
UNIT 13 CLASSIFICATION OF DATA

Structure

- 13.1 Introduction
 - Objectives
- 13.2 Classification of Data
- 13.3 Tabulation of Data
- 13.4 Summary
- 13.5 Solutions/Answers

13.1 INTRODUCTION

In Unit 12 of Block-3 of this course, we have discussed some methods of data collection whether the target population from where the information collected was small or large. After collection of data, next step is to classify the data in such a manner that it becomes ready for proper presentation.

The need for proper presentation arises because of the fact that statistical data in their raw form are almost defy comprehension. When data are presented in easy-to-read form, it can help the reader to acquire knowledge in much shorter period of time and also facilitate statistical analysis.

A statistical table is a presentation of numbers in a logical arrangement, with some brief explanation to show what they are. However, before tabulating data, it is often necessary to first classify them. So, the concept of classification is described in Sec. 13.2 of the unit and that of tabulation is discussed in Sec. 13.3.

Objectives

After studying this unit, you should be able to:

- classify a data set according to the nature of the data;
- construct a discrete frequency distribution for a discrete type of data;
- construct a continuous frequency distribution for a continuous type of data;
- classify the collected data according to the class intervals; and
- arrange the data into a suitable form of a table.

13.2 CLASSIFICATION OF DATA

This unit is a combination of classification and presentation in tabular form of given data. After collection, classification is the next step in processing collected data. Classification means grouping of related facts into different classes. Information in one class differs from those of other class with respect to some characteristics. Sorting particulars according to one basis of classification and then on another basis is called cross-classification. This process can be repeated as many times as the possible sources of classification are there. Classification of data is a function very similar to that of sorting letters in a post office. Let us explain it further by considering a situation where university receives applications of candidates for filling up some posts for its various departments or disciplines. The applications received for the posts in the university are sorted according to the departments or disciplines to which they pertain. It is well known that the applications collected in an office are

sorted into different lots, department or discipline wise, i.e. in accordance with their destinations as Social Sciences, Engineering, Basic Sciences, etc. They are then put in separate belongings each containing applications with a common characteristic, viz, having the same discipline. Classification of statistical data is comparable to the categorisation process. The process of classification gives distinction to important information gathered, while dropping unnecessary facts, enables comparison and a statistical treatment of the material collected. Now the question may arise in your mind that how collected data is classified. The answer of this question is given under the heading types of classification as discussed below.

13.2.1 Types of Classification

Broadly, data can be classified under following categories:

- (i) Geographical classification
- (ii) Chronological classification
- (iii) Qualitative classification
- (iv) Quantitative classification

Let us discuss these one by one:

(i) Geographical Classification

In geographical classification, data are classified on the basis of location, region, etc. For example, if we present the data regarding production of sugarcane or wheat or rice, in view of the four main regions in India, this would be known as geographical classification as given below in Table-13.1.

Geographical classification is usually listed in alphabetical order for easy reference. Items may also be listed by size to emphasis the magnitude of the areas under consideration such as ranking the states based on population. Normally, in reference tables, the first approach (i.e. listing in alphabetical order) is followed.

Table -13.1: Classification of Production of Wheat

| Region | Production of Wheat (in .000 kg.) |
|-----------------|-----------------------------------|
| Eastern Region | 2873 |
| Northern Region | 1646 |
| Southern Region | 2059 |
| Western Region | 986 |

(ii) Chronological Classification

Classification of data observed over a period of time is known as chronological classification. For example, let us consider the profit figures of a company as shown below for the year from 2001 to 2010.

Table -13.2: Profits of the Company from Year 2001 to 2010

| Year | Profit (in crores of rupees) | Year | Profit (in crores of rupees) |
|------|------------------------------|------|------------------------------|
| 2001 | 20 | 2006 | 12 |
| 2002 | 21 | 2007 | 25 |
| 2003 | 10 | 2008 | 14 |
| 2004 | 18 | 2009 | 19 |
| 2005 | 15 | 2010 | 23 |

Time series data are usually listed in chronological order, normally in ascending order of time, like 2001, 2002,... .When the major emphasis falls on the most recent events, a reverse time order may be used.

(iii) Quantitative Classification

Quantitative classification refers to the classification of data according to some characteristics that can be measured numerically such as height, weight, income, age, sales, etc. For example, the employees of an institute may be classified according to their pay scales as follows:

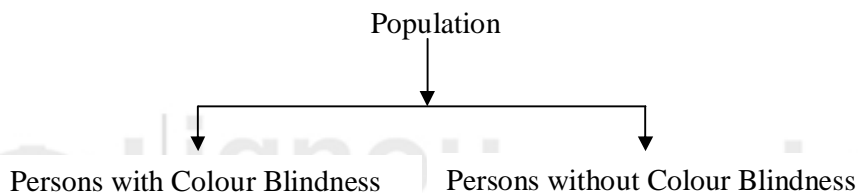
Table-13.3: Quantitative Classification of 840 Employees According to their Pay Scales

| Scale of Pay | Number of Employees |
|---------------|---------------------|
| 9300 - 34800 | 467 |
| 15600 - 39100 | 215 |
| 37400 - 67000 | 158 |
| Total | 840 |

The quantitative classification is a combination of two elements, namely “Variable, i.e. the pay scale” and “the frequency (the number of employees in each class)” in the above example. There are 467 employees getting salary according to the pay scale 9300-34800, 215 employees are getting salary according to the pay scale 15600-39100 and so on. The quantitative classification gives birth to a frequency distribution which is discussed in subsection 13.2.2.

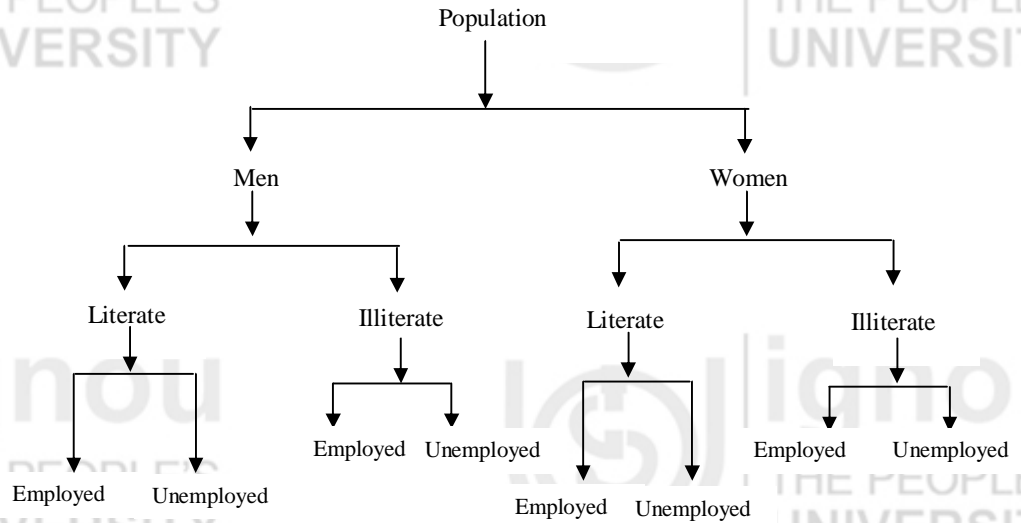
(iv) Qualitative Classification

In qualitative classification, data are classified on the basis of some attributes or qualitative characteristics such as sex, colour of hair, literacy, religion, etc. You should note that in this type of classification the attribute under study cannot be measured quantitatively. One can only count it according to its presence or absence among the individuals of the population under study. For example, in case of colour blindness, we may find out as how many persons are colour blind in a given population. It is not possible to measure the degree of colour blindness in each case. Thus, when only one attribute is studied, two classes are formed – one for possessing the attribute and the other for not possessing it. This type of classification is known as simple classification. For example, the population under study may be divided into two categories based on the characteristic ‘Colour blindness’ as follows:



In a similar manner, we may classify population of a colony on the basis of education qualification, employment, sex, etc. This type of classification where two by two classes are formed is called two fold or dichotomous classification. If, instead of forming only two classes, we further divide the data on the basis of some other attributes within those attributes is known as manifold classification. For example, we may first divide the population into ‘men’ and ‘women’ on the basis of the attribute ‘sex’. Each of these classes may be further subdivided into literate and illiterate on the basis of the attribute

'literacy'. Further classification can be done on the basis of some other attribute say employment. Such type of classification is known as manifold classification and is shown as follows:



Now, you can try the following exercises:

- E1)** The amount of production of wheat (in ,000 kg.) are 230, 376, 136, 583 for the cities Bhopal, Agra, Mumbai and Chandigarh respectively. Classify the data.
- E2)** If a company is manufacturing a product from 2001 to 2010 and earning the profits (in crores of rupees) as 10, 15, 13, 17, 12, 16, 17, 21, 20, 18 for the last 10 years respectively. Classify the given data.

13.2.2 Frequency Distribution

When observations, whether they are discrete or continuous, available on a single characteristic of a large number of individuals, it becomes necessary to condense the data as far as possible without losing any information of interest. Let us consider the ages of 30 students selected at random from among those studying in a certain class.

20, 22, 25, 22, 21, 22, 25, 24, 23, 22, 21, 20, 21, 22, 23, 25, 23, 24, 22, 24, 21, 20, 23, 21, 22, 21, 20, 21, 22, 25.

This presentation of the data is not considered as good since for large number of observations it is not easy to handle the data in this form. A better way to express the figures is shown in Table 13.4 below:

Table-13.4: Frequency Distribution of 30 Students According to their Age

| Age of students | Tally Mark | Frequency |
|-----------------|------------|-----------|
| 20 | | 04 |
| 21 | | 07 |
| 22 | | 08 |
| 23 | | 04 |
| 24 | | 03 |
| 25 | | 04 |
| Total | | 30 |

A bar (|) called tally mark is put against the number when it occurs. After putting this mark four times against the value, a cross tally is put on these 4 tallies for the fifth mark as shown in the above table. From the sixth mark onwards, we start afresh in the similar manner. This technique facilitates easy counting of the tally marks at the end. The presentation of the data as given in Table 13.4 is known as frequency distribution.

A frequency distribution refers to the data which are classified on the basis of some variables that can be measured such as wages, age of children, etc. A variable refers to the characteristic that varies in magnitude in a frequency distribution. It may be either discrete or continuous. A discrete variable is that which generally takes integer values. For example, the number of students, the number of books, etc. A continuous variable can take integer or fractional values within the range of possibilities, such as the height or weight of individuals. Generally speaking, continuous data are obtained through measurements while discrete data are derived by counting. A series described by a continuous variable is called continuous series. Similarly, series represented by a discrete variable is called discrete series.

According to the nature of the variable, the frequency distribution may be of two types, i.e. discrete frequency distribution and continuous frequency distribution. Let us discuss them one by one.

Discrete Frequency Distribution

A frequency distribution in which the information is distributed in different classes on the basis of a discrete variable is known as discrete frequency distribution. For example, frequency distribution of number of children in 20 families is discrete frequency distribution as shown in Table–13.5.

Table –13.5: Frequency Distribution of the Number of Children in 20 Families

| No. of children | Tally Mark | Frequency |
|-----------------|------------|-----------|
| 0 | | 3 |
| 1 | | 4 |
| 2 | | 6 |
| 3 | | 4 |
| 4 | | 3 |
| Total | | 20 |

Continuous Frequency Distribution

A distribution in which the information is distributed in different classes on the basis of a continuous variable is known as continuous frequency distribution.

There may be some variables which have integer values as well as fractional values. Frequency distribution of such variables is called continuous frequency distribution. An example of a continuous frequency distribution is given below in Table-13.6.

Table 13.6: Frequency Distribution of Heights of 50 Persons

| Heights (cm) | Tally Mark | Frequency |
|--------------|------------|-----------|
| 120 -130 | | 3 |
| 130 -140 | | 5 |
| 140- 150 | | 10 |
| 150 -160 | | 14 |
| 160 -170 | | 12 |
| 170-180 | | 4 |
| 180-190 | | 2 |
| Total | | 50 |

After discussing the discrete and continuous frequency distributions let us discuss the Relative and Cumulative frequency distributions which are of the similar importance as analysis point of view of data is considered.

Relative Frequency Distribution

A relative frequency corresponding to a class is the ratio of the frequency of that class to the total frequency. The corresponding frequency distribution is called relative frequency distribution. If we multiply each relative frequency by 100, we get the percentage frequency corresponding to that class and the corresponding frequency distribution is called “Percentage frequency distribution”. Let us take an example in which both relative and percentage frequency distributions are prepared.

Example 1: A frequency distribution of marks of 50 students in a subject is as given below:

Class (Marks): 0-10 10-20 20-30 30-40 40-50

Frequency: 6 10 14 18 2

Prepare relative and percentage frequency distributions.

Solution: The relative and percentage frequency distributions can be formed as given in the following table:

| Class (Marks) X | Frequency (f) | Relative frequency (f/N) | Percentage Frequency (f/N) × 100 |
|--------------------|-------------------|-----------------------------|-------------------------------------|
| 0-10 | 6 | $6/50 = 0.12$ | $0.12 \times 100 = 12\%$ |
| 10-20 | 10 | $10/50 = 0.20$ | $0.20 \times 100 = 20\%$ |
| 20-30 | 14 | $14/50 = 0.28$ | $0.28 \times 100 = 28\%$ |
| 30-40 | 18 | $18/50 = 0.36$ | $0.36 \times 100 = 36\%$ |
| 40-50 | 2 | $2/50 = 0.04$ | $0.04 \times 100 = 4\%$ |
| Total | $\sum f = N = 50$ | 1.00 | 100 |

Cumulative Frequency Distribution

The cumulative frequency of a class is the total of all the frequencies up to and including that class. A cumulative frequency distribution is a frequency distribution which shows the observations ‘less than’ or ‘more than’ a specific value of the variable.

The number of observations less than the upper class limit of a given class is called the less than cumulative frequency and the corresponding cumulative frequency distribution is called less than cumulative frequency distribution.

Similarly, the number of observations corresponding to the value of more than the lower class limit of a given class is called more than cumulative frequency and the corresponding cumulative frequency distribution is called 'more than' cumulative frequency distribution. Following is an example, wherein 'less than' and 'more than' cumulative frequency distributions have been obtained.

Example 2: For the following frequency distribution of marks of 50 students in a subject, form both types of cumulative frequency distributions.

| | | | | | |
|------------------------|------|-------|-------|-------|-------|
| Class (Marks) | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 |
| No. of Students | 7 | 11 | 15 | 12 | 5 |

Solution: Cumulative frequency distributions are formed as given in the following table:

| Given Frequency Distribution | | Less Than Cumulative Frequency Distribution | | More Than Cumulative Frequency Distribution | |
|-------------------------------------|------------------------|--|------------------------|--|-----------------------|
| Classes | No. of Students | Marks Less than | No. of students | Marks More than | No of students |
| 0-10 | 07 | 10 | 07 | 0 | 50 |
| 10-20 | 11 | 20 | 18 | 10 | 43 |
| 20-30 | 15 | 30 | 33 | 20 | 32 |
| 30-40 | 12 | 40 | 45 | 30 | 17 |
| 40-50 | 05 | 50 | 50 | 40 | 05 |
| Total | 50 | | | | |

Now, you can try the following exercises.

E3) Construct a discrete frequency distribution for 25 students studying in a class having the following ages (in years):

20, 21, 19, 18, 20, 20, 19, 18, 21, 19, 22, 21, 18, 19, 21, 22, 19, 18, 20, 19, 20, 22, 20, 21, 20.

E4) Construct a continuous frequency distribution for the 50 students studying in a class having the following heights (in cm): 146, 156, 152, 167, 178, 180, 172, 162, 148, 153, 161, 173, 163, 174, 147, 179, 148, 151, 168, 172, 165, 173, 172, 180, 175, 145, 153, 154, 162, 164, 170, 172, 160, 161, 158, 152, 163, 165, 170, 168, 158, 149, 155, 160, 150, 149, 167, 176, 169, 159.

After discussing the frequency distributions we now discuss how the concept of frequency distribution can be used to classify the data according to the class intervals in the next subsection.

13.2.3 Classification According to Class Intervals

To make data understandable, data are divided into number of homogeneous groups or sub groups. In classification, according to class intervals, the observations are arranged systematically into a number of groups called classes. Such classification is most popular in practice. But before this discussion we have to define some terms which will be used in the above classification.

(i) Class Limits

The class limits are the lowest and the highest values of a class. For example, let us take the class 10-20. The lowest value of this class is 10 and the highest

20. The two boundaries of a class are known as the lower limit and upper limit of the class. The lower limit of a class is the value below which there can not be any value in that class. The upper class limit of a class is the value above which no value can belong to that class.

(ii) Class Intervals

The class interval of a class is the difference between the upper class limit and the lower class limit. For example, in the class 10-20 the class interval is 10 (i.e. 20 minus 10). This is valid in the case of exclusive method discussed in this subsection later on. If the inclusive frequency distribution (discussed later on in this subsection) is given then first it is converted to exclusive form and then class interval is calculated. The size of the class interval is determined by number of classes and the total range of data.

(iii) Range of Data

The range of data may be defined as the difference between the lower class limit of the first class interval and the upper class limit of the last class interval.

(iv) Class Frequency

The number of observations corresponding to the particular class is known as the frequency of that class or the class frequency. In the given frequency distribution (Table -13.7), the frequency of the class 10-20 is 12 which implies that there are 12 persons having ages between 10-20. If we add together the frequencies of all individual classes, we obtain the total frequency.

Table-13.7: Frequency Distribution of 50 Persons having Ages between 0-50 Years.

| Classes | Frequencies |
|---------|-------------|
| 0-10 | 08 |
| 10-20 | 12 |
| 20-30 | 15 |
| 30-40 | 10 |
| 40-50 | 05 |
| Total | 50 |

(v) Class Mid Value

It is the value lying half way between the lower and upper class limits of class-interval, mid-point or mid value of a class is defined as follows:

$$\text{Mid Value of a Class} = \frac{\text{Upper class limit} + \text{Lower class limit}}{2}$$

For the purpose of further calculations in statistical analysis, mid value of each class is taken to represent that class.

Now we are in position to discuss the two methods of classification according to class intervals, namely “**Exclusive Method**” and “**Inclusive Method**”. Let us discuss these two methods one by one:

Exclusive Method

Under this method, a class interval is such that each upper class limit is excluded from the class interval. Here in this method, class intervals are so fixed that the upper limit of one class is the lower limit of the next class. In the

following example there are 24 students who have secured the marks between 0 and 50. A student who secured 20 marks would be included in class 20-30, not in 10–20. This method is widely followed in practice.

Example 3: 24 students appeared in an entrance test where all questions are objective type with 25% –ve marking. The marks obtained out of 50 maximum marks are as follows:

17, 16, 7, 30, 21, 42, 44, 36, 22, 22, 25, 31, 31, 34, 30, 36, 35, 45, 25, 15, 20, 42, 40, 30

Prepare a frequency distribution by using exclusive method.

Solution: Frequency distribution of marks obtained by above 24 students is given below in table 13.8 using exclusive method as follows:

Table 13.8: Frequency Distribution of 24 Students by Exclusive Method

| Classes | Tally bar | No. of Students |
|---------|-----------|-----------------|
| 0-10 | | 1 |
| 10-20 | | 3 |
| 20-30 | | 6 |
| 30-40 | | 9 |
| 40-50 | | 5 |
| Total | | 24 |

Inclusive Method

Under the inclusive method of classification both lower class limit as well as the upper limit of a class is included in that class itself. Following frequency distribution is formed using inclusive method for the data of Example 3 given above.

Table 13.9: Frequency Distribution of 24 Students by Inclusive Method

| Class | Tally bar | No. of Students |
|-------|-----------|-----------------|
| 0-9 | | 1 |
| 10-19 | | 3 |
| 20-29 | | 6 |
| 30-39 | | 9 |
| 40-49 | | 5 |
| Total | | 24 |

That means if data are classified in such a way that the lower as well as the upper class limits are included in the same class interval, it is called inclusive class interval.

For converting data from inclusive form to exclusive form, first of all we find the half of the difference of lower limit of that class and upper limit of the preceding class. This value is then subtracted from lower limit of each class and added to the upper limit of each class. In the above example, this can be easily understood as $(10-9)/2 = 0.5$. So, the class intervals are as $-0.5- 9.5, 9.5-19.5, \dots, 39.5-49.5$. If all the observations of data are positive then the lower limit of first class can be taken 0. Therefore, in this case the class intervals are as $0-9.5, 9.5-19.5, \dots, 39.5-49.5$.

Remark

- (i) Lower limit of a class interval is always included in the class in both the method discussed above.
- (ii) In exclusive method upper limit of a class is not included in the class. That is why the name exclusive.
- (iii) In inclusive method upper limit of a class is also included in the class. That is why the name inclusive.

13.2.4 Principles of Classification

It is difficult to formulate any hard and fast rule for classifying the data. However, the following general considerations may be considered for ensuring meaningful classification of data:

- (1) The whole data should be preferably divided into number of classes between 5 and 15. However, there is no rigidity about it. The classes can be more than 15 depending upon the total number of observations and variations between them and the details required for given data, but they should not be less than 5 because in that case the classification may not reveal the essential characteristics.

To determine the approximate number of classes (K) the following formula is suggested by “Struges”:

$$K = 1 + 3.322 \text{ Log } N, \text{ where } K = \text{the approximate number of classes}$$

N = total number of observations

Log = the natural logarithm

However, the appropriate number of classes to be taken for a given data depends upon the personal judgment and other considerations such as range of data, total number of observations, etc.

- (2) One should avoid odd values of class intervals as far as possible, e.g. 3, 7, 11, 26, 39, etc. One should prefer 5 or 10 or multiple of 5 or 10 as class intervals such as 5, 10, 20, 25, 100, etc, because the human mind is accustomed more to think in terms of certain multiples of 5 or 10.
- (3) The lower class limit of the first class of a frequency distribution should either be zero or 5 or multiple of five. For example if the lowest value of the data is 26 and we have taken a class interval of 5, then the first class should be 25-30, instead of 26-31. Similarly if the lowest value of the series is 43 and the class interval is 5 then the first class should be 40-45 inspite of 43-48.
- (4) To maintain continuity and to get correct class interval, we should adopt exclusive method of classification. However, where ‘inclusive’ method has been adopted it is necessary to make an adjustment to determine the correct class interval and to maintain continuity.

How the adjustment is made when data are given by inclusive method explained in the previous sub Sec. 13.2.4. The same adjustment has been done in the frequency distribution given in Table 13.9, which is given in Table 13.10 as shown on the next page:

Table 13.10: Frequency Distribution of 24 Persons by Inclusive Method

| Classes | No. of Students |
|-----------|-----------------|
| – 0.5-9.5 | 01 |
| 9.5-19.5 | 03 |
| 19.5-29.5 | 06 |
| 29.5-39.5 | 09 |
| 39.5-49.5 | 05 |
| Total | 24 |

- (5) The intervals of all the classes should be of the same size, because if the class intervals are not of the same width, it is difficult to make meaningful comparison between classes. Sometimes the data may require the inclusion of so many class intervals that the frequency distribution will become large. Then the classification may be done as follows:

below 10
10-20
20-30
30-40
above 40

These classes are called open end classes and the distribution is known as open end frequency distribution.

It may be noted that the frequency distributions, like other types of data presentation, are always constructed to serve some specific purpose. The technical requirements outlined above must be supplemented by sound subjective judgments if proper frequency distributions are to be formed.

After learning so much about classification of data, you have got/realised the importance of classification. So before move to next section, let us just highlight/outline some of the main points related to the importance of classification:

- It is preliminary for further statistical analysis,
- It facilitates comparison and make conclusion easy,
- It facilitates tabulation.

Now, you can try the following exercises.

E5) The marks of 30 students in statistics are given below:

10, 12, 25, 32, 27, 32, 38, 43, 39, 55, 29, 38, 57, 08, 06, 13, 27, 25, 29, 53, 55, 45, 35, 48, 47, 59, 15, 19, 48, 55

Classify the above data by taking a suitable class interval.

E6) Present the following data of the profits (in crores of Rs.) of the 60 companies in the years 2009-10:

41, 17, 83, 63, 55, 92, 60, 58, 70, 06, 67, 82, 33, 44, 57, 49, 34, 73, 54, 63, 36, 52, 32, 75, 60, 33, 09, 79, 28, 30, 42, 93, 43, 80, 03, 32, 57, 67, 84, 64, 63, 11, 35, 28, 10, 23, 08, 41, 60, 32, 72, 53, 92, 88, 62, 55, 60, 33, 40, 57

Classify data by inclusive method.

E7) Use the data given in the E6 to present the same using principle of adding and subtracting the correction factor.

13.3 TABULATION OF DATA

One of the simplest and most revealing devices for summarising and presenting data in a meaningful arrangement is statistical table. We can also define a statistical table as the logical listing of quantitative data in columns and rows of numbers with sufficient explanatory statements. The statements may be given in the form of titles, headings and notes to make clear the full meaning of data and their origin.

In other words, a table is a systematic arrangement of statistical data in columns and rows. Rows are horizontal arrangements, whereas columns are vertical ones. A table can solve the purpose of the presentation and facilitate comparison. The simplification results from the clear-cut and systematic arrangement, which enables the reader to quickly locate the desired information. Comparison is facilitated by bringing related items of information close together.

13.3.1 Components of a Table

The various components of a table may vary case to case depending upon the given data. But a good table must contain at least the following components:

1. Table Number
2. Table Heading
3. Caption
4. Stub
5. Body of Table
6. Head Note
7. Foot Note

Let us throw some light on these components one by one:

1. Table Number

A statistical table should be numbered. There are different ways with regard to the place where table number is to be given. The table number may be shown either in the centre at the top above the title or in the left hand side of the table at the top. When there are many columns, it is desirable to number each column so that easy reference to it is possible.

2. Table Heading

A good table should have a suitable heading. The heading is a brief description of the contents of the table. It should be placed above the table. It should answer the following questions:

- (a) What categories of statistical data are shown?
- (b) Where the data occurred?
- (c) When the data occurred?

In other words the heading of the table should be clear, brief and self-explanatory, but some times long title may have to be used for the sake of clarity. The title should be so worded that it permits one and only one interpretation.

3. Caption

Caption refers to the column heading, and explains what information column presents. It may consist of one or more column headings, i.e. under a column

heading there may be two or more sub headings. The caption should be clearly defined and placed at the middle of the column. If the different columns are expressed in different units, the unit should be specified along with the captions.

4. Stub

The stubs are row headings. They are placed at the extreme left of the table and perform the same function for the horizontal rows in the table as the captions do for the vertical columns.

5. Body

The body of the table is the central part of table that contains the numerical information presented in table. This is the most vital part of the table.

6. Head Note

Head note is a brief explanatory statement applying to all or a major part of the material presented in the table and is placed below the title entered and enclosed in brackets. It is used to explain certain points relating to the whole table that have not been included in the title nor in the captions or stubs. For example, the unit of measurement is frequently written as the head note such as “in thousands” or “million tons” or “in crores”, etc.

7. Foot Note

Anything in a table which the reader supposed to find difficult to understand should be explained in footnotes. Footnotes may be placed directly below the body of the table. The footnotes are generally used for the following purposes:

- (a) Any special circumstances affecting the data, for example, strike, fire, etc.
- (b) To clarify any thing in the table.
- (c) To give the source in case of the secondary data. If any information in the table obtained from some journal, its name, date of publication, page number, table number, etc. should be mentioned so that if the user wishes to check the data from the original source, he could know where to look for the information.

After discussing the parts of a table, let us discuss different kinds of tables, through which we can represent or arrange the different types of informations.

13.3.2 Types of Tables

Tables may broadly be classified into following two categories.

1. Simple and Complex Tables
2. General Purpose and Special Purpose Tables

1. Simple and Complex Tables

The simple and complex tables can be differentiated on the basis of number of characteristics presented and studied. If the data based on one characteristic is presented, the table is known as simple table. The simple table is also known as one way table. On the other hand, in a complex table, two or more characteristics are presented. The complex tables are frequently used in practice because they facilitate to incorporate full information and a proper consideration of all related facts. If the data are tabulated on the basis of only two characteristics then the table is known as two way table. If three

Presentation of Data

characteristics are arranged in a table then the table is known as treble table. When four or more characteristics are simultaneously presented it is known as manifold tabulation.

The following table presenting the distribution of marks obtained by 100 students in a test is an illustration of a simple table:

Table-13.11: Distribution of Marks Obtained by 100 Students in Statistics

| Marks | No. of Students |
|----------|-----------------|
| Below 10 | 5 |
| 10-20 | 8 |
| 20-30 | 12 |
| 30-40 | 10 |
| 40-50 | 15 |
| 50-60 | 18 |
| 60-70 | 17 |
| 70-80 | 13 |
| Above 80 | 02 |
| Total | 100 |

Two Way Table

Two way table shows two characteristics and is formed when either the stub or the caption or both are divided into two categories. In the following example the nature of such a table is given and is an illustration of two-way table (a complex table):

Table -13.12: Number of Persons Living in a Colony According to Age and Sex.

| Age | Persons Living in the Colony | | Total |
|--------------|------------------------------|---------|-------|
| | Males | Females | |
| Below 15 | 12 | 6 | 18 |
| 15-25 | 20 | 12 | 32 |
| 25-35 | 42 | 27 | 69 |
| 35-45 | 25 | 18 | 43 |
| 45-55 | 10 | 8 | 18 |
| 55-65 | 8 | 5 | 13 |
| 65 and Above | 5 | 2 | 07 |
| Total | 122 | 78 | 200 |

Higher Order Table

When three or more characteristics are represented in the table then such a table is called higher order table. The need for such a table arises when we are interested in presenting three or more characteristics simultaneously.

It should be remembered that as the number of characteristics increases, the table becomes more and more condensing. It is advised normally not more than four characteristics should be represented in the same table. When more than four characteristics are to be represented we should form more than one table depicting relationship between different attributes.

2. General Purpose and Special Purpose Tables

General purpose tables, also known as reference tables or repository tables, and provide the information for general use or reference. They usually contain detailed information and are not used for specific discussion. In other words, these tables serve as a repository of information and are arranged for easy reference such as the tables published by government agencies, the tables contained in the statistical abstract of the Indian Union, tables in the census reports, etc.

The general tables tell facts which are not for particular discussion. If general tables are used by a researcher, they are usually placed in the form of appendix at the end of the report for easy reference.

Special purpose tables, also known as summary tables or analytical tables, provide information for particular discussion. These tables are also called derivative tables since they are often derived from general tables. A special purpose table should be designed in such a way that a reader may easily refer to the table for comparison, analysis or emphasis concerning the specific discussion.

Now, you can try the following exercises.

E8) In a sample survey study about the drinking habits in two cities, it is observed that, in city X 57% are male, 22% are drinkers, and 14% are male drinkers, whereas in city Y 52% are male, 28% are drinkers and 21% are male drinkers. Tabulate the above information.

E9) Present the following information in a suitable tabular form:

In 2009 out of a total 2000 employees in a company 1550 were members of a trade union. The number of women employees was 250, out of which 200 did not belonging to any trade union. While in 2010 the number of union employees was 1725 out of which 1600 were men. The number of none union employees was 380 among which 155 were women.

13.4 SUMMARY

In this unit we have covered the concepts of classification and tabulation of data. That is we have discussed:

- 1) Classification of a data set according to the nature of data.
- 2) The methods of construction of a frequency distribution.
- 3) The methods of construction of discrete and continuous frequency distributions.
- 4) Fundamentals of classification of data according to the class intervals.
- 5) The methods of construction of relative and cumulative frequency distributions.
- 6) Parts of a table.
- 7) Types of the tables and presenting data into a suitable form of a table.

13.5 SOLUTIONS/ANSWERS

E1) The classification of the data for the production of wheat according to the given cities can be done in the following way:

Table 13.13: Geographical Classification of the Production of Wheat

| Region | Production of Wheat (in .000 kg.) |
|------------|---------------------------------------|
| Agra | 376 |
| Bhopal | 230 |
| Chandigarh | 583 |
| Mumbai | 136 |

E2) Classification of the profits of a company from 2001 to 2010 can be done in the following way:

Table 13.14: Chronological Classification of Profits from 2001 to 2010

| Year | Profits (in crores of rupees) | Year | Profits (in crores of rupees) |
|------|-------------------------------------|------|-------------------------------------|
| 2001 | 10 | 2006 | 16 |
| 2002 | 15 | 2007 | 17 |
| 2003 | 13 | 2008 | 21 |
| 2004 | 17 | 2009 | 20 |
| 2005 | 12 | 2010 | 18 |

E3) Discrete frequency distribution for the given information can be constructed in the following way:

Table 13.15: Discrete Frequency Distribution of 25 Students According to their Age

| Age of the students | Tally Mark | No. of the students |
|---------------------|------------|---------------------|
| 18 | | 04 |
| 19 | | 06 |
| 20 | /// | 07 |
| 21 | /// | 05 |
| 22 | | 03 |
| Total | | 25 |

E4) The continuous frequency distribution for the given information can be constructed in the following way:

Table 13.16: Continuous Frequency Distribution of 50 Students According to their Heights

| Heights (cm) | Tally Mark | Frequency |
|--------------|------------|-----------|
| 145-150 | /// | 07 |
| 150-155 | /// | 07 |
| 155-160 | /// | 05 |
| 160-165 | /// | 09 |
| 165-170 | /// | 07 |
| 170-175 | /// | 09 |
| 175-180 | | 04 |
| 180-185 | | 02 |
| Total | | 50 |

E5) Let us determine the suitable class interval with the help of the following formula:

$$i = \frac{\text{Range}}{1 + 3.322 \log N}$$

$$\text{Range} = 59 - 06 = 53, N = 30$$

$$i = \frac{53}{1 + 3.322 \log 30} = \frac{53}{1 + 4.91} = 8.97 \approx 9$$

Since values like 3, 7, 9 etc., should be avoided and therefore, we will take 10 as the class interval and hence let us take the first class as 5-15 and thus the following table is formed:

Table 13.17: Continuous Frequency Distribution of 30 Students According to their Heights

| Heights (cm) | Tally Mark | Frequency |
|--------------|------------|-----------|
| 05-15 | | 5 |
| 15-25 | | 2 |
| 25-35 | | 8 |
| 35-45 | | 5 |
| 45-55 | | 5 |
| 55-65 | | 5 |
| Total | | 30 |

E6) As the least value is 3 and the highest value is 93, so using

$$i = \frac{\text{Range}}{1 + 3.322 \log N} = \frac{93 - 3}{1 + 3.322 \log 60} \approx 13.03 \approx 13$$

since, values like 3, 7, 9, 11, 13 etc., should be avoided and therefore, we will take 14 as class interval and hence let us take the first class as 0-14 and thus the following table is formed.

Table 13.18: Continuous Frequency Distribution of 60 Students According to their Heights

| Heights (cm) | Tally Mark | Frequency |
|--------------|------------|-----------|
| 0-14 | | 06 |
| 15-29 | | 04 |
| 30-44 | | 16 |
| 45-59 | | 10 |
| 60-74 | | 14 |
| 75-89 | | 07 |
| 90-104 | | 03 |
| Total | | 60 |

E7) Table 13.19 given on next page illustrates the way of classification of data according to the exclusive method and principle of correction factor in classification.

Table 13.19: Continuous Frequency Distribution of 60 Students According to their Heights

| Heights (cm) | Tally Mark | Frequency |
|--------------|------------|-----------|
| -0.5-14.5 | | 06 |
| 14.5-29.5 | | 04 |
| 29.5-44.5 | | 16 |
| 44.5-59.5 | | 10 |
| 59.5-74.5 | | 14 |
| 74.5- 04.5 | | 07 |
| 89.5-94.5 | | 03 |
| Total | | 60 |

E8) The following table is the representation of the data for the given information's regarding the drinkers in city X and city Y.

Table13.20: Presentation of Data regarding the Drinkers in City X and City Y in the form of Two Way Table

| Attributes | City X | | Total | City Y | | Total |
|---------------------|--------|---------|-------|--------|---------|-------|
| | Males | Females | | Males | Females | |
| Drinkers | 14 | 8 | 22 | 21 | 7 | 28 |
| Non-drinkers | 43 | 35 | 78 | 31 | 41 | 72 |
| Total | 57 | 43 | 100 | 52 | 48 | 100 |

E9) The following table is showing the trade union membership.

Table 13.21: Presentation of Data regarding the Trade Union Membership in the Year 2009 and 2010 in the form of Two Way Table

| Category | 2009 | | | 2010 | | Total |
|--------------|---------------------|--------------------|-------|---------------------|--------------------|-------|
| | Trade Union Members | None Union Members | Total | Trade Union Members | None Union Members | |
| Men | 1500 | 250 | 1750 | 1600 | 225 | 1825 |
| Women | 50 | 200 | 250 | 125 | 155 | 280 |
| Total | 1550 | 450 | 2000 | 1725 | 380 | 2105 |