# Classification and Tabulation of Data

SESSION 2

Structure



## 2.1 INTRODUCTION

**Prerequisite**

- Lab Session 1 of MSTL-001 (Basic Statistics Lab).
- Unit 13 of MST-001 (Foundation in Mathematics and Statistics).

In Lab Session 1, we have familiarised you with the basic features and functions of MS Excel 2007. So far, you have a basic idea of how Excel 2007 looks and works. You will now learn how to use Excel 2007 for classification of data.

In Unit 13 of MST-001 (Foundation in Mathematics and Statistics), you have learnt that we arrange raw data properly to get the desired information from the data and to facilitate the statistical analysis of data. Raw data do not give a clear picture, so we need to classify the data by dropping unnecessary details.

Classification may be defined as a process of arranging observations in groups or classes according to their resemblances and affinities. In this process, we group related facts into one class. It is the next step in processing the data after data collection. Classification can be done based on qualitative and quantitative aspects. We use qualitative classification for the attributes or characteristics that cannot be measured but can be classified into different categories, e.g., sex, religion, region, occupation, etc. We use quantitative classification for the measurable characteristics or variables. If a variable takes only exact (integer) values, we classify the data into discrete frequency distribution. When the given values of the variable are integers as well as fractional and large in number, we form a continuous frequency distribution. After classification, we can arrange data in the form of a table (in rows and columns). This process is called tabulation and makes it easy for us to understand and compare data.

In this lab session, you will learn about qualitative and quantitative classification of data. In quantitative classification, you will learn how to construct discrete and continuous frequency distributions using MS Excel 2007 along with relative, percentage and cumulative frequency distributions.

## Objectives

After performing the activities of this session, you should be able to:

- prepare the spreadsheet in MS Excel 2007;
- classify qualitative data;
- construct a discrete frequency distribution; and
- construct a continuous frequency distribution.

## 2.2 PROBLEM DESCRIPTION

In this lab session, we state three problems to illustrate different kinds of classification of data:

1. We first consider the problem related to the choice of different flavours of packed juice. A survey was conducted among 100 consumers about their preference for the flavour of packed juice. Classify the raw data recorded in Table 1.

**Table 1: Preference of consumers regarding flavour of juice**

| S. No. | Flavour of Juice | S. No. | Flavour of Juice | S. No. | Flavour of Juice |
|---|---|---|---|---|---|
| 1 | Mixed | 35 | Orange | 69 | Mixed |
| 2 | Pineapple | 36 | Mixed | 70 | Mixed |
| 3 | Litchi | 37 | Litchi | 71 | Litchi |
| 4 | Apple | 38 | Mixed | 72 | Mixed |
| 5 | Apple | 39 | Litchi | 73 | Mixed |
| 6 | Apple | 40 | Pineapple | 74 | Apple |
| 7 | Litchi | 41 | Mixed | 75 | Mixed |
| 8 | Apple | 42 | Pineapple | 76 | Mixed |
| 9 | Litchi | 43 | Mixed | 77 | Apple |
| 10 | Pineapple | 44 | Orange | 78 | Orange |
| 11 | Mixed | 45 | Mixed | 79 | Mixed |
| 12 | Pineapple | 46 | Orange | 80 | Apple |
| 13 | Mixed | 47 | Orange | 81 | Orange |
| 14 | Litchi | 48 | Litchi | 82 | Orange |
| 15 | Mixed | 49 | Mixed | 83 | Orange |
| 16 | Mixed | 50 | Mixed | 84 | Orange |
| 17 | Orange | 51 | Orange | 85 | Orange |
| 18 | Mixed | 52 | Orange | 86 | Apple |
| 19 | Orange | 53 | Orange | 87 | Mixed |
| 20 | Apple | 54 | Orange | 88 | Pineapple |
| 21 | Litchi | 55 | Apple | 89 | Mixed |
| 22 | Litchi | 56 | Mixed | 90 | Orange |
| 23 | Litchi | 57 | Apple | 91 | Mixed |
| 24 | Litchi | 58 | Orange | 92 | Litchi |
| 25 | Mixed | 59 | Orange | 93 | Litchi |
| 26 | Litchi | 60 | Orange | 94 | Mixed |
| 27 | Apple | 61 | Pineapple | 95 | Orange |
| 28 | Pineapple | 62 | Mixed | 96 | Mixed |
| 29 | Mixed | 63 | Orange | 97 | Litchi |
| 30 | Pineapple | 64 | Apple | 98 | Mixed |
| 31 | Apple | 65 | Orange | 99 | Orange |
| 32 | Mixed | 66 | Apple | 100 | Mixed |
| 33 | Mixed | 67 | Litchi | | |
| 34 | Litchi | 68 | Orange | | |

2. A survey was conducted on the usage of mobile phones. A sample of 120 families was selected for this purpose. The data on the number of mobile phones in each family are recorded in Table 2. Construct a discrete frequency distribution.

**Table 2: Number of mobile phones per family**

| Family No. | Number of Mobile Phones | Family No. | Number of Mobile Phones | Family No. | Number of Mobile Phones |
|---|---|---|---|---|---|
| 1 | 5 | 41 | 4 | 81 | 3 |
| 2 | 5 | 42 | 5 | 82 | 7 |
| 3 | 1 | 43 | 6 | 83 | 6 |
| 4 | 2 | 44 | 2 | 84 | 8 |
| 5 | 6 | 45 | 5 | 85 | 2 |
| 6 | 1 | 46 | 4 | 86 | 5 |
| 7 | 5 | 47 | 6 | 87 | 2 |
| 8 | 4 | 48 | 7 | 88 | 5 |
| 9 | 2 | 49 | 4 | 89 | 3 |
| 10 | 6 | 50 | 4 | 90 | 8 |
| 11 | 3 | 51 | 7 | 91 | 3 |
| 12 | 6 | 52 | 5 | 92 | 5 |
| 13 | 1 | 53 | 4 | 93 | 4 |
| 14 | 6 | 54 | 8 | 94 | 2 |
| 15 | 6 | 55 | 8 | 95 | 5 |
| 16 | 3 | 56 | 6 | 96 | 2 |
| 17 | 3 | 57 | 3 | 97 | 5 |
| 18 | 4 | 58 | 6 | 98 | 5 |
| 19 | 6 | 59 | 3 | 99 | 2 |
| 20 | 4 | 60 | 5 | 100 | 7 |
| 21 | 4 | 61 | 4 | 101 | 1 |
| 22 | 2 | 62 | 3 | 102 | 2 |
| 23 | 3 | 63 | 6 | 103 | 7 |
| 24 | 5 | 64 | 4 | 104 | 4 |
| 25 | 1 | 65 | 2 | 105 | 4 |
| 26 | 3 | 66 | 5 | 106 | 2 |
| 27 | 4 | 67 | 2 | 107 | 7 |
| 28 | 3 | 68 | 7 | 108 | 5 |
| 29 | 4 | 69 | 5 | 109 | 3 |
| 30 | 6 | 70 | 4 | 110 | 4 |
| 31 | 2 | 71 | 4 | 111 | 3 |
| 32 | 3 | 72 | 4 | 112 | 2 |
| 33 | 3 | 73 | 5 | 113 | 6 |
| 34 | 1 | 74 | 3 | 114 | 3 |
| 35 | 5 | 75 | 7 | 115 | 6 |
| 36 | 4 | 76 | 7 | 116 | 3 |
| 37 | 8 | 77 | 4 | 117 | 5 |
| 38 | 3 | 78 | 4 | 118 | 4 |
| 39 | 3 | 79 | 5 | 119 | 3 |
| 40 | 6 | 80 | 5 | 120 | 2 |

3. A sample of 150 bulbs was selected to check the life span of the bulbs. The data on the life span of bulbs are recorded in Table 3.

**Table 3: Life span of bulbs**

| S. No. | Life of Bulb (in hours) | S. No. | Life of Bulb (in hours) | S. No. | Life of Bulb (in hours) |
|---|---|---|---|---|---|
| 1 | 1087 | 51 | 975 | 101 | 578 |
| 2 | 1289 | 52 | 1278 | 102 | 754 |
| 3 | 876 | 53 | 526 | 103 | 678 |
| 4 | 725 | 54 | 1296 | 104 | 758 |

| S. No. | Life of Bulb (in hours) | S. No. | Life of Bulb (in hours) | S. No. | Life of Bulb (in hours) |
|---|---|---|---|---|---|
| 5 | 900 | 55 | 607 | 105 | 754 |
| 6 | 1080 | 56 | 832 | 106 | 708 |
| 7 | 952 | 57 | 776 | 107 | 598 |
| 8 | 741 | 58 | 697 | 108 | 834 |
| 9 | 1000 | 59 | 972 | 109 | 678 |
| 10 | 900 | 60 | 945 | 110 | 904 |
| 11 | 957 | 61 | 991 | 111 | 848 |
| 12 | 745 | 62 | 856 | 112 | 1145 |
| 13 | 926 | 63 | 950 | 113 | 1044 |
| 14 | 781 | 64 | 678 | 114 | 867 |
| 15 | 1120 | 65 | 862 | 115 | 1063 |
| 16 | 1208 | 66 | 768 | 116 | 928 |
| 17 | 1138 | 67 | 961 | 117 | 1022 |
| 18 | 1298 | 68 | 622 | 118 | 875 |
| 19 | 862 | 69 | 658 | 119 | 934 |
| 20 | 711 | 70 | 754 | 120 | 985 |
| 21 | 886 | 71 | 754 | 121 | 1033 |
| 22 | 1066 | 72 | 1074 | 122 | 694 |
| 23 | 938 | 73 | 678 | 123 | 584 |
| 24 | 727 | 74 | 847 | 124 | 1187 |
| 25 | 991 | 75 | 689 | 125 | 1114 |
| 26 | 501 | 76 | 871 | 126 | 890 |
| 27 | 697 | 77 | 568 | 127 | 834 |
| 28 | 731 | 78 | 578 | 128 | 1131 |
| 29 | 912 | 79 | 1140 | 129 | 1030 |
| 30 | 687 | 80 | 614 | 130 | 947 |
| 31 | 721 | 81 | 859 | 131 | 698 |
| 32 | 866 | 82 | 1050 | 132 | 914 |
| 33 | 762 | 83 | 1187 | 133 | 674 |
| 34 | 857 | 84 | 981 | 134 | 935 |
| 35 | 848 | 85 | 750 | 135 | 920 |
| 36 | 650 | 86 | 768 | 136 | 824 |
| 37 | 514 | 87 | 1076 | 137 | 1019 |
| 38 | 854 | 88 | 1060 | 138 | 680 |
| 39 | 1070 | 89 | 789 | 139 | 784 |
| 40 | 846 | 90 | 833 | 140 | 798 |
| 41 | 790 | 91 | 915 | 141 | 1065 |
| 42 | 1087 | 92 | 857 | 142 | 1049 |
| 43 | 986 | 93 | 754 | 143 | 598 |
| 44 | 824 | 94 | 564 | 144 | 822 |
| 45 | 1005 | 95 | 737 | 145 | 867 |
| 46 | 870 | 96 | 600 | 146 | 846 |
| 47 | 964 | 97 | 845 | 147 | 889 |
| 48 | 1172 | 98 | 675 | 148 | 837 |
| 49 | 876 | 99 | 1186 | 149 | 1115 |
| 50 | 637 | 100 | 967 | 150 | 575 |

For the data in Problems 3,

- compute suitable width for the class intervals, and
- construct the continuous, relative, percentage and cumulative frequency distributions.

## 2.3 QUALITATIVE CLASSIFICATION

You have learnt about the classification of qualitative data in Unit 13 of MST-001. When we have qualitative data, which is not measurable, we classify the data based on two or more categories of the qualitative characteristic. We briefly mention the main steps as follows:

**Step 1:** In qualitative classification, we first arrange all categories of the qualitative characteristic in a column of the table.

**Step 2:** We count the number of values, which correspond to that particular category of the qualitative characteristic among the given data. For example, we can classify the data based on literacy. Here we can form two groups (literate or non-literate) and count the number of persons belonging to each group.

## Steps in Excel

In Problem 1, the data is qualitative. We are given five flavours of juice, i.e., we have five categories of a qualitative characteristic. We now describe the procedure of classifying the qualitative data of Table 1 in Excel 2007. In order to classify the given qualitative data, we follow the steps given below:

**Step 1:** We enter the data given in Table 1 in Excel 2007 spreadsheet. We start by entering the heading of the data in Row 1 and the data itself from Row 2 in Excel sheet. So the entry will go up to Row 101. For the given data, the spreadsheet looks as shown in Fig. 2.1.

| | A | B |
|---|---|---|
| 1 | S. No. | Flavour of Juice |
| 2 | 1 | Mixed |
| 3 | 2 | Pineapple |
| 4 | 3 | Litchi |
| 5 | 4 | Apple |
| 6 | 5 | Apple |
| 7 | 6 | Apple |
| 8 | 7 | Litchi |
| 9 | 8 | Apple |
| 10 | 9 | Litchi |
| 11 | 10 | Pineapple |

Qualitative

**Fig. 2.1: Partial screenshot of the spreadsheet for the given data.**

**Step 2:** We type the title, "Flavour of Juice", in Cell D1. Since five different flavours (orange, apple, mixed, litchi and pineapple) are given in the data of Table 1, we type these flavours in Cells D2:D6 as shown in Fig. 2.2.

| | D | E |
|---|---|---|
| 1 | Flavour of Juice | |
| 2 | Orange | |
| 3 | Apple | |
| 4 | Mixed | |
| 5 | Litchi | |
| 6 | Pineapple | |
| 7 | | |

**Fig. 2.2**

**Step 3:** We type the title, "Number of Consumers", in Cell E1. For determining the number of consumers who like the orange flavour, we

1. select Cell E2,
2. click on the ***Formulas*** tab, and
3. click on the ***More Functions*** → ***Statistical*** → ***Countif*** as shown in Fig. 2.3.

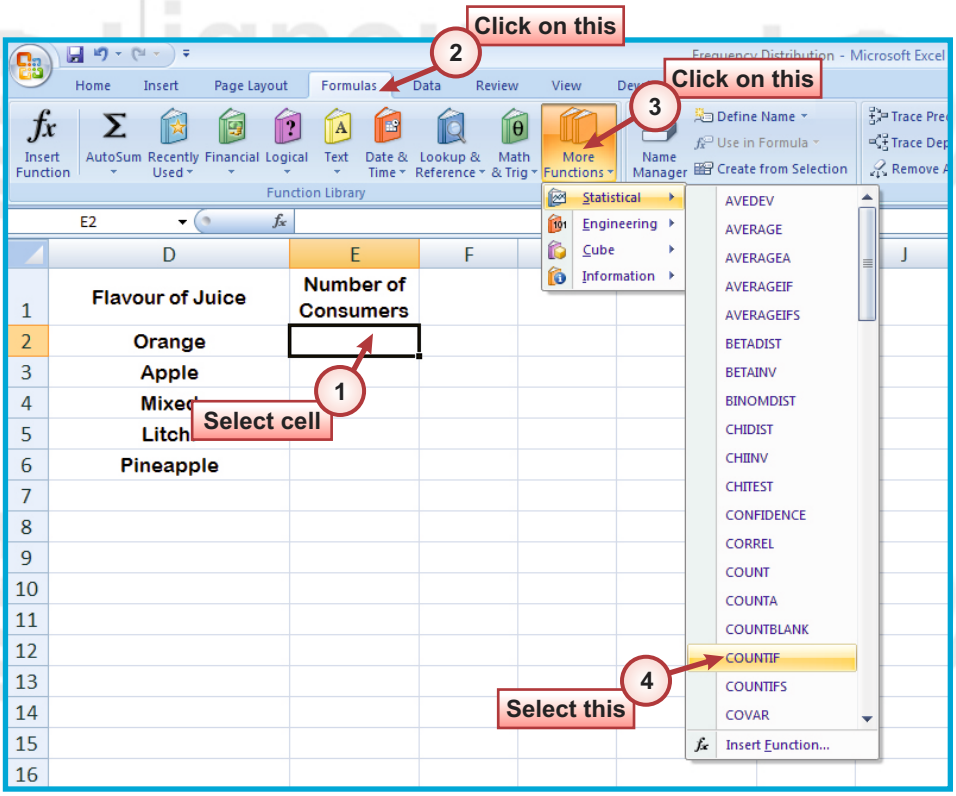



**Fig. 2.3**

A new dialog box opens as shown in Fig. 2.4.

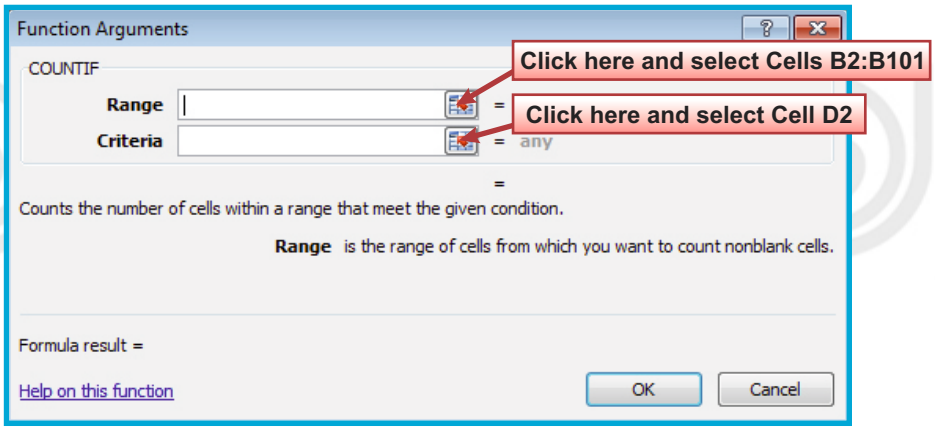



**Fig. 2.4**

**Step 4:** Since we are counting the orange flavour, we select or type ***Range*** as Cells B2:B101 and ***Criteria*** as Cell D2 as shown in Fig. 2.5a. Then we click on ***OK*** or press ***Enter*** and get the result (Fig. 2.5b).

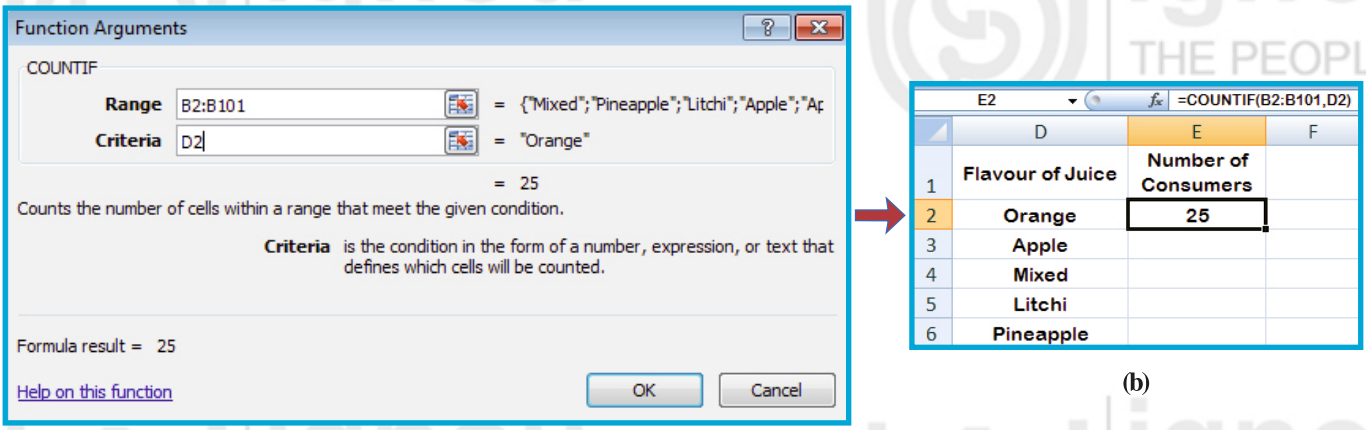



**Fig. 2.5**

**Step 5:** We now fix the range given in ***Range***, i.e., Cells B2:B101 by putting the "$" sign, i.e., $B$2:$B$101 as shown in Fig. 2.6. Note that it helps us in copying the formula in Cells E3:E6.

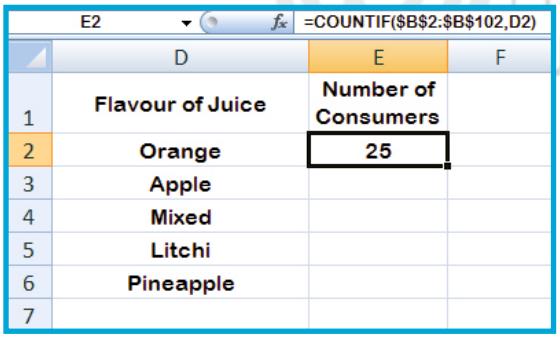


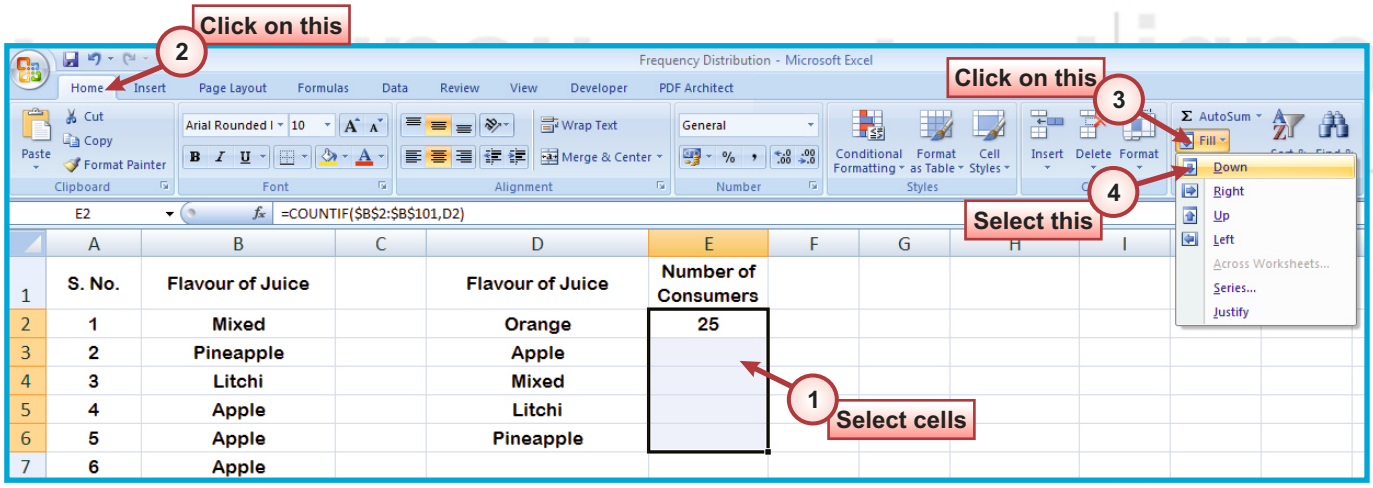

**Fig. 2.6**

From Fig. 2.6, the value in Cell E2 is 25. It means that in the given data there are 25 consumers who like orange flavour.

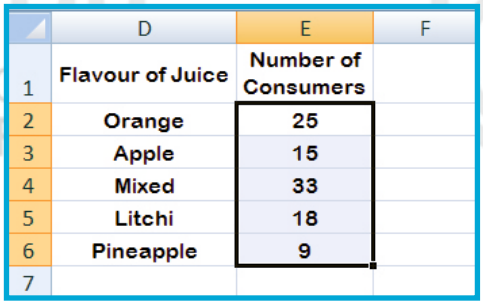
**Step 6:** To compute the number of consumers who like the remaining flavours, refer to Fig. 2.7a. We

1. select Cells E2:E6,
2. click on the ***Home*** tab, and
3. click on ***Down*** under the ***Fill*** option.

We get the result shown in Fig. 2.7b.




(a)




(b)

**Fig. 2.7**

**Step 7:** To calculate the total number of consumers, we select Cell E7 and click on the ∑ ***AutoSum*** option shown by the arrow in Fig. 2.8a. It gives the sum function in Cell E7 as shown in Fig. 2.8b.

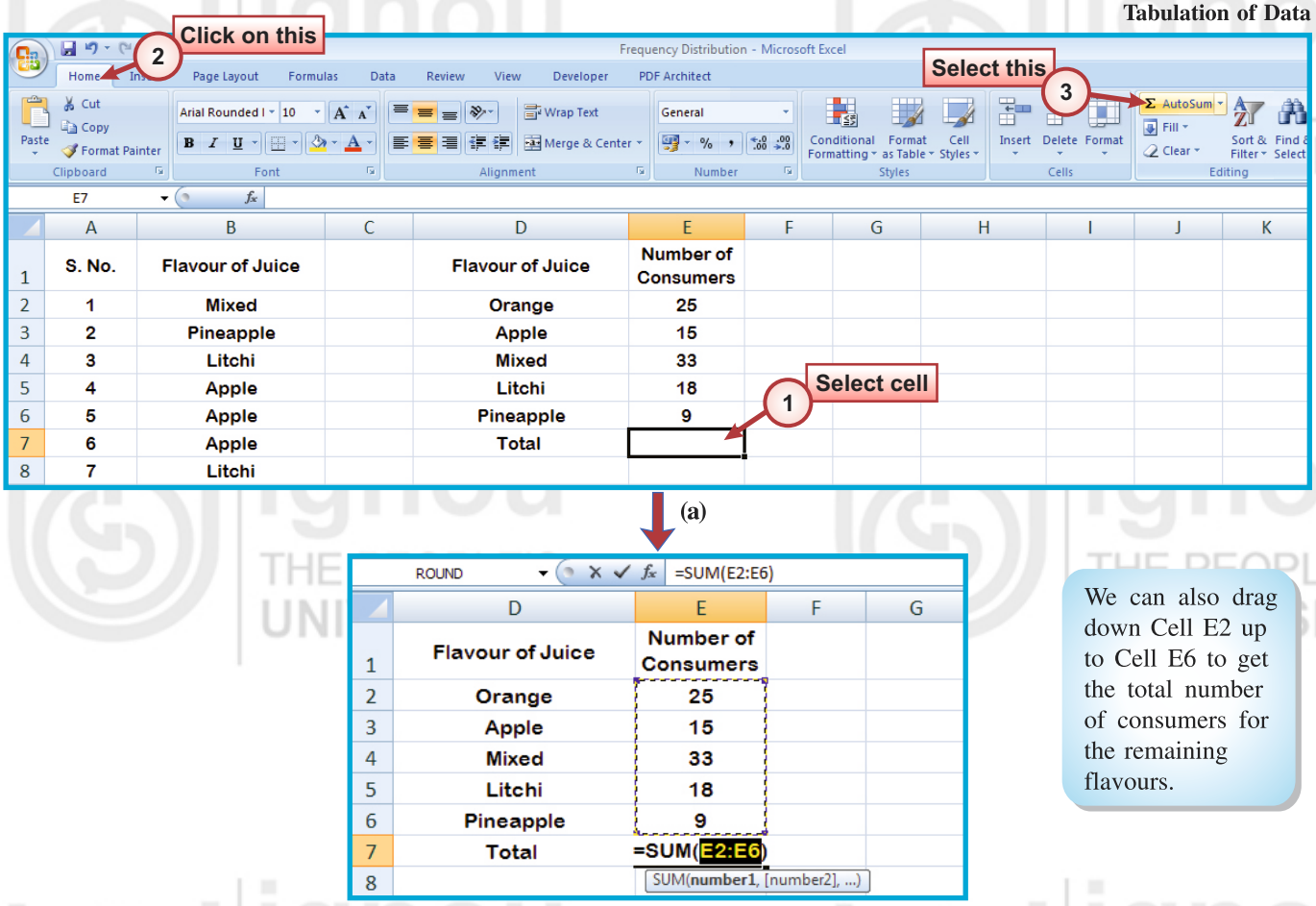



Fig. 2.8

We can also drag down Cell E2 up to Cell E6 to get the total number of consumers for the remaining flavours.

**Step 8:** When we press ***Enter***, we get the output, i.e., the total number of consumers as shown in Fig. 2.9.

| | D | E | F |
|---|---|---|---|
| 1 | Flavour of Juice | Number of Consumers | |
| 2 | Orange | 25 | |
| 3 | Apple | 15 | |
| 4 | Mixed | 33 | |
| 5 | Litchi | 18 | |
| 6 | Pineapple | 9 | |
| 7 | Total | 100 | |

Fig. 2.9

Fig. 2.9 shows the qualitative classification of data. From the data in Table 1 on consumer choice, we note that the highest number of consumers, i.e., 33 consumers like the mixed flavour and only 9 like the pineapple flavour.

## 2.4 DISCRETE FREQUENCY DISTRIBUTION

You have learnt about the classification of quantitative data in Unit 13 of MST-001. We can construct two types of frequency distributions: (i) discrete and (ii) continuous. We shall discuss the continuous frequency distribution in Sec. 2.5. When a variable takes exact (integer) values and the given values of the variable are less in number, we form a discrete frequency distribution. We briefly mention the main steps for discrete frequency distribution as follows:

**Step 1:** For forming a discrete frequency distribution, we first arrange all given distinct values of the variable in a column of the table.

**Step 2:** We compute the number of times a particular value occurs in the data. It is called the frequency of that value of data. In this way, we compute the frequencies for all values of the given data.

## Steps in Excel

In Problem 2, the data of the number of mobile phones in 120 families is quantitative and has only eight integer values. So we can form the discrete frequency distribution for the data given in Table 2. Here we describe the procedure of forming a discrete frequency distribution of quantitative data in MS Excel 2007. In order to classify the given data, we follow the steps given below:

| | A | B |
|---|---|---|
| 1 | Family No. | Number of Mobile Phones |
| 2 | 1 | 5 |
| 3 | 2 | 5 |
| 4 | 3 | 1 |
| 5 | 4 | 2 |
| 6 | 5 | 6 |
| 7 | 6 | 1 |
| 8 | 7 | 5 |
| 9 | 8 | 4 |
| 10 | 9 | 2 |
| 11 | 10 | 6 |

**Fig. 2.10: Partial screenshot of the spreadsheet for the given data.**

**Step 1:** We enter the data given in Table 2 in Excel 2007 spreadsheet as explained in Step 1 of Sec. 2.3 (Fig. 2.10).

**Step 2:** We type the text, "Number of Mobile Phones", in Cell D1. In the given data, the minimum number of mobile phones in a family is 1 while the maximum number of mobile phones is 8. So we type the number of mobile phones, i.e., 1 to 8 in Cells D2:D9 as shown in Fig. 2.11.

| | D |
|---|---|
| 1 | Number of Mobile Phones |
| 2 | 1 |
| 3 | 2 |
| 4 | 3 |
| 5 | 4 |
| 6 | 5 |
| 7 | 6 |
| 8 | 7 |
| 9 | 8 |

**Fig. 2.11**

**Step 3:** We type the text, "No. of Families", in Cell E1. We now compute the frequency, i.e., count the number of families corresponding to each number of mobile phones. Since the data is quantitative, we can use the ***Frequency*** function as shown in Fig. 2.12. For this, we

1. select eight cells corresponding to the eight values of mobile phones, i.e., Cells E2:E9,
2. click on the ***Formulas*** tab, and
3. click on ***More functions*** → ***Statistical*** → ***Frequency***.

We can also use ***Countif*** function instead of ***Frequency*** function to compute the number of families.

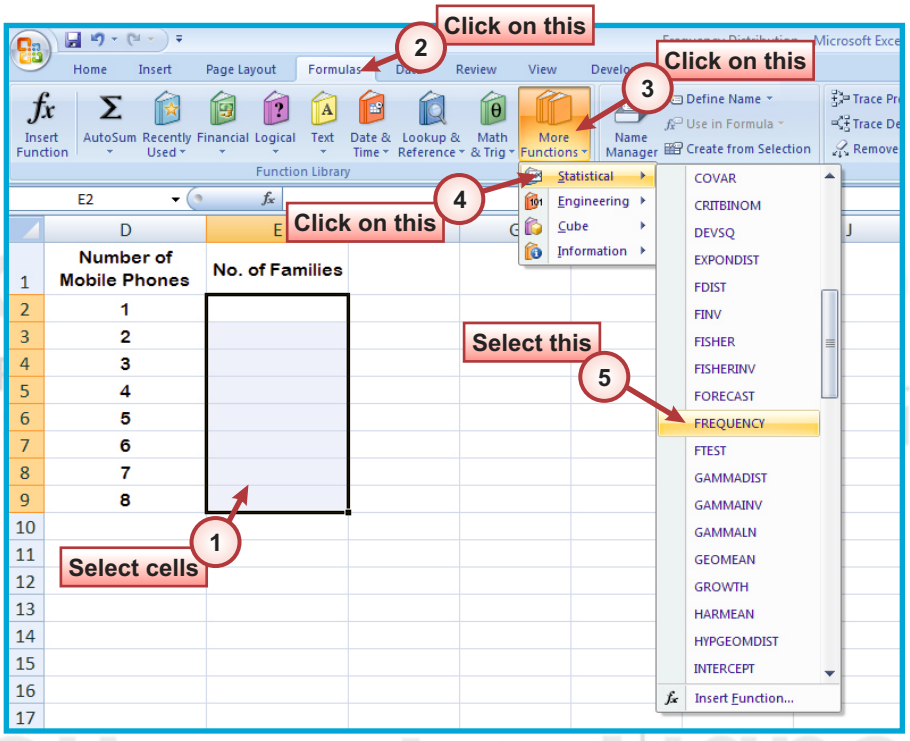



**Fig. 2.12**

A new dialog box opens (Fig. 2.13). This dialog box requires the value of *Data_array* and *Bin_array*.

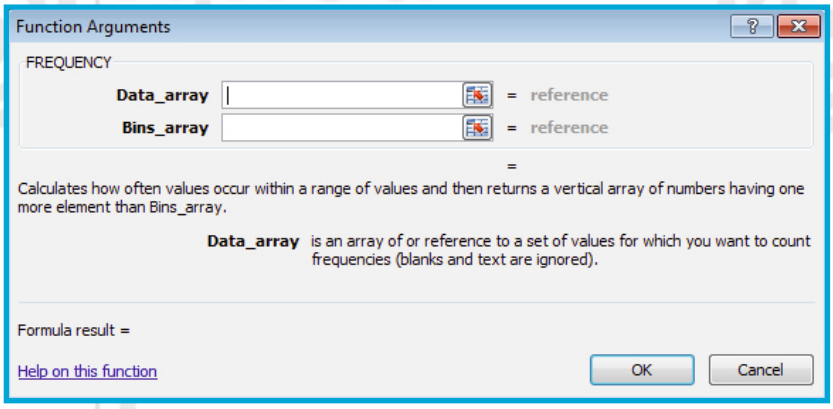


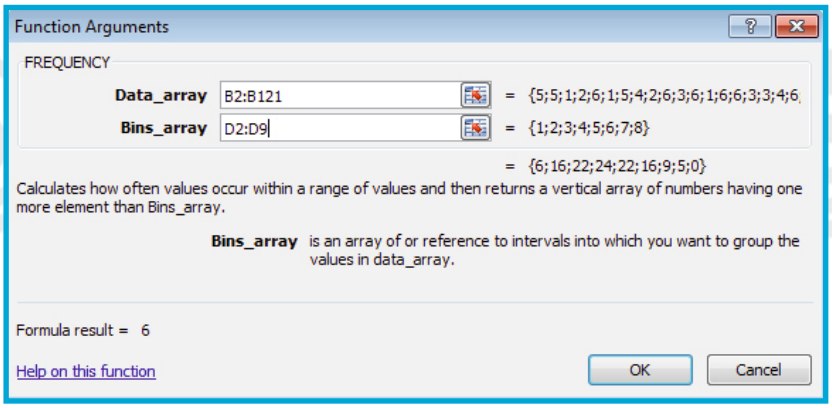

**Fig. 2.13**

The frequency formula is an array formula. To input an array formula, we first select the range of cells in which we want to get results. We type or use formula in the first cell after selecting range and then press *Ctrl* + *Shift* + *Enter* to get the result.

**Step 4:** The *Data_array*, i.e., values of the data for which we wish to count the frequencies and the *Bin_array*, i.e., reference into which we wish to group or classify the data given in Cells B2:B121 and Cells D2:D9, respectively.

We select or type Cells B2:B121 as *Data_array* and Cells D2:D9 as *Bins_array* as shown in Fig. 2.14a. We now press *Ctrl* + *Shift* + *Enter* keys together instead of simply pressing *Enter* key as the *Frequency* function works as an array. We can also click *OK* while pressing *Ctrl* + *Shift* keys. The result is shown in Fig. 2.14b.

The *Frequency* formula counts the frequency of the corresponding value provided in *Bin_array* but it is important to note that if we provide the single reference (value) in Bin-array, it counts the total number of values which are less than and equal to the value in Bin-array.




(a)

If we type, say, 10, 20, 30 and so on as *Bins_array*, it considers it as an interval, i.e., ≤ 10, 11-20, 21-30, and so on. For the given problem, we consider *Bins_array* 1, 2,..., 8 (Fig. 2.12). Since these values are all distinct values given in the data, this formula will compute the discrete frequency distribution.

CTRL+SHIFT+ENTER

E2 {=FREQUENCY(B2:B121,D2:D9)}

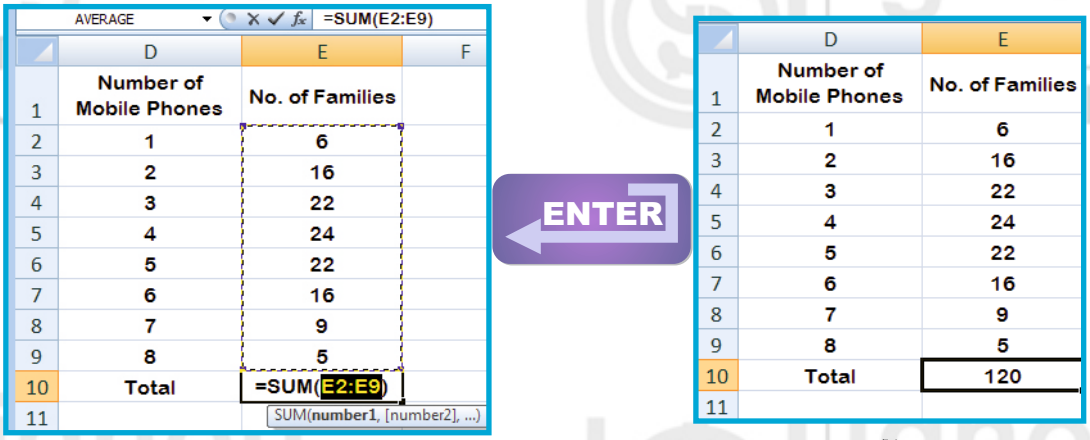
| | D | E | F | G |
|---|---|---|---|---|
| 1 | Number of Mobile Phones | No. of Families | | |
| 2 | 1 | 6 | | |
| 3 | 2 | 16 | | |
| 4 | 3 | 22 | | |
| 5 | 4 | 24 | | |
| 6 | 5 | 22 | | |
| 7 | 6 | 16 | | |
| 8 | 7 | 9 | | |
| 9 | 8 | 5 | | |

(b)

**Fig. 2.14**

Fig. 2.14 shows the number mobile phones and its corresponding frequency, i.e., number of families having the respective number of mobile phones.

**Step 5:** In Cell E10, we determine the total number of families as explained in Steps 7 and 8 of Sec. 2.3 and as shown in Fig. 2.15.




(a) (b)

**Fig. 2.15**

Fig. 2.15b represents the discrete frequency distribution of the given data on mobile phones.

## 2.5 CONTINUOUS FREQUENCY DISTRIBUTION

You have learnt about the classification of quantitative data to form the continuous frequency distribution in Unit 13 of MST-001. When the given values of the variable are integers as well as fractional and large in number, we form a continuous frequency distribution. Here we briefly mention the main steps as under:

**Step 1:** We determine the minimum and maximum values of the data.

**Step 2:** We calculate the number of class intervals using Sturges' rule given by

$$K = 1+3.322 \log_{10} N \quad \ldots (1)$$

where K – number of class intervals, and

N – total number of observations.

**Step 3:** We now compute the width of the class interval as follows:

$$i = \frac{\text{Range}}{1+3.322\log_{10} N} \quad \ldots (2)$$

where i – width of class interval, and

Range = Maximum value – Minimum value.

Note that we prefer the class intervals in multiples of 5 or 10.

**Step 4:** In the inclusive method, we include both the lower limit and the upper limit of each class interval.

**Step 5:** In the exclusive method, we exclude the upper limit of each class interval. The upper limit of one class is the lower limit of the next class.

**Step 6:** We compute the relative frequency as a ratio of the frequency of that class to the total frequency, i.e., total number of observations given by

$$\text{Relative frequency} = \frac{f}{N} \quad \ldots (3)$$

where f – frequency, and

N – total number of observations.

The corresponding frequency distribution is called the **relative frequency distribution**.

**Step 7:** We multiply each relative frequency by 100 to get the percentage frequency of that class as

$$\text{Percentage frequency} = \frac{f}{N} \times 100 \quad \ldots (4)$$

The corresponding frequency distribution is called the **percentage frequency distribution**.

**Step 8:** The number of observations less than the upper class limit is known as **less than cumulative frequency** of that class. The corresponding distribution is called **less than cumulative frequency distribution**.

**Step 9:** Similarly, the number of observations more than and equal to the lower class limit of a class is known as **more than cumulative frequency** of that class. The corresponding distribution is called **more than cumulative frequency distribution**.

**Step 10:** The first **less than cumulative frequency** is the same as the first **frequency** and we calculate the remaining frequencies by adding the frequency of the respective class in the cumulative frequency of previous class.

**Step 11:** The first **more than cumulative** is the same as the **total** of the frequencies and we calculate the remaining frequencies by subtracting the previous class frequency from the previous class cumulative frequency.

## Steps in Excel

In Problem 3, the data of life of bulbs is quantitative. Since the values of data given in Table 3 are integers but large in number, we form a continuous frequency distribution. Here we describe the procedure of forming a continuous frequency distribution of quantitative data using MS Excel 2007. In order to classify the given data, we follow the steps given below:

**Step 1:** We enter the data given in Table 3 in Excel 2007 spreadsheet as shown in Fig. 2.16.

| | A | B | C |
|---|---|---|---|
| 1 | S.No. | Life of Bulb (in hours) | |
| 2 | 1 | 1087 | |
| 3 | 2 | 1289 | |
| 4 | 3 | 876 | |
| 5 | 4 | 725 | |
| 6 | 5 | 900 | |
| 7 | 6 | 1080 | |
| 8 | 7 | 952 | |
| 9 | 8 | 741 | |
| 10 | 9 | 1000 | |
| 11 | 10 | 900 | |

**Fig. 2.16: Partial screenshot of the spreadsheet for the given data.**

**Step 2:** We type "Min. =" in Cell C2. We compute the minimum value of the given data in Cell C3 as shown in Fig. 2.17. For this, we

1. select Cell C3,
2. click on the ***Formulas*** tab,

3. click on the ***More Functions***, and
4. choose ***Min*** function under ***Statistical***.

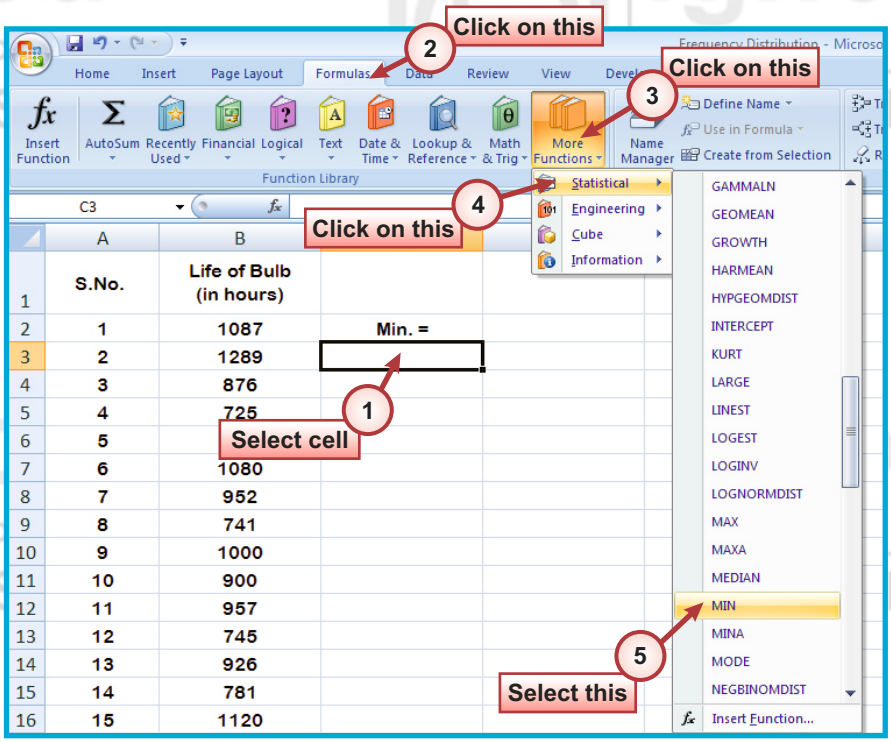



**Fig. 2.17**

**Step 3:** When we click on ***Min*** function shown in Fig. 2.17, a new dialog box opens. We select or type Cells B2:B151 in the dialog box as shown in Fig. 2.18a and click on ***OK***. The result is shown in Fig. 2.18b.

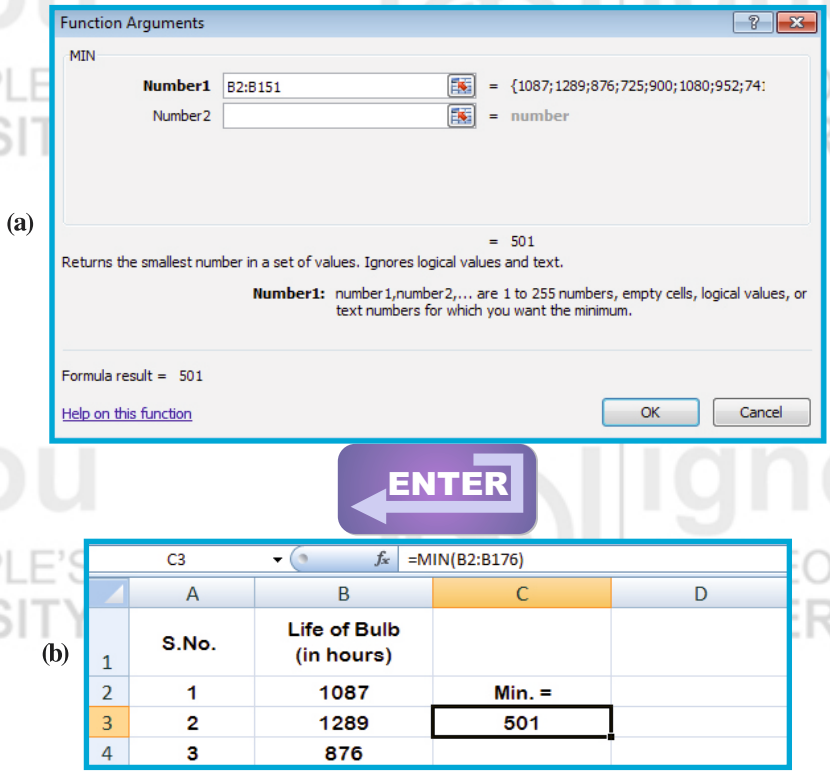



**Fig. 2.18**

**Step 4:** We type "Max. =" in Cell C4. We now determine the maximum value of the given data in Cell C5 as shown in Fig. 2.19. For this, we

1. select Cell C5,
2. click on the ***Formulas*** tab,

3. click on the ***More Functions***,
4. choose ***Max*** function under ***Statistical*** (Fig. 2.19a),
5. select or type Cells B2:B151 in the dialog box that opens (Fig. 2.19b), and
6. click on ***OK*** and get the output shown in Fig. 2.19c.

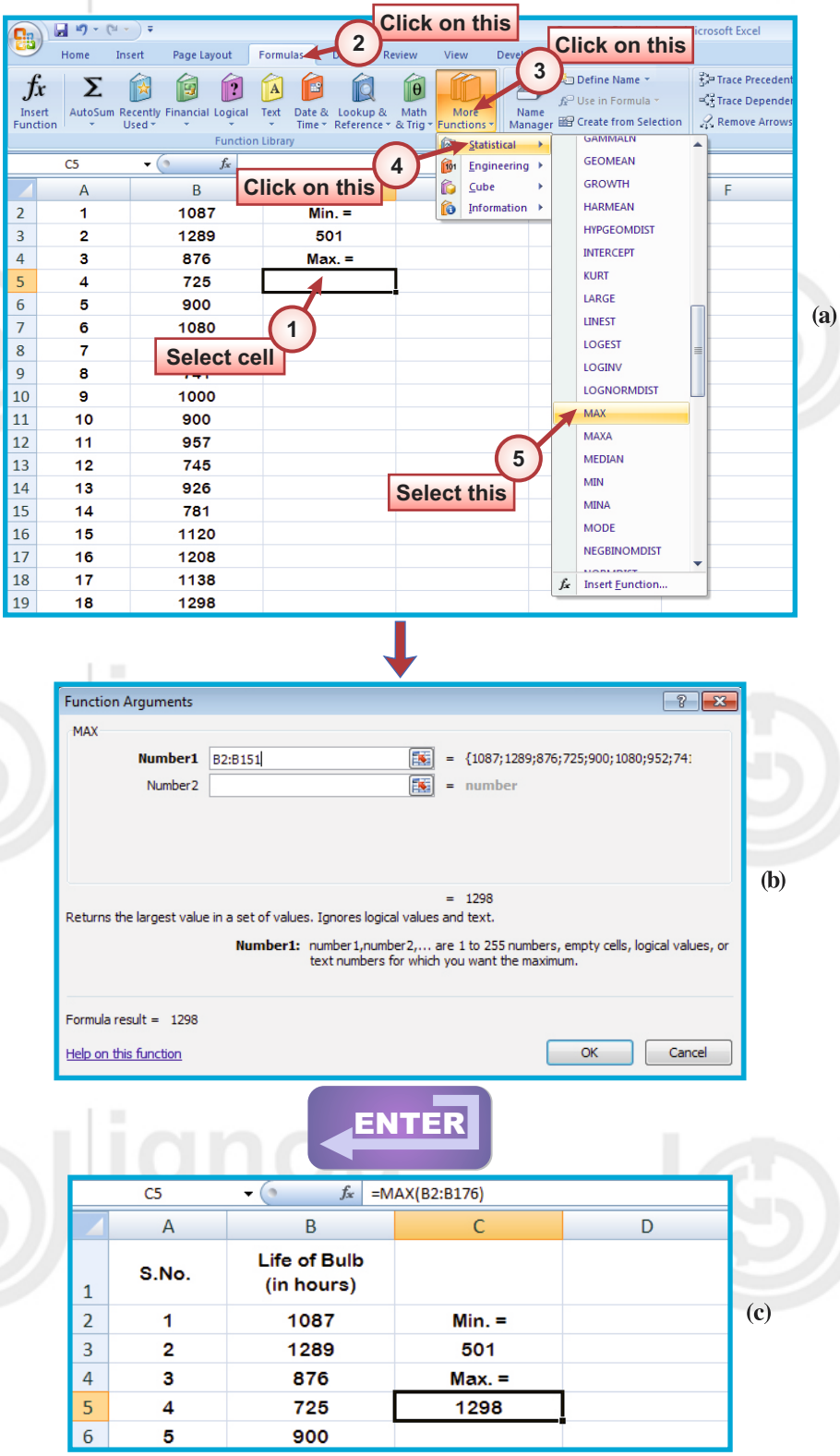
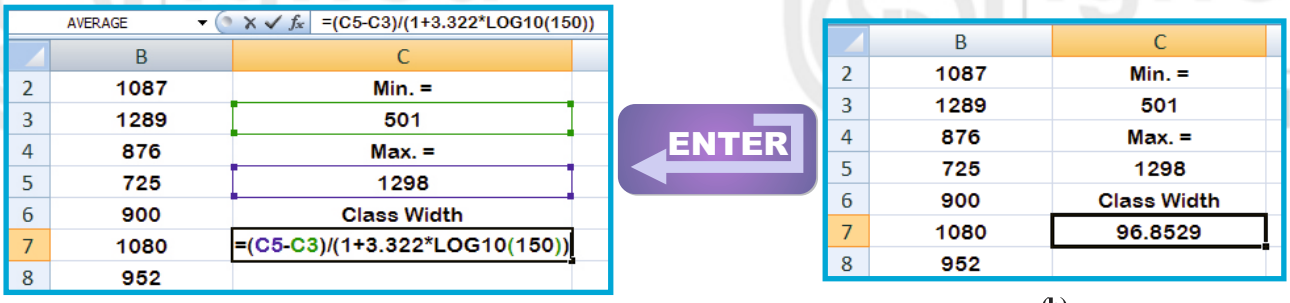



**Fig. 2.19**

**Step 5:** We type the text, ''Class Width'', in Cell C6. We consider the formula given in equation (2) in this section for computing the width of the class intervals. Note that the range is the difference of the maximum value given in Cell C5 and the minimum value given in Cell C3. Moreover,

N = 150. So we type "=(C5–C3)/(1+(3.322*Log 10(150)))" in Cell C7 as shown in Fig. 2.20a.




| | B | C |
|---|---|---|
| 2 | 1087 | Min. = |
| 3 | 1289 | 501 |
| 4 | 876 | Max. = |
| 5 | 725 | 1298 |
| 6 | 900 | Class Width |
| 7 | 1080 | =(C5-C3)/(1+3.322*LOG10(150)) |
| 8 | 952 | |

(a)

| | B | C |
|---|---|---|
| 2 | 1087 | Min. = |
| 3 | 1289 | 501 |
| 4 | 876 | Max. = |
| 5 | 725 | 1298 |
| 6 | 900 | Class Width |
| 7 | 1080 | 96.8529 |
| 8 | 952 | |

(b)

**Fig. 2.20**

When we press ***Enter***, we get the number of classes as 96.8529 (Fig. 2.20b). As you have learnt in Unit 13 of MST-001, we should avoid these types of values and prefer class intervals in multiples of 5 or 10. So we shall approximate 96.8529 to 100 and take it as the class interval. Hence, we take the classes as 500-600, 600-700, and so on.

**Step 6:** We type "Class Interval (Exclusive)" and "Class Interval (Inclusive)" in Cells D1 and E1, respectively, as shown in Fig. 2.21. The class intervals given in Cells D2:D9 are formed using the exclusive method and the class intervals given in Cells E2:E9 are formed using the inclusive method.

| | C | D | E |
|---|---|---|---|
| 1 | | Class Interval (Exclusive) | Class Interval (Inclusive) |
| 2 | Min. = | 500-600 | 500-599 |
| 3 | 501 | 600-700 | 600-699 |
| 4 | Max. = | 700-800 | 700-799 |
| 5 | 1298 | 800-900 | 800-899 |
| 6 | | 900-1000 | 900-999 |
| 7 | | 1000-1100 | 1000-1099 |
| 8 | | 1100-1200 | 1100-1199 |
| 9 | | 1200-1300 | 1200-1299 |
| 10 | | | |

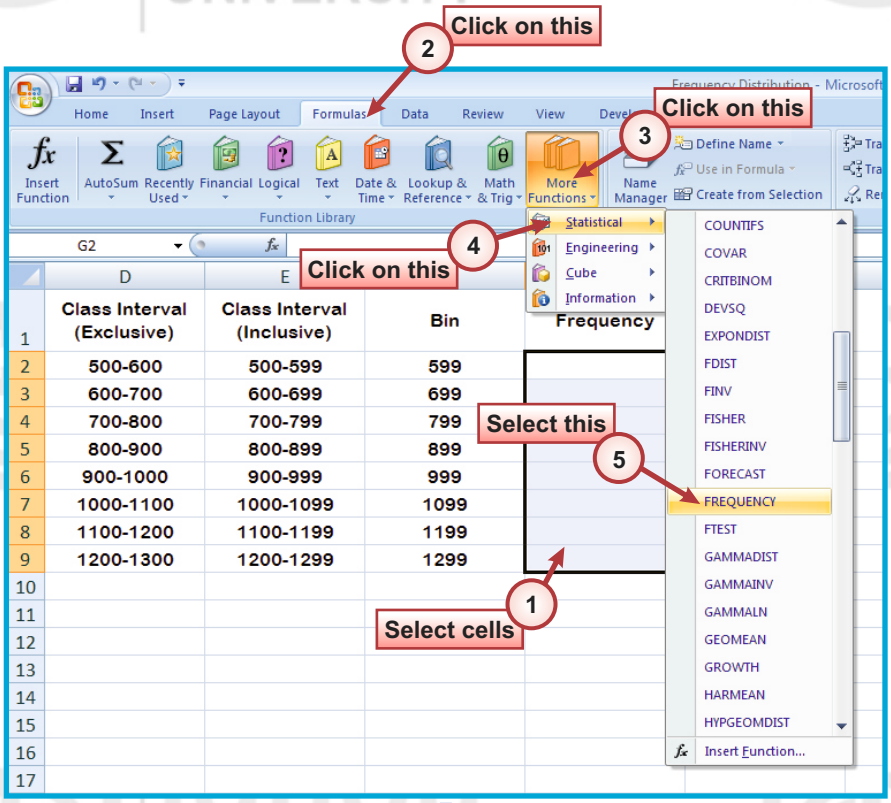
**Fig. 2.21**

**Step 7:** For calculating the frequencies in Excel 2007, we have to formulate the ***Bin_array***, which give the frequencies of the corresponding class intervals. Here we consider the first ***Bin*** value as 599 because the ***Frequency*** formula computes the total number of values (or frequency), which are less than or equal to the ***Bin*** value. So we take ***Bin_array*** equal to the upper limit of inclusive class interval and type "Bin" in Cell F1 and the values of ***Bin_array*** in Cells F2:F9 as shown in Fig. 2.22.

| | D | E | F |
|---|---|---|---|
| 1 | Class Interval (Exclusive) | Class Interval (Inclusive) | Bin |
| 2 | 500-600 | 500-599 | 599 |
| 3 | 600-700 | 600-699 | 699 |
| 4 | 700-800 | 700-799 | 799 |
| 5 | 800-900 | 800-899 | 899 |
| 6 | 900-1000 | 900-999 | 999 |
| 7 | 1000-1100 | 1000-1099 | 1099 |
| 8 | 1100-1200 | 1100-1199 | 1199 |
| 9 | 1200-1300 | 1200-1299 | 1299 |

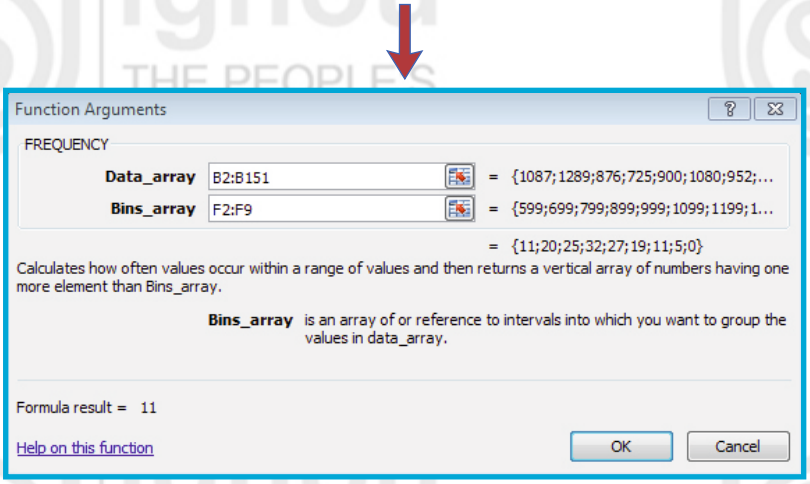
**Fig. 2.22**

**Step 8:** We type "No. of Bulbs (Frequency)" in Cell G1. We select Cells G2:G9 and follow the method explained in Steps 3 and 4 of Sec. 2.4. (Fig. 2.23a). Here we select ***Data_array*** as Cells B2:B151 and ***Bin_array*** as F2:F9 (Fig. 2.23b). We press ***Ctrl*** + ***Shift*** + ***Enter*** keys together. The output is shown in Fig. 2.23c.




| | D | E | Bin | Frequency |
|---|---|---|---|---|
| 1 | Class Interval (Exclusive) | Class Interval (Inclusive) | | |
| 2 | 500-600 | 500-599 | 599 | |
| 3 | 600-700 | 600-699 | 699 | |
| 4 | 700-800 | 700-799 | 799 | |
| 5 | 800-900 | 800-899 | 899 | |
| 6 | 900-1000 | 900-999 | 999 | |
| 7 | 1000-1100 | 1000-1099 | 1099 | |
| 8 | 1100-1200 | 1100-1199 | 1199 | |
| 9 | 1200-1300 | 1200-1299 | 1299 | |
| 10 | | | | |
| 11 | | | | |
| 12 | | | | |
| 13 | | | | |
| 14 | | | | |
| 15 | | | | |
| 16 | | | | |
| 17 | | | | |

(a)




(b)

CTRL+SHIFT+ENTER

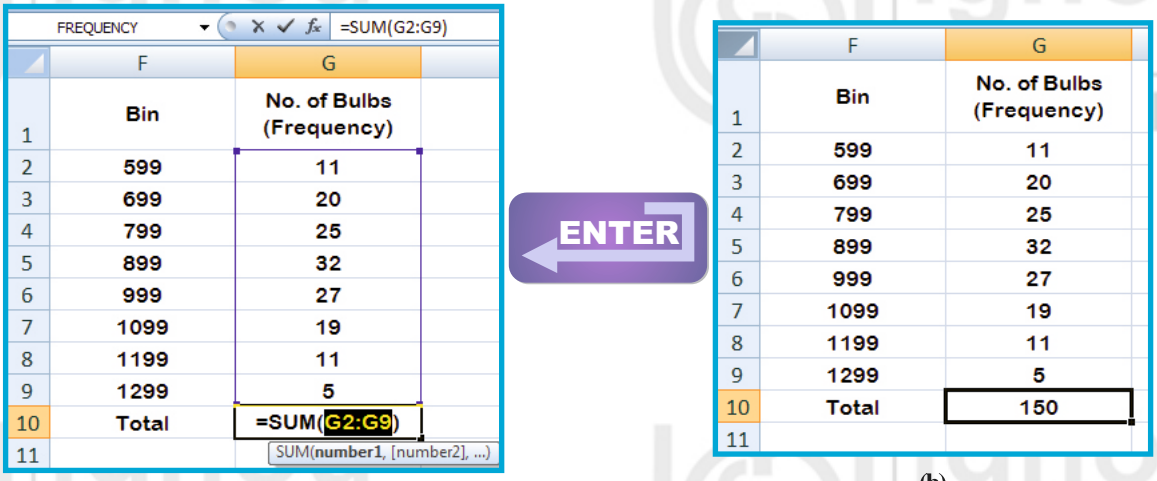
| | D | E | F | G |
|---|---|---|---|---|
| 1 | Class Interval (Exclusive) | Class Interval (Inclusive) | Bin | No. of Bulbs (Frequency) |
| 2 | 500-600 | 500-599 | 599 | 11 |
| 3 | 600-700 | 600-699 | 699 | 20 |
| 4 | 700-800 | 700-799 | 799 | 25 |
| 5 | 800-900 | 800-899 | 899 | 32 |
| 6 | 900-1000 | 900-999 | 999 | 27 |
| 7 | 1000-1100 | 1000-1099 | 1099 | 19 |
| 8 | 1100-1200 | 1100-1199 | 1199 | 11 |
| 9 | 1200-1300 | 1200-1299 | 1299 | 5 |

(c)

Fig. 2.23

If the given data have values up to k decimal places, we consider the values of **Bin** up to (k+1) decimal places. For example, if the data have the values up to 3 decimal places. We consider the values of ***Bin_array*** up to 4 decimal place, say, 9.9999, 29.9999, etc., to avoid the misclassification.

**Step 9:** In Cell G10, we determine the total number of bulbs as explained in Steps 7 and 8 of Sec. 2.3 and shown in Fig. 2.24.



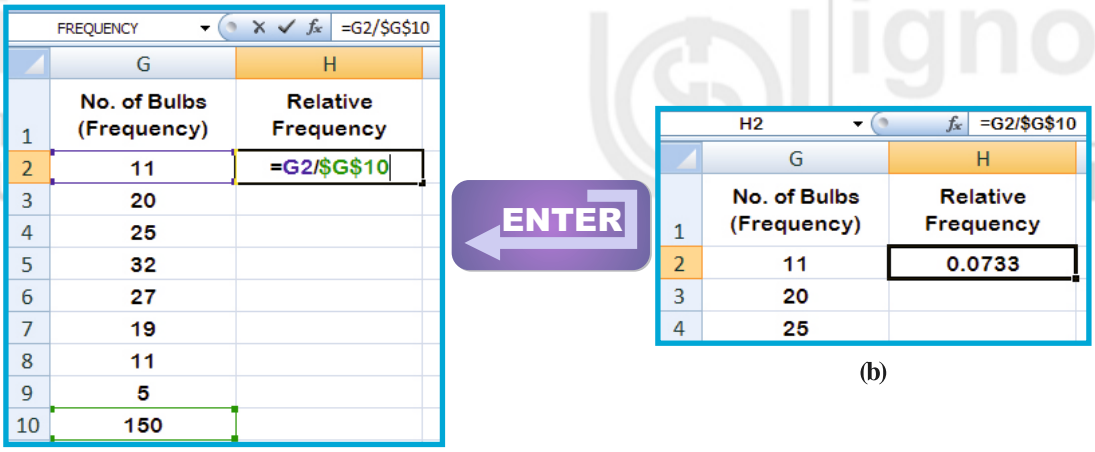

(a) (b)

Fig. 2.24

The formula with dollar sign ($) is used for an absolute references.

The values in Columns D and G taken together (Figs. 2.23c and 2.24b) form the **frequency distribution of bulbs** by the **exclusive method**. The values in columns E and G (Figs. 2.23c and 2.24b) taken together form the **Frequency distribution of bulbs** by the **inclusive method**.

**Step 10:** To prepare the relative frequency distribution, we type "Relative Frequency" in Cell H1 and "=G2/$G$10" in Cell H2 and press ***Enter*** as shown in Figs. 2.25a and b.




(a) (b)

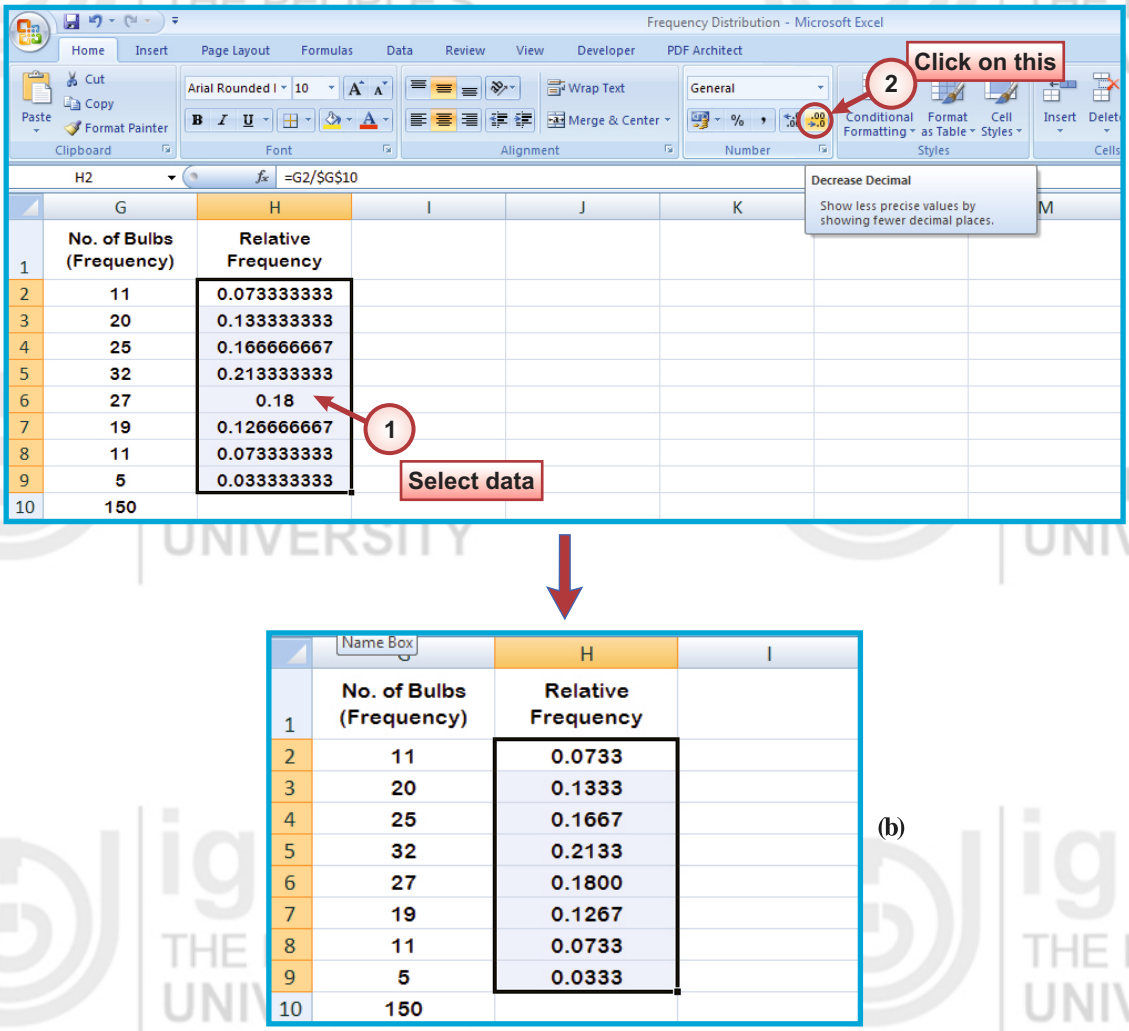
Fig. 2.25

**Step 11:** We can now determine the relative frequencies for the remaining class intervals as explained in Step 6 of Sec. 2.3 (Fig. 2.26).

| | G | H |
|---|---|---|
| 1 | No. of Bulbs (Frequency) | Relative Frequency |
| 2 | 11 | 0.073333333 |
| 3 | 20 | 0.133333333 |
| 4 | 25 | 0.166666667 |
| 5 | 32 | 0.213333333 |
| 6 | 27 | 0.18 |
| 7 | 19 | 0.126666667 |
| 8 | 11 | 0.073333333 |
| 9 | 5 | 0.033333333 |
| 10 | 150 | |

Fig. 2.26

**Step 12:** We can also decrease the number of decimal places by clicking on ***Decrease Decimal*** option as shown in Fig. 2.27a. Here we decrease it up to 4 decimal places as shown in Fig. 2.27b.




(a)

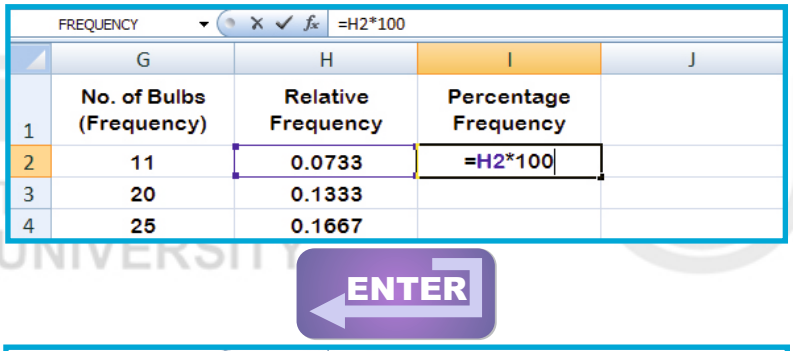
| | G | H | I | J | K |
|---|---|---|---|---|---|
| 1 | No. of Bulbs (Frequency) | Relative Frequency | | | |
| 2 | 11 | 0.073333333 | | | |
| 3 | 20 | 0.133333333 | | | |
| 4 | 25 | 0.166666667 | | | |
| 5 | 32 | 0.213333333 | | | |
| 6 | 27 | 0.18 | | | |
| 7 | 19 | 0.126666667 | | | |
| 8 | 11 | 0.073333333 | | | |
| 9 | 5 | 0.033333333 | | | |
| 10 | 150 | | | | |

(b)

| | G | H | I |
|---|---|---|---|
| 1 | No. of Bulbs (Frequency) | Relative Frequency | |
| 2 | 11 | 0.0733 | |
| 3 | 20 | 0.1333 | |
| 4 | 25 | 0.1667 | |
| 5 | 32 | 0.2133 | |
| 6 | 27 | 0.1800 | |
| 7 | 19 | 0.1267 | |
| 8 | 11 | 0.0733 | |
| 9 | 5 | 0.0333 | |
| 10 | 150 | | |

Fig. 2.27

**Step 13:** To compute the percentage frequency, we type "Percentage Frequency" in Cell I1 and "=H2*100" in Cell I2 as shown in Fig. 2.28a. When we press ***Enter***, we get the result (Fig. 2.28b).




(a)

| | G | H | I | J |
|---|---|---|---|---|
| 1 | No. of Bulbs (Frequency) | Relative Frequency | Percentage Frequency | |
| 2 | 11 | 0.0733 | =H2*100 | |
| 3 | 20 | 0.1333 | | |
| 4 | 25 | 0.1667 | | |

(b)

| | G | H | I | J |
|---|---|---|---|---|
| 1 | No. of Bulbs (Frequency) | Relative Frequency | Percentage Frequency | |
| 2 | 11 | 0.0733 | 7.3333 | |
| 3 | 20 | 0.1333 | | |
| 4 | 25 | 0.1667 | | |

Fig. 2.28

**Step 14:** We determine the remaining percentage frequencies as explained in Step 6 of Sec. 2.3. The result is shown in Fig. 2.29.

| | G | H | I |
|---|---|---|---|
| 1 | No. of Bulbs (Frequency) | Relative Frequency | Percentage Frequency |
| 2 | 11 | 0.0733 | 7.3333 |
| 3 | 20 | 0.1333 | 13.3333 |
| 4 | 25 | 0.1667 | 16.6667 |
| 5 | 32 | 0.2133 | 21.3333 |
| 6 | 27 | 0.1800 | 18.0000 |
| 7 | 19 | 0.1267 | 12.6667 |
| 8 | 11 | 0.0733 | 7.3333 |
| 9 | 5 | 0.0333 | 3.3333 |
| 10 | 150 | | |

**Fig. 2.29**

**Step 15:** We can compute the total of relative and percentage frequencies in Cells H10 and I10, respectively, as explained in Steps 7 and 8 in Sec. 2.3 (Fig. 2.30).

| | G | H | I |
|---|---|---|---|
| 1 | No. of Bulbs (Frequency) | Relative Frequency | Percentage Frequency |
| 2 | 11 | 0.0733 | 7.3333 |
| 3 | 20 | 0.1333 | 13.3333 |
| 4 | 25 | 0.1667 | 16.6667 |
| 5 | 32 | 0.2133 | 21.3333 |
| 6 | 27 | 0.1800 | 18.0000 |
| 7 | 19 | 0.1267 | 12.6667 |
| 8 | 11 | 0.0733 | 7.3333 |
| 9 | 5 | 0.0333 | 3.3333 |
| 10 | 150 | 1 | 100 |
| 11 | | | |

**Fig 2.30**

The class intervals along with the relative frequencies and percentage frequencies given in Columns H and I, constitute the **relative** and **percentage frequency distributions**, respectively.

**Step 16:** We type "No. of Bulbs (Less Than Cumulative Frequency)" in Cell J1. Since the first **less than cumulative frequency** is the same as the first frequency given in Cell G2, we type "=G2" in Cell J2 as shown in Fig. 2.31.

The less than cumulative frequency distribution corresponds to the upper limit of the exclusive type class intervals given in Cells D2:D9.

J2 $f_x$ =G2

| | J | K |
|---|---|---|
| 1 | No. of Bulbs (Less Than Cumulative Frequency) | |
| 2 | 11 | |
| 3 | | |
| 4 | | |
| 5 | | |
| 6 | | |

**Fig. 2.31**

**Step 17:** To compute the second frequency, we type "=G3+J2" in Cell J3 (Fig. 2.32a) and drag it down up to Cell J9 as shown in Fig. 2.32b.

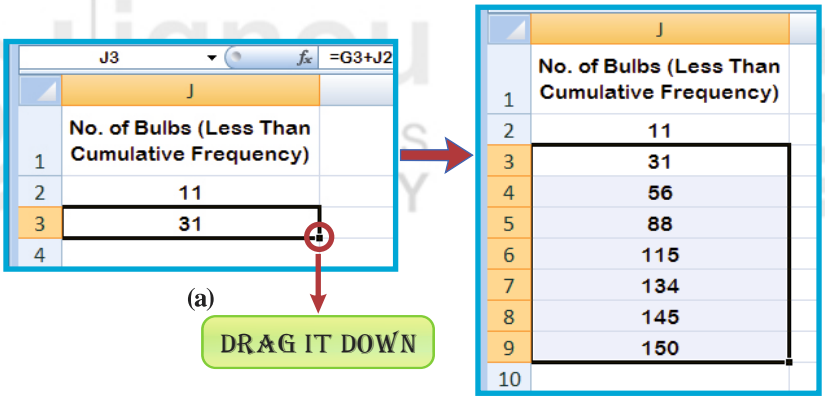



Fig. 2.32

**Step 18:** We type "No. of Bulbs (More Than Cumulative Frequency)" in Cell K1. The first **more than type cumulative frequency** is equal to the total frequency given in Cell G10. So we type "=G10" in Cell K2 (Fig. 2.33)

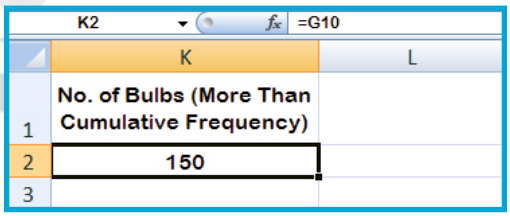



Fig. 2.33

**Step 19:** To compute the second more than type cumulative frequency, we type "=K2-G2" in Cell K3 (Fig. 2.34a) and drag it down up to Cell K9 as shown in Fig. 2.34b.

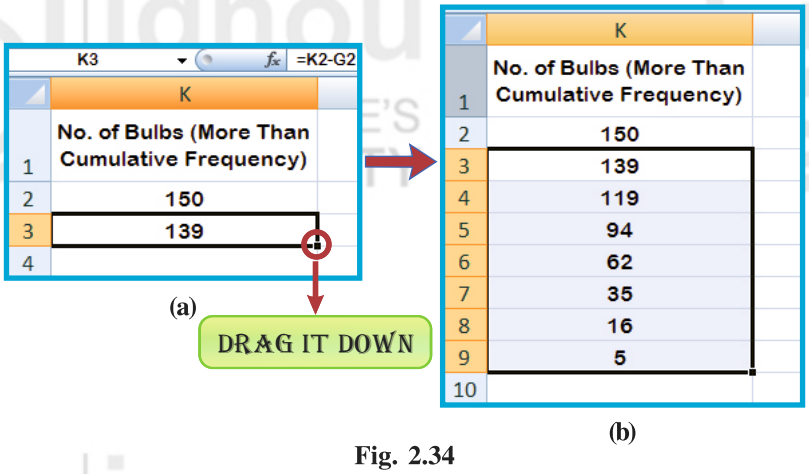



Fig. 2.34

You should now apply this method to other problems for practice.

## Activity

Classify the data suitably with the help of MS Excel 2007 and interpret the results for

**A1)** Example 1 given in Unit 13 of MST-001, and

**A2)** Exercises E3 to E6 given in Unit 13 of MST-001.

Match the results with the manual calculation done in Unit 13 of MST-001.

## Continuous Assessment 2

1. A survey was conducted on the preference of cooking oil in a family. Data were collected for 100 families. Their preferences are recorded in Table 4. Classify the data recorded in Table 4.

**Table 4: Choice of families regarding cooking oils**

| S. No. | Cooking Oil | S. No. | Cooking Oil | S. No. | Cooking Oil |
|---|---|---|---|---|---|
| 1 | Olive | 35 | Ghee | 69 | Cottonseed |
| 2 | Ghee | 36 | Cottonseed | 70 | Others |
| 3 | Mustard | 37 | Mustard | 71 | Mustard |
| 4 | Soybean | 38 | Olive | 72 | Cottonseed |
| 5 | Soybean | 39 | Mustard | 73 | Others |
| 6 | Soybean | 40 | Olive | 74 | Soybean |
| 7 | Mustard | 41 | Ghee | 75 | Ghee |
| 8 | Soybean | 42 | Olive | 76 | Olive |
| 9 | Mustard | 43 | Others | 77 | Soybean |
| 10 | Ghee | 44 | Sunflower | 78 | Sunflower |
| 11 | Others | 45 | Cottonseed | 79 | Cottonseed |
| 12 | Ghee | 46 | Sunflower | 80 | Soybean |
| 13 | Cottonseed | 47 | Sunflower | 81 | Sunflower |
| 14 | Mustard | 48 | Mustard | 82 | Ghee |
| 15 | Others | 49 | Cottonseed | 83 | Sunflower |
| 16 | Ghee | 50 | Cottonseed | 84 | Ghee |
| 17 | Sunflower | 51 | Sunflower | 85 | Sunflower |
| 18 | Olive | 52 | Sunflower | 86 | Soybean |
| 19 | Sunflower | 53 | Ghee | 87 | Cottonseed |
| 20 | Soybean | 54 | Sunflower | 88 | Olive |
| 21 | Mustard | 55 | Soybean | 89 | Cottonseed |
| 22 | Mustard | 56 | Olive | 90 | Sunflower |
| 23 | Mustard | 57 | Soybean | 91 | Cottonseed |
| 24 | Mustard | 58 | Sunflower | 92 | Mustard |
| 25 | Cottonseed | 59 | Sunflower | 93 | Mustard |
| 26 | Mustard | 60 | Sunflower | 94 | Cottonseed |
| 27 | Soybean | 61 | Olive | 95 | Ghee |
| 28 | Olive | 62 | Cottonseed | 96 | Cottonseed |
| 29 | Cottonseed | 63 | Sunflower | 97 | Mustard |
| 30 | Olive | 64 | Soybean | 98 | Cottonseed |
| 31 | Soybean | 65 | Sunflower | 99 | Sunflower |
| 32 | Cottonseed | 66 | Soybean | 100 | Others |
| 33 | Others | 67 | Mustard | | |
| 34 | Mustard | 68 | Ghee | | |

2. Data of egg consumption in 120 families were collected. The number of eggs consumed in each family is recorded in Table 5. Construct a discrete frequency distribution for the data.

**Table 5: Consumption of eggs per family**

| Family No. | Consumption of Eggs | Family No. | Consumption of Eggs | Family No. | Consumption of Eggs |
|---|---|---|---|---|---|
| 1 | 7 | 41 | 7 | 81 | 3 |
| 2 | 5 | 42 | 5 | 82 | 7 |
| 3 | 1 | 43 | 6 | 83 | 6 |
| 4 | 9 | 44 | 2 | 84 | 8 |
| 5 | 6 | 45 | 5 | 85 | 2 |
| 6 | 6 | 46 | 4 | 86 | 5 |
| 7 | 5 | 47 | 6 | 87 | 2 |
| 8 | 6 | 48 | 7 | 88 | 5 |
| 9 | 9 | 49 | 4 | 89 | 9 |
| 10 | 6 | 50 | 4 | 90 | 8 |
| 11 | 7 | 51 | 7 | 91 | 3 |
| 12 | 6 | 52 | 5 | 92 | 5 |
| 13 | 1 | 53 | 4 | 93 | 4 |
| 14 | 6 | 54 | 8 | 94 | 2 |
| 15 | 6 | 55 | 8 | 95 | 5 |
| 16 | 3 | 56 | 6 | 96 | 2 |
| 17 | 9 | 57 | 3 | 97 | 5 |
| 18 | 8 | 58 | 6 | 98 | 5 |
| 19 | 6 | 59 | 3 | 99 | 2 |
| 20 | 4 | 60 | 5 | 100 | 7 |
| 21 | 8 | 61 | 4 | 101 | 1 |
| 22 | 7 | 62 | 3 | 102 | 10 |
| 23 | 9 | 63 | 6 | 103 | 7 |
| 24 | 5 | 64 | 4 | 104 | 4 |
| 25 | 7 | 65 | 2 | 105 | 4 |
| 26 | 3 | 66 | 5 | 106 | 10 |
| 27 | 7 | 67 | 2 | 107 | 7 |
| 28 | 9 | 68 | 7 | 108 | 5 |
| 29 | 6 | 69 | 5 | 109 | 10 |
| 30 | 6 | 70 | 4 | 110 | 4 |
| 31 | 2 | 71 | 4 | 111 | 8 |
| 32 | 8 | 72 | 4 | 112 | 8 |
| 33 | 3 | 73 | 5 | 113 | 6 |
| 34 | 8 | 74 | 3 | 114 | 3 |
| 35 | 5 | 75 | 7 | 115 | 6 |
| 36 | 6 | 76 | 7 | 116 | 3 |
| 37 | 8 | 77 | 4 | 117 | 5 |
| 38 | 3 | 78 | 4 | 118 | 4 |
| 39 | 9 | 79 | 5 | 119 | 3 |
| 40 | 6 | 80 | 5 | 120 | 10 |

3. Data of rainfall were collected to study the rainfall patterns in 120 different areas of a state and are given in Table 6.

**Table 6: Rainfall data**

| Rainfall (in mm) | Rainfall (in mm) | Rainfall (in mm) | Rainfall (in mm) | Rainfall (in mm) |
|---|---|---|---|---|
| 187.2 | 230.7 | 180.8 | 251.1 | 195.6 |
| 207.9 | 186.3 | 208.4 | 207.6 | 221.8 |
| 171.5 | 229.2 | 240.6 | 211.9 | 206.2 |
| 217.7 | 230.1 | 208.2 | 187.2 | 235.9 |
| 199.5 | 220.1 | 187.2 | 207.9 | 218.5 |
| 185.7 | 189.4 | 234.7 | 171.5 | 168.5 |
| 268.8 | 188.4 | 206.4 | 217.7 | 183.6 |
| 201.5 | 179.1 | 227.8 | 199.5 | 213.0 |
| 173.1 | 199.2 | 188.4 | 185.7 | 180.8 |
| 192.7 | 228.8 | 226.3 | 268.8 | 208.4 |
| 157.3 | 190.7 | 178.7 | 201.5 | 240.6 |
| 244.4 | 251.1 | 228.8 | 173.1 | 208.2 |
| 182.2 | 207.6 | 182.6 | 192.7 | 187.2 |
| 241.5 | 211.9 | 230.7 | 157.3 | 234.7 |
| 232.4 | 228.3 | 186.3 | 244.4 | 206.4 |
| 182.4 | 204.4 | 229.2 | 182.2 | 227.8 |
| 182.7 | 195.6 | 230.4 | 241.5 | 228.3 |
| 151.2 | 221.8 | 220.1 | 232.4 | 204.4 |
| 191.6 | 206.6 | 189.4 | 182.4 | 195.6 |
| 188.4 | 235.9 | 188.4 | 182.7 | 221.8 |
| 226.3 | 218.5 | 179.1 | 151.2 | 206.4 |
| 178.7 | 168.4 | 199.2 | 191.6 | 235.9 |
| 228.8 | 183.6 | 228.8 | 228.3 | 218.5 |
| 182.6 | 213.2 | 190.4 | 204.4 | 168.7 |

For this data,

- compute the suitable width for class intervals.
- construct continuous, relative, percentage and cumulative frequency distributions.

## Home Work: Do it Yourself

1) Follow the steps explained in Secs. 2.3, 2.4 and 2.5 to classify the data of Tables 1, 2 and 3. Take the final screenshots and keep them in your record book.

2) Develop the spreadsheets for the exercises of "Continuous Assessment 2" as explained in this lab session. Take screenshots of the final spreadsheets.

3) **Do not forget** to keep all screenshots in your record book as these will contribute to your continuous assessment in the Laboratory.