**Classification & Tabulation**

Once the collected data are scrutinized and the errors therein are removed, it should be made suitable to be used for investigation. In order to render the scrutinized data suitable to be used for investigation, *they are* ***classified on the basis of likeness of their nature and properties and the classified data are shown in the form of tables******in order to present them precisely.*** The process of classification and tabulation is of great importance for statistical operations because from a mere heap of collected information no results can be arrived at.

**Classification**: The process of arranging data into homogenous groups or classes according to some common characteristics present in the data is called classification.

**For example, d**uring the process of sorting letters in a post office, the letters are classified according to the cities and further arranged according to streets.

## ****Object of Classification****

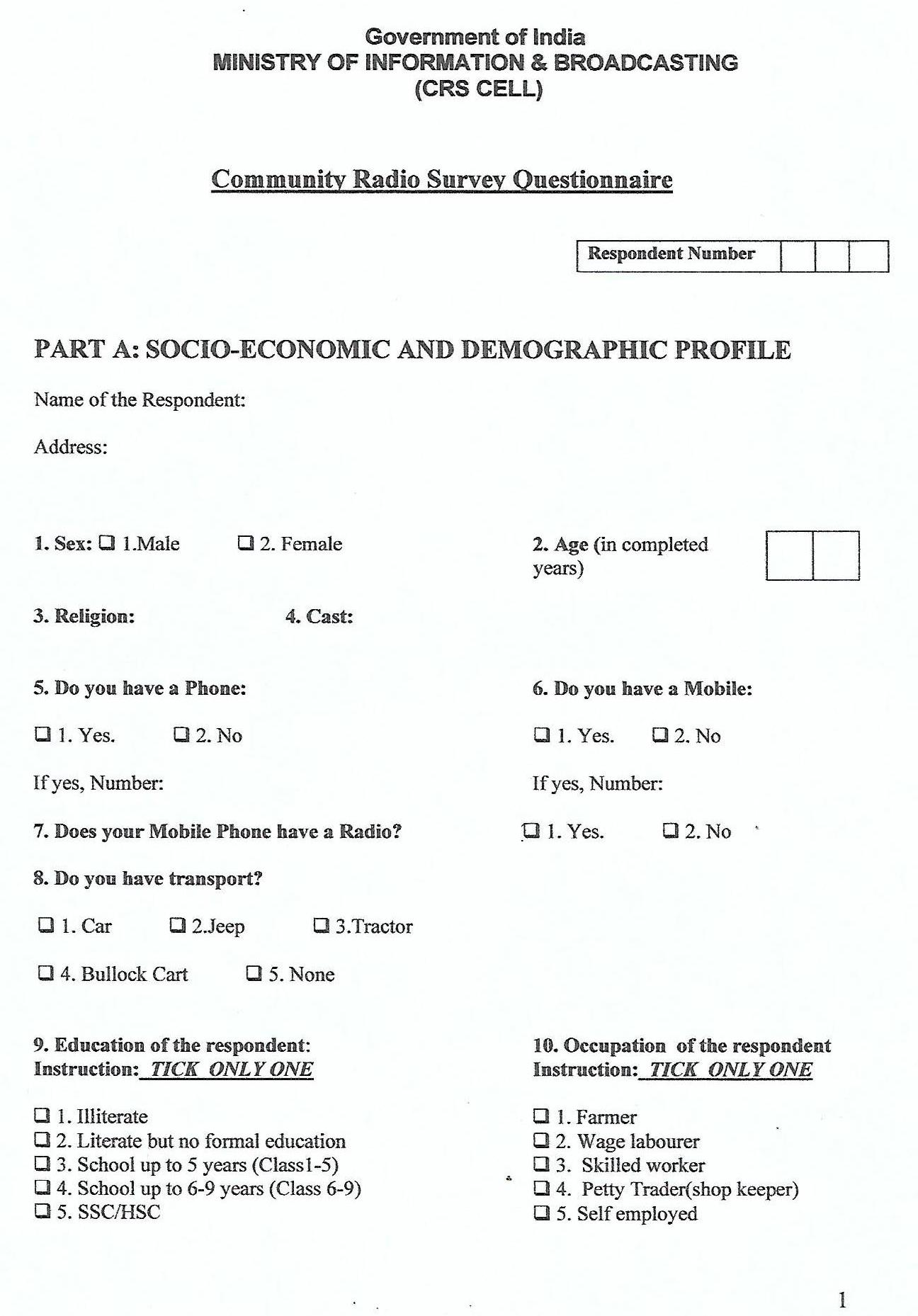
* **To simplify facts:** Classified data are in simple form. Classified data may be easily understood and may be put in short form. Therefore, the object of classification is to simplify the data.
* **To express similarity and dissimilarity:** One of the objects of classification of data is to arrange them according to their similarity and dissimilarity so that they may be easy to understand, for example, classification of population according to education, marital status, literacy, etc.
* **To help in comparison:** Unless the data are classified, they will not be comparable. But after classification the data can be compared. For example, if we want to compare the standard of different colleges of Calcutta University, we will have to make the data comparable by classifying them in different classes.
* **To express relationship:** This is also the object of classification to express relationship of data by classification so that the cause and effect may be understood.
* **To increase the utility of data:** When the data are classified, they can be used by other persons also. Hence, classification increases the utility of data.
* **To form the basis of tabulation:** The first step for tabulation is the classification; hence, it forms the basis of tabulation.

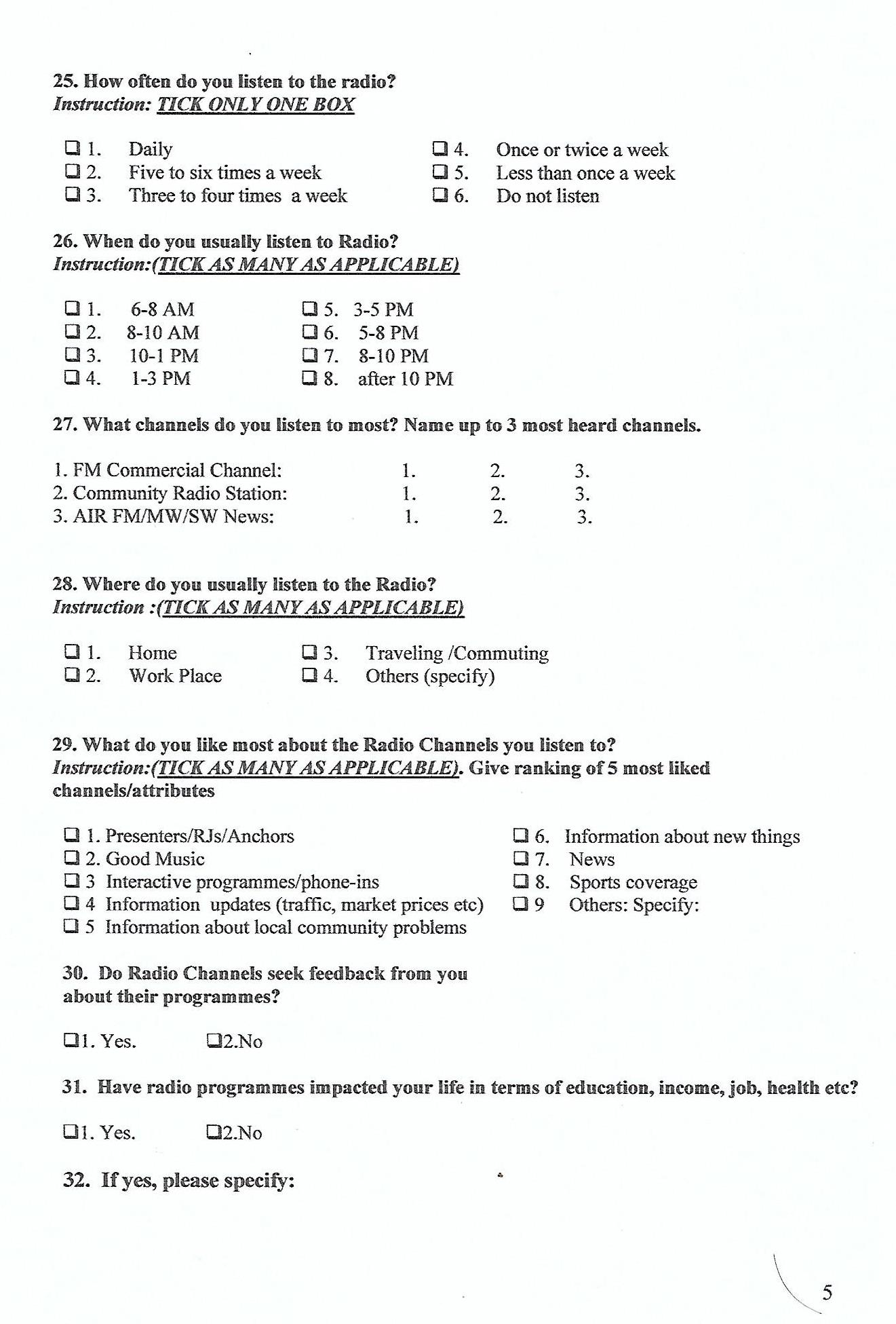
#### ****Characteristics of Good Classification****

* **Comprehensiveness:** Classification should be as much comprehensive as possible. It should include all the units in it.
* **Separate Classes:** Classification should be such that every different unit is put in a separate class, and the question as to where to place a particular unit should not arise. In other words, classification should be clear and simple.
* **Stability:** Every classification should have a quality of stability. Once a classification is done on a particular basis, that basis should not be changed again and again. Otherwise, the data will not be comparable.
* **Conformity:** Classification should be in conformity with the object of investigation. If we want to study the standard of students, the classification should be according to marks.
* **Homogeneity:** In classification every unit of homogeneous quality should be placed in one class so that correct conclusions may be drawn.
* **Elasticity:** Classification should have a characteristic of elasticity or flexibility so that it may be modified as per changed circumstances.

In an investigation pertaining to the population of a place, data in the beginning about male, female. adult, child, educated, uneducated, employed and unemployed persons of each family will be collected through schedules and questionnaires. But in order to get result from these data it will be necessary that data about all the categories mentioned are placed in different groups. It is only then that the number of men, women, major and minor persons of that place can be known. Likewise, educated, uneducated, employed and unemployed constituting the population can also be known.

These various groups can be put in a table which gives a clear picture of the whole universe. Therefore, facts not classified and properly arranged don’t serve any purpose. Let us take into consideration the **Survey data on radio listening behaviour of people** for a better understanding about the need for classification.





**Exercises:** (a) Prepare a blank table showing the occupation of the respondents classified into gender. (b) Prepare a blank table showing the respondents to Question no. 28 classified into occupation and gender.

**Types of Classification**

Broadly speaking, there are four types of classification. They are:

#### **Geographical/Spatial Classification**

#### **Chronological/Temporal Classification**

#### **Qualitative Classification**

#### **Quantitative Classification**

#### ****Geographical/Spatial Classification****

Under this type of classification, the data are ***classified on the basis of area or place***, and as such, this type of classification is also known as areal or spatial classification. The areas may be in terms of countries, states, districts, or zones according as the data are distributed. For the purpose of ready reference and ranking, the different classes form under the classification should be arranged in order of their alphabets or size of the frequencies respectively. Generally, in case of reference tables, alphabetical arrangements are made while in case of summary tables, ranking arrangements are made.

This type of classification is suitable for those data which are distributed geographically relating to a phenomenon viz. population, mineral resources, production, sales, students of universities etc.

#### ****Chronological/Temporal Classification****

Under this type of classification, the ***data collected are classified on the basis of time of their occurrence.*** As such, the series obtained under this classification is purely known as a time series.

This type of classification is suitable for those data which take place in course of time viz. population, production, sales, results etc. The different classes obtained under this classification are arranged in order of the time which may begin either with the earliest, or the latest period.

#### ****Qualitative Classification****

Under this type of classification, ***the data obtained are classified on the basis of certain descriptive character or qualitative aspect of a phenomenon*** viz. sex, beauty, literacy, honesty, intelligence, religion, eye-sight etc.

As such, this sort of classification is also otherwise known as ‘descriptive classification’. Such type of classifications are usually dichotomous in nature in which the whole data are divided into two groups viz, a group with the absence of the attribute such as blind and not-blind, or deaf and not-deaf etc.

#### ****Quantitative Classification****

Under this type of classification, ***the collected data are classified on the basis of certain variable*** viz. mark, income, expenditure, profit, loss, height, weight, age, price, production etc. This type of classification is also known as ‘classification by variables’.

Data may be further classified as frequency data and non-frequency data. The qualitative as well as quantitative data belong to the frequency group whereas time series data and geographical data belong to the non-frequency group.

**Exercises**

1. Prepare a blank table showing the average wages of males and females classified into two age-groups of eighteen years and over, and under eighteen years in five industries in two different months.
2. Prepare a blank table showing the respondents to the question no. 31, classified into sex and the transport used by them.
3. Represent the information contained in the following passage in a suitable tabular form:

The total number of accidents in Southern Railway in 1960 was 3500, and it decreased by 300 in 1961 and by 700 in 1962. The total number of accidents in Metre Gauge section showed a progressive increase from 1960 to 1962. In was 245 in 1960; 346 in 1961; and 428 in 1962. In the Metre gauge section, “Not compensated” cases were 49 in 1960, 77 in 1961 and 108 in1962. “Compensated” cases in the broad gauge section were 2867, 2587 and 2152. From the above report, prepare a table as per the rules of tabulation.

1. Present the following information in a concise tabular form and indicate which type of lamp shows the greatest wastage during manufacture:

Lamps are rejected at several manufacturing stages for different faults. 12000 glass tubes are supplied to make 40-watt, 60-watt and 100-watt lamps in the ratio 1:2:3. At the stage I, 10 % of 40-watt, 4% of the 60-watt and 5% of the 100-watt bulbs are broken. At the stage II, about 1% of the remainder of the lamps have broken filaments. At the stage III, 100 100-watt lamps have badly soldered caps, and half as many have crooked caps; twice as many 40-watt and 60-watt lamps have these faults. At the stage IV, about 3% are rejected for bad type-marking and 1 in every 100 are broken in the packing which follows.

## Presentation of data

It is necessary to present the classified data into proper form so that they may be statistically analyzed and properly understood by a common man. Usually three modes are used for presentation of statistical data. These are:

* Textual presentation
* Tabular presentation or Tabulation.
* Graphical presentation.

**Textual presentation:** This method comprises presenting data with the help of a paragraph or a number of paragraphs. The official report of an inquiry commission is usually made by textual presentation.

**Example:** In 1999, out of a total of five thousand workers of a factory, four thousand and two hundred were members of a Trade Union. The number of female workers was twenty per cent of the total workers out of which thirty per cent were members of the Trade Union.

In 2000, the number of workers belonging to the trade union was increased by twenty per cent as compared to 1999 of which four thousand and two hundred were male.

The number of workers not belonging to trade union was nine hundred and fifty of which four hundred and fifty were females.

The merit of this mode of presentation lies in its simplicity and even a layman can present data by this method. The observations with exact magnitude can be presented with the help of textual presentation. Furthermore, this type of presentation can be taken as the first step towards the other methods of presentation. Textual presentation, however, is not preferred by a statistician simply because, it is dull, monotonous and comparison between different observations is not possible in this method.

**Tabulation:**Tabulation may be defined as the logical and systematic organization of statistical data in rows and columns, designed to simplify the presentation and facilitate comparisons.

We may consider the following guidelines for tabulation:

|  |  |
| --- | --- |
| 1. | A statistical table should be allotted a serial number along with a self-explanatory title. |
| 2. | The table under consideration should be divided into caption, Box-head, Stub and Body.  Different parts of a table  C:\Users\SUSANTA KUMAR GAURI\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.Word\Scan.jpg   * **Caption:** Caption is the upper part of the table, describing the columns and sub-columns, if any. * **Box-Head:** The Box-head is the entire upper part of the table which includes columns and sub-column numbers, unit(s) of measurement along with caption. * **Stub:** Stub is the left part of the table providing the description of the rows. * **Body:** The body is the main part of the table that contains the numerical figures. |
| 3. | The table should be well-balanced in length and breadth. |
| 4. | The data must be arranged in a table in such a way that comparison(s) between different figures are made possible without much labour and time.  Also the row totals, column totals, the units of measurement must be shown. |
| 5. | The data should be arranged intelligently in a well-balanced sequence and the presentation of data in the table should be appealing to the eyes as far as practicable. |
| 6. | Source of the data and Footnote, if necessary, for bringing clarity about any rows or columns should be shown at the bottom of the table. |

The same data of textual presentation can be represented using tabular representation as follows:

**Table 1.** Status of the workers of the factory on the basis of their trade union membership for 1999 and 2000.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Status | Member of TU | | | Non-member | | | Total | | |
| Year | M | F | T | M | F | T | M | F | T |
| 1999 | 3900 | 300 | 4200 | 300 | 500 | 800 | 4200 | 800 | 5000 |
| 2000 | 4200 | 840 | 5050 | 500 | 450 | 950 | 4700 | 1290 | 5990 |

**Source:** Bulletin of labour ministry, WB, January, 2001

**Footnote:** TU, M, F and T stand for trade union, male, female and total respectively.

## ****Objects of Tabulation****

* **To simplify complex data:** When the data are tabulated then unnecessary spread over and duplication is removed. The data are presented in a systematic way in columns and rows so that it becomes simpler to understand them. Thus, it saves our time also.
* **To facilitate comparison:** By tabulation, comparison becomes easy because the table is distributed among different parts and we make totals of each part also. It facilitates comparisons.
* **To give an identity to the data:** When the data are presented in the form of a table, they can be used for other purposes also according to their similarity.
* **To reveal patterns:** The trend of data can easily be understood by tabulation which is difficult without it.
* **Statistical analysis:** By, tabulating the data, they can be easily analyzed and interpreted and other statistical measures can also be applied.

## ****Difference between Classification and Tabulation****

Both the processes of classification and tabulation are important in a statistical investigation. By both of them the data are presented in concise form. Even then following are the differences between the two:

* There is a difference in sequence. The data are classified first and tabulated only thereafter.
* In classification data are classified according to their similarity and dissimilarity but in tabulation classified facts are presented in columns and rows.
* Classification is a method of analysis while tabulation is a method of presentation of data.