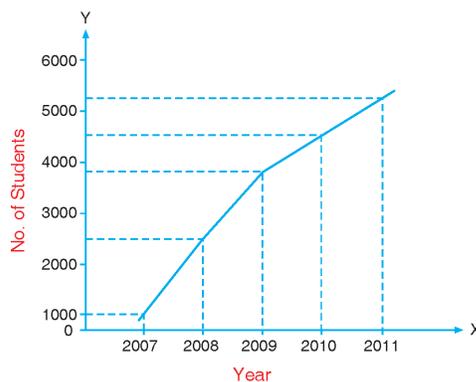
Illustration 2

The number of students in a school for five years is given below:

Table 17.13

Year	2007	2008	2009	2010	2011
No. of students	1000	2500	3800	4500	5200

We can present this data in the form of a graph

Fig. 17.2 Enrolment of student P (2007-2011)

WHAT YOU HAVE LEARNT

- Data means any quantitative information about income, population, prices etc.
- Statistics/Data are the aggregate of facts, affected to a marked extent by multiplicity of cause, numerically expressed, having reasonable standard of accuracy, collected for predetermined purpose and placed in relation to each other.
- Data are important in economic planning, for determination of national income, in forming fiscal and monetary policies and assist central bank of a country.
- Data which are originally collected for the first time for the purpose of the survey, are called primary data.
- When we use the data which have already been collected by others, the data are called secondary data.
- Primary data can be collected by: (i) Direct personal investigation (ii) Indirect investigation (iii) through correspondent (iv) by mailed questionnaire (v) through schedules
- Sources of secondary data may be in the form of published or unpublished data.
- Data can be presented in the form of classification individual series, discrete series and continuous series; graphs and diagrams.
- Data can be presented in the form of simple bar diagram

TERMINAL QUESTIONS

1. Define data. How are primary data collected?
2. What is the difference between primary and secondary data?
3. Distinguish between variable and a an attribute.
4. Explain the following (a) Classification
5. Explain the following methods of presentation of data:
   (a) Tabulation   (b) Diagram
6. Construct a simple bar diagram from the data given below:

State	Number of management colleges
Rajasthan	200
Punjab	400
Gujarat	150

ANSWERS TO INTEXT QUESTIONS

Intext Questions 17.1

(i) No   (ii) No   (iii) No   (iv) Yes

Intext Questions 17.2

(i) Variable   (ii) Attribute   (iii) Attribute   (iv) Variable   (v) Variable