PART A

Representation of Data

REPRESENTATION OF DATA

The first part of this lab course comprises Sessions 1 to 5 which will enable you to develop practical skills for classifying raw data as well as constructing diagrams or graphs in MS Excel 2007.

In Block 4 of MST-001 (Foundation in Mathematics and Statistics), you have learnt that it is difficult to comprehend and get information from raw data. So we need to classify raw data in tabular form to acquire information in much shorter period of time which helps in summarising the raw data. You have also learnt how to construct some commonly used diagrams for representing tabulated data, graphs for frequency distributions and time series data.

Session 1 serves as an introductory overview of Excel 2007. In this session, you will learn how to work with Excel 2007 by creating and formatting the spreadsheets. We shall explain how to use charts, built-in functions and the ***Data Analysis ToolPak*** in Excel 2007.

In **Session 2**, you will learn how to classify qualitative data on the basis of qualitative characteristics, i.e., attributes such as sex, region, religion, etc., using Excel 2007. This type of classification is known as qualitative classification. We shall also explain how to classify quantitative data by constructing discrete and continuous frequency distributions using Excel 2007 for discrete and continuous data, respectively.

We have explained that we can also represent tabulated data with the help of different kinds of commonly used diagrams. **Session 3** deals with the commonly used diagrams such as the simple bar diagram, multiple bar diagram, sub-divided bar diagram and pie diagram in Excel 2007 while the graphs used to represent the frequency distributions are discussed in **Session 4**.

In **Session 5**, we demonstrate the construction of graphs for time series data, such as line, range and band graphs using Excel 2007. We shall also consider the construction of box plot.

All sessions of this part have been designed in such a way that after successfully completing them, you will be able to classify qualitative as well as quantitative data, represent data diagrammatically or graphically using MS Excel 2007.

Before studying the lab sessions of Part A, we expect that you have studied Block 4 of the course MST-001 entitled Foundation in Mathematics and Statistics thoroughly. All concepts and formulae used there will be applied here.

Introduction to MS Excel 2007

SESSION 1

Structure

1.1 Introduction
Objectives
1.2 How to Start Excel 2007
1.3 Basics of Excel 2007
1.4 The Ribbon
1.5 Working with Excel 2007
1.6 Entering Data in an Excel Sheet
Typing Data in a New Excel Worksheet
Copying Data from an Existing Excel Worksheet
Creating New Data by Transforming the Existing Data
Copying a Formula in Different Cells
1.7 Working with Formula
Mathematical Operations
Writing a Formula
1.8 Editing and Formatting the Worksheet
1.9 Data Analysis ToolPak

1.1 INTRODUCTION

The Microsoft Excel is a part of the **Microsoft Office** suite. It is the most widely used spreadsheet application for organising and analysing data, creating tables and charts, etc. Microsoft Excel is also commonly used to perform different types of complicated numerical calculations for the given data.

This spreadsheet application is used to store information in columns and rows, which can then be organised and processed. Spreadsheets are designed to work well with numbers but often include text. Excel organises your work into workbooks and each workbook can contain many worksheets. These worksheets are used to list and analyse data. Excel is generally available on all public access PCs (e.g., in the Library and PC Labs).

In this lab session, we introduce MS Excel 2007. We also explore the basic concepts, functions, editing and formatting, etc., in Excel which are useful for this lab course. You will study the applications of various statistical tools using Excel 2007 in the subsequent lab sessions.

Objectives

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

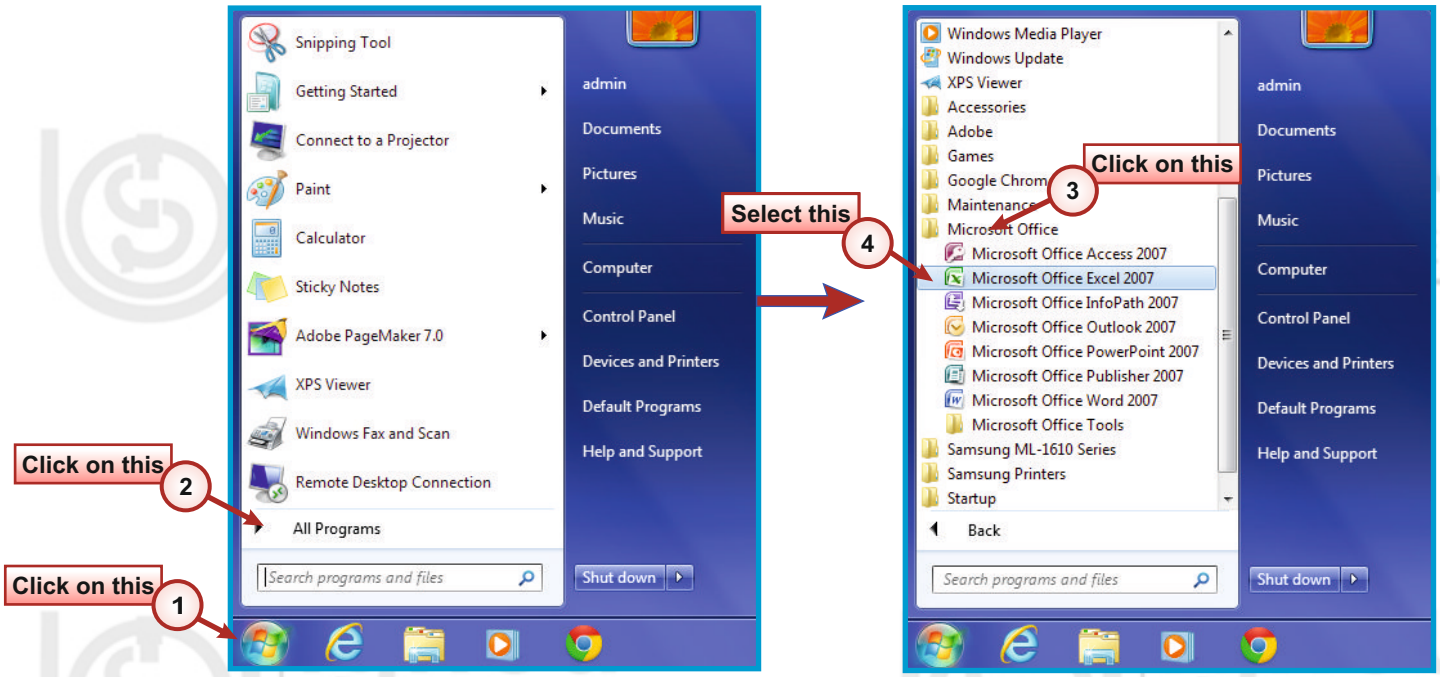
- start Excel 2007 and enter data in the worksheet;
- create and type the formulae;

- edit and format the worksheet;
- develop the spreadsheet in MS Excel 2007; and
- work with MS Excel 2007.

1.2 HOW TO START EXCEL 2007

The Excel program is located in the **Microsoft Office** folder. To use MS Excel, we first start the computer. We can open MS Excel by double clicking on the MS Excel icon (shortcut) which is on the desktop or using the windows button. If the PC has Windows 7, we open MS Excel using the Windows button by selecting ***Windows*** → ***All programs*** → ***Microsoft Office*** → ***Microsoft Office Excel 2007***. As shown in Fig. 1.1, we

1. click on the ***Windows*** icon ( ) (which is at the lower left corner of the monitor screen) as shown in Fig. 1.1a,
2. point or click on ***All Programs*** which displays the list of all programs,
3. click on the ***Microsoft Office*** folder from the sub-menu (which displays the list of ***Microsoft Office*** applications) as shown in Fig. 1.1b, and
4. click on the ***Microsoft Office Excel 2007*** under the ***Microsoft Office*** folder (Fig. 1.1b). The Microsoft Excel window shown in Fig. 1.3 appears on the monitor screen.



(a) (b)

Fig. 1.1

If the shortcut icon does not appear on the desktop, we can also create a shortcut of MS Excel on the desktop. As shown in Figs. 1.1a, b and 1.2a, we

1. click on the ***Windows*** icon (Fig. 1.1a),
2. click on ***All Programs*** (Fig. 1.1a),
3. click on the ***Microsoft Office*** folder (Fig. 1.1b),
4. right-click on the ***Microsoft Office Excel 2007*** (Fig. 1.2a), and
5. click on the ***Desktop (create shortcut)*** option under ***Send to*** menu (Fig. 1.2a). The Excel shortcut shown in Fig. 1.2b appears on the desktop.

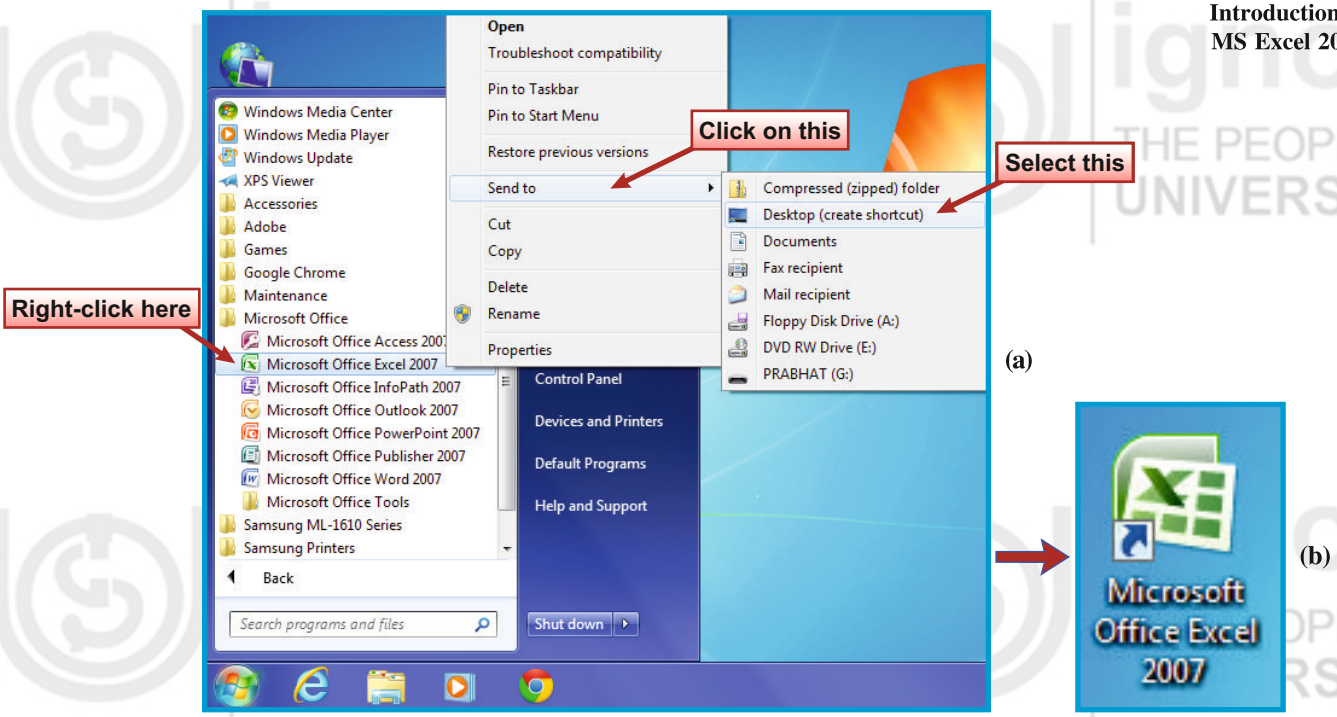


Fig. 1.2

1.3 BASICS OF EXCEL 2007

When we start Excel, the Excel spreadsheet appears within a few moments on our monitor screen. It looks similar to Fig. 1.3. Before we start working on the Excel spreadsheet, we briefly discuss various parts of the Excel spreadsheet shown in Fig. 1.3.

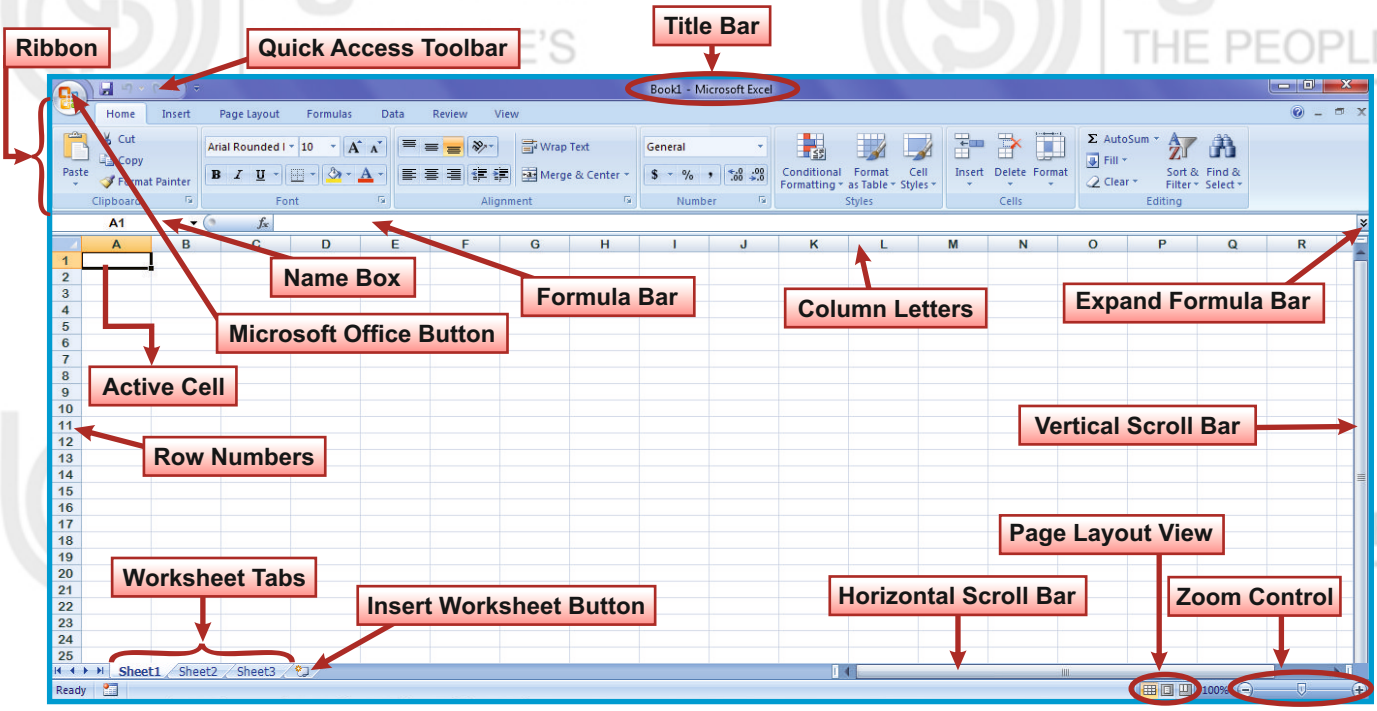


Fig. 1.3: Various parts of the Excel spreadsheet.

1. Microsoft Office Button

The ***Microsoft Office*** sign at the upper-left corner (Fig. 1.3) of the Excel window is called the ***Microsoft Office Button***. It is basically used for handling files. A close-up view of the ***Microsoft Office Button*** is shown in Fig. 1.4a.

When we click on this button, it provides several options such as ***New***, ***Save***, ***Print***, ***Close***, etc., as shown in Fig. 1.4b.

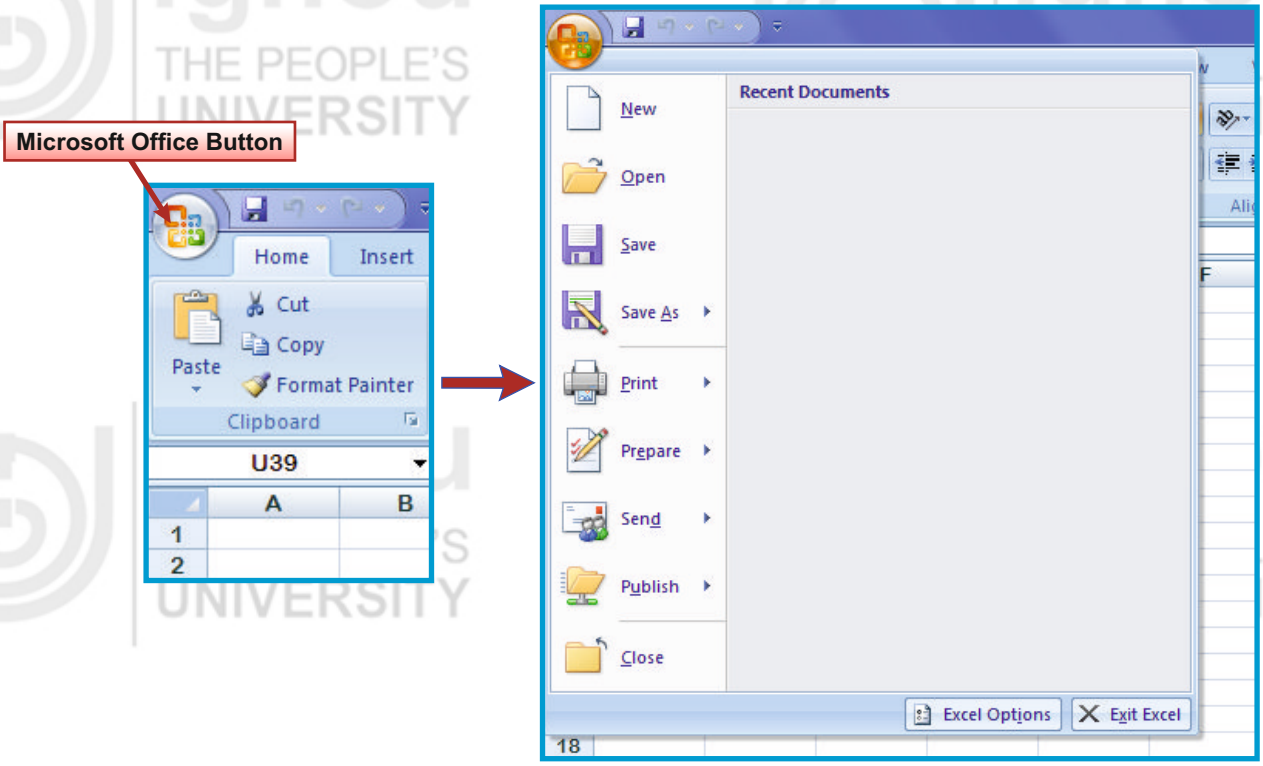


Fig. 1.4

2. Workbook

By default, each new workbook contains three sheets.

In Fig. 1.5, **Sheet 1** is an active sheet.

Notice from Fig. 1.3 that the title bar is labelled "**Book 1 Microsoft Excel**". This MS Excel document is called a **workbook**. Each workbook comprises one or more **worksheets**. By default, it contains three worksheets (Fig. 1.5).

3. Worksheet

The lower left corner of the Excel window (Fig. 1.5) initially has three worksheets labelled **Sheet 1**, **Sheet 2**, and **Sheet 3**, which are called the ***Worksheet*** or ***Sheet*** tabs. We can say that a workbook is like a file into which we can place as many worksheets as memory allows. By clicking on these ***Sheet*** tabs with the mouse, we can display the contents of that particular worksheet. The current worksheet is called the **Active sheet**. Each worksheet is made up of multiple grids of cells in the form of rows and columns.

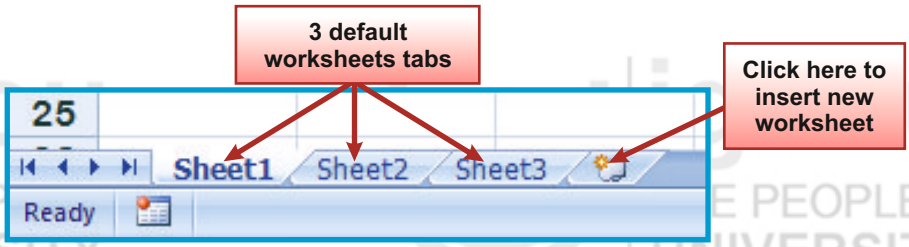


Fig. 1.5

4. Columns

We have **alphabetically** titled columns ranging from **A** to **XFD**, i.e., there are a total of **16,384** columns in a worksheet. We can select the entire column by clicking on the column title.

5. Rows

We have **numerically** titled rows ranging from **1** to **1,048,576**, i.e., there are a total of **1,048,576** rows in a worksheet. We can select the entire row by clicking on the row title.

6. Cells

Cells are the basic rectangular building blocks of a spreadsheet. They are assigned an address, generally referred to as a cell reference, according to their column and row labels (e.g., the cell in Column H at Row 11 is referenced as Cell H11). In this way, we have a total of **17,179,869,184** cells in an Excel sheet. Each cell can contain a **label**, **value** or **formula**. Labels can contain any combination of letters, numbers, or symbols. Values are the numbers and only values (numbers) can be used in calculations. A value can also be a date or time.

A group of cells is called **range**. We write a range address by specifying its upper left cell and its lower right cell, separated by a colon. For example, B10:B20 is used to refer to the range of 11 cells in Column B from Rows 10 to 20.

7. Active Cell

A cell having dark border (outline) is called the **active cell**. In Fig. 1.6, Cell A4 is the active cell. A small black square called the **fill handle** is displayed at the bottom right corner of the active cell.

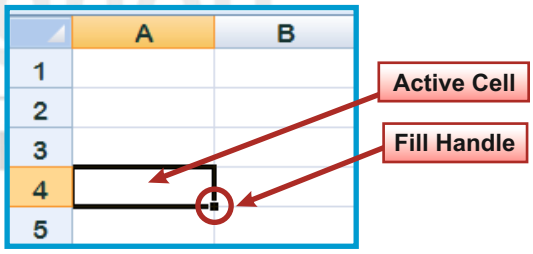


Fig. 1.6

8. Excel Close Button

We can close the Excel application window by clicking on the ***Excel Close*** button as shown in Fig. 1.7.

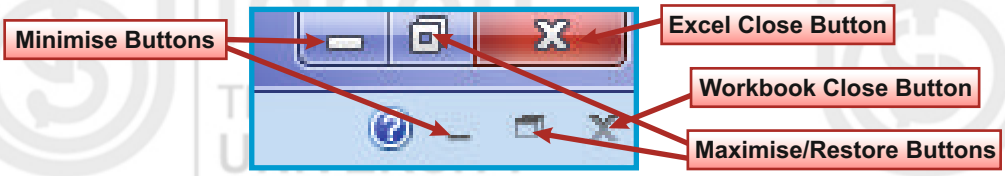


Fig. 1.7

9. Workbook Close Button

We can close the active workbook by clicking on the ***Workbook Close*** button as shown in Fig. 1.7.

10. Minimise and Maximise/Restore Buttons

We can minimise Excel/Workbook by clicking on the ***Excel/Workbook Minimise*** button. We use ***Excel/Workbook Maximise/Restore*** button to maximise or restore (unmaximise) the Excel/Workbook size (Fig. 1.7).

11. Ribbon

It contains the commands in Excel and has three parts: tabs, groups and commands. We shall elaborate on it in Sec. 1.4.

12. Horizontal and Vertical Scrollbars

We can scroll the sheet horizontally and vertically, respectively, using the ***Horizontal and Vertical Scrollbars***.

13. Quick Access Toolbar

We can customise the most commonly used commands on the quick access toolbar as shown in Fig. 1.8.

When we have committed an error, we can reverse the action using the ***Undo*** command or by pressing ***Ctrl + Z*** keys. We can reverse the last 100 actions which we have performed using the ***Undo*** command before closing the file. The ***Redo*** command or ***Ctrl + Y*** keys perform in the reverse direction of the ***Undo*** command. It restores the actions for which we have used the ***Undo*** command.

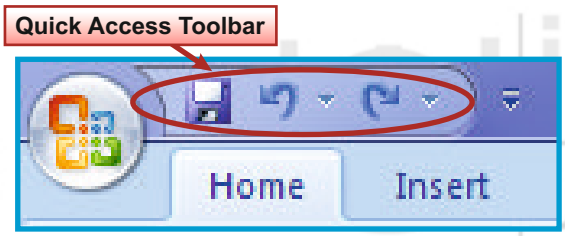


Fig. 1.8

In Fig. 1.8, the Quick Access Toolbar consists of ***Save File*** ( ), ***Undo*** ( ) and ***Redo*** ( ) icons.

14. Formula Bar

When we enter or type a formula and other information in any cell, it appears in the formula bar. For example, if we type **Statistics** in any cell, say, Cell B2, it also appears in the formula bar (Fig. 1.9).

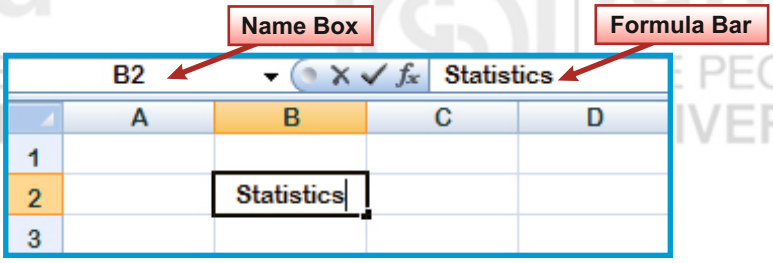


Fig. 1.9

15. Name Box

The name box is on the left side of the formula bar. If we click on a cell, we can see that its name appears on the name box. If we click on a different cell, the name will change accordingly (Fig. 1.9).

1.4 THE RIBBON

The Ribbon is a panel that contains eight tabs. Each tab is related to a specific task and contains several groups of related commands, menus or lists. Note that some of these groups have an arrow ( ) in the bottom right corner of that group. We get a dialog box which contains all commands of that group when we click on it. We briefly elaborate on the look of the ribbon for all eight tabs, groups within each tab and commands within the groups as follows:

1. Home Tab

The home tab contains the following groups within the tab and commands within the groups (Fig. 1.10):

i) Clipboard (cut, copy, paste, etc.)

ii) Font (font type, font size, etc.)

iii) Alignment (left, right, justified, etc.)

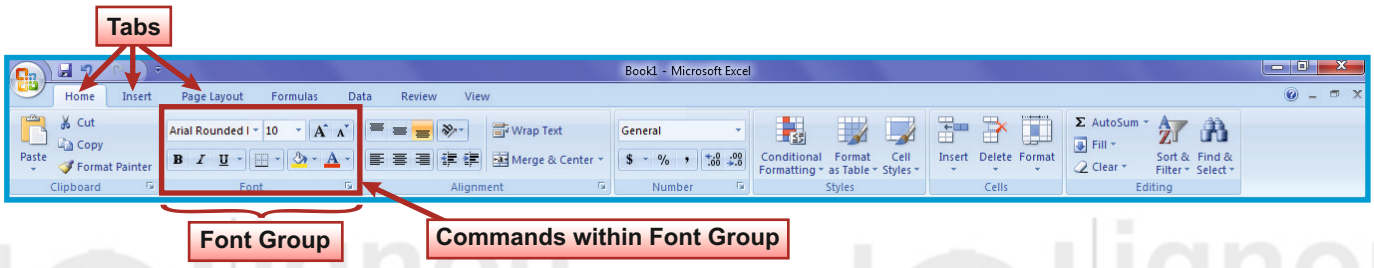


Fig. 1.10

iv) Number ($, %, decimal, etc.)

v) Styles (format tables, cells, etc.)

vi) Cells (insert, delete cell(s), rows or columns, etc.)

vii) Editing (fill, find & select, etc.)

Remember, the look of the ribbon may vary with the screen resolution.

2. Insert Tab

It is used to create tables, charts, etc., and contains the following groups within the tab and commands within the groups (Fig. 1.11):

i) Tables (pivot table and tables)

ii) Illustrations (picture, clip art, etc.)

iii) Charts (column, line, pie, scatter, etc.)

iv) Links (hyperlink)

v) Text (text box, object, symbol, wordart, etc.)

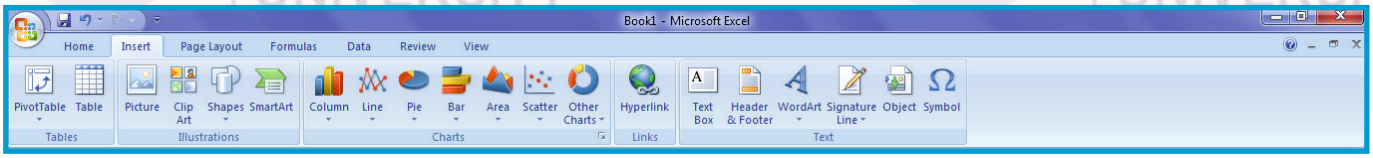


Fig. 1.11

3. Page Layout Tab

It is used for setting the page layout and contains the following groups within the tab and commands within the groups (Fig. 1.12):

i) Themes (colors, fonts, etc.)

ii) Page Setup (margins, orientation, background, etc.)

iii) Scale to Fit (scale, height, etc.)

iv) Sheet Options (gridlines, heading, etc.)

v) Arrange (selection pane, align, etc.)

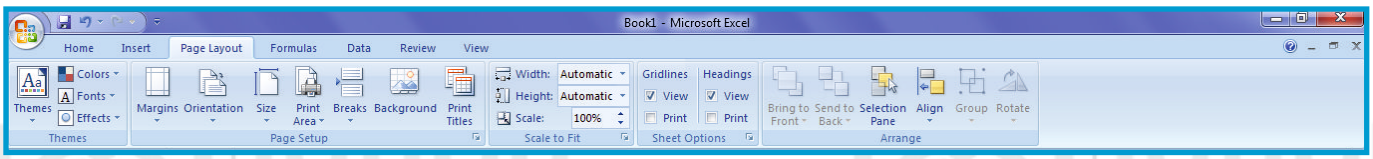


Fig. 1.12

4. Formulas Tab

It is used to apply a formula for numerical calculations and contains the following groups within the tab and commands within the groups (Fig. 1.13):

i) Function Library (insert function, autosum, math & trig, etc.)

ii) Defined Names (name manager, define name, etc.)

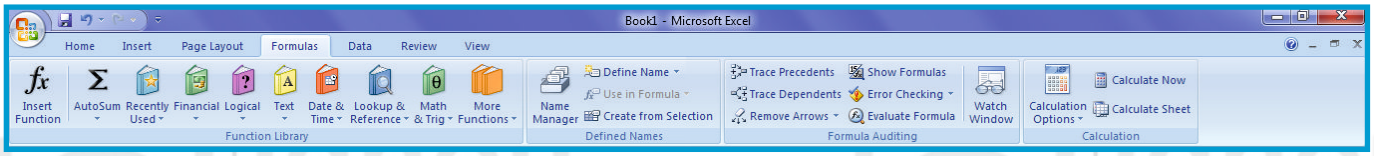


Fig. 1.13

iii) Formula auditing (show formulas, remove arrows, etc.)

iv) Calculation (calculation options, calculate now, etc.)

5. Data Tab

It contains the following data related groups within the tab and commands within the groups (Fig. 1.14):

i) Get External Data (from text, web, etc.)

ii) Connections (connections, refresh all, etc.)

iii) Sort & Filter (sort, filter, etc.)

iv) Data Tools (text to columns, data validation, etc.)

v) Outline (group, ungroup, etc.)

vi) Analysis (data analysis)

Note: If the ***Data Analysis*** command under ***Data*** tab (far right) does not appear on your computer, you need to activate it (see Sec.1.9).

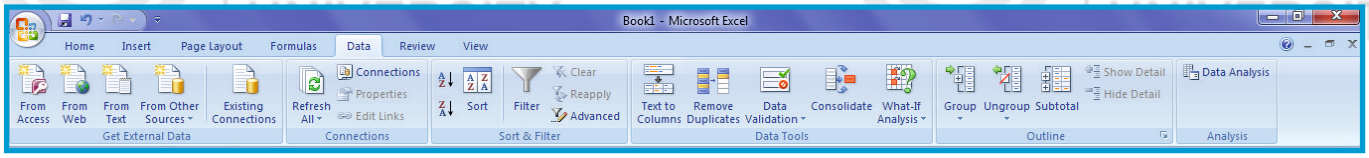


Fig. 1.14

6. Review Tab

This tab contains the following groups within the tab and commands within the groups (Fig.1.15):

i) Proofing (spelling, thesaurus, etc.)

ii) Comments (new comments, previous, etc.)

iii) Changes (protect sheet, protect workbook, etc.)

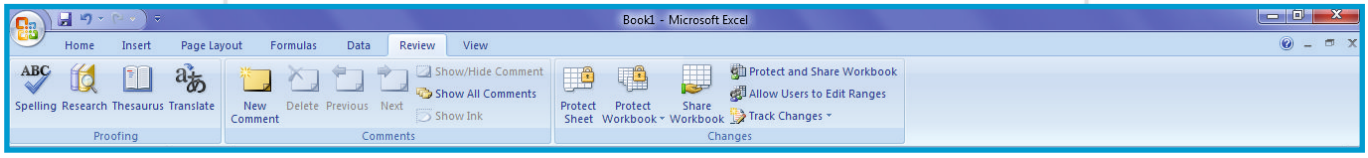


Fig. 1.15

7. View Tab

The view tab contains the following groups within the tab and commands within the groups that control various aspects of how a sheet is viewed (Fig.1.16):

i) Workbook Views (normal, page layout, etc.)

ii) Show/Hide (gridlines, ruler, etc.)

iii) Zoom (zoom, 100%, etc.)

iv) Window (new window, split, hide, etc.)

v) Macros (macros)

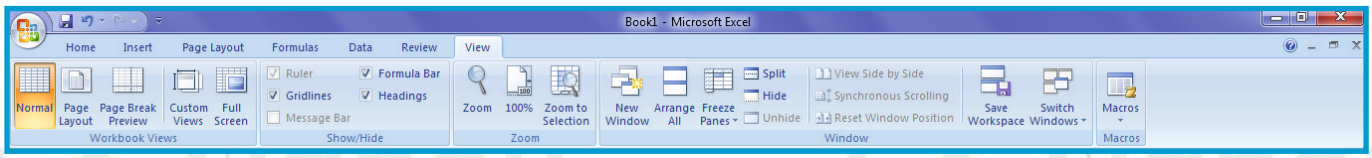


Fig. 1.16

8. Developer Tab

This tab is generally not visible by default. It contains commands that are useful for programmers (Fig. 1.17). To display the ***Developer*** tab, we click on ***Excel Options*** under ***Office button*** and then select ***Popular***. We check the box next to ***Show Developer Tab in the Ribbon***.

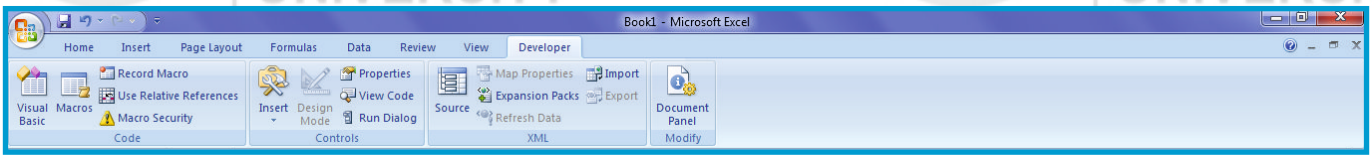
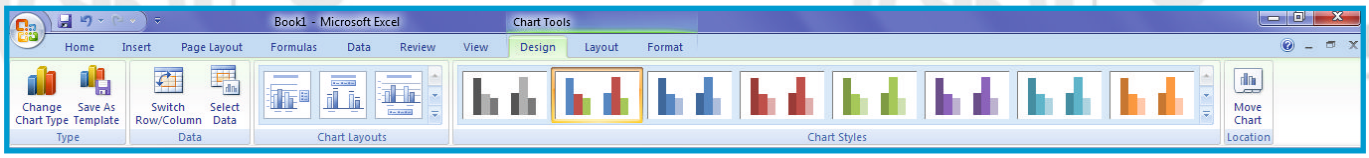
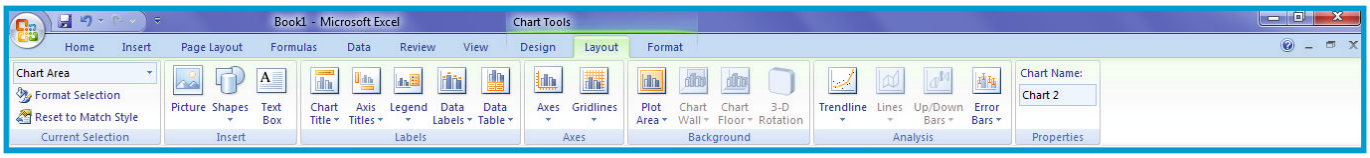


Fig. 1.17

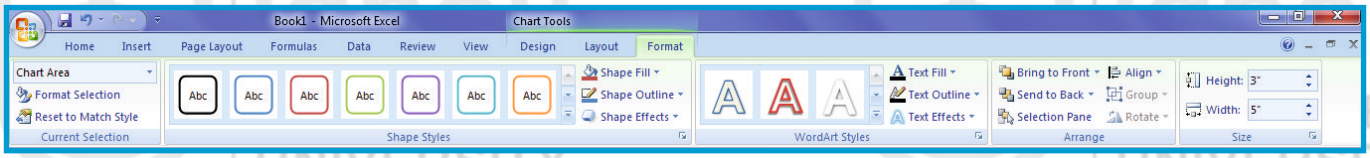
Besides these tabs, you will also come across three chart tool tabs: ***Design***, ***Layout*** and ***Format*** when you create any chart. These tabs are useful to deal with charts (Fig. 1.18). You can access these tabs by clicking on the chart you have created.



(a)



(b)



(c)

Fig. 1.18

1.5 WORKING WITH EXCEL 2007

While working with Excel 2007, you must keep the following points in mind:

1. File Format

The most common data formats used to store data in Excel files are as follows:

- Excel workbook with extension **.xlsx** means that the file is in Excel 2007 format. This may not be recognised by earlier versions of Excel such as Excel 97 or Excel 2003. We can also access an Excel file in lower version of Excel if we save it in a compatible mode.
- Excel workbook with extension **.xls** means that the file is in a version lower than Excel 2007.

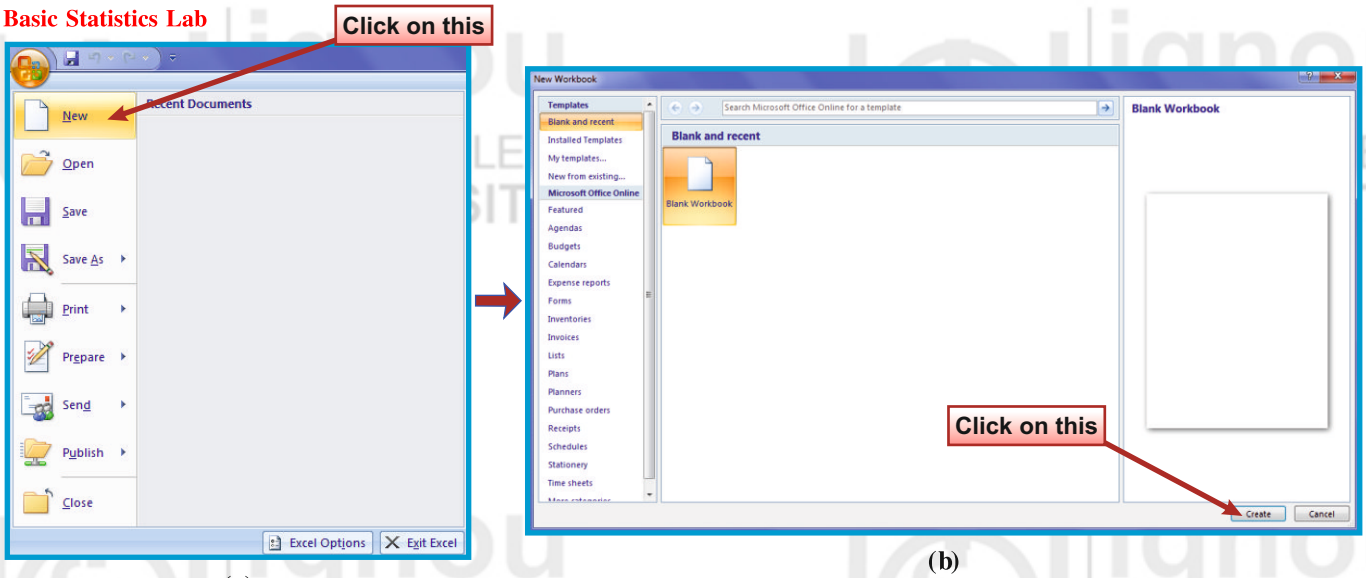
2. Creating a New File

When we start Excel application, we get a blank workbook (Fig. 1.3) by default. We can also open a new workbook whenever we desire. We can create a new workbook by clicking on the ***Office*** button and selecting ***New*** from the menus as shown in Fig. 1.19a. A new dialog box appears. We select ***Blank Workbook*** option and click on ***Create*** as shown in Fig. 1.19b. This gives us a new screen with a blank workbook area.

When we open Excel 2007, it creates a workbook, which contains 3 worksheets by default.

Open the workbook to start your work on Excel worksheet. To create a new workbook, click on ***File*** → ***New*** → ***Blank Document*** (or press ***Ctrl+N***).

Click on ***File*** → ***Open*** (or press ***Ctrl+O***) to open/ retrieve an existing workbook.

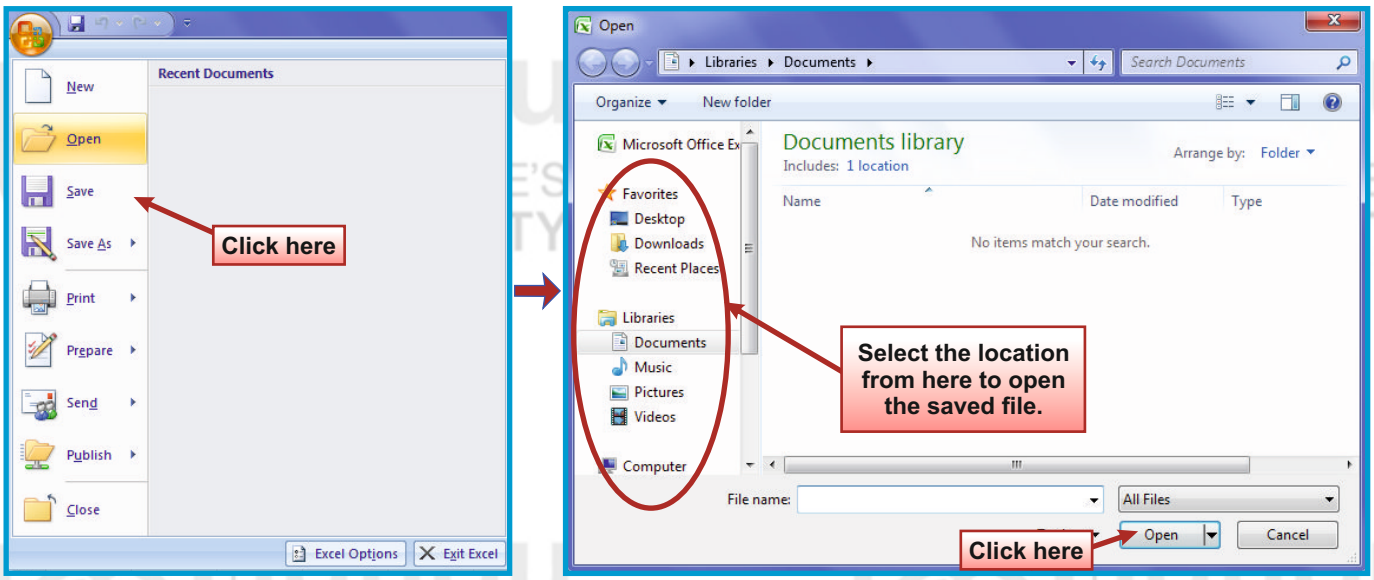


(a) (b)

Fig. 1.19

3. Opening an Existing File

If we want to use or continue working on a workbook that we have created earlier, we click on the ***Office*** button and then on the ***Open*** button as shown in Fig. 1.20a. A new dialog box appears (Fig. 1.20b). Then we select the path of the folder where we have saved the file, select the required file and click on the ***Open*** button to open the file. This will open the required file. We shall explain this option with an example in Sec. 1.6.



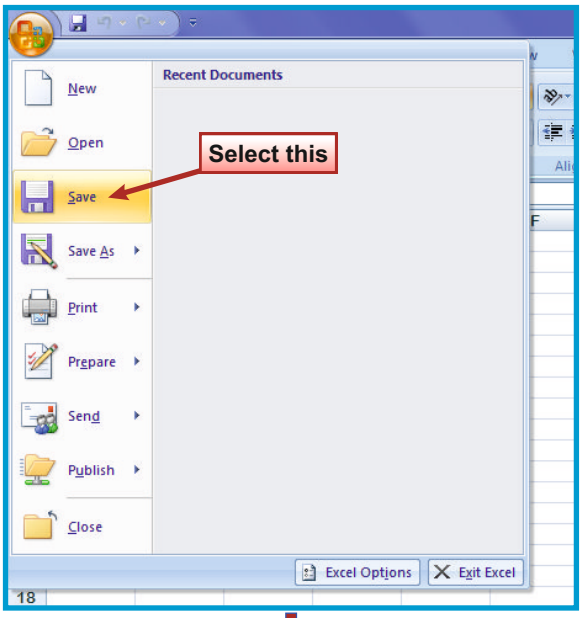
(a) (b)

Fig. 1.20

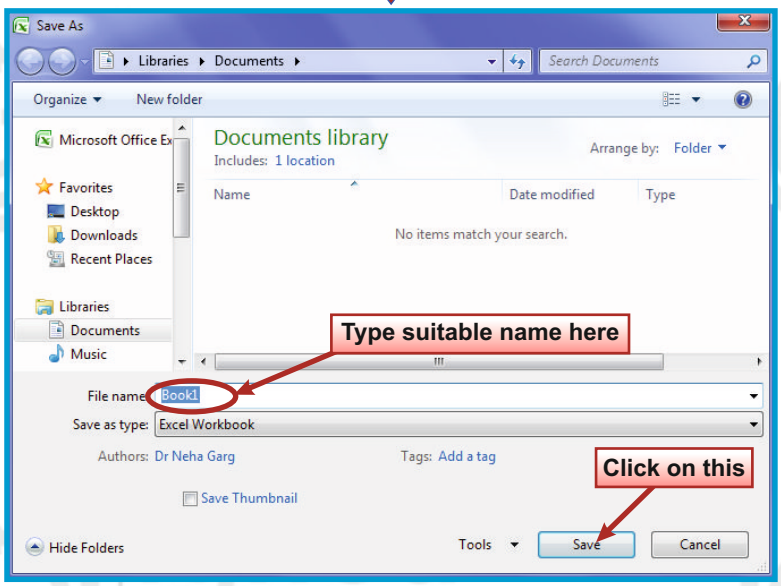
4. Saving a File

We need to make a secure copy of the workbook in the form of a file to save the work. We click the ***Office*** button and select the ***Save*** option as shown in Fig. 1.21a. We can also click on the ***Save*** button (Fig. 1.8) on the ***Quick Access Toolbar*** to save the file. If we are saving the file for the first time, a dialog box opens (Fig. 1.21b). In this dialog box, we type a suitable name for the workbook file in the ***File Name*** box and specify the location where we want to save it from the left pane shown in Fig. 1.21b. The location can be any folder on the hard disk. Initially, Excel offers the ***Documents*** folder. But we can change this by selecting a new drive or folder.

Note that you should first save the workbook on the hard disk with a suitable name, which will help you identify the files of the lab sessions, e.g., for **Lab Session 1** of **Basic Statistics Lab**, you can use the file name **"BS_LS 1"** or any other name of your choice. We click on the ***Save*** button to save a file after typing its name. All subsequent saves will update the existing file without asking for a file name. You can also use the separate sheet within this workbook for each activity, problems and continuous assessment.



(a)



(b)

Fig. 1.21

If you wish to save an existing file again with another name or make a copy of the file, choose the ***Save As*** option instead of ***Save***. It is very important that **you should save results frequently to avoid losing a file**.

5. Closing a File

When we have finished working on a document, we close it. To close a file, we click on the ***Office*** button and select ***Close*** from the menu (Fig. 1.22), or click on the ***Close*** button in the workbook window (Fig. 1.7). If you have made any changes since saving the file earlier, Excel will ask you to save it. To close Excel, we click on the ***Office*** button and click on the ***Exit Excel*** option (Fig. 1.22) or click on the ***Close*** button in Excel window (Fig. 1.7).

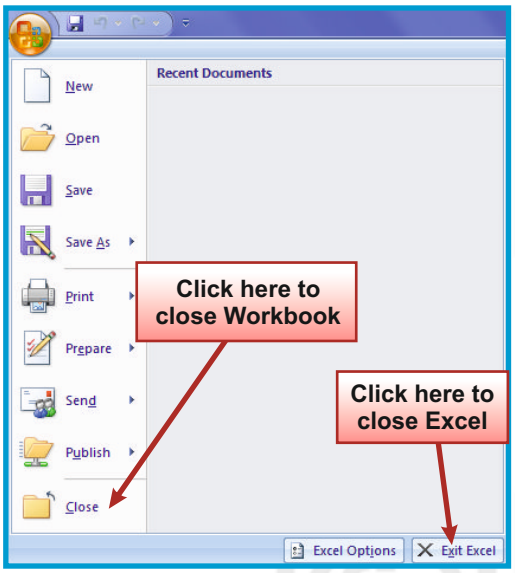


Fig. 1.22

6. Shortcut Mouse Commands

In addition to the ***Ribbon*** commands explained in Sec. 1.4, Excel also provides shortcut menu options for some commands. We get these shortcut menus when we press the right button of the mouse, i.e., right-click the mouse. We can also use these shortcut mouse commands to edit the cell(s)/row(s)/column(s). We select the cell(s)/row(s)/column(s) and right-click the mouse button. It opens a shortcut of the list of commands at the mouse pointer position (Fig. 1.23).

For example, when we select Cell A4 and right-click the mouse, certain shortcut mouse commands appear as shown in Fig. 1.23. This shortcut menu shows all options that we can use for the cell(s)/row(s)/column(s).

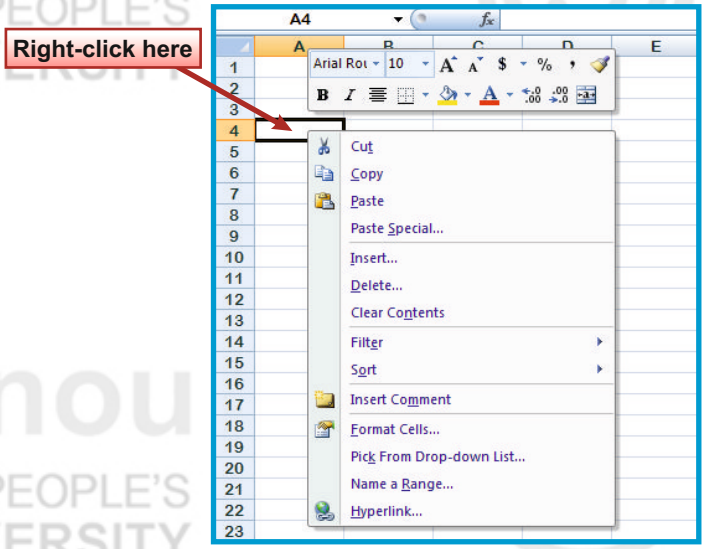


Fig. 1.23

7. Renaming a Worksheet

As shown in Fig. 1.24, to rename a worksheet, we

1. click on the ***Format*** option in the ***Cells*** group under the ***Home*** tab,
2. click on the ***Rename Sheet***, and
3. type a suitable name.

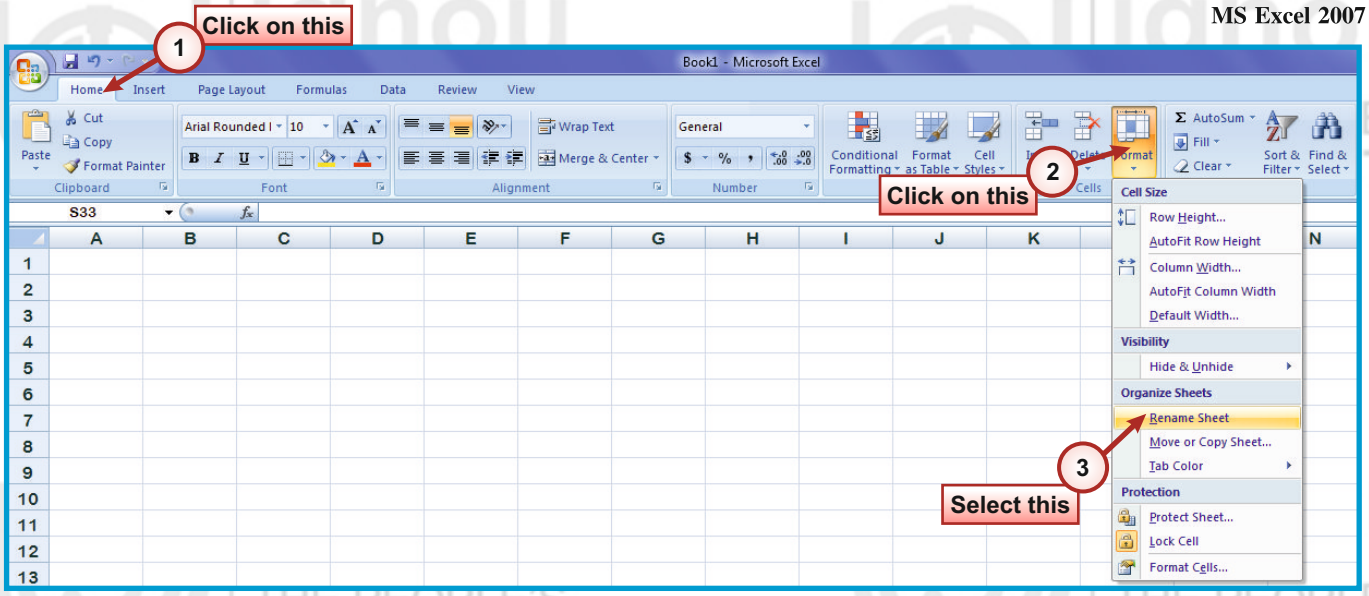
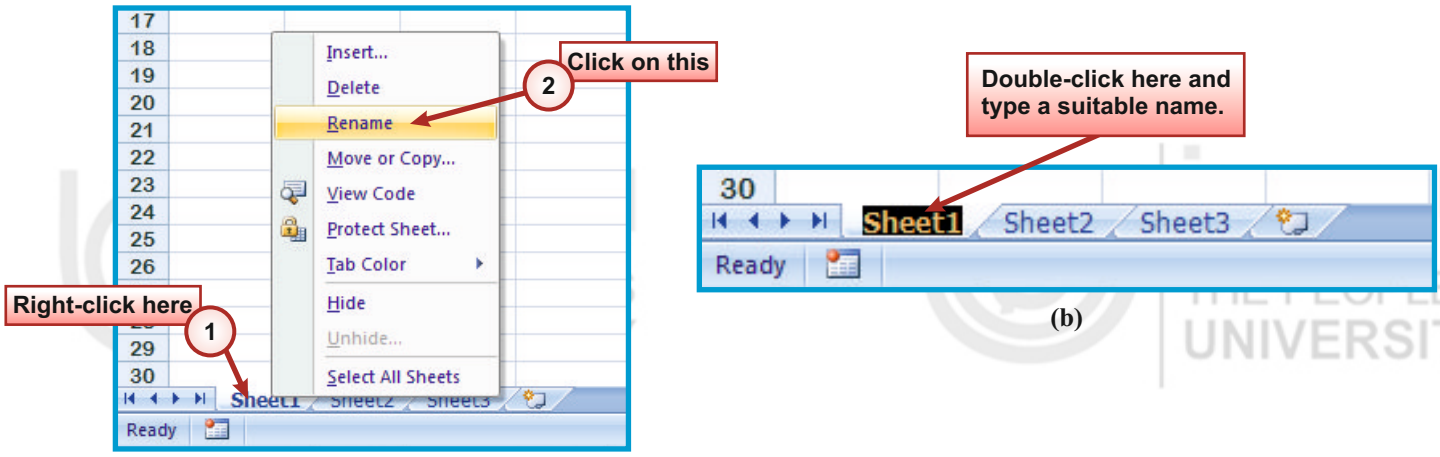


Fig. 1.24

We can also rename a worksheet using shortcut mouse commands. We right-click on the existing sheet name, select ***Rename*** or double-click on the worksheet name and type a new name as shown in Fig. 1.25.



(a) (b)

Fig. 1.25

8. Adding a New Worksheet

As shown in Fig. 1.26, to add a new worksheet, we

1. select the ***Sheet*** tab before which we want to insert a new sheet,
2. click on the ***Insert*** option in the ***Cells*** group under the ***Home*** tab, and
3. click on the ***Insert Sheet*** option. It will give us a new sheet named **"Sheet 4"** in this case (Fig. 1.27c).

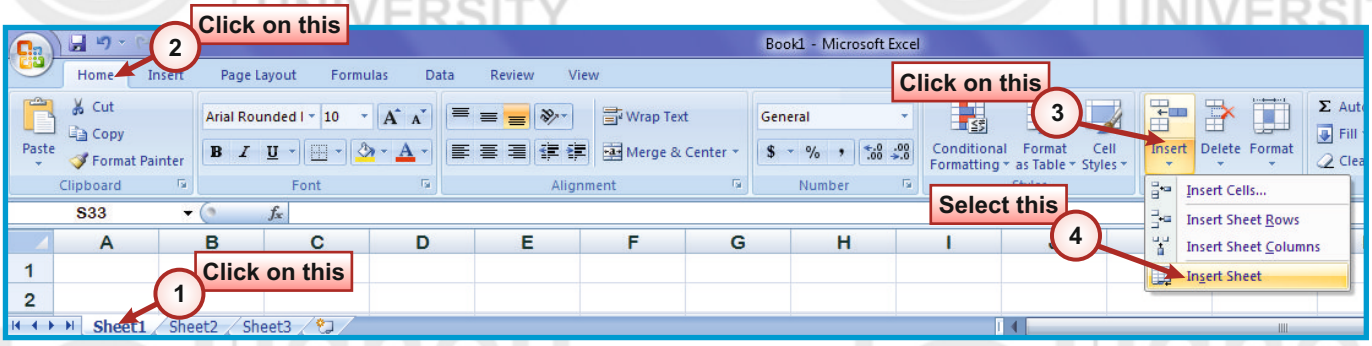


Fig. 1.26

We can also add a new worksheet to the workbook by clicking on the ***Insert Worksheet*** tab at the bottom of the workbook (to the right of the ***Sheet 3*** tab) or we can right-click on one of the existing ***Worksheet*** tabs and select ***Insert*** as shown in Fig. 1.27a. A new dialog box appears, and we select the ***Worksheet*** option and click on ***OK*** as shown in Fig. 1.27b. It gives us a new sheet named **"Sheet 4"** in this case (Fig. 1.27c).

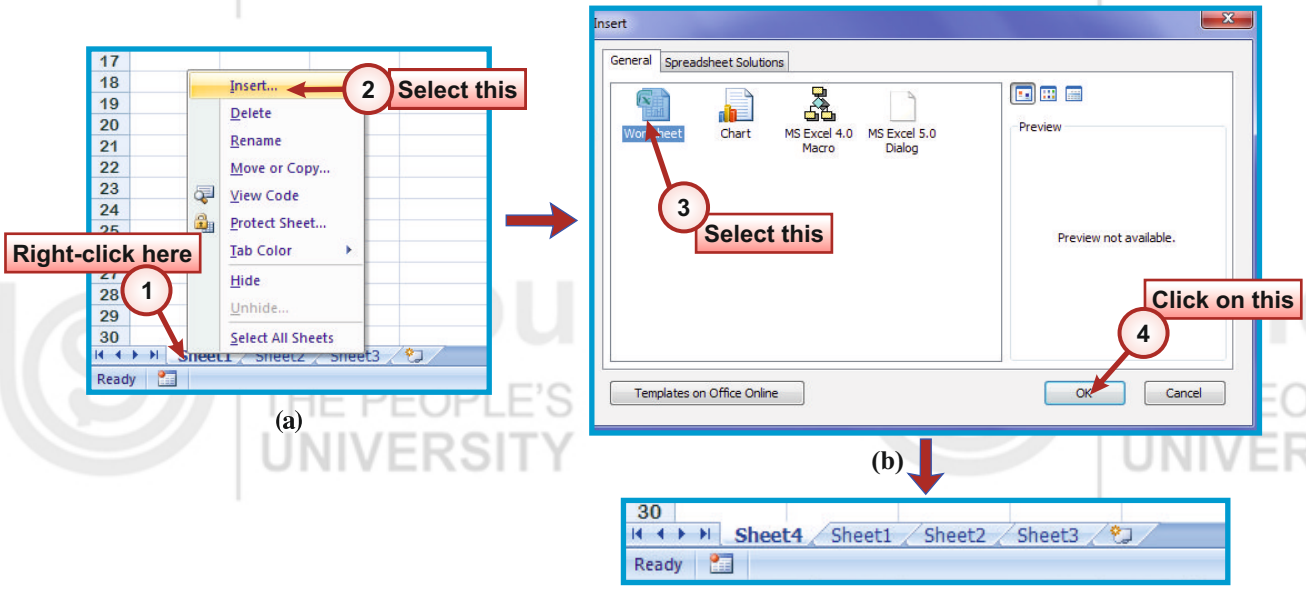


Fig. 1.27

9. Deleting a Worksheet

As shown in Fig. 1.28, to delete a worksheet, we

1. select the ***Sheet*** tab that we wish to delete,
2. click on the ***Delete*** option in the ***Cells*** group under the ***Home*** tab, and
3. click on the ***Delete Sheet*** option.

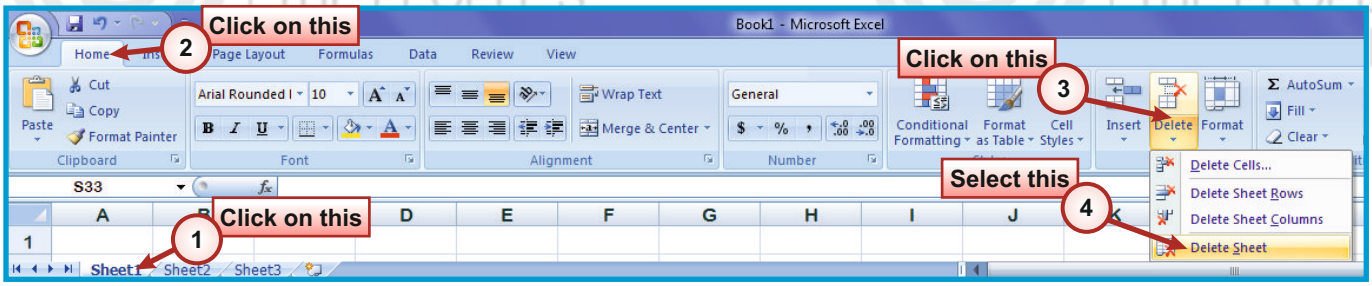


Fig. 1.28

We can also delete a worksheet by right-clicking on its tab and then selecting ***Delete*** as shown in Fig. 1.29.

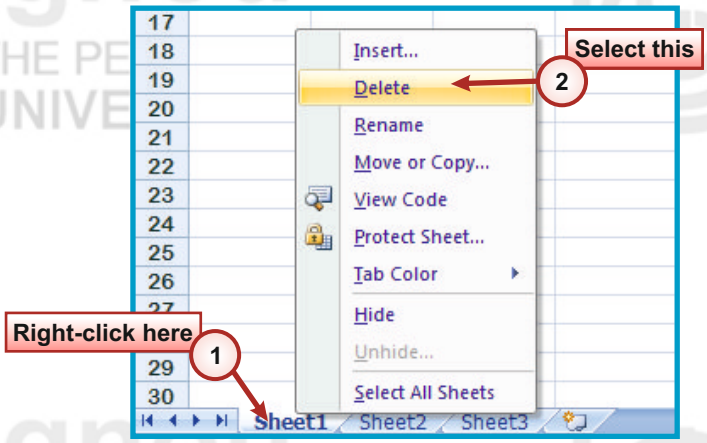
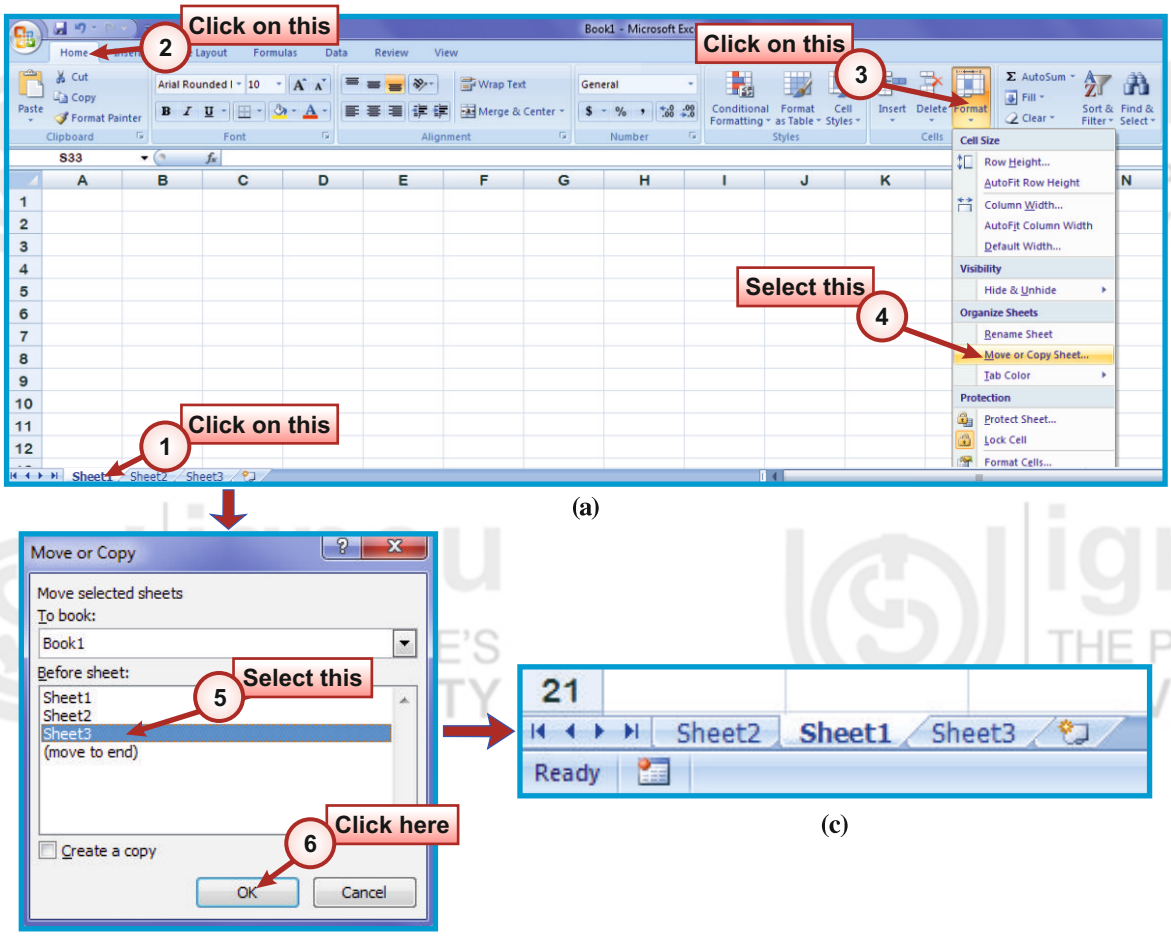


Fig. 1.29

10. Re-arranging the Worksheets

We can re-arrange the worksheets to put them at the desired position. As shown in Fig. 1.30, to re-arrange the worksheets, we

1. select the ***Sheet*** tab we wish to re-arrange,
2. click on the ***Format*** option in the ***Cells*** group under the ***Home*** tab,
3. click on the ***Move or Copy Sheet*** option shown in Fig. 1.30a. This opens a new dialog box as shown in Fig. 1.30b, and
4. select the location before which we wish to place the sheet in ***Before sheet*** options. Here we choose **Sheet 3** and click on ***OK***. Fig. 1.30c shows that **Sheet 1** is now moved to a position before **Sheet 3**.



(a) (b) (c)

Fig. 1.30

We can also use shortcut mouse commands to re-arrange the worksheets. We right-click on the ***Sheet*** tab that we wish to arrange, say, **Sheet 1** and choose ***Move or Copy*** option as shown in Fig. 1.31. It will open the same dialog box as shown in Fig. 1.30b.

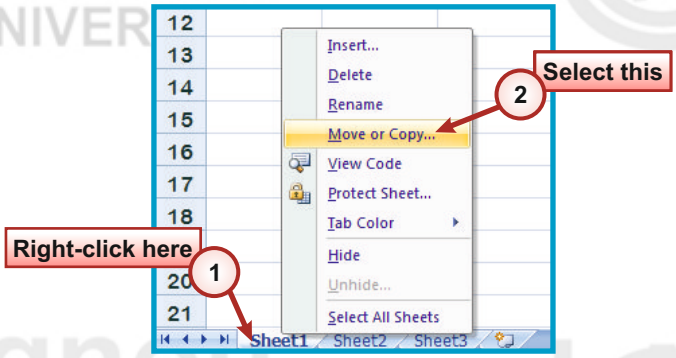


Fig. 1.31

We can also move the sheet by clicking on the ***Worksheet*** tab that we wish to move and dragging it to the desired location without leaving the left button on the mouse. When we drag the mouse, we see a small arrow. We release the left button on the mouse when we see that the arrow is at the desired position. In Fig. 1.32, the arrow is shown between **Sheets 2** and **3**. When we release the left button of the mouse at this point, **Sheet 1** is placed between **Sheets 2** and **3** as shown in Fig. 1.30c.

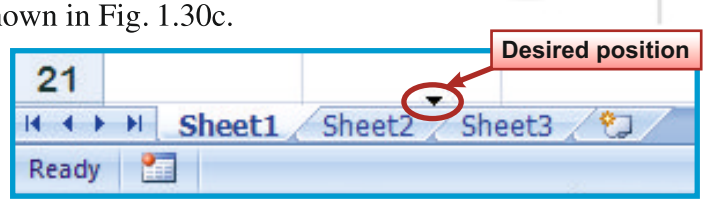
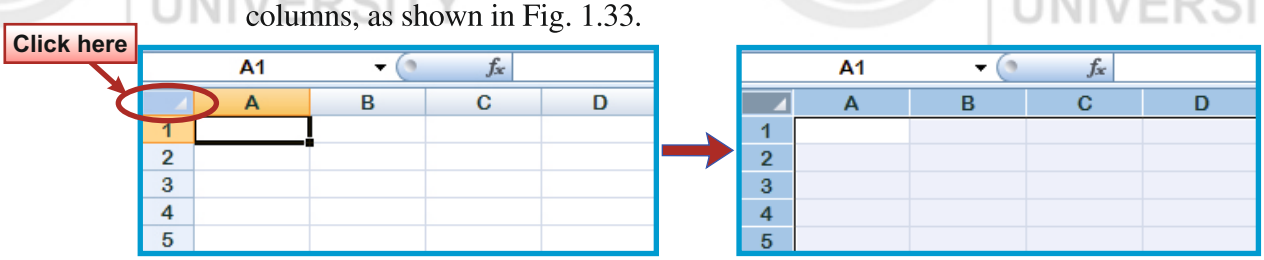


Fig. 1.32

11. Duplicating (Copying) the Entire Sheet

We can also make a copy of the entire sheet in two different ways: (i) by ribbon commands and (ii) by shortcut mouse commands. We first explain how to copy the entire sheet using the ***Ribbon*** commands. For this, we select all cells of the sheet by pressing ***Ctrl+A*** or clicking at the top-left intersection of rows and columns, as shown in Fig. 1.33.



(a) **(b)**

Fig. 1.33

We click on the ***Copy*** option under the ***Home*** tab as shown in Fig. 1.34 or press ***Ctrl+C***.

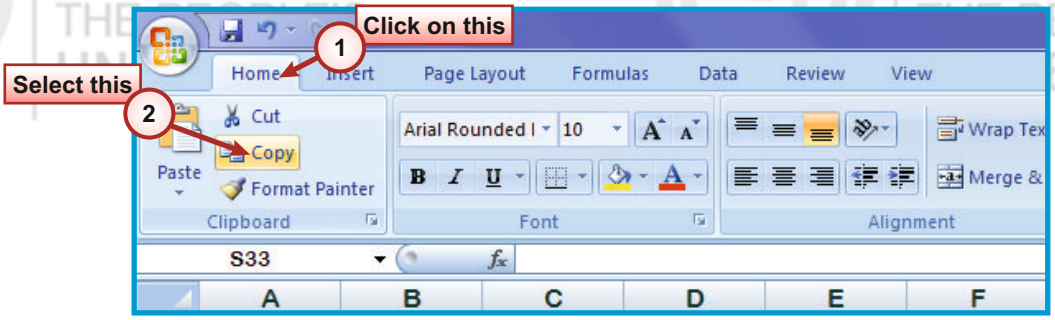


Fig. 1.34

We select another sheet or workbook in which we wish to copy that sheet and then select Cell A1 in that sheet. Then we click on the ***Paste*** option under the ***Home*** tab or press ***Ctrl+V*** as shown in Fig. 1.35.

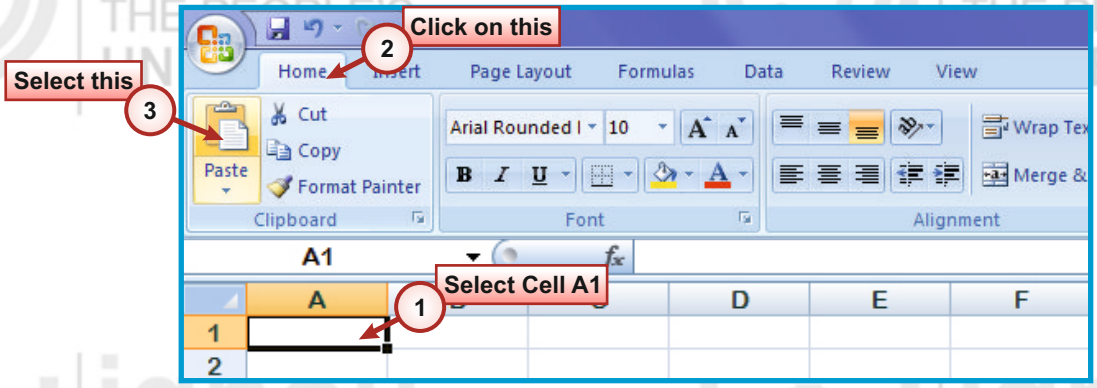
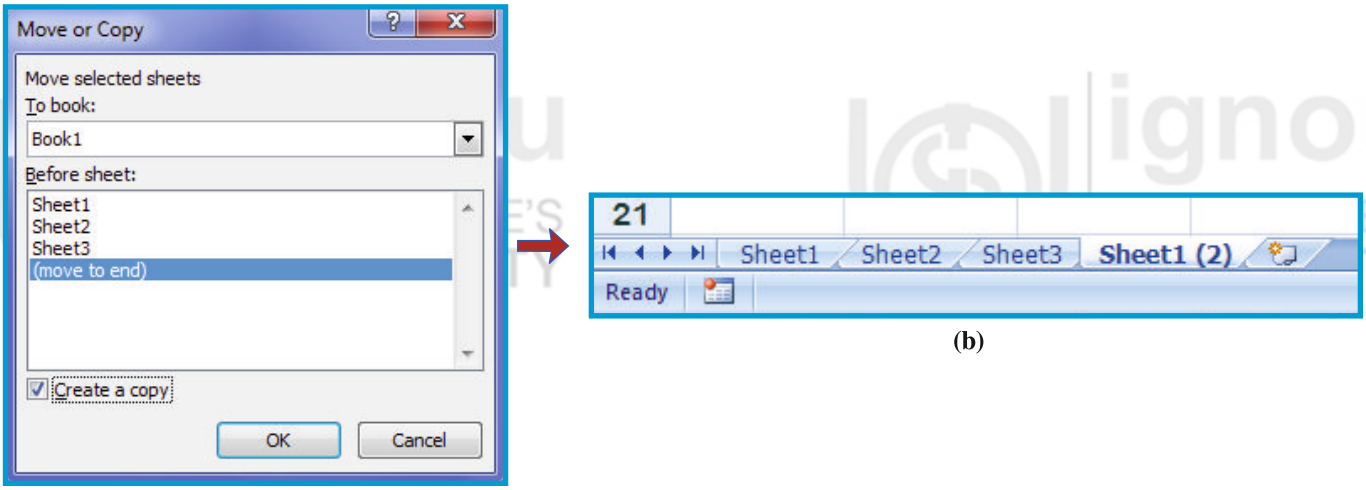


Fig. 1.35

We can also use another method to copy the entire sheet. For this, we right-click on the ***Sheet*** tab of the sheet that we wish to copy. From the shortcut mouse commands, we select ***Move or Copy*** option (Fig. 1.31a). The ***Move or Copy*** dialog box enables us to copy the sheet either to a different location in the current workbook or to a different workbook. You should remember to mark the ***Create a copy*** checkbox (Fig.1.36a). A copy of **Sheet 1** as **Sheet 1 (2)** is shown in Fig. 1.36b.

- If we wish to copy data in the existing workbook (by default), we can choose the sheet before which we wish to place this sheet or we can also choose the ***(move to end)*** option in the ***Before Sheet*** options.



(a) (b)

Fig. 1.36

- We can choose the ***(new book)*** option in ***To book*** to copy in a new workbook. In all cases, we select the ***Create a Copy*** option.

We give the new worksheet a suitable name so that it is easy to recognise.

12. Moving around the Worksheets and Cells in a Worksheet

To move from one worksheet to another, we click on the ***Sheet*** tabs. The name of the active sheet is shown in bold. It is important to be able to move around the cells of a worksheet effectively because we can type or change data only in the active cell. We can move around the cells using the arrow keys or moving the mouse pointer to the required cell and clicking on it to make the cell active. In Fig. 1.3, **Sheet 1** is an active sheet and **Cell A1** is an active cell.

We can also use different keyboard shortcuts to move the active cell:

Up Arrow (↑) – To move the active cell one row up.

Down Arrow (↓) – To move the active cell one row down.

Left Arrow (←) – To move the active cell one column to the left.

Right Arrow (→) **or Tab** – To move the active cell one column to the right.

Ctrl + Down Arrow – To move the active cell automatically to the bottom.

Ctrl + Up Arrow – To move the active cell to the top.

Ctrl + Right Arrow – To move the active cell to the right end.

Ctrl + Left Arrow – To move the active cell to the left end.

To move between the cells on a worksheet, we can also directly click on any cell or use the arrow keys. To see the remaining area of a sheet, we can use either vertical or horizontal scroll bars.

13. Deleting Row(s) or Column(s)

If we select the row or column and just press the ***Delete*** key, then only the content in it disappears but the blank row or column remains. There may be many situations when we want to **delete** the entire row(s) or column(s). To delete the row(s) or column(s), we select the row(s) or column(s) that we wish to delete, then click on ***Delete*** under ***Home*** tab and choose ***Delete Sheet Rows*** or ***Delete Sheet Columns***, respectively, as shown in Fig. 1.37.

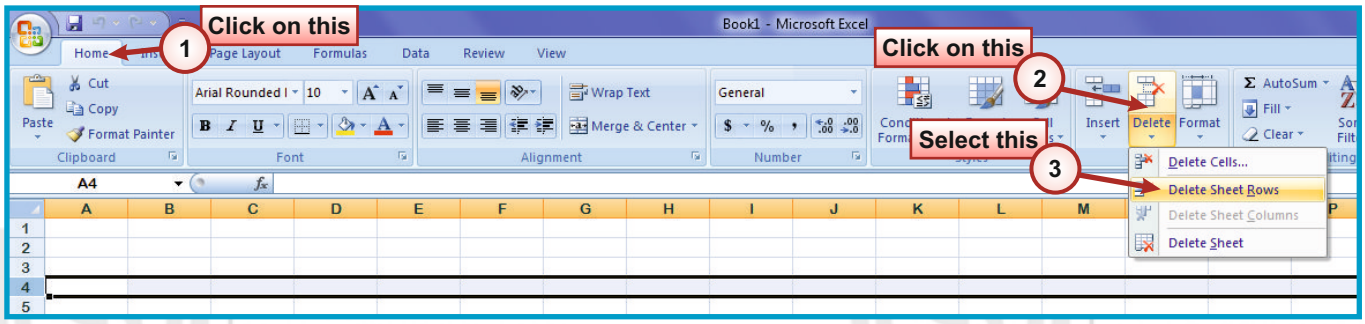


Fig. 1.37

We can also delete the row(s) or column(s) using the short-cut mouse commands. For this, we select the row(s) or column(s) that we wish to delete, right-click the mouse on the selected row number or column letter, and then choose ***Delete*** as shown in Fig. 1.38.

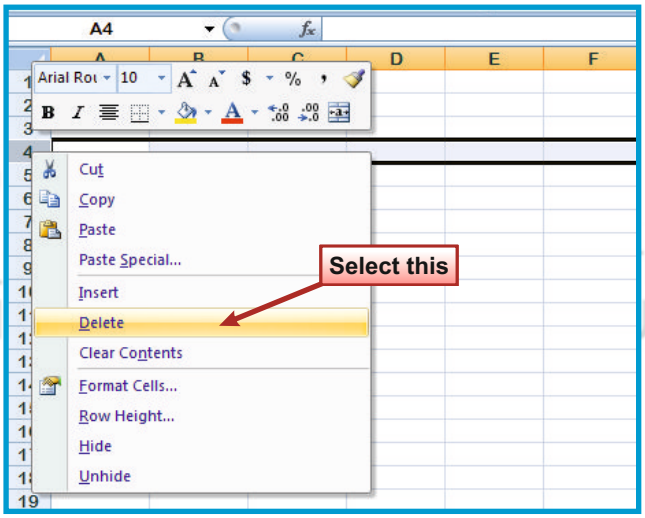


Fig. 1.38

14. Inserting Row(s) or Columns(s)

While working on the spreadsheet, we need to insert row(s) or column(s) many times. To insert row(s) or column(s), we select the row(s) or column(s) before which we wish to insert the empty row(s) or column(s) and then click on ***Insert Sheet Rows*** or ***Insert Sheet Columns*** in ***Insert*** command under the ***Home*** tab as shown in Fig. 1.39.

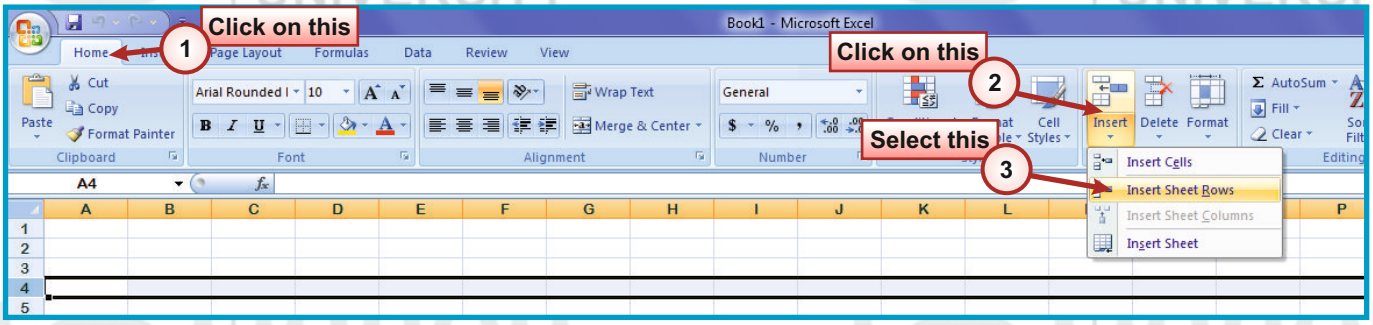


Fig. 1.39

We can also insert row(s) or column(s) using the short-cut commands. We select the row(s) or column(s) before which we wish to place the extra row(s) or column(s), right-click the mouse and then select ***Insert*** option as shown in Fig. 1.40.

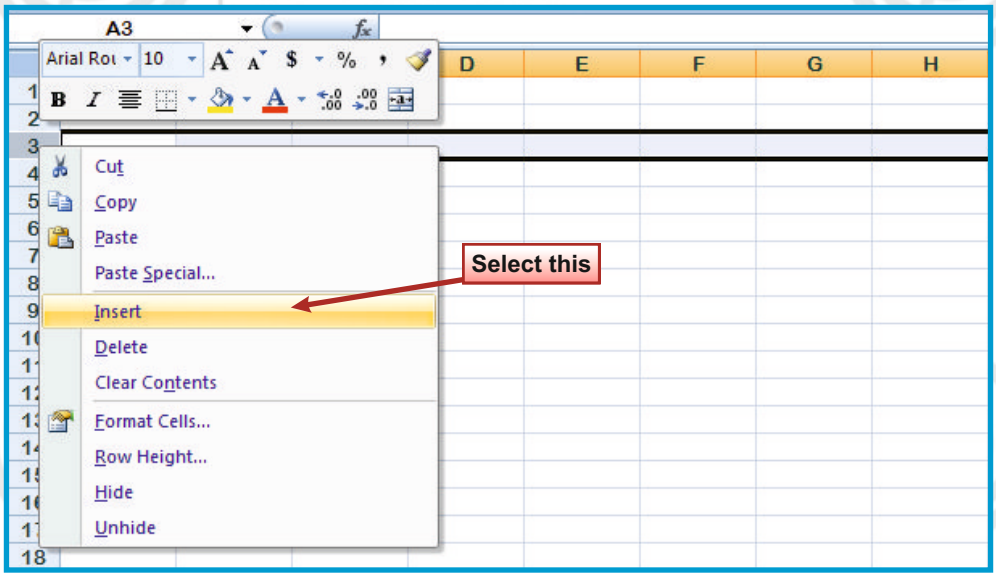


Fig. 1.40

15. Cell Reference

As explained in Sec. 1.3, a new worksheet is a grid of **rows** and **columns**. The rows are labelled with numbers and the columns with letters. Each intersection of a row and column is a **cell**. Each cell has an **address**, which is the combination of the column letter and the row number.

- Refer to Fig. 1.6. Note that Cell A4 is the active cell, which is highlighted with a thick border around it. This cell corresponds to Column A and Row 4. So it is labelled as Cell A4.
- If we are considering entries from Row 4 of Column A up to Row 11 of Column H, we write them as Cells A4:H11. It is the range, i.e., **reference** for these cells.
- Cell references are of two types: (i) relative and (ii) absolute. Most often, cell references are relative but these can also be absolute. Absolute cell references have the prefix $. For example, A4:H11 is a relative cell reference while $A$4:$H$11 is an absolute cell reference.
- **Relative Cell References:** These references can change when the cell references are copied to a new location. For example, if we type "=A2-B2" in Cell C2 and copy Cell C2 to C3 (i.e., move down one cell), the new cell (Cell C3) will be "=A3-B3". You will learn how to copy the cell references in Sec. 1.6.
- **Absolute cell references:** These references do not change when the cell references are copied to a new location. For example, if we type "=$A$2-$B$2" in Cell D2 and copy Cell D2 to Cell D3 (i.e., move down one cell), the new cell (Cell D3) will be "=$A$2-$B$2".
- Cell references can also be part relative and part absolute. Sometimes we need to change the row reference when the references are copied to keep the same column reference or vice versa. In such situations, absolute and relative references are combined. For example, $A4 is absolute Column A and

relative Row 4, while A$4 is relative Column A and absolute Row 4. We can also use the ***F4*** key to create the absolute and mixed references. If we select the reference of Cell A4, we get $A$4, A$4, $A4 and A4 each time we press the ***F4*** key.

16. Selecting the Data

When we work in Excel, we need to select the data on a worksheet many times. We can do this in various ways and you can use any one of them. We have explained that a cell must be active to enter information into it. To select a cell, we click on it. To select a number of cells, we can use one of the following ways:

i) We click on the first entry of the given data, press the ***Shift*** key and use the ***Arrow*** keys to scroll up to the last entry of the data that we wish to select without leaving the ***Shift*** key. For example, if we wish to select the data given in Cells A4:H11, we click on Cell A4, press the ***Shift*** key and use the ***Arrow*** keys to select the data up to Cell H11 without leaving the ***Shift*** key as shown in Fig. 1.41.

ii) We click on the first value of the data, press the ***Shift*** key and then click on the last entry without leaving the ***Shift*** key. For example, if we wish to select the data given in Cell A4:H11, we click on Cell A4, press the ***Shift*** key and then click on Cell H11 without leaving the ***Shift*** key. It will select all cells between and including A4 and H11 (Fig. 1.41).

iii) We can also select the data using the mouse. We left-click on Cell A4, hold the left button on the mouse and drag up to the required last cell without leaving the left button. Fig. 1.41 shows the highlighted Cells A4:H11, which we have selected.

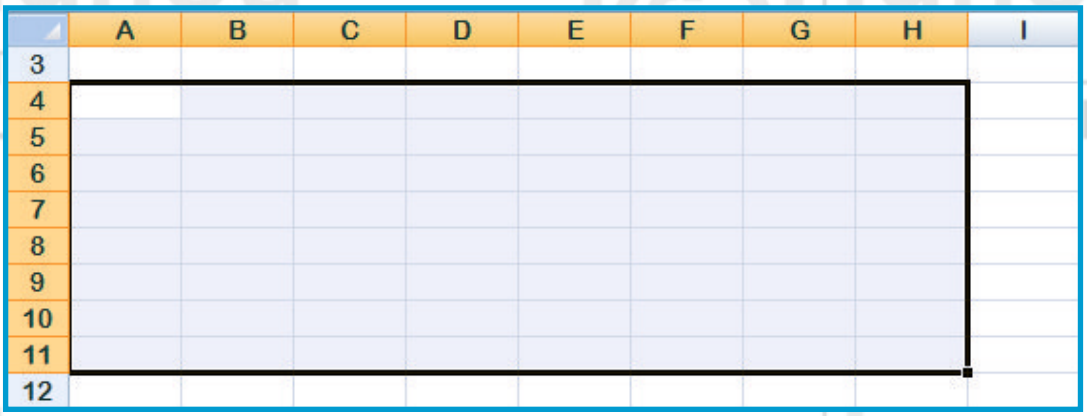


Fig. 1.41

To select several cells, which are not adjacent, we press the ***Ctrl*** key and click on the cells, which we wish to select. Fig. 1.42 shows the selected Cells A2, B3, C4 and D5.

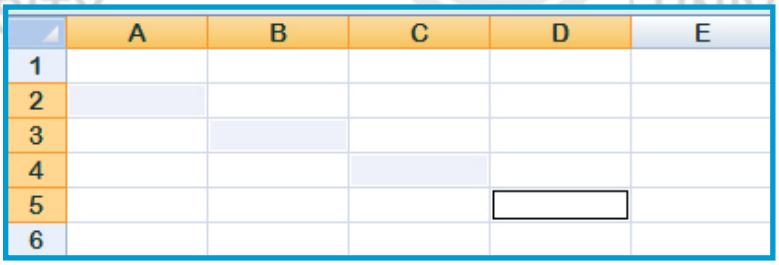


Fig. 1.42

iv) We can also select the entire row or column by clicking on the row number and column label.

17. Inserting a Chart

We can use Excel to construct different types of charts to represent the data diagrammatically. Various chart options are available in the ***Charts*** group under the ***Insert*** tab (Fig. 1.43). We shall discuss in detail about the construction of various charts in Lab Sessions 3, 4 and 5.

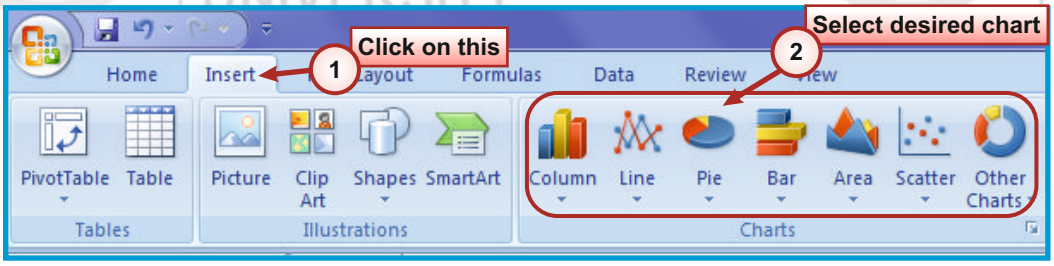


Fig. 1.43

18. Inserting a Symbol

We can also insert symbols, e.g., μ, σ, α, etc., in a cell of Excel spreadsheet. We first select the cell in which we wish to insert the symbol. To insert a symbol, we click on ***Symbol*** in ***Text*** group under the ***Insert*** tab as shown in Fig. 1.44a. A new dialog box appears as shown in Fig. 1.44b. We click on the required symbol, click on ***Insert*** or we can also double click on the symbol to insert it and then click on ***Cancel*** to close the box.

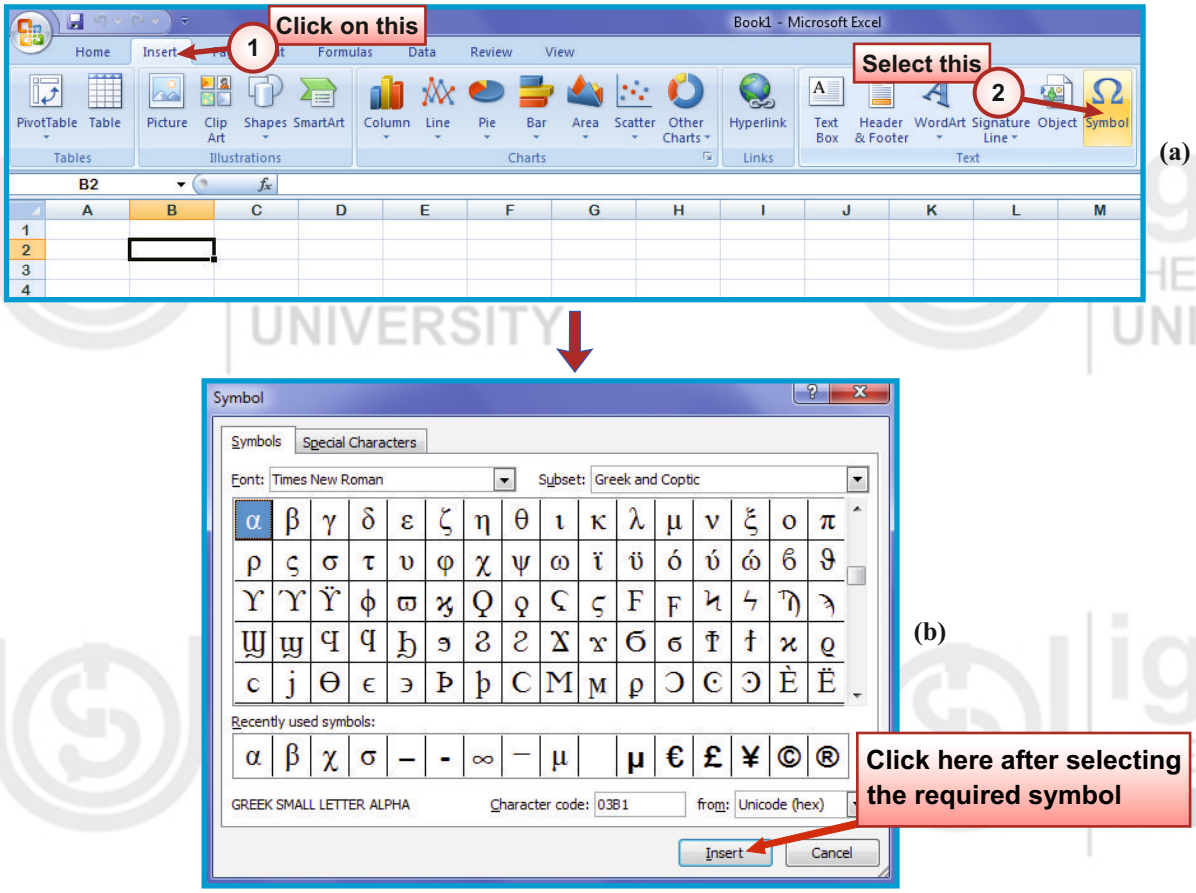


Fig. 1.44

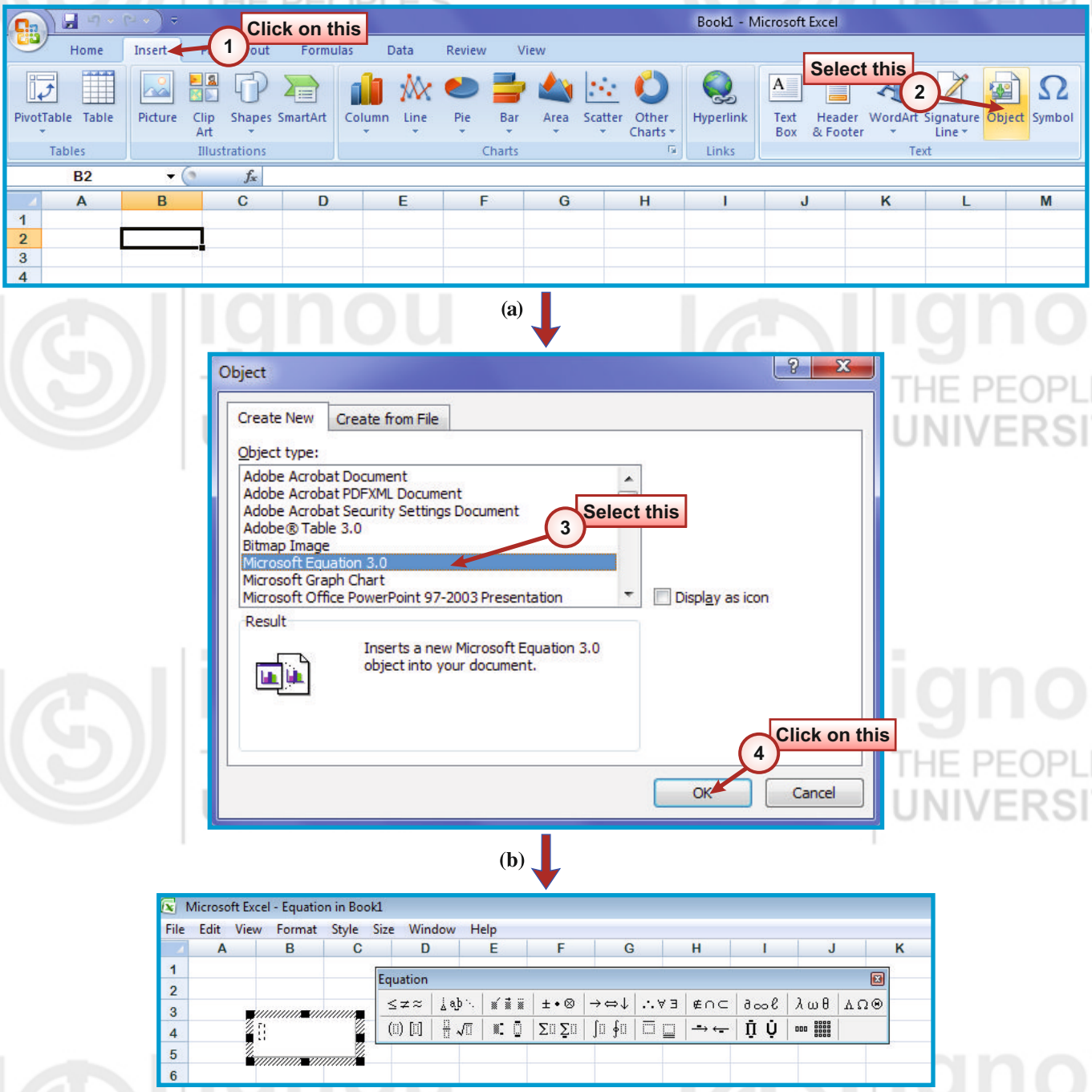
19. Inserting a Mathematical Equation

We can also type mathematical equations such as $\bar{x}$, σ_1^2, $\sum_{i=1}^{n} x_i^2$, etc., in Excel. For this, we

1. click on the ***Object*** option in ***Text*** group under the ***Insert*** tab as shown in Fig. 1.45a so that a new dialog box appears (Fig. 1.45b),

2. click on the ***Microsoft Equation 3.0*** option and click on ***OK*** as shown in Fig.1.45b, and

3. type the required equation using the various options shown in Fig.1.45c.



(c)

Fig. 1.45

1.6 ENTERING DATA IN AN EXCEL SHEET

We can enter data in an Excel sheet in many ways. We can

i) type data in a new Excel worksheet,

ii) copy data from an existing Excel worksheet, and

iii) create new data by transforming the existing data.

Besides the above mentioned ways used commonly, we can also use data from other file formats. We are not discussing it here as it is out of the scope of this course.

1.6.1 Typing Data in a New Excel Worksheet

To enter any information in a cell, we first select the cell, say, Cell A1 and begin typing with the help of the keyboard. When we finish entering the information (label or value) in that cell, we

- press ***Enter*** to move to the next cell below (in this case, Cell A2) or
- press ***Tab*** to move to the next cell to the right (in this case, Cell B1) or
- click on any cell to select it.

Note that Excel assumes any number typed as positive. If we want to enter a negative number, we use a minus sign "-" or enclose the number in the parentheses "( )".

We consider a simple example to explain how to enter data. Suppose we have the data of maximum and minimum temperature for 5 days as recorded in Table1.

Table 1: Maximum and minimum temperature data

Maximum Temperature (in ^{o}C)	Minimum Temperature (in ^{o}C)
40	18
41	20
38	16
36	15
43	21

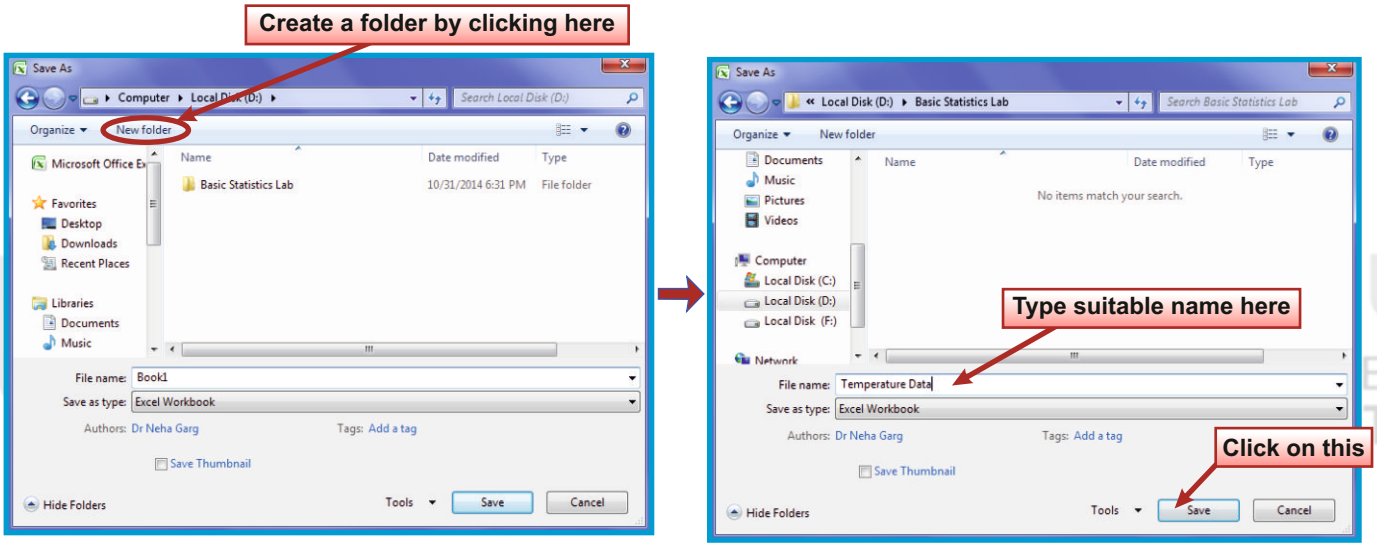
The data given in Table 1 consist of two columns of information. The first and second columns of the table contain the maximum and minimum temperatures, respectively, for five days. We now start entering this data in an Excel sheet. For this, we

1. click on Cell A1 to make it active,
2. type "Maximum Temperature" in Cell A1, and press ***Enter*** or press ***Down Arrow*** (↓) key to select Cell A2,
3. type "40", the maximum temperature for day 1 in Cell A2. In the same way, we can type the remaining values up to Cell A6,
4. click on Cell B1 to enter the data of the second column of Table 1,
5. type "Minimum Temperature" in Cell B1,
6. select Cell B2 and type "18", the minimum temperature for day 1. In the same way, we can type the remaining values of minimum temperature up to Cell B6. The resulting Excel sheet with the data is shown in Fig. 1.46.

	A	B	C
1	Maximum Temperature	Minimum Temperature	
2	40	18	
3	41	20	
4	38	16	
5	36	15	
6	43	21	
7			

Fig. 1.46

Notice from Fig. 1.46 that the labels of the data given in Table 1, i.e., "Maximum Temperature" and "Minimum Temperature" are in Row 1, i.e., Cells A1 and B1, respectively. The values of data are shown in Rows 2 to 6, i.e., Cells A2:B6. We now save the file as explained in Sec. 1.4. We type the name of the file as "**Temperature Data**" and save it in a folder named "**Basic Statistics Lab**" in **D** drive as shown in Fig.1.47.



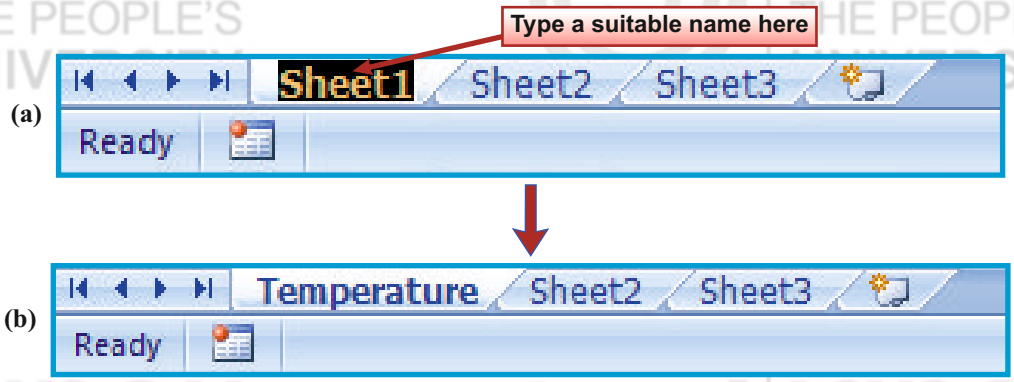
(a) (b)

Fig. 1.47

Usually the file name will have extension .xls or .xlsx. Excel 2007 reads all previous versions of Excel files, though not vice-versa. For example, the Excel file **"Temperaure Data.xlsx"** is an Excel 2007 file.

Changing the Worksheet Name

We can also assign a suitable name to the worksheet, which contains this data. We select the ***Sheet*** tab at the bottom, say, "**Sheet 1**", right-click the mouse and select ***Rename*** option (Fig. 1.29) or double click on the ***Sheet*** tab. Then we give the worksheet a suitable name, say, "**Temperature**" as shown in Fig. 1.48.



(a)

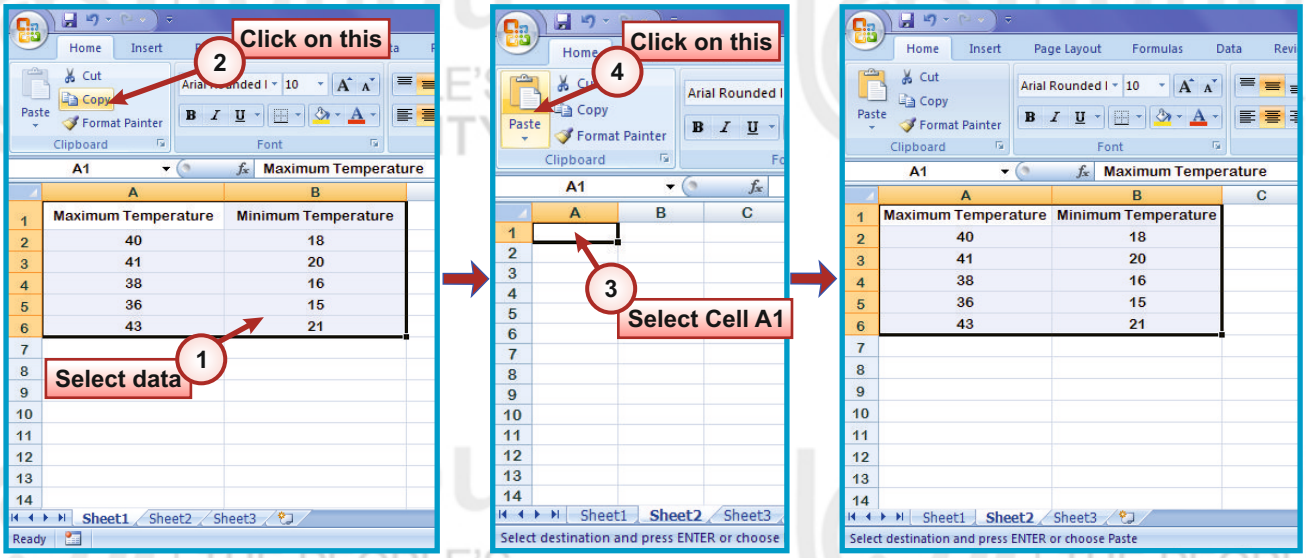
(b)

Fig. 1.48

1.6.2 Copying Data from an Existing Excel Worksheet

We consider the file "**Temperature Data.xlsx**", which is an Excel 2007 workbook created in Sec. 1.6.1. Suppose we wish to copy the data on the sheet named "**Temperature**" and paste it on the another sheet, say, "**Sheet 2**". For this, as shown in Fig. 1.49, we

1. select data, i.e., Cells A1:B6 given on the sheet **"Temperature"**,
2. click on ***Copy*** option under the ***Home*** tab,
3. click on "**Sheet 2**" tab,
4. select Cell A1 (or any other cell where we wish to paste this data) of **Sheet 2**, and
5. click on ***Paste*** option under the ***Home*** tab.



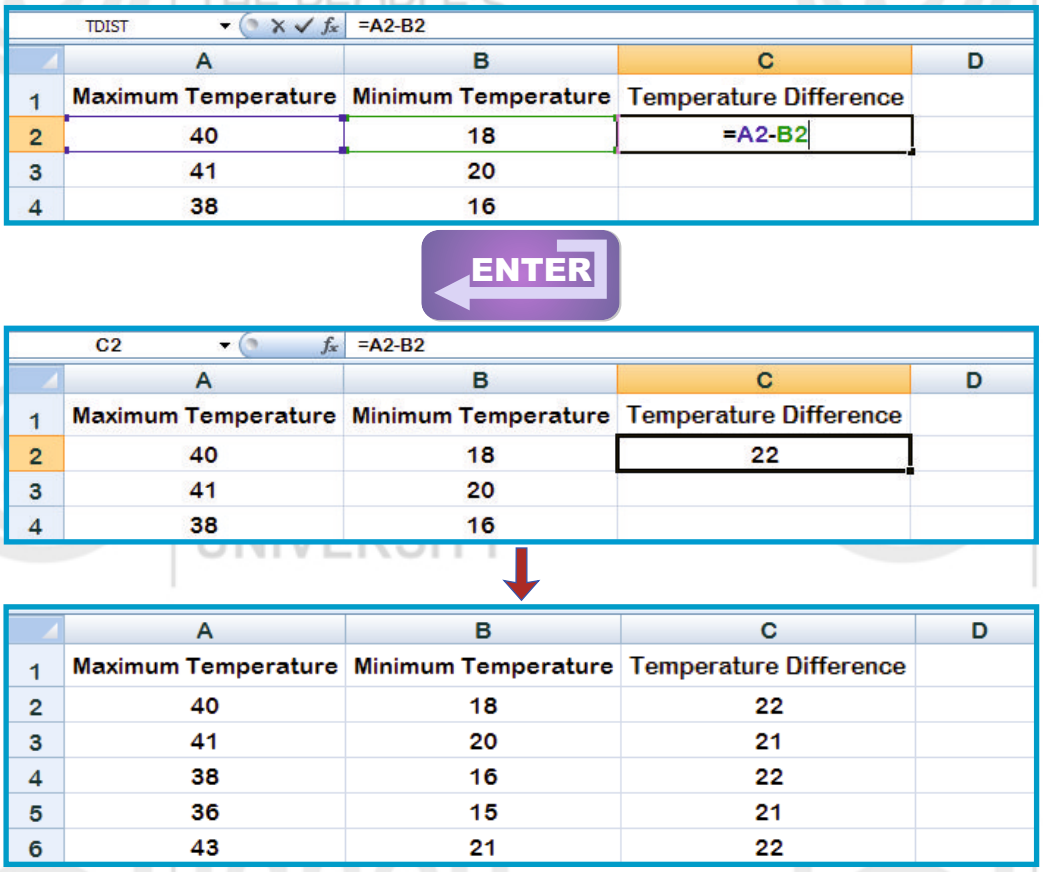
(a) (b) (c)

Fig. 1.49

1.6.3 Creating New Data by Transforming the Existing Data

This option of data creation in Excel is very attractive as it makes calculations simple and easy. We can create new data by transforming the existing data. Suppose we open the file named "**Temperature Data.xlsx**" created in Sec. 1.6.1. We can now create the data of **temperature difference** using the existing data, i.e., the maximum and minimum temperatures. We shall explain in detail how to type a formula in Excel in Sec. 1.7. For computing the **temperature difference**, we

1. type "Temperature Difference" in Cell C1,



(a)

(b)

(c)

Fig. 1.50

2. type "=A2-B2" in Cell C2 as shown in Fig. 1.50a [It will compute the value by subtracting the value of Cell B2 from Cell A2 (Fig. 1.50b)], and

3. compute the temperature difference for the remaining days in the same way.

We get the data of temperature difference in Cells C2:C6 (Fig. 1.50c).

These calculations can be done in a simpler way. We now explain how to apply a formula in different cells to get the result with minimum effort.

1.6.4 Copying a Formula in Different Cells

When we have to use the same formula in multiple adjacent horizontal or vertical cells, we need to type the formula only in the first cell. We can then copy it into the remaining cells. We now explain how to copy the formula of Cell C2, i.e., "=A2-B2" up to Cell C6 (Fig. 1.51).

Here we discuss two different ways of copying the formula of the first cell in the other adjacent cells:

i) Dragging Down the Formula

We can copy the formula using the mouse. For this, we

1. select Cell C2,
2. put the cursor at the lower right corner, i.e., drag handle, of the Cell C2 (as a result plus sign (+) appears as shown in Fig. 1.51a), and
3. left-click the mouse and drag the cursor down up to Cell C6 without leaving the left button (Fig. 1.51b). As we release the left button on the mouse, we get the result shown in Fig. 1.51c.

(a)

	A	B	C	D
1	Maximum Temperature	Minimum Temperature	Temperature Difference	
2	40	18	22	
3	41	20		
4	38	16		
5	36	15		

Drag it down with left click

(b)

	A	B	C	D
1	Maximum Temperature	Minimum Temperature	Temperature Difference	
2	40	18	22	
3	41	20		
4	38	16		
5	36	15		
6	43	21		
7				

Release left button now

(c)

	A	B	C	D
1	Maximum Temperature	Minimum Temperature	Temperature Difference	
2	40	18	22	
3	41	20	21	
4	38	16	22	
5	36	15	21	
6	43	21	22	
7				

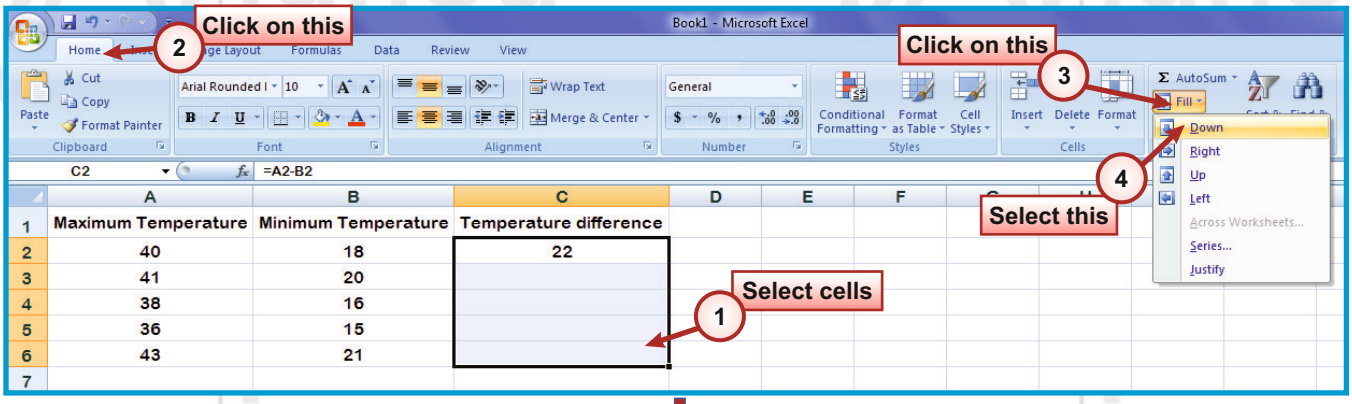
Fig. 1.51

In this way, we copy the formula "=A2-B2" of Cell C2 up to Cell C6. Note that the formula "=A2-B2" has an relative reference, so it will change accordingly as we drag it down as shown in Fig. 1.51. You can also explore the drag right, up or left options to copy the formula in the right, upper or left cells, respectively.

ii) Fill Option

We can also copy the formula using the ***Fill*** option available under the ***Home*** tab. For this, as shown in Fig. 1.52, we

1. select the cells starting from the cell that contains the formula and then all cells in which we wish to copy the formula, i.e., Cells C2:C6,
2. click on the ***Fill*** option in the ***Editing*** group under the ***Home*** tab (Fig. 1.52a), and
3. choose ***Down*** option which will copy the formula given in Cell C2 down up to Cell C6 as shown in Fig. 1.52b.



(a)

	A	B	C	D
1	Maximum Temperature	Minimum Temperature	Temperature difference	
2	40	18	22	
3	41	20	21	
4	38	16	22	
5	36	15	21	
6	43	21	22	
7				

(b)

Fig. 1.52

You can also explore the ***Right, Up*** or ***Left*** options to copy the formula in the right, upper or left cells, respectively.

Note that you can also copy the values other than formulae, i.e., year name, month name, series of numbers, etc., using these two options for copying.

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

Activity 1

Create a worksheet in Excel as follows:

i) Type "Sunday" in Cell A1 and drag it down up to Cell A7.

ii) Enter minimum and maximum temperature data for the past one week in your city. You may get this data from the newspapers. Obtain the temperature differences as explained in Sec. 1.6. Give a suitable name to your worksheet.

1.7 WORKING WITH FORMULA

A mathematical formula is used to calculate a result based on data from one or more cells. It consists of some combinations of the standard mathematical operators (e.g., +, -, *, /), but may also include functions (e.g., ***Sqrt( )***, ***Abs( )***, etc.). When we type a formula in a cell, that cell will generally display the result obtained by the formula, rather than the formula itself. Formulae automatically do calculations on the values in other specified cells and display the result in the cell in which we enter the numbers or type the formula. For example, when we type "=A2+B2" in Cell C2, it means that Cell C2 contains the sum of the values given in Cells A2 and B2. The value displayed in Cell C2 will then be a function of the numbers entered in Cells A2 and B2.

We first discuss the mathematical operations and their order before explaining how to type a formula in Excel.

1.7.1 Mathematical Operations

Excel applies the mathematical method of carrying out the calculations known as **BODMAS** rule, which defines the order of mathematical operations as follows:

Brackets	( )
of (order)	^ (power, square root, etc.)
Division	/
Multiplication	*
Addition	+
Subtraction	–

For example, if we consider the formula "=4+2*5", it gives the answer 14. You must have learnt this rule while studying arithmetic in school. Excel also does the multiplication before addition.

If we consider the formula "=(4+2)*5", it gives the answer 30. Excel also does the addition in the brackets before multiplication.

There are many situations in which we need to use several brackets to type a formula in Excel. In such situations, Excel carries out the calculations of the innermost brackets first and then goes to the outer brackets.

If we consider the formula "=(4+2)*5+3/2", it gives the answer 31.5. In this example, Excel first adds the numbers within the brackets, carries out multiplication and division in the formula and then adds the results as per BODMAS.

If we consider the formula "=((4+2)*5+3)/2, it gives the result 16.5.

In this example, Excel adds the numbers within the inner brackets, then multiplies and adds the numbers within the outer brackets. Lastly, it divides as per BODMAS.

To perform calculations in Excel, you should keep two things in mind:

i) You should type the formula in the cell in which you wish the result to appear.

ii) All **Excel formulae** should start with the equal to (=) sign.

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

Activity 2

Solve the following exercises with the help of MS Excel 2007 and interpret the results:

i) $$\left(8\times\frac{7}{2}\right)+25-\frac{20}{4}\times\frac{\left(15-\frac{31}{1+7\times3}\right)}{3}$$

ii) $$\left(\frac{4\times3+25}{5+41}-\frac{75}{3}\right)\times3-15$$

iii) $$\frac{2+3-\frac{9}{3}+3+\frac{\left(\frac{8}{2}+6\times2\right)}{7}\times(2+3-4)+12-2\times5}{5}$$

1.7.2 Writing a Formula

We can use more than one mathematical operation in a formula typed in Excel and it will calculate the result using the BODMAS Rule.

However, we can also use brackets to change the order since any part of the formula which is contained within the brackets will be calculated first. Note that we can either press the ***Enter*** key or click the ***tick*** (✓) sign in the ***Formula Bar*** to get the results when we complete typing a formula (Fig. 1.53).

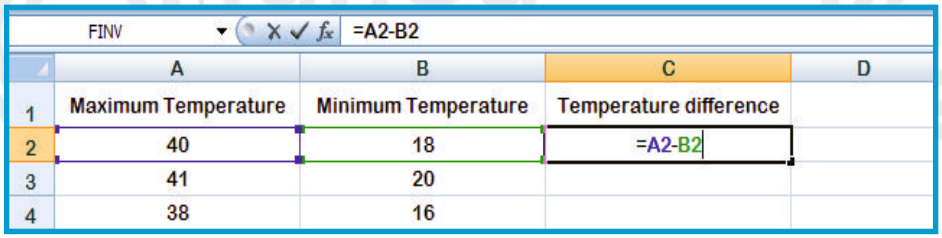


Fig. 1.53

We can also edit the formula later as follows:

1. We can double-click on a formula to edit it.
2. We can also edit the formula in the ***Formula Bar***, but we must first click on the cell that contains the formula, to make it active.

Note that we can also use cell references to create a formula, which use the values contained in those cells for the calculation. For example, if we type "=A2-B2" in Cell C2, it will subtract the values of Cells A2 and B2 and display the result in Cell C2.

We can also use various functions of Excel, which are **built-in** functions such as ***Average( )***, ***Sum( )***, ***Var( )***, ***Sqrt( )***, etc. To apply these functions, we can choose the desired function available in ***Formulas*** group under ***Insert*** tab. We shall explain the use of these functions in the subsequent lab sessions.

We can also directly type these functions in the cell in which we wish to display the output.

Auto Fill Formula/Formula Auto Complete

We can directly type a built-in formula in the desired cell. Excel 2007 also

exhibits an **Auto Complete** feature, which displays a continually updated drop-down list of matching formulae (Fig. 1.54). We can choose the required formula from this drop down list. For example, if we wish to use the ***Average( )*** function in a cell, when we start typing "=Average", we get the drop-down list as shown in Fig. 1.54.

Note that formula which we type in Excel is not case sensitive, we can use upper as well as lower case. For example, if we wish to use ***Average*** function, we can type "=Average( )", "=average( )"or "=AVERAGE( )".

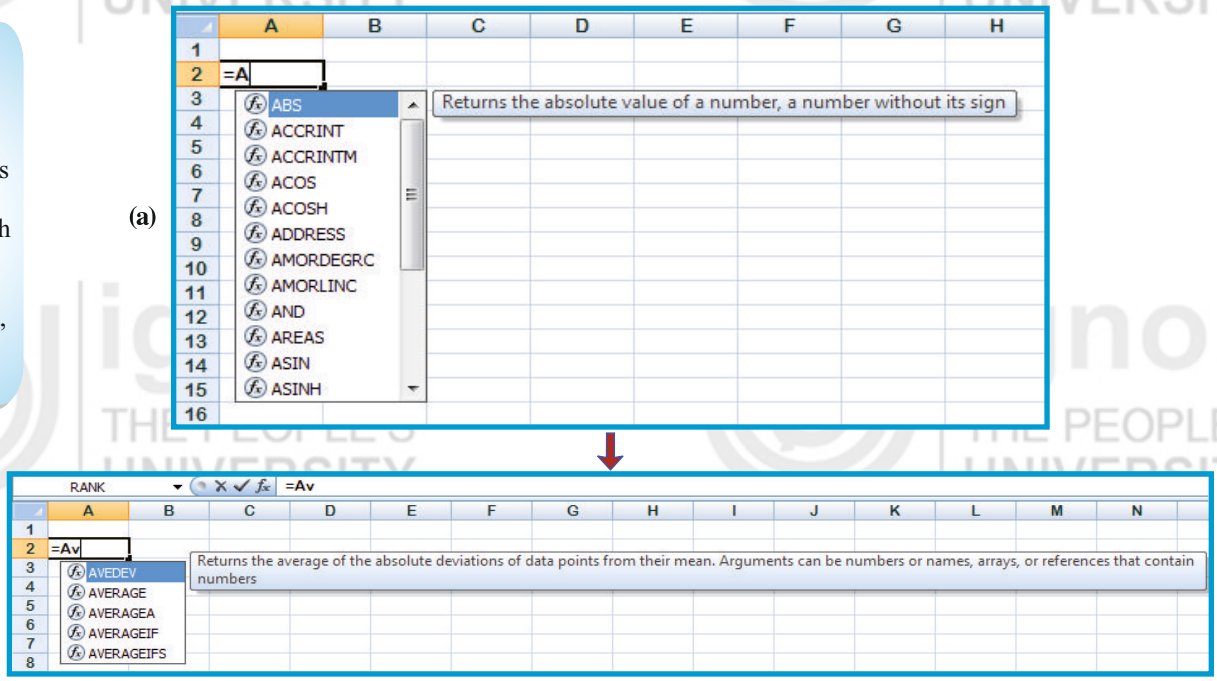


Fig. 1.54

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

Activity 3

Type the following data in an Excel sheet and compute the per capita income for each family:

No. of Members in a Family	Family Income (in ₹)
5	45000
3	40000
4	65000
5	55000

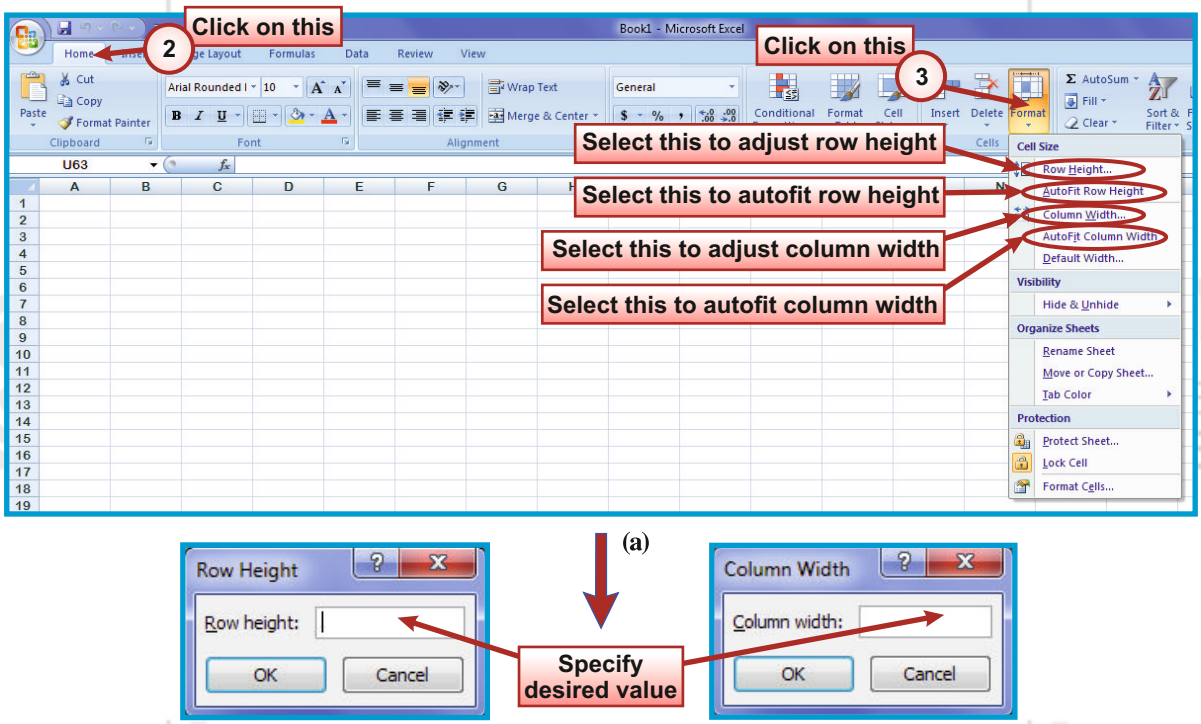
1.8 EDITING AND FORMATTING THE WORKSHEET

You have learnt how to create a worksheet, i.e., type data, apply formulae, etc., using Excel. We now explain how to edit and format the worksheet so that it becomes easier to handle the information stored in it and also makes it look better.

i) Changing the Row Heights and Column Widths

While working on the worksheet, we also need to change the row height/ column width many times. For example, we may wish to decrease the row height/column width to accommodate more information on a page, or increase it to accommodate the long value or text in a single cell. We can change the row height and column width using any one of the following methods:

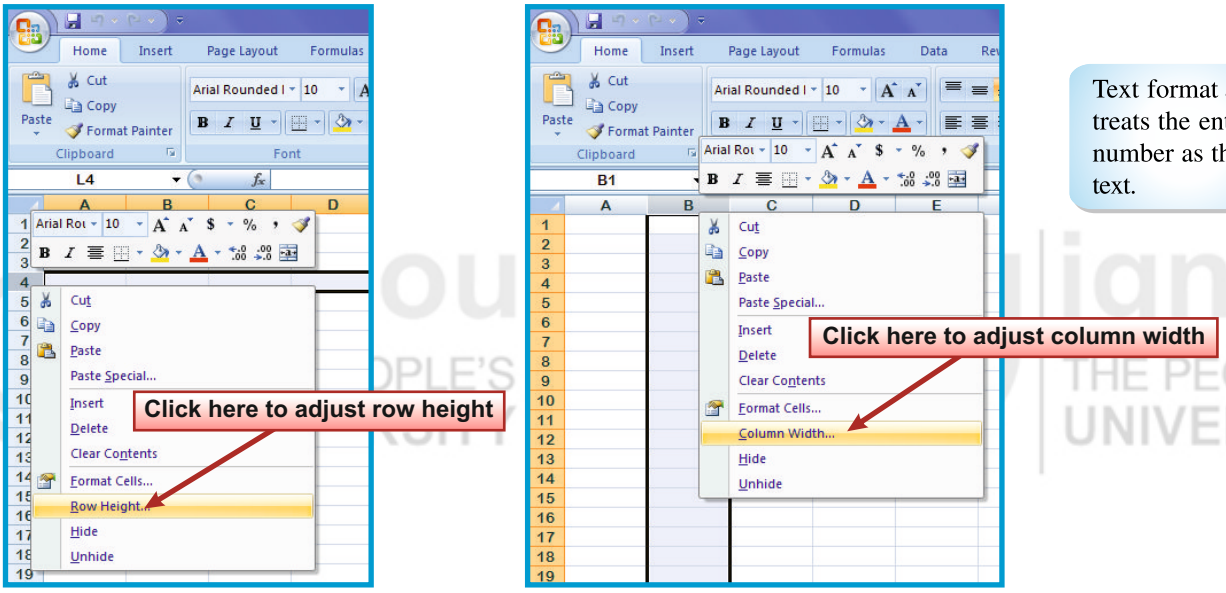
a) We select the desired row(s) or olumn(s) to adjust the **row height/column width**, click on the ***Format*** option in ***Cells*** group under the ***Home*** tab on the ribbon and choose ***Row Height/Column Width*** option. We get the dialog box shown in Fig. 1.55b or c, respectively, if we select row(s) or column(s). We type a desired value of row height/column width in the dialog box as shown in Fig. 1.55.



(a) (b) (c)

Fig. 1.55

b) We can also use the shortcut mouse commands by right-clicking on the selected row(s) or column(s) and choose ***Row Height*** or ***Column Width*** option as shown in Fig. 1.56. We get the dialog box shown in Fig. 1.55b or c, respectively. We enter the desired value to adjust it according to the requirement.

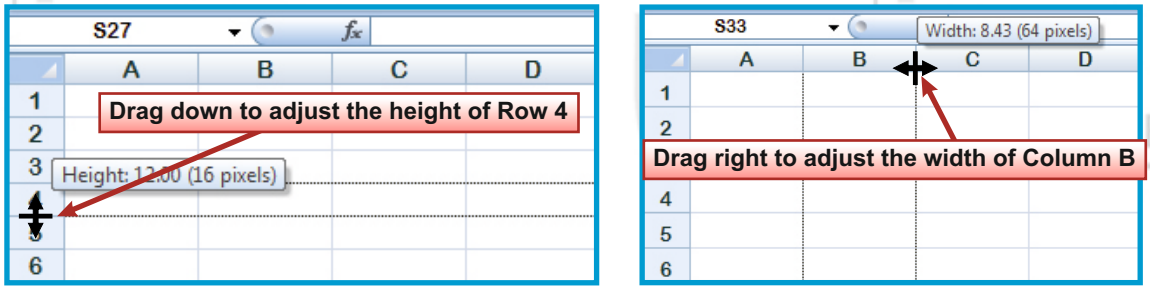


Text format also treats the entered number as the text.

(a) (b)

Fig. 1.56

c) We can simply adjust the **Row height/Column width** by dragging the lower border of the row (Fig. 1.57a) or right border of the column (Fig. 1.57b) with left-click of the mouse until we get the desired row height or column width.



(a) (b)

Fig. 1.57

d) We can also choose ***Home*** → ***Cells*** → ***Format*** → ***AutoFit Row Height*** or ***AutoFit Column Width*** (Fig. 1.55a) to adjust the selected row height or column width so that the widest entry in the column fits. Rather than selecting an entire column, we can just select cells in the column and the column is adjusted based on the widest entry in our selection.

e) We can double-click the right border of the column or lower border of the row to set the column width or row height automatically to adjust the wider entry in the column or row (Fig. 1.58).

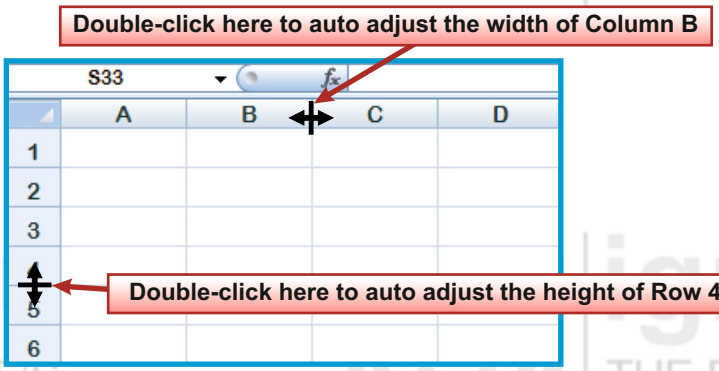
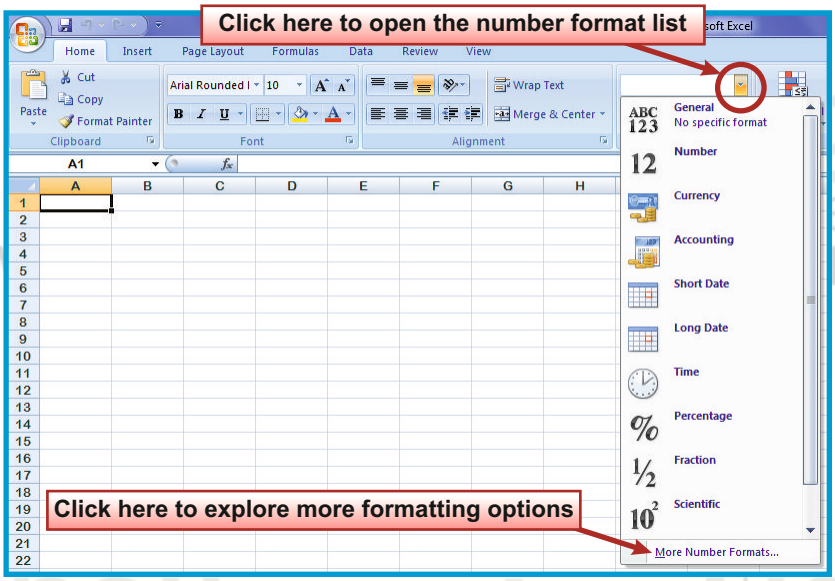


Fig. 1.58

ii) Changing Number Format

We can also use various number formatting options available in Excel to represent or change the appearance of the data in the worksheet. Excel offers 12 different number categories options such as **number**, **currency**, **date**, **time**, **percentage**, etc., as shown in Fig. 1.59.



To apply a specific number format, we

1. select the cell or range of data which we wish to format,
2. click on the arrow next to the ***Number*** format which opens the drop down ***Number format list*** as shown in Fig. 1.59, and
3. click on the suitable number category.

We can also click on the ***More Number Formats*** options shown in Fig. 1.59 to get the detailed formatting options. It opens the ***Format Cells*** dialog box (see Fig. 1.60).

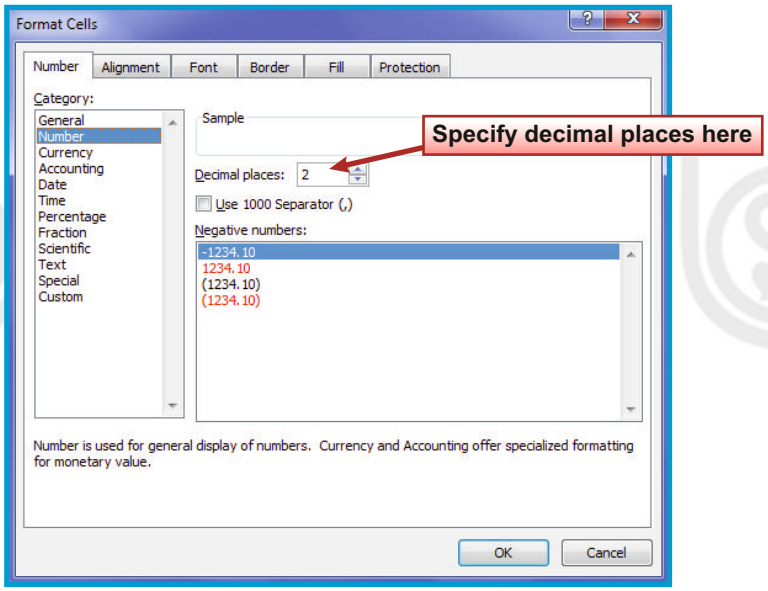


Fig. 1.60

In this dialog box, we can choose the desired number formatting option. When we wish to display data exactly as entered, we use the text format. For example, if we type the class interval "1-2" in Excel sheet, it will change to "01-Feb" by default. In this situation, we use the ***Text*** format to display it as "1-2".

Text format also treats the entered number as the text.

You can also explore other formatting options available in Excel.

iii) Increasing and Decreasing Decimal Places

For the numbers that are already entered on a worksheet, we can increase or decrease the number of places that are displayed after the decimal point for the numeric data using the ***Increase Decimal*** and ***Decrease Decimal*** buttons.

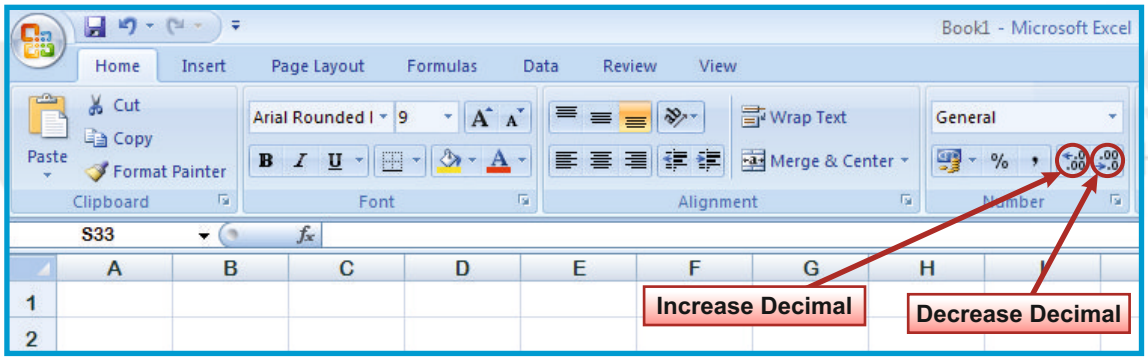


Fig. 1.61

If we wish to increase or decrease the number of decimal places shown in a cell, or in the result of a formula, we

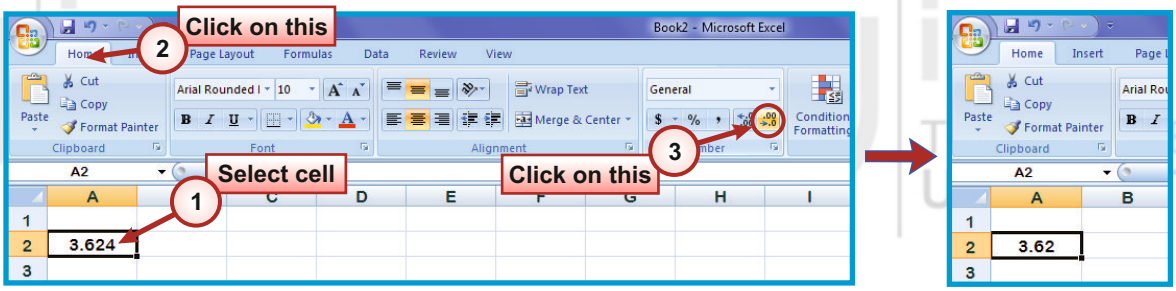
1. select the cell or range that contains the numbers for which we wish to change the number of places after decimal, and

2. click on the ***Increase Decimal*** or ***Decrease Decimal*** buttons in the ***Number*** group under the ***Home*** tab to display one more digit or one less digit after the decimal point.

Note that ***Increase Decimal*** or ***Decrease Decimal*** options add or remove only one digit after the decimal at a time. To add or remove another digit, we have to click on these options again.

For example, suppose we type the value "3.624" in Cell A2. If we wish to decrease the number of places after decimal, i.e., display 3.624 as 3.62, we

1. select cell A2 that we wish to format,
2. click on the ***Decrease Decimal*** button ( ) to decrease the number of places after the decimal as shown in Fig. 1.62a. The result is shown in Fig. 1.62b.



(a) (b)

Fig. 1.62

By default, Excel displays 2 decimal places when we use a built-in ***Number*** format, such as ***Number, Currency, Percentage,*** etc. to the data. We can also change the number of decimal places after applying the number format.

In the same way, we can increase the number of places after decimal by clicking on the ***Increase Decimal*** button ( ) on the ***Number*** group under the ***Home*** tab.

We can also directly specify the decimal places. For this, we click on the arrow at the lower right corner on the ***Number*** group under the ***Home*** tab as shown in Fig. 1.59. The resulting dialog box is shown in Fig. 1.60. The default setting is 2 decimal places.

We can also use the shortcut mouse commands. For this, we select the cell or range, right-click the mouse and choose the ***Format Cells*** option as shown in Fig. 1.23. It opens the same dialog box as shown in Fig. 1.60.

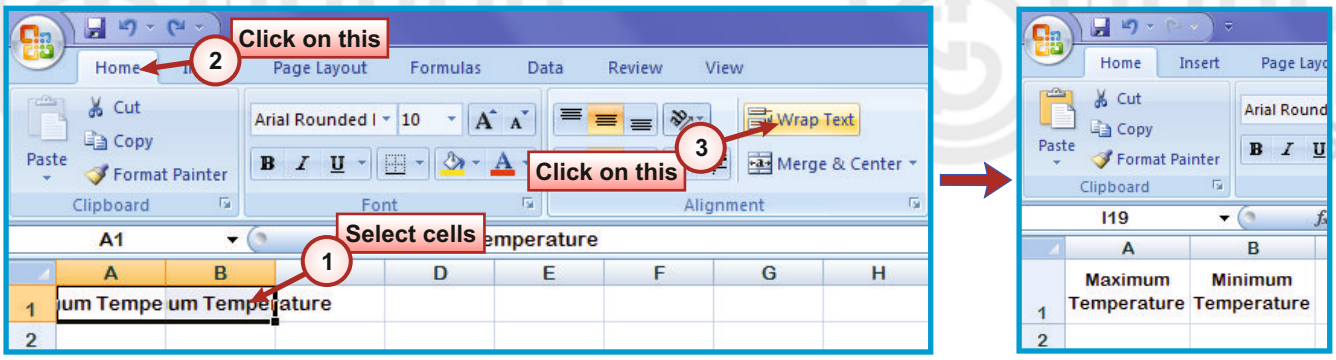
Note that the increase and decrease options **only change the appearance** of the number and **do not change the actual number**. For example, suppose we have number 3.624 and if we decrease one digit after decimal, it appears as 3.62 but the number is still 3.624 and Excel uses 3.624 and not 3.62 for the calculations.

iv) Wrapping the Text

In many situations, the line of the text entered in a cell may be wider than the cell itself. In such situations, the text may be hidden beyond the edges of the cell. Although we can resize the cell or column, we can also use the **wrapping the text** option. To wrap text within the cell, we

1. select the cell or range of cells we wish to format, and
2. click on the ***Wrap Text*** button on the ***Alignment*** group under the ***Home*** tab as shown in Fig. 1.63a.

The wrap text option wraps the text within a cell. It also increases the height of the row automatically. If it does not happen, we adjust the row height as explained in this section. This option shifts the text down according to the column width as shown in Fig. 1.63b.



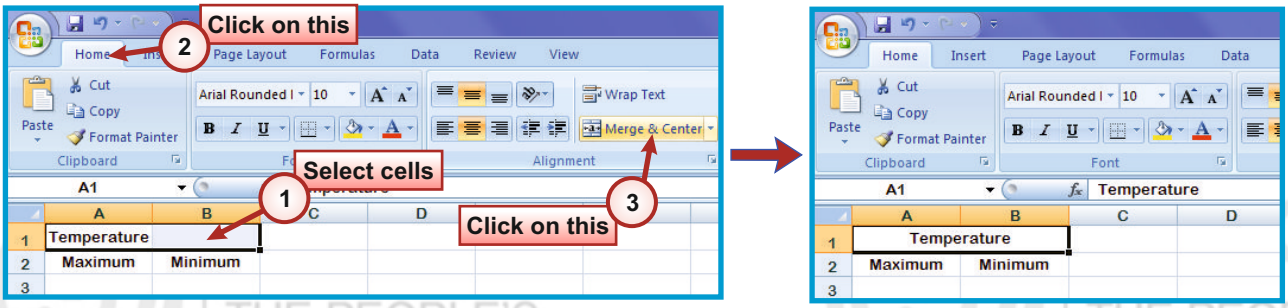
(a) (b)

Fig. 1.63

v) Merging the Cells

We can also merge two or more adjacent cells in Excel. When we merge two or more adjacent horizontal or vertical cells, we get one large cell. To merge cells, we

1. select the cells or range, which we want to merge, and
2. click on the ***Merge & Center*** option on the ***Alignment*** group under the ***Home*** tab as shown in Fig. 1.64a.



(a) (b)

Fig. 1.64

We can also click on the arrow located next to the ***Merge & Center*** button and choose the required merging option from the options shown in Fig. 1.65.

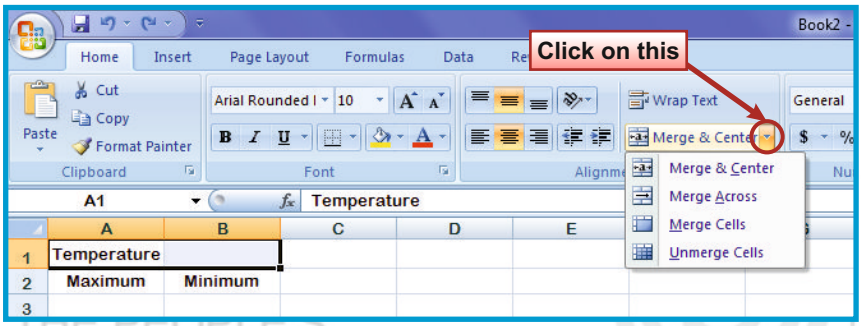


Fig. 1.65

vi) Applying a Background Colour

We can change the background colour of the cells using the ***Fill Color*** option available in Excel to highlight or show that the information given in those cells is important. As shown in Fig. 1.66, to apply a background colour, we

1. select the cell or range on which we wish to apply a background colour,
2. click on the arrow next to the ***Fill Color*** button on the ***Font*** group under the ***Home*** tab (which provides a collection of different background colours), and
3. select the colour we wish to apply on the background of the selected cells.

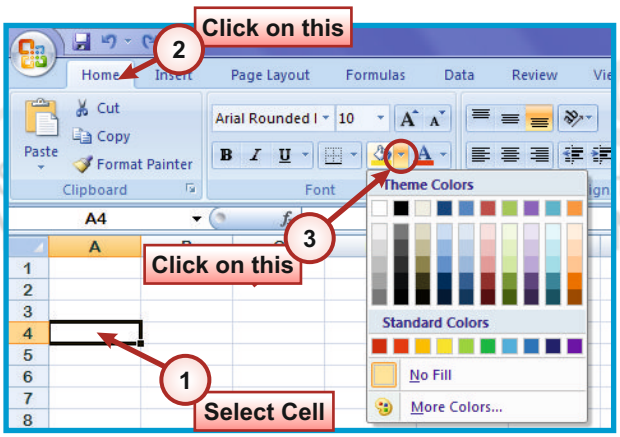


Fig. 1.66

vii) Print Screen Option to Capture a Screenshot

In each session of the lab course, you will need to capture the screen of the required results and charts, which is a snapshot of the monitor screen. The ***Print Screen*** key often abbreviated as ***Print Scrn***, ***Prnt Scrn***, ***Prt Scn***, ***Prt Scr***, ***Prt Sc*** or ***Pr Sc***, etc., is generally placed near the ***Insert*** key on the keyboard and also sometimes shares a key with **System Request** (***SysRq***).

When we press ***Print Screen***, it creates an image of the screen in the ***clipboard***. We then open a word document and click on the ***Paste*** option under the ***Home*** tab of the Word document. It pastes the screenshot in this word document.

1.9 DATA ANALYSIS TOOLPAK

If you are not able to locate ***Analysis Toolpak*** under ***Add-Ins*** category (Fig. 1.68), you should run the setup program for **Excel** or **Microsoft Office** from the CD and reinstall it. The ***Analysis Toolpak*** will appear in ***Add-Ins*** box available after you restart Excel. You may need to activate it.

Since we shall apply many techniques for data analysis in Excel worksheet using the ***Data Analysis ToolPak***, it is necessary to activate it on your computer. The ***Analysis ToolPak*** is an **add-in** of the Microsoft Office Excel which is available when we install Microsoft Office or Excel. ***Data Analysis ToolPak*** is a collection of 19 data analysis tools used for statistical analysis. The analysis tools included in the ***Data Analysis ToolPak*** enable you to analyse the worksheet data using ANOVA, F-test, t-test, regression, correlation, descriptive statistics, etc.

This ***Data Analysis ToolPak*** comes with Excel 2007. But when we use it for the first time, we may have to install it or activate it. To activate the ***Data Analysis ToolPak***, we click on the ***Microsoft Office*** button and then click on ***Excel Options*** button as shown in Fig. 1.67.

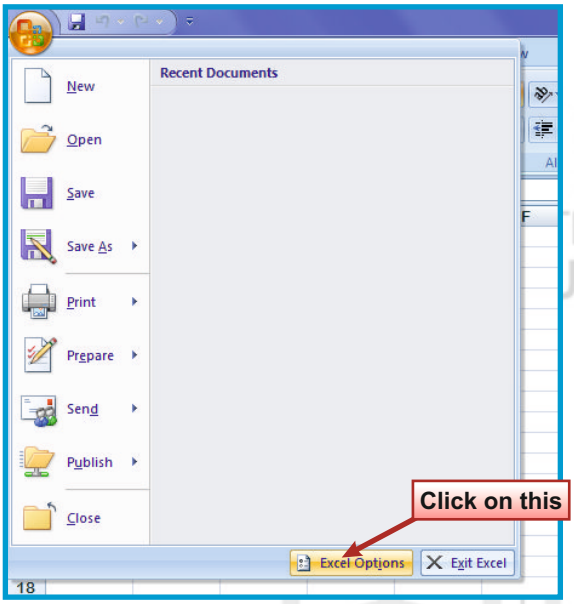


Fig. 1.67

The ***Excel Options*** opens a new dialog box as shown in Fig. 1.68. We click on the ***Add-Ins*** option in the left pane. The ***Add-Ins*** tab contains a list of all add-in programs installed on the computer.

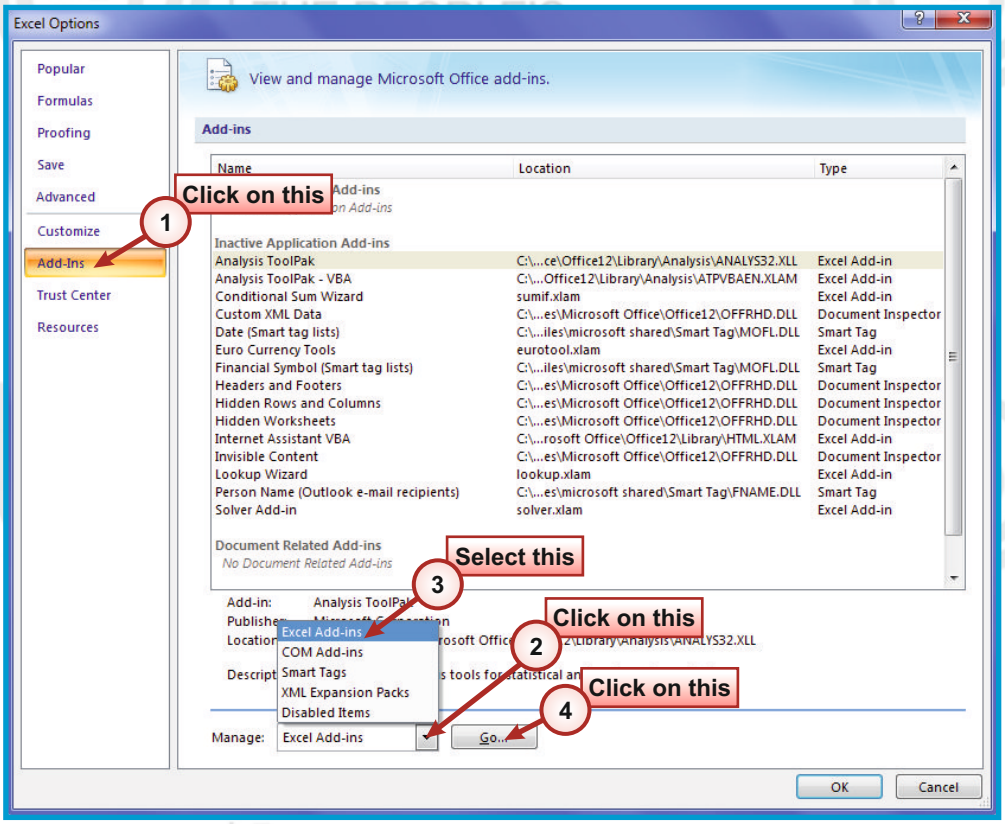


Fig. 1.68

We select ***Excel Add-ins*** from the ***Manage*** box at the bottom and click on ***Go***. It opens a new dialog box as shown in Fig. 1.69.

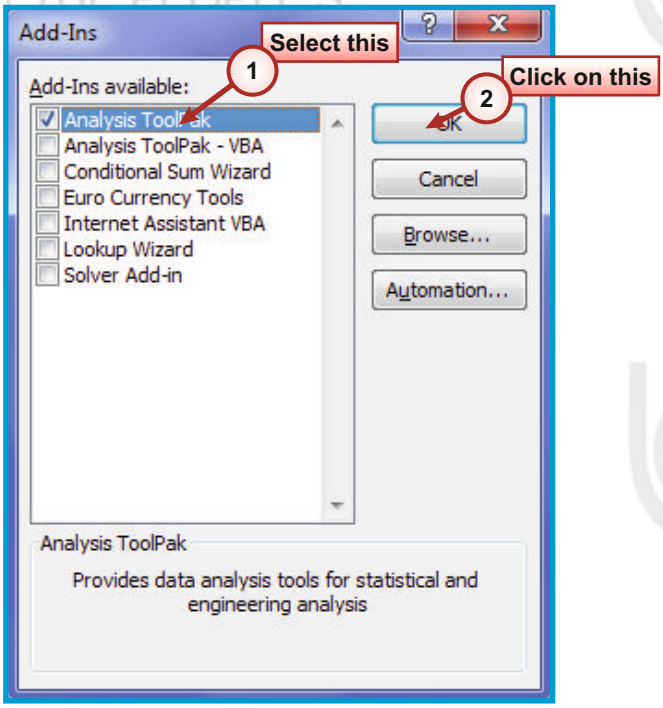
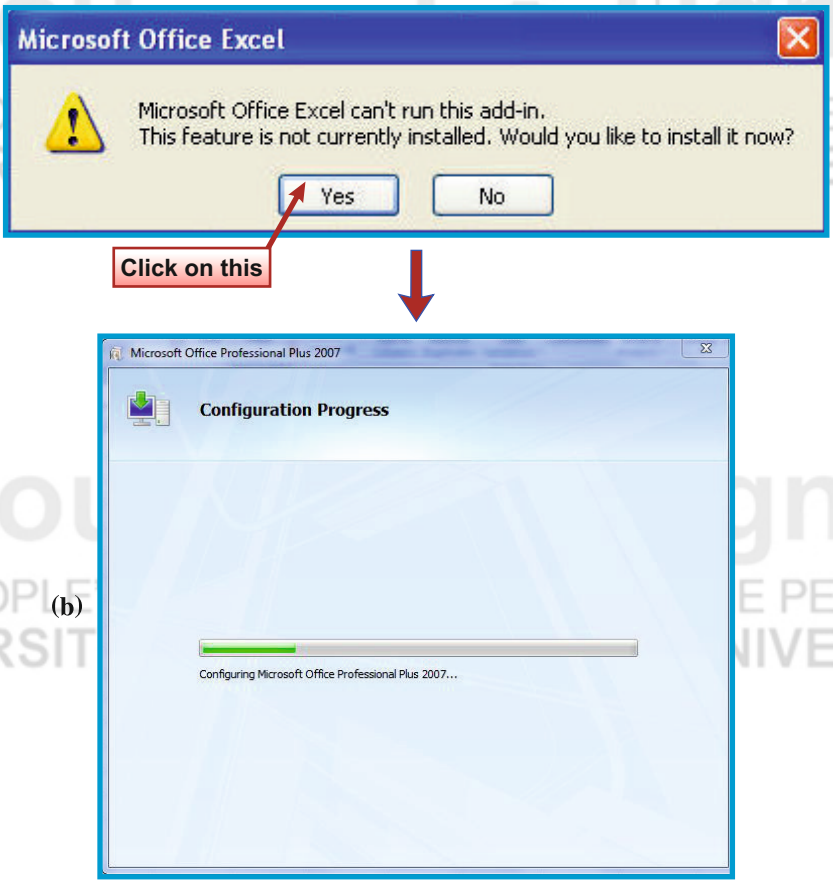


Fig. 1.69

We select the ***Analysis ToolPak*** box in the available ***Add-Ins*** list and then click on the ***OK*** button. If an alerting dialog box appears asking you to install the add-in, click on ***Yes*** and wait for Excel to configure the add-in as shown in Fig. 1.70.

(a)



(b)

Fig. 1.70

After we activate the ***Analysis ToolPak***, the ***Data Analysis*** command will be available in the ***Analysis*** group under the ***Data*** tab as shown in Fig. 1.14.

Continuous Assessment 1

A researcher wished to study the percentage of male and female students in a college. For this, she collected the number of male and female students enrolled in a particular programme of the college for 20 years from 1991 to 2010. The data are recorded in Table 2.

Table 2: Number of male and female students

S.No.	Year	Number of Students	
		Male Students	Female Students
1	1991	321	758
2	1992	296	836
3	1993	375	886
4	1994	402	874
5	1995	331	557
6	1996	387	564
7	1997	442	564
8	1998	378	668
9	1999	435	602
10	2000	498	612
11	2001	425	705
12	2002	447	738

S.No.	Year	Number of Students	
		Male Students	Female Students
13	2003	521	725
14	2004	435	654
15	2005	465	775
16	2006	568	684
17	2007	563	757
18	2008	498	834
19	2009	556	806
20	2010	621	819

- Type the data of Table 2 in Excel in the same format as given.
- Use merging option for typing the text "Number of Students".
- Use the wrapping text option to wrap the phrase "Male Students" and the phrase "Female Students".
- Type 1 and 2 in Cells A3 and A4, respectively, select Cells A3:A4 and drag them down or fill down, up to Cell A22 to fill the serial numbers.
- Apply **Fill down** or **Drag down** option to enter the years.
- Calculate the total number of students for each year.
- Insert a column after "Male Students" and calculate the percentage of male students for each year in the inserted column.
- Insert a column after "Female Students" and calculate the percentage of female students for each year in the inserted column.
- Save this file with the name "**Continuous Assessment 1**" in **D** drive of your computer.
- Rename the sheet as "**Students' Data**".
- Decrease the decimal up to 2 places for the percentage of male and female students.
- Highlight the minimum and maximum percentage of male and female students with light orange and purple colour, respectively.

Home Work: Do It Yourself

1) Follow the steps explained in Secs. 1.2 to 1.9 to learn how to work with MS Excel 2007. Take the final screenshots of the tasks and activities done in Secs. 1.2 to 1.9 and keep them in your record book.

2) Develop the spreadsheet for the "Continuous Assessment 1" as explained in this lab session. Take the screenshots of the final spreadsheet.

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

Classification and Tabulation of Data

SESSION 2

Structure

2.1 Introduction
Objectives
2.2 Problem Description
2.3 Qualitative Classification
2.4 Discrete Frequency Distribution
2.5 Continuous Frequency Distribution

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.

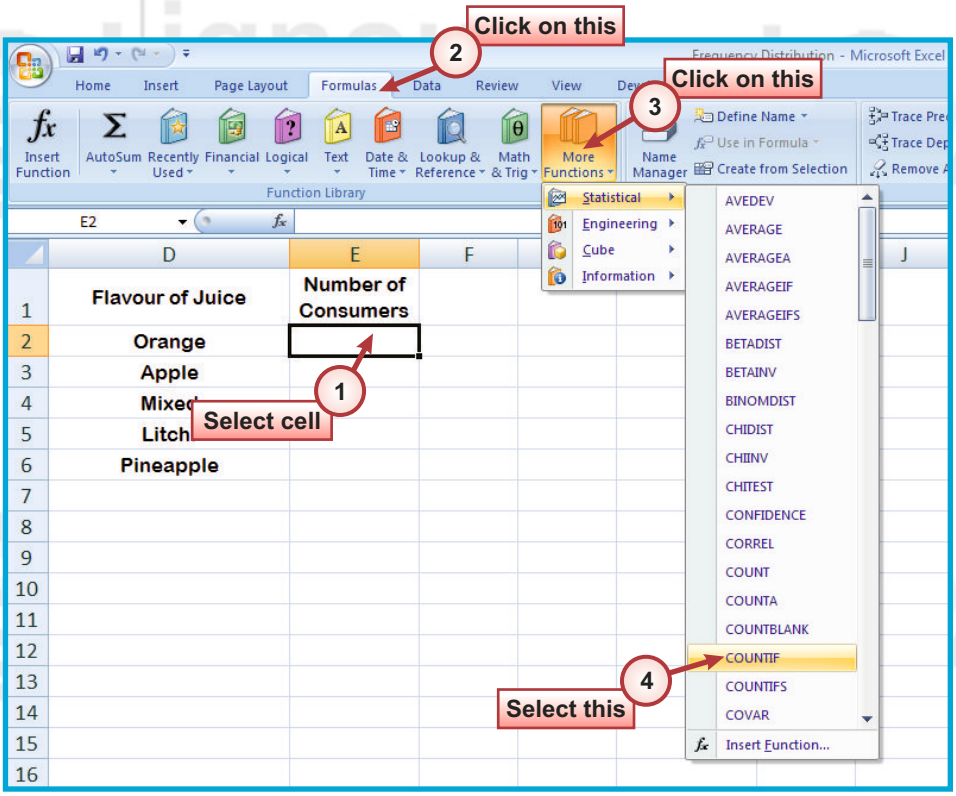


Fig. 2.3

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

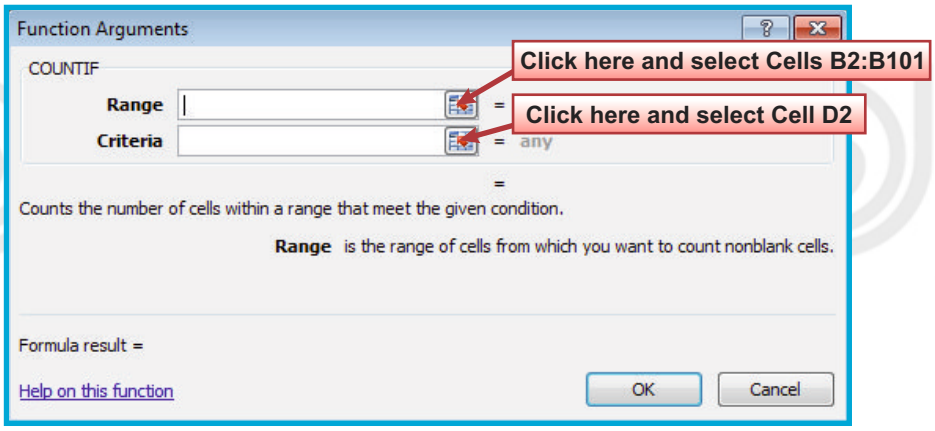
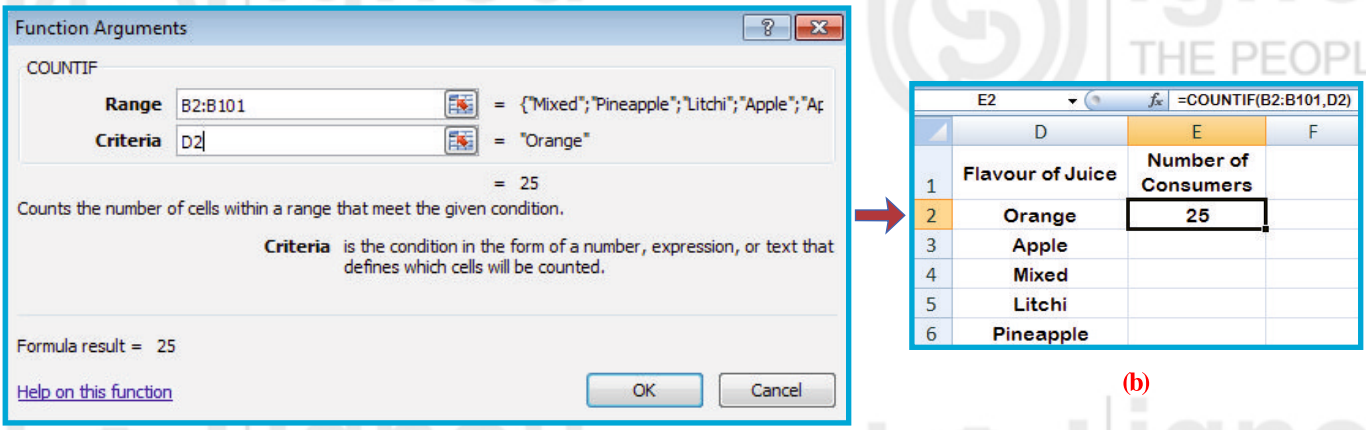


Fig. 2.4

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).



(a) (b)

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.

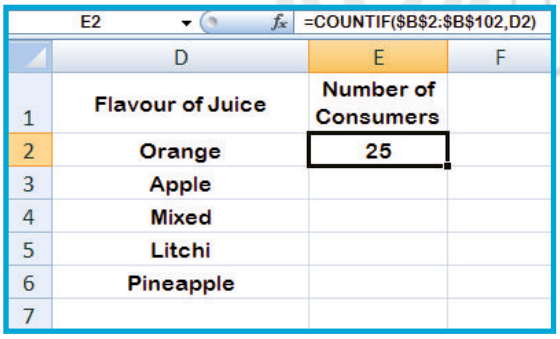


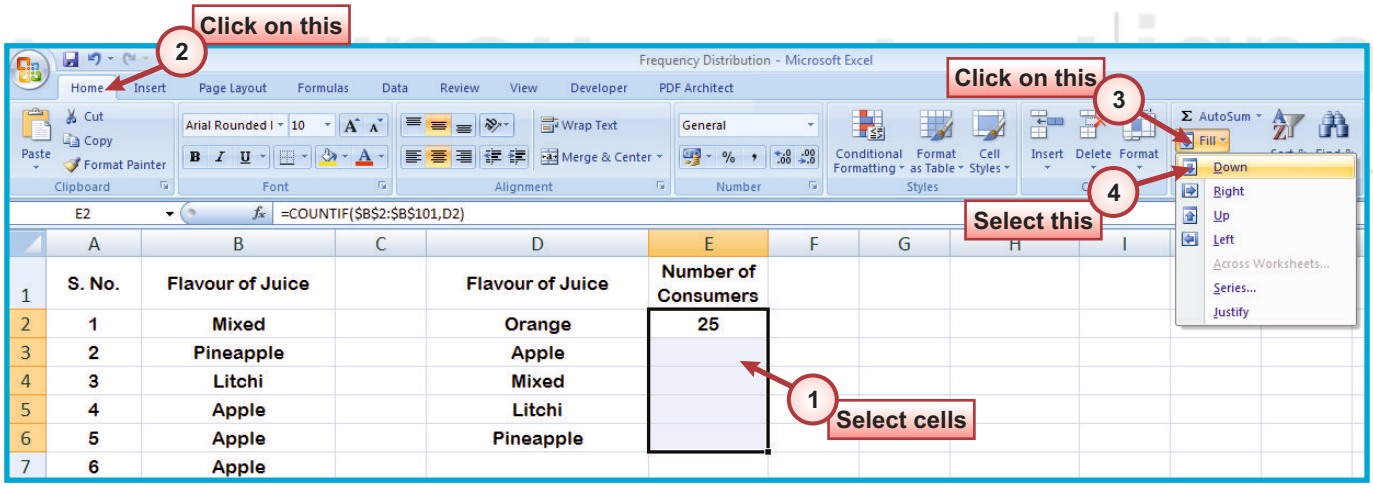
Fig. 2.6

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.

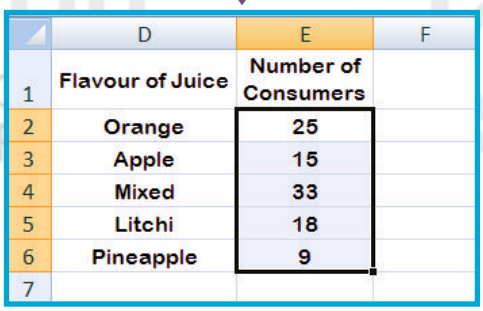
Step 6: To compute the number of consumers who like the remaining flavours, refer to Fig. 2.7a. We

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.



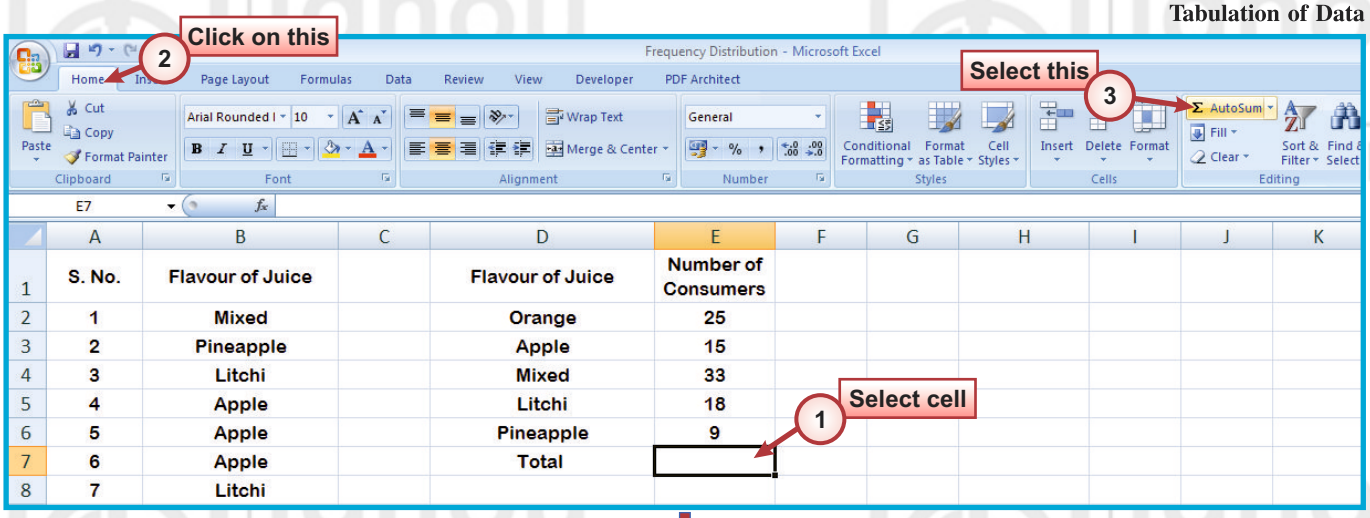
(a)



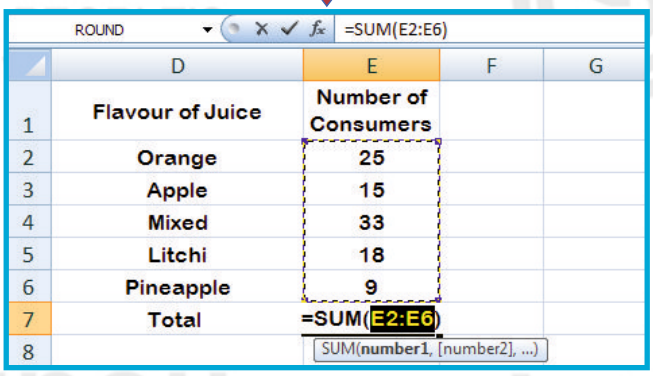
(b)

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.



(a)



(b)

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.

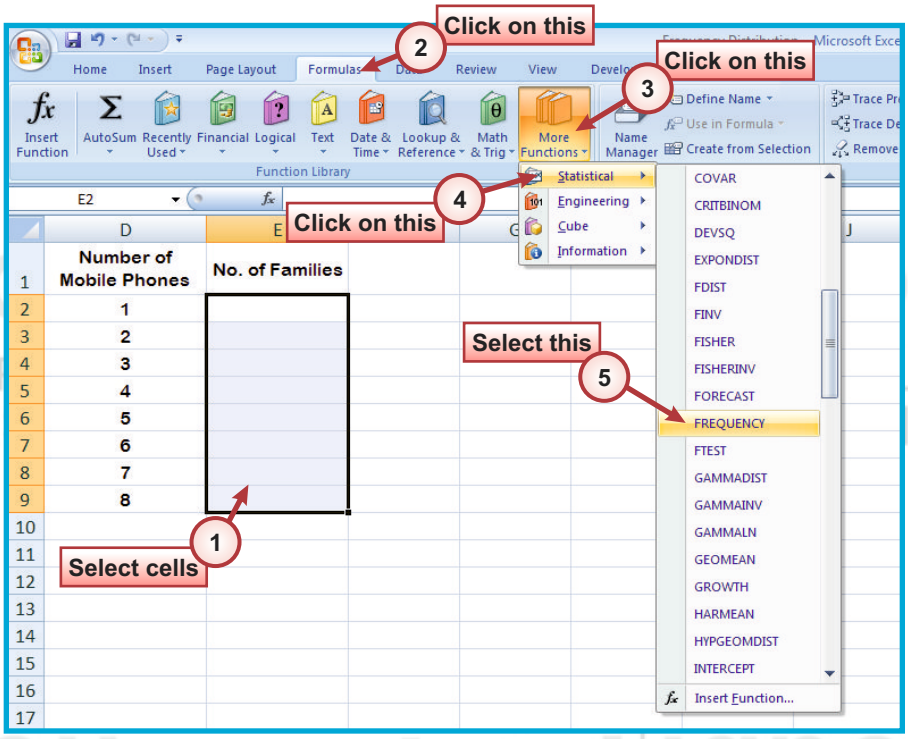


Fig. 2.12

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

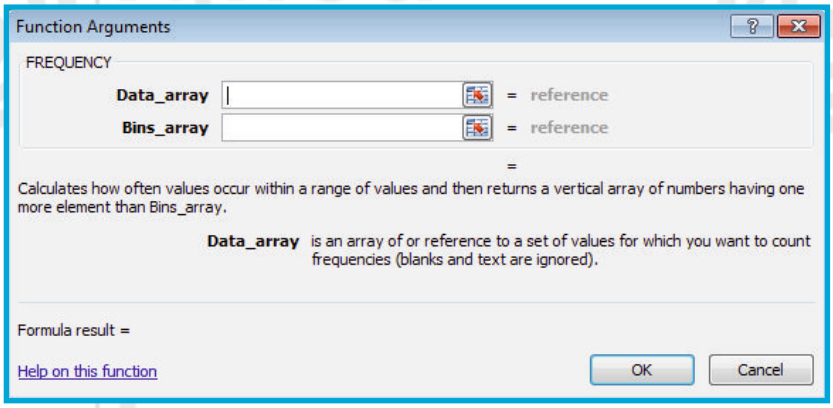


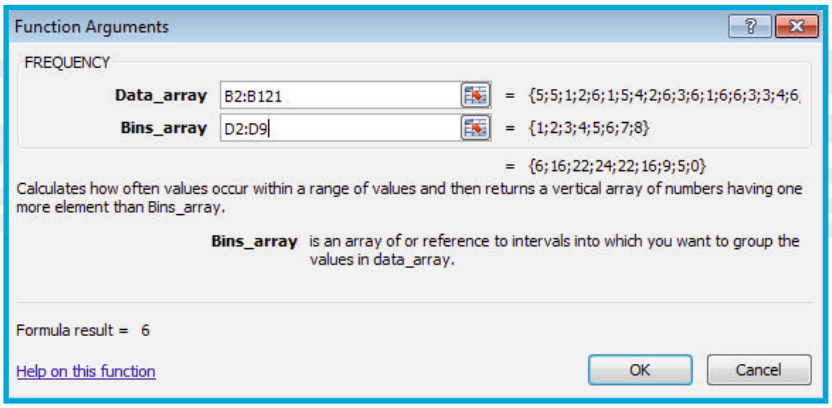
Fig. 2.13

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)}

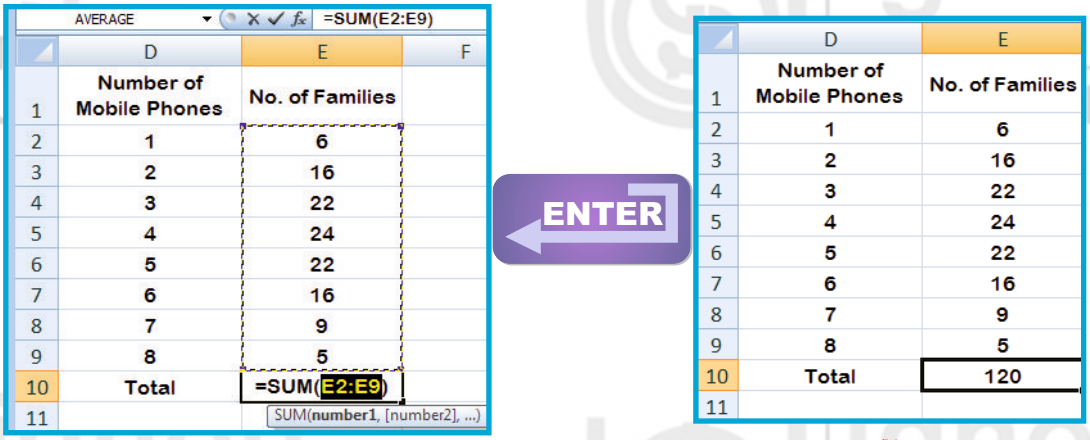
	D	E	F	G
1	Number of Mobile Phones	No. of Families		
2	1	6		
3	2	16		
4	3	22		
5	4	24		
6	5	22		
7	6	16		
8	7	9		
9	8	5		

(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

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

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:

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

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

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

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

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

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***.

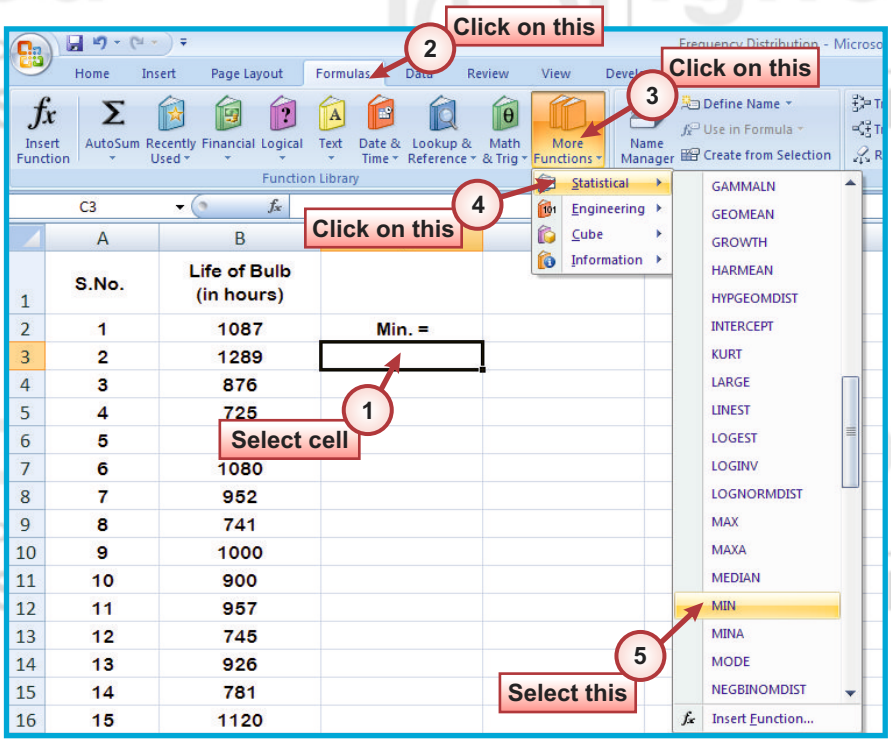
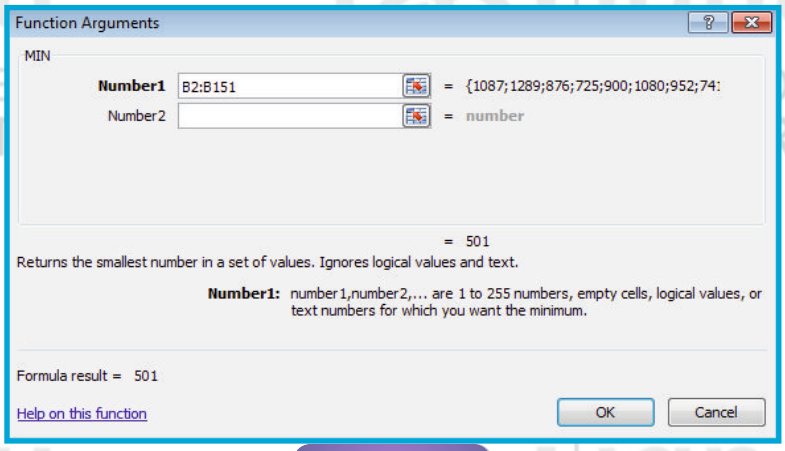


Fig. 2.17

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.

(a)



(b)

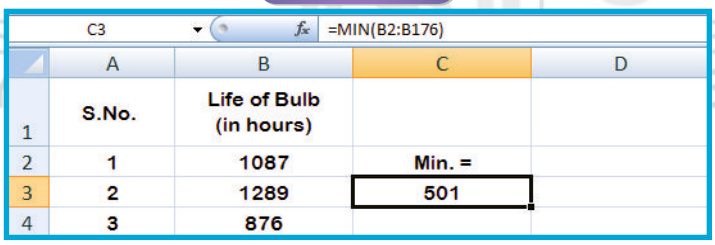


Fig. 2.18

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.

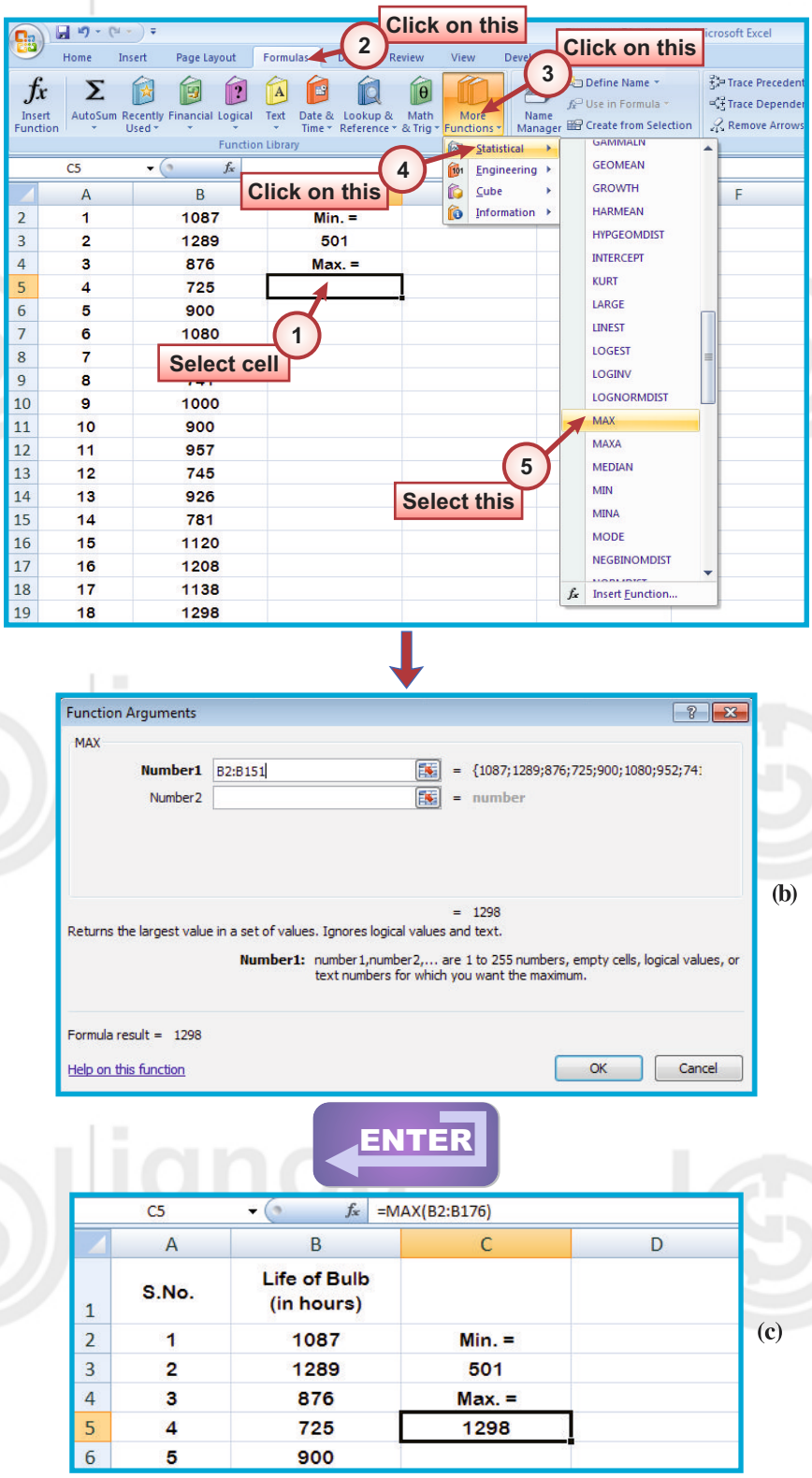


Fig. 2.19

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.

AVERAGE =(C5-C3)/(1+3.322*LOG10(150))

	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)

ENTER

	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			

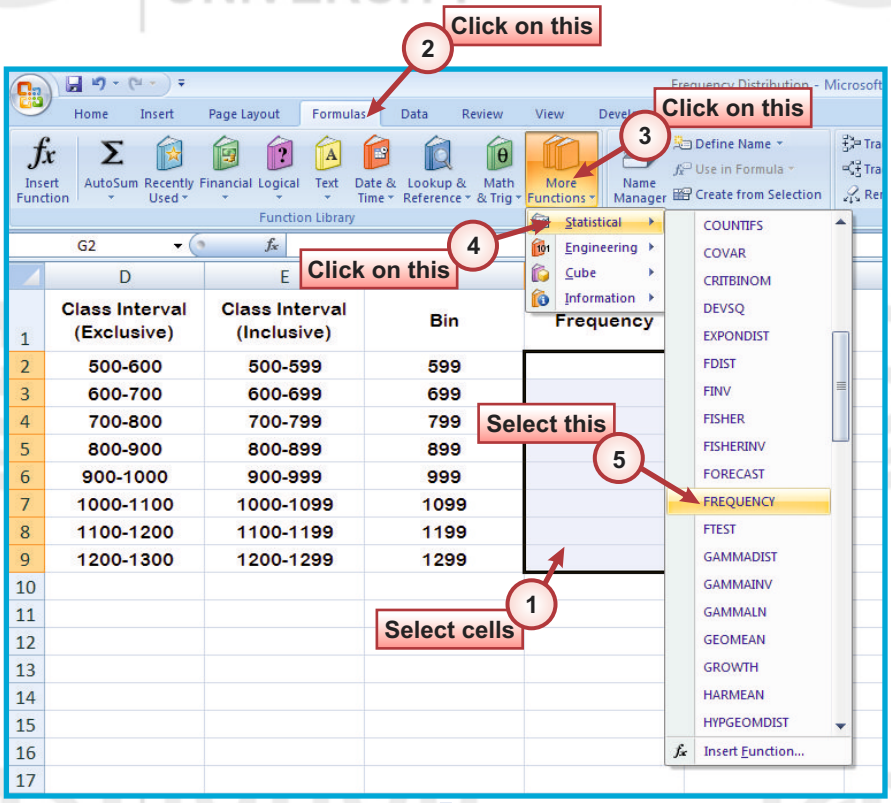
Fig. 2.21

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

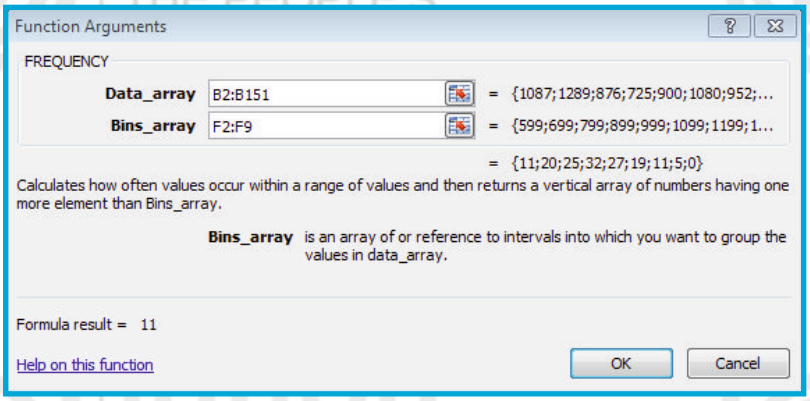
Fig. 2.22

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.



(a)

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.



(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

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.

FREQUENCY | =SUM(G2:G9)

	F	G
1	Bin	No. of Bulbs (Frequency)
2	599	11
3	699	20
4	799	25
5	899	32
6	999	27
7	1099	19
8	1199	11
9	1299	5
10	Total	=SUM(G2:G9)
11		SUM(number1, [number2], ...)

(a)

ENTER

	F	G
1	Bin	No. of Bulbs (Frequency)
2	599	11
3	699	20
4	799	25
5	899	32
6	999	27
7	1099	19
8	1199	11
9	1299	5
10	Total	150
11		

(b)

Fig. 2.24

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**.

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

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.

FREQUENCY | =G2/$G$10

	G	H
1	No. of Bulbs (Frequency)	Relative Frequency
2	11	=G2/$G$10
3	20	
4	25	
5	32	
6	27	
7	19	
8	11	
9	5	
10	150	

(a)

ENTER

H2 | =G2/$G$10

	G	H
1	No. of Bulbs (Frequency)	Relative Frequency
2	11	0.0733
3	20	
4	25	

(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.

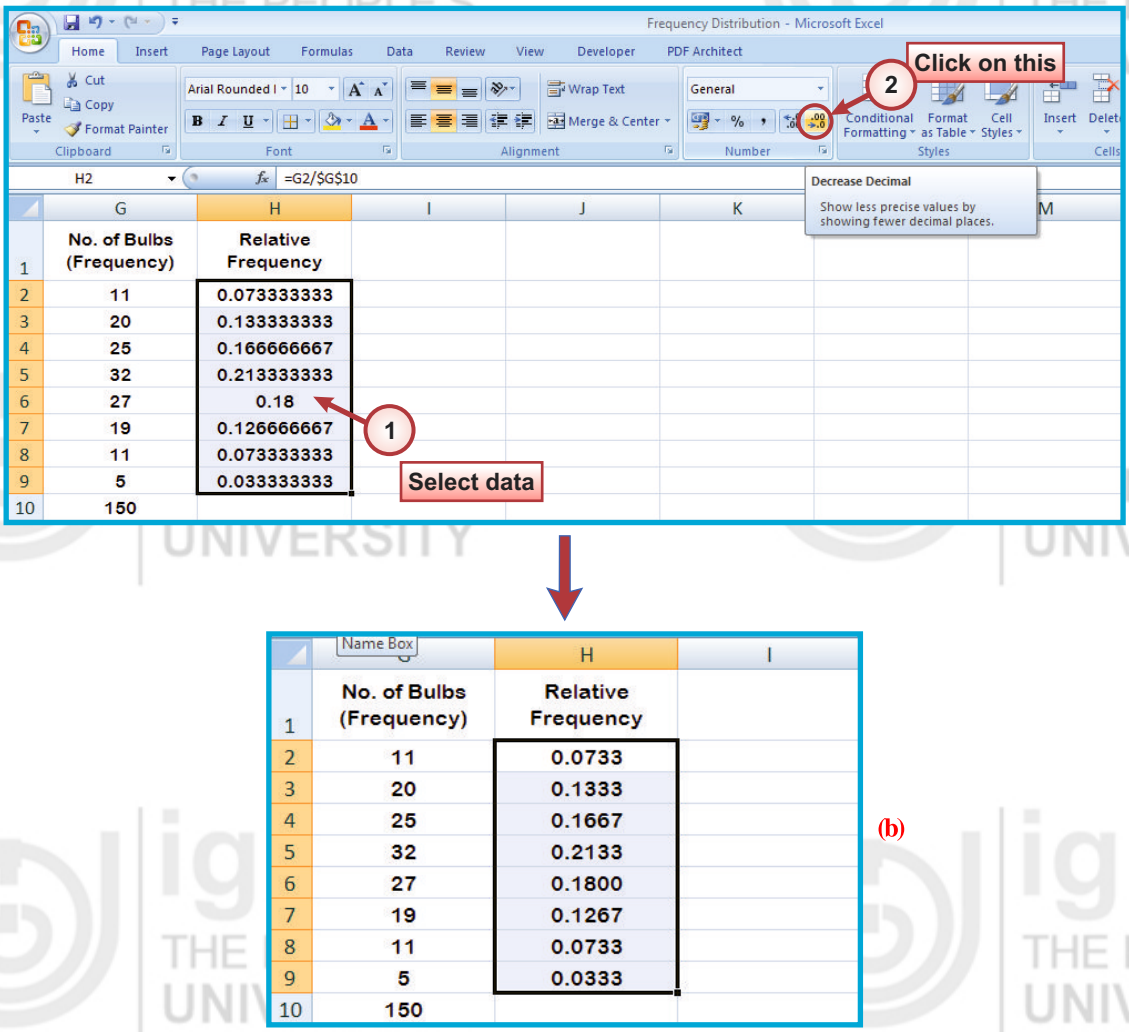


Fig. 2.27

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).

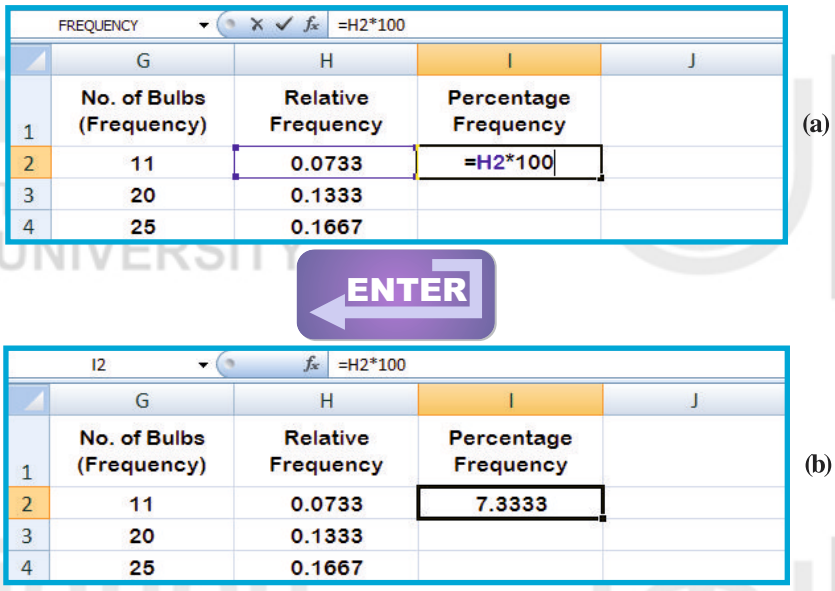


Fig. 2.28

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.

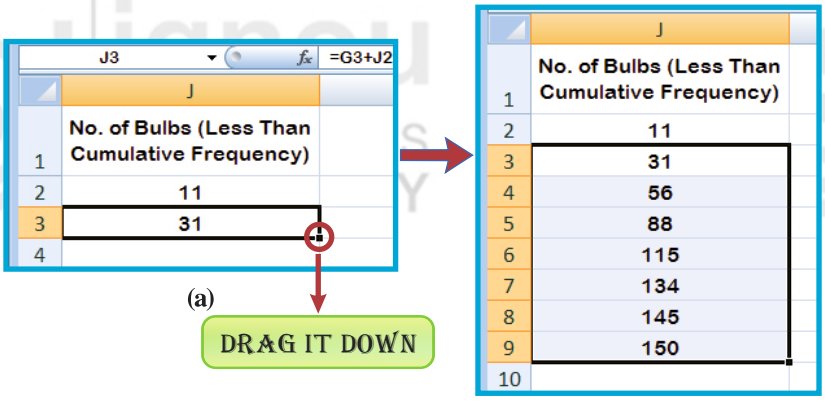


Fig. 2.32 (b)

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)

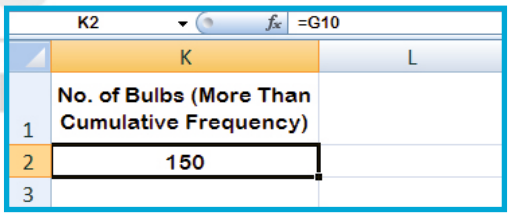
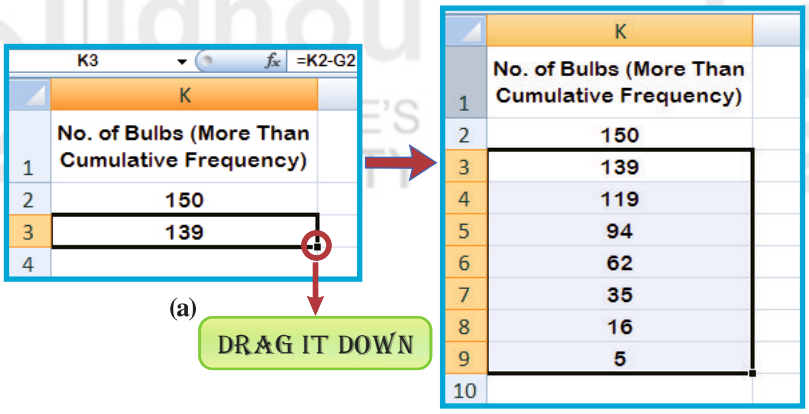


Fig. 2.33

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.



(b)

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.

Diagrammatic Presentation of Data

SESSION 3

Structure

3.1 Introduction
Objectives
3.2 Problem Description
3.3 Simple Bar Diagram
3.4 Multiple Bar Diagram
3.5 Sub-divided Bar Diagram
3.6 Percentage Bar Diagram
3.7 Pie Diagram

3.1 INTRODUCTION

In Lab Session 2, you have learnt the classification of qualitative data, formation of discrete and continuous frequency distributions to classify the raw data so that we can extract the maximum information from the data. In Unit 14 of MST-001 (Foundation in Mathematics and Statistics), you have learnt how to construct different types of bar diagrams (such as simple, multiple and sub-divided bar diagrams). You have also learnt about the pie diagram, which can be used to represent data in an attractive and easily understandable way.

Prerequisite

- Lab Session 1 of MSTL-001 (Basic Statistics Lab).
- Unit 14 of MST-001 (Foundation in Mathematics and Statistics).

You will agree with the saying that a picture is worth a thousand words. Diagrammatic representation of data conveys information better than numerical data and helps in understanding facts at a glance. Therefore, in this lab session, you will learn how to construct different types of bar diagrams and pie diagram using Excel 2007 for pictorial representation of data.

In the next lab session, we shall discuss graphical representation of data in the form of frequency distributions.

Some authors use the word "chart" instead of "diagram". You may use whichever term you like.

Objectives

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

- prepare the spreadsheet in MS Excel 2007;
- construct simple, multiple and sub-divided bar diagrams;
- construct pie diagram; and
- format the bar and pie diagrams.

3.2 PROBLEM DESCRIPTION

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

1. A survey was conducted among 100 consumers to ascertain which flavour of juice was preferred by them. The data are recorded in Table 1.

Table 1: Choice of consumers regarding flavour of juice

Flavour of Juice	Number of Consumers
Orange	25
Apple	15
Mixed Fruits	33
Litchi	18
Pineapple	9

Draw a bar diagram for the given data.

2. A survey was conducted among 100 consumers to find out which brand (A, B or C) was preferred by them for various flavours of juices. The data are recorded in Table 2.

Table 2: Choice of consumers regarding brand and flavour of juice

Flavour of Juice	No. of Consumers		
	A	B	C
Orange	25	17	28
Apple	15	13	8
Mixed Fruits	33	23	14
Litchi	18	22	14
Pineapple	9	5	6

Draw the multiple and sub-divided bar diagrams to compare the three brands of different flavours of juices.

3. Crops of cereals, pulses, vegetables, fruits, flowers and other plants are planted in two fields A and B of 1000 sq.ft. each but the area under each crop is different for both fields. The data for area under each crop for both fields are given in Table 3. Draw the percentage bar and pie diagrams to compare the cropping patterns of the fields.

Table 3: Area under crops

Crops	Area under Crop (in sq. ft.)	
	A	B
Cereals	235	95
Pulses	180	50
Vegetables	320	360
Fruits	150	250
Flowers	80	165
Others	35	80

3.3 SIMPLE BAR DIAGRAM

You have learnt in Unit 14 of MST-001 that the simple bar diagram is the most commonly used diagram. We can represent the data based on only one variable using a bar diagram. The bars can be drawn either vertically or horizontally. We briefly mention the main steps as follows:

Step 1: We draw the horizontal (X) axis and the vertical (Y) axis.

Step 2: We take the categories of the variable on the X-axis and the magnitude of the data along the Y-axis if we draw the bars vertically.

Step 3: We draw the vertical bars of the same width and the height of the bars represents the magnitude of the corresponding category. These bars are separated from each other by equal intervals.

We take the magnitude of the data along the horizontal axis (X -axis) if we draw the bars horizontally.

Steps in Excel

You have learnt about the manual plotting of simple bar diagram in Unit 14 of MST-001. Here we explain how MS Excel 2007 is used to plot a simple bar diagram to represent the data given in Problem 1 as follows:

Step 1: We enter the data given in Table 1 in Excel 2007 spreadsheet as shown in Fig. 3.1.

	A	B	C
1	Flavour of Juice	No. of Consumers	
2	Orange	25	
3	Apple	15	
4	Mixed Fruits	33	
5	Litchi	18	
6	Pineapple	9	

Fig. 3.1: Partial screenshot of the spreadsheet for the given data.

Step 2: To obtain the bar chart, we

1. select Cells A2:B6 (see Fig. 3.2),
2. click on the *Insert* tab,
3. click on the *Column* option in the *Charts* group, and
4. select a chart subtype *Clustered Column* that we wish to use.

If the value of the variable is negative, the bar goes below the axis (other side of the axis).

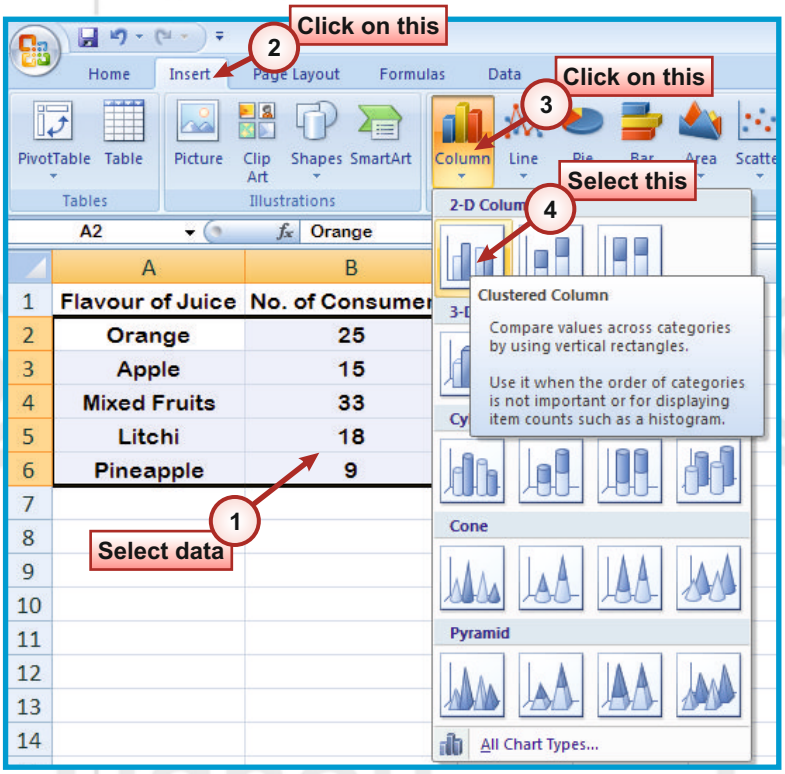


Fig. 3.2

Step 3: The resulting chart is shown in Fig. 3.3 and is called the bar chart. The five flavours given in Column A of Fig. 3.1 provide the five bars while the values in Column B present the height of each corresponding bar (Fig. 3.3).

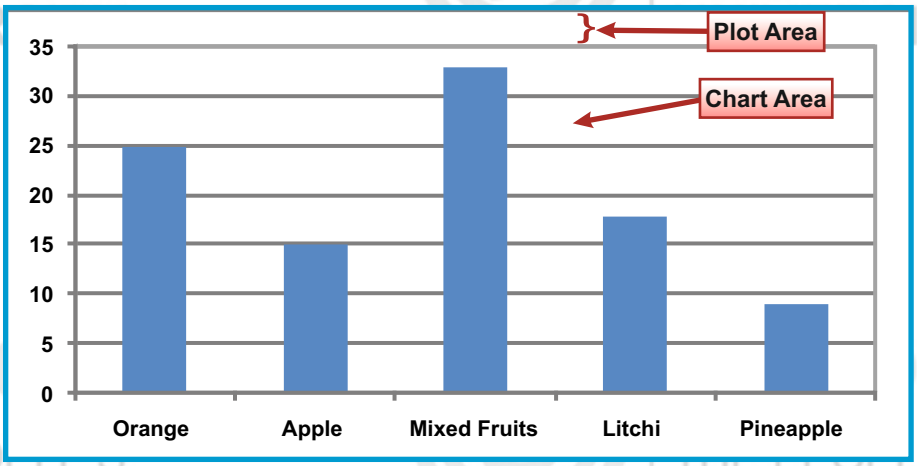


Fig. 3.3

Step 4: Fig. 3.3 shows the chart that we get by default using Excel. Using different features of Excel 2007 to format the chart, you can change the look of the chart. You may try to explore these features yourself to represent the chart in different and attractive ways. For example, you may like to

- ✓ eliminate the grid lines and the border around the chart,
- ✓ change the colour of the bars,
- ✓ change the background colour of the plot area and chart area,
- ✓ add the horizontal and vertical axis titles, etc., as desired.

We discuss some of these formats one at a time. Refer to Fig. 3.4. To eliminate the gridlines, you should

1. click on the chart area shown in Fig. 3.4,
2. click on the ***Layout*** tab under ***Chart Tools*** (see Fig. 3.4), and
3. click on the ***Gridlines*** → ***Primary Horizontal Gridlines*** → ***None*** as shown in Fig. 3.4.

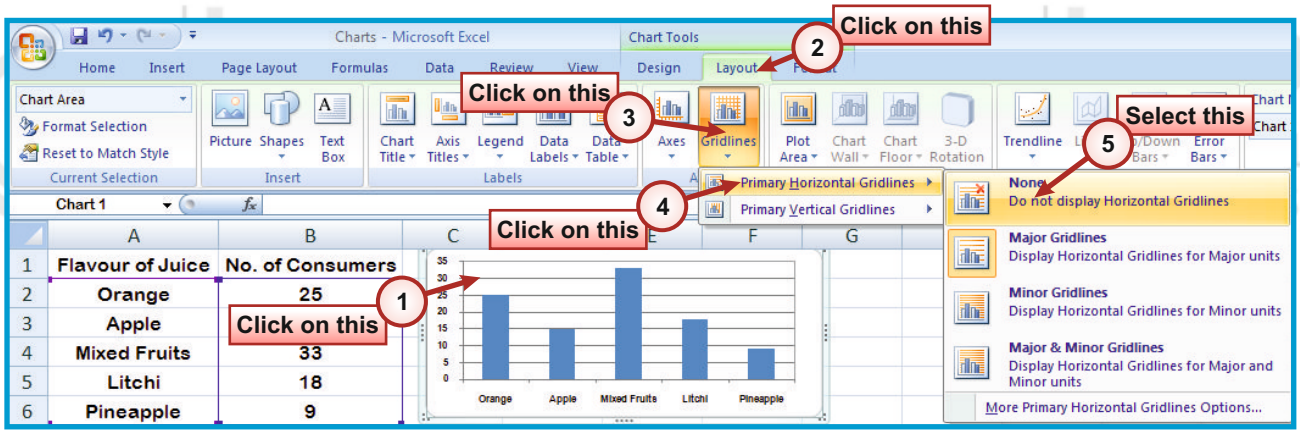


Fig. 3.4

Step 5: To change the colour of the bars, refer to Fig. 3.5. You should

1. click on the bars,

2. click on the *Format* tab under *Chart Tools*,
3. click on the *Shape Fill*, and
4. choose the colour you want. Here we have chosen the purple colour.

For exploring more colours, you can also choose the option *More Fill Colors* (Fig. 3.5).

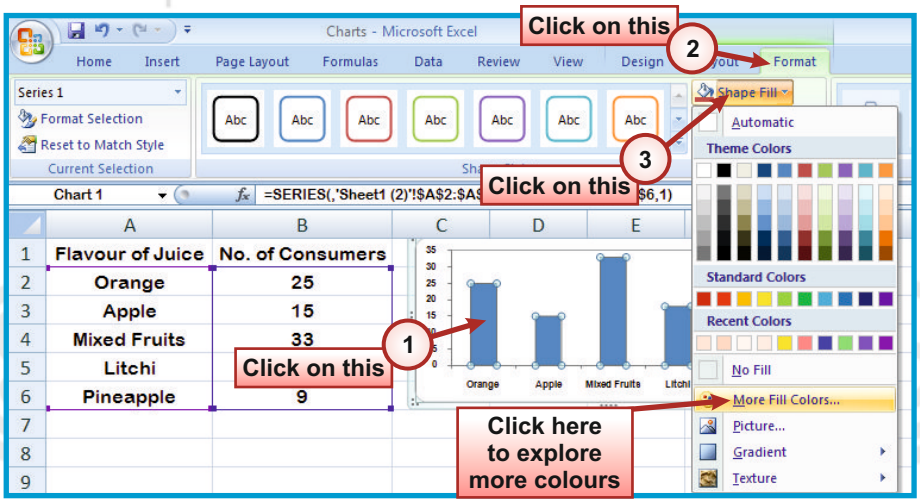


Fig. 3.5

Step 6: To change the colour of the plot area and the chart area, you should select the plot area and chart area of the chart one at a time and format them as explained in Step 5.

Step 7: To eliminate the border around the chart, refer to Fig. 3.6. You should

1. click on the chart,
2. click on the *Format* tab under *Chart Tools*,
3. click on the *Shape Outline*, and
4. choose *No Outline* option.

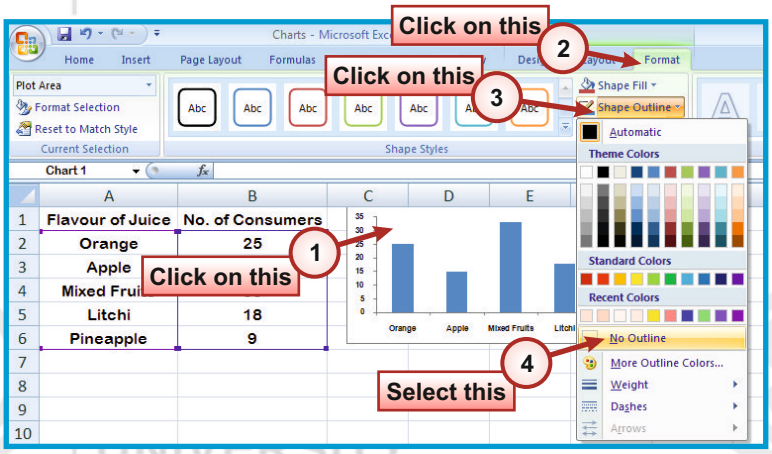


Fig. 3.6

Step 8: To add the horizontal axis title, you should

1. click on the chart,
2. click on the *Layout* tab under *Chart Tools* as shown in Fig. 3.7, and
3. click on the *Axis Titles* → *Primary Horizontal Axis Title* → *Title Below Axis*. You can type "**Flavour of Juice**" as the title of the horizontal axis or some other title of your choice.

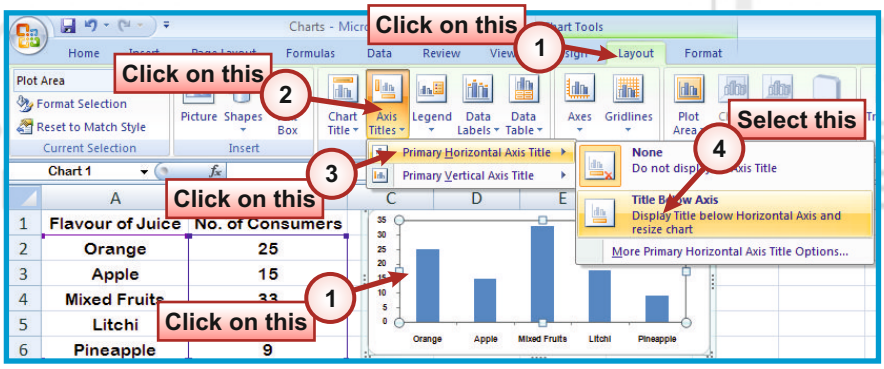


Fig. 3.7

Step 9: To add the vertical axis title, you should

1. click on the chart,
2. click on the *Layout* tab under *Chart Tools* (Fig. 3.8), and
3. click on the *Axis Titles* → *Primary Vertical Axis Title* → *Rotated Title*. You can type "**Number of Consumers**" as the title of the vertical axis or some other title of your choice.

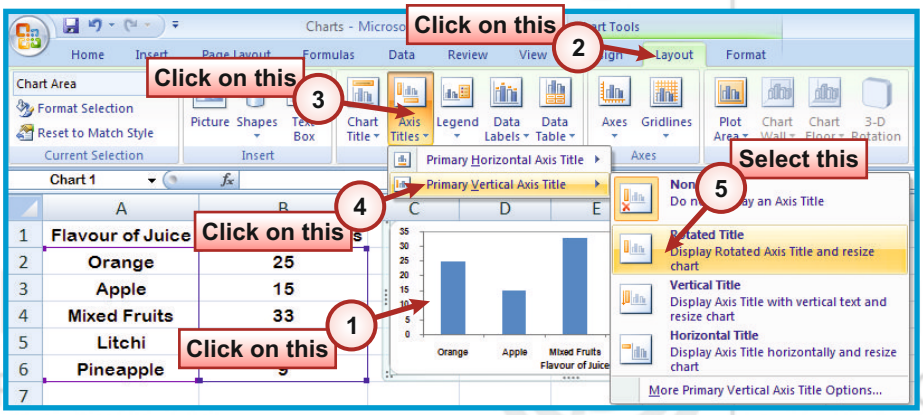


Fig. 3.8

Step 10: There are many other options in Excel 2007, which you may like to explore. If we format the chart as described in Steps 4 to 9, we get the simple bar diagram shown in Fig. 3.9.

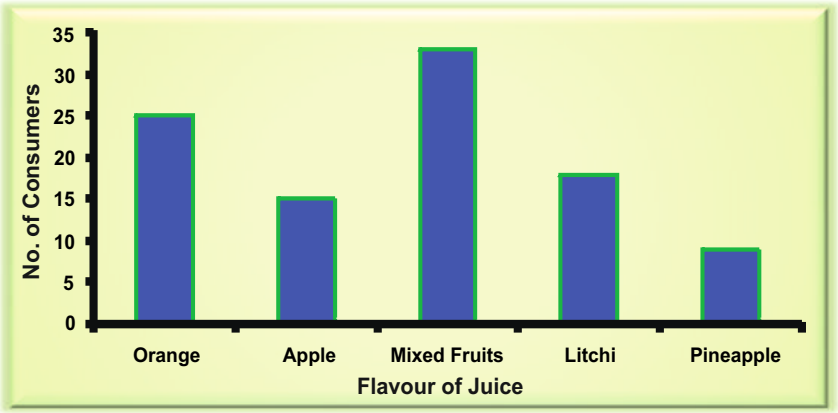


Fig. 3.9: Simple bar diagram

Step 11: Excel 2007 also provides the bar diagram in the 3D format. To obtain the 3D bar diagram, refer to Fig. 3.10. You should

1. select the data given in Cells A2:B6,
2. click on the *Insert* tab,
3. click on the *Column* option in the *Charts* group, and
4. select a chart subtype *3-D Clustered Column*.

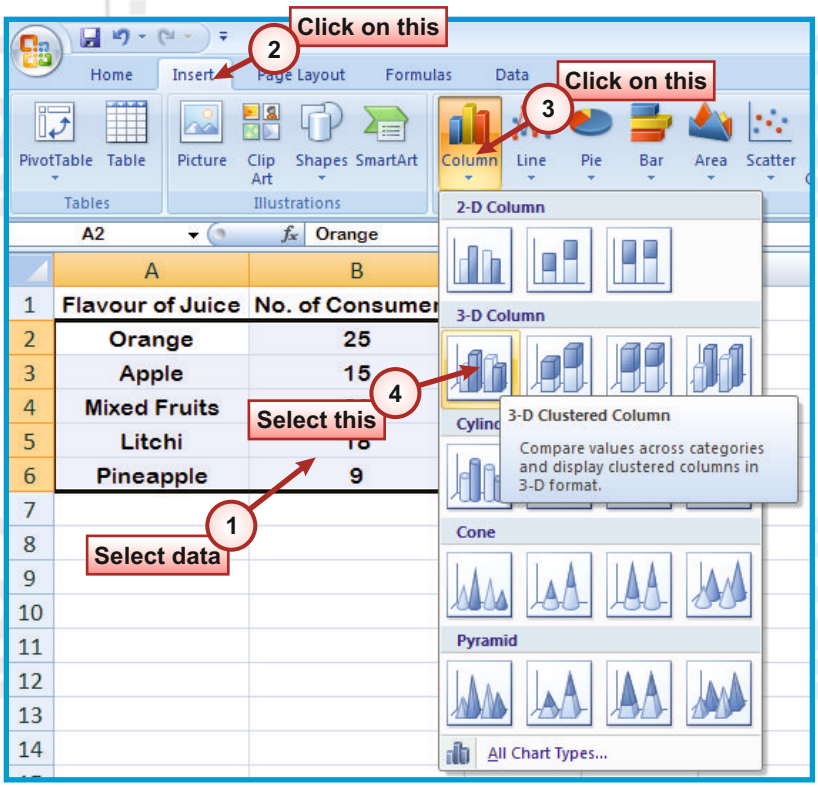


Fig. 3.10

Step 12: If we format the 3D simple bar diagram as explained in Steps 4 to 9, we get the pictorial representation shown in Fig. 3.11.

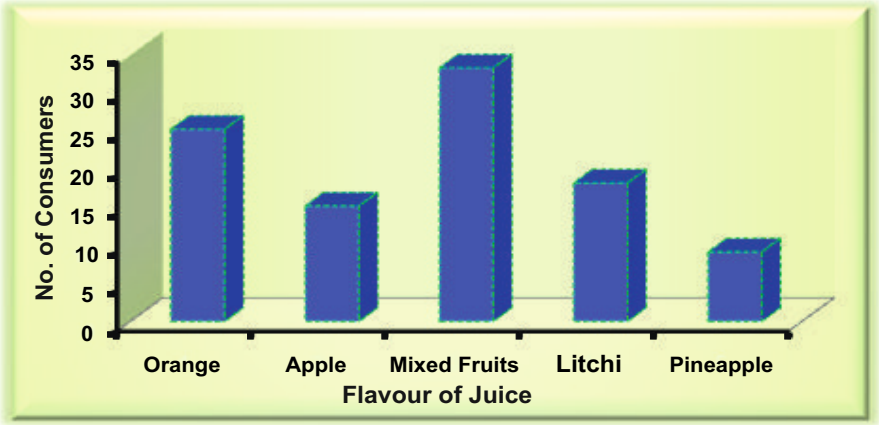


Fig. 3.11

3.4 MULTIPLE BAR DIAGRAM

You have learnt in Unit 14 of MST-001 that we use the multiple bar diagram to compare two or more interrelated variables and also the magnitudes of the variables in two or more aspects or categories. The method of plotting the multiple bar diagram is the same as the one for a simple bar diagram. Here we have separate adjacent bars to represent the categories of each variable. We briefly mention the main steps as follows:

Step 1: We draw the horizontal (X) and the vertical (Y) axes.

Step 2: We take the categories of the variables on the X-axis and the range of the data on the Y-axis.

Step 3: For each category of all variables, we draw vertical bars corresponding to their data ranges. In multiple bar diagram, the bars

corresponding the same category of the variables are placed adjacent to each other.

Step 4: We use different colours, patterns, etc., to distinguish the bars representing different variables.

Steps in Excel

In Problem 2, we have three brands of juice (A, B and C) and each brand has five categories of flavours. Therefore, we can represent the data of Table 2 to compare three brands of juice using a multiple bar diagram. You have learnt about the manual plotting of multiple bar diagram in Unit 14 of MST-001. We now describe the procedure of plotting the multiple bar diagram using MS Excel 2007.

Step 1: We enter the data given in Table 2 in Excel 2007 spreadsheet as shown in Fig. 3.12.

	A	B	C	D	E
1		No. of Consumers			
2	Flavour of Juice	A	B	C	
3	Orange	25	17	28	
4	Apple	15	13	8	
5	Mixed Fruits	33	23	14	
6	Litchi	18	22	14	
7	Pineapple	9	5	6	
8					

Fig. 3.12: Screenshot of the spreadsheet for the given data.

Step 2: The steps used to obtain the multiple bar diagram are shown in Fig. 3.13. We

1. select the data given in Cells A2:D7,
2. click on the *Insert* tab,
3. click on the *Column* option in the *Charts* group, and
4. select a chart subtype *3-D Clustered Column* that we wish to use.

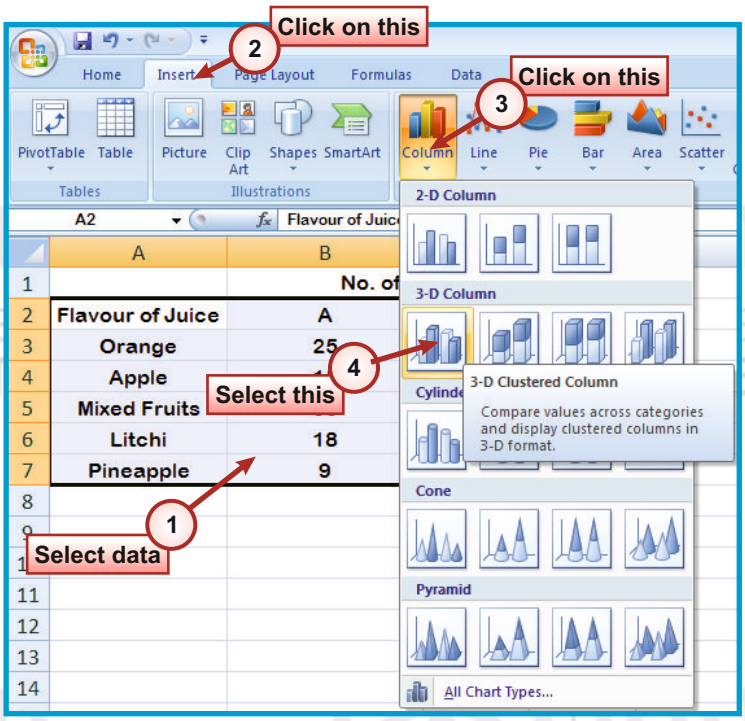


Fig. 3.13

Step 3: We format the chart as explained in Steps 4 to 9 of Sec. 3.3 and obtain the multiple bar diagram shown in Fig. 3.14.

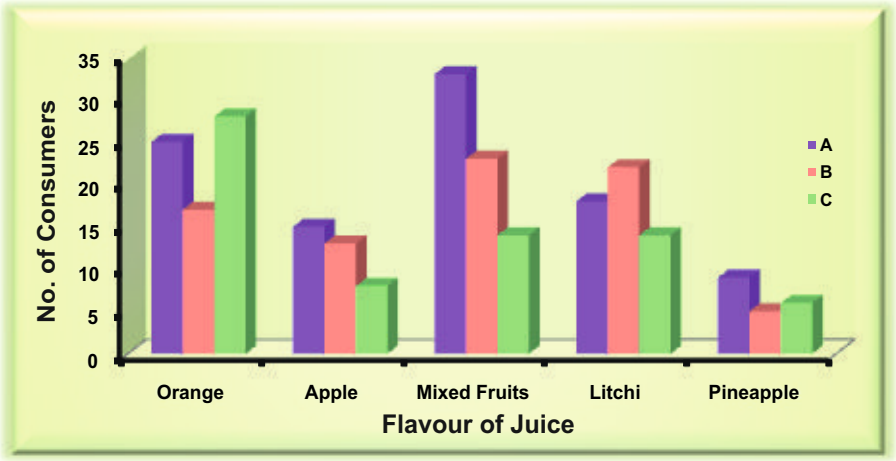


Fig. 3.14: Multiple bar diagram.

3.5 SUB-DIVIDED BAR DIAGRAM

You have learnt in Unit 14 of MST-001 that the sub-divided bar diagram is another way of representing and comparing two or more variables as well as the breakup of one variable into several components. It is also known as **component bar diagram**. We briefly mention the main steps as follows:

Step 1: We draw the horizontal (X) and the vertical (Y) axes.

Step 2: We take the categories of the variables on the X-axis and the range of the data on the Y-axis.

Step 3: We draw the vertical bars for each category of variables. In the sub-divided bar diagram, the height of each bar represents the total magnitude or value of the variable.

Step 4: We divide each bar in several parts or components according to the share of each component of the variable.

Step 5: We use different colours, patterns, etc., to distinguish each part of the bar representing different components of the variable.

Steps in Excel

In Problem 2, we have three brands of juice (A, B and C) and each brand has five categories of flavours. We can also represent the data of Table 2 to compare the breakup of each flavour into three brands using a sub-divided bar diagram. You have learnt about the manual plotting of sub-divided bar diagram in Unit 14 of MST-001. Here we describe the procedure of plotting the sub-divided bar diagram in MS Excel 2007:

Step 1: The steps used to obtain the sub-divided bar diagram are shown in Fig. 3.15. We

1. select the data given in Cells A2:D7 (Fig. 3.15),
2. click on the *Insert* tab,
3. click on the *Column* option in the *Charts* group, and
4. select a chart subtype *Stacked Column in 3-D* that we wish to use.

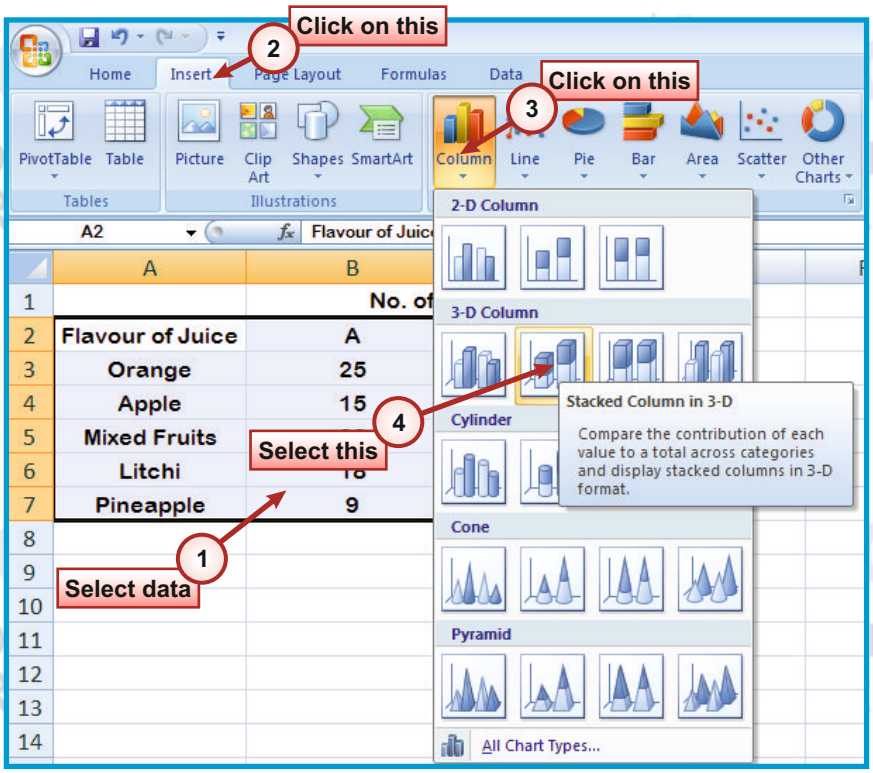


Fig. 3.15

Step 2: We format the chart as explained in Steps 4 to 9 of Sec. 3.3. The resulting sub-divided bar diagram is shown in Fig. 3.16.

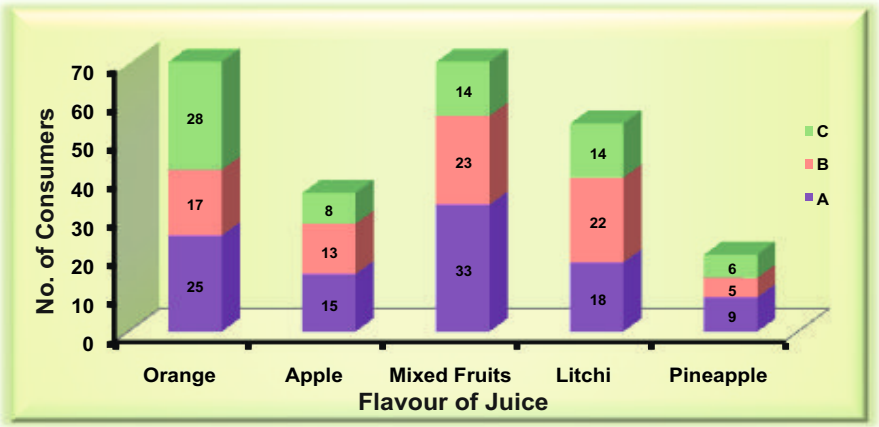


Fig. 3.16: Sub-divided bar diagram.

3.6 PERCENTAGE BAR DIAGRAM

You have learnt in Unit 14 of MST-001 that the percentage bar diargram is similar to the sub-divided bar diagram. The main difference between these two diagrams is that in percentage bar diagram, we transform the given magnitude or value of each component of the variable into percentages of the total magnitude of that variable instead of considering only magnitude. We divide these bars in terms of percentages of the components. We briefly mention the main steps as under:

Step 1: We draw the horizontal (X) and the vertical (Y) axes.

Step 2: We take the categories of the variables on the X-axis. We consider percentage (0% to 100%) as labels on the vertical (Y) axis.

Step 3: In percentage bar diagram, we draw vertical bars of equal heights, which represent the total 100%.

Step 4: We divide each bar into several parts or components, which are transformed into percentages of the total according to the share of each component of the variable.

Step 5: We use different colours, patterns, etc., to distinguish each part of the bar representing different components of the variable.

Steps in Excel

In Problem 3, we have two fields (A and B) and each field has 6 categories of crops. We can represent the data of Table 3 to compare the breakup of each crop into two fields using a percentage bar diagram. You have learnt about the manual plotting of percentage bar diagram in Unit 14 of MST-001. Here we describe the procedure of plotting the percentage bar diagram using MS Excel 2007:

Step 1: We enter the data given in Table 3 in MS Excel 2007 spreadsheet as shown in Fig. 3.17.

	A	B	C	D
1		Area under Crop (in Sq ft)		
2	Crop	A	B	
3	Cereals	235	95	
4	Pulses	180	50	
5	Vegetables	320	360	
6	Fruits	150	250	
7	Flowers	80	165	
8	Others	35	80	

Fig. 3.17: Screenshot of the spreadsheet for the given data.

Step 2: The steps used to obtain the percentage bar diagram are shown in Fig. 3.18. We

1. select the data given in Cells A2:C8,
2. click on the *Insert* tab,
3. click on the *Column* option in the *Charts* group, and
4. select a chart subtype *100% Stacked Column in 3-D* that we wish to use.

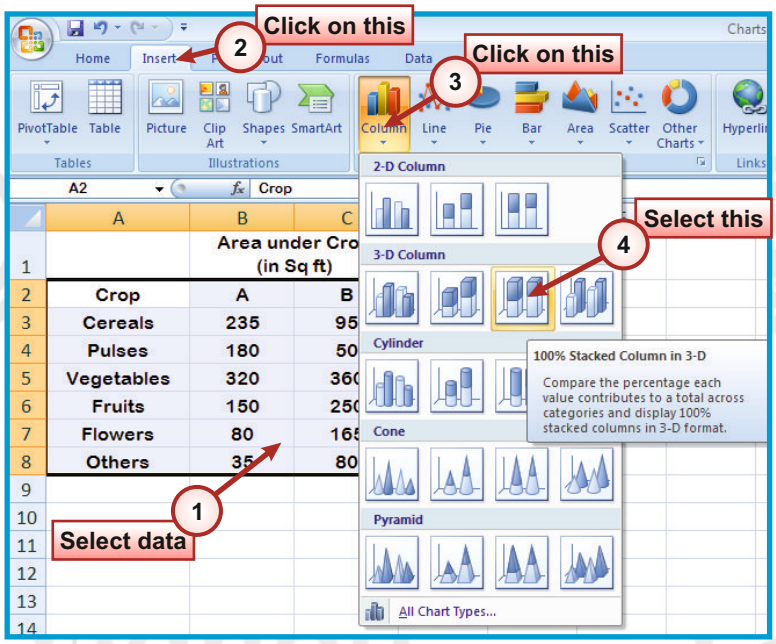


Fig. 3.18

Step 3: We get the percentage bar diagram shown in Fig. 3.19 in which each bar with colours blue and red shows the fields A and B, respectively, corresponding to different types of crops.

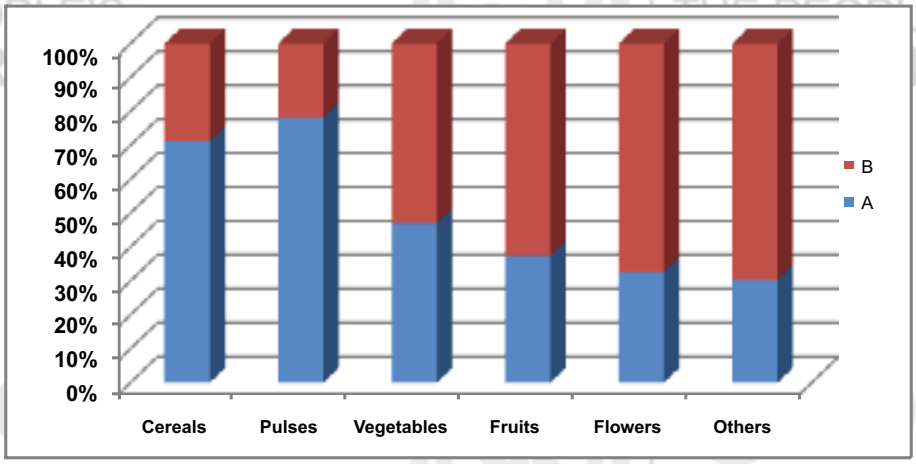


Fig. 3.19

Step 4: Since we wish to compare the area under different crops in the fields A with B, each bar should represent the percentage of area under each crop in fields A and B. If we wish to show the fields along the X-axis instead of the crops, we follow the steps shown in Fig. 3.20 and

1. click on the chart shown in Fig. 3.19,
2. click on the ***Design*** tab under the ***Chart Tools***, and
3. select the ***Switch Row/Column*** option.

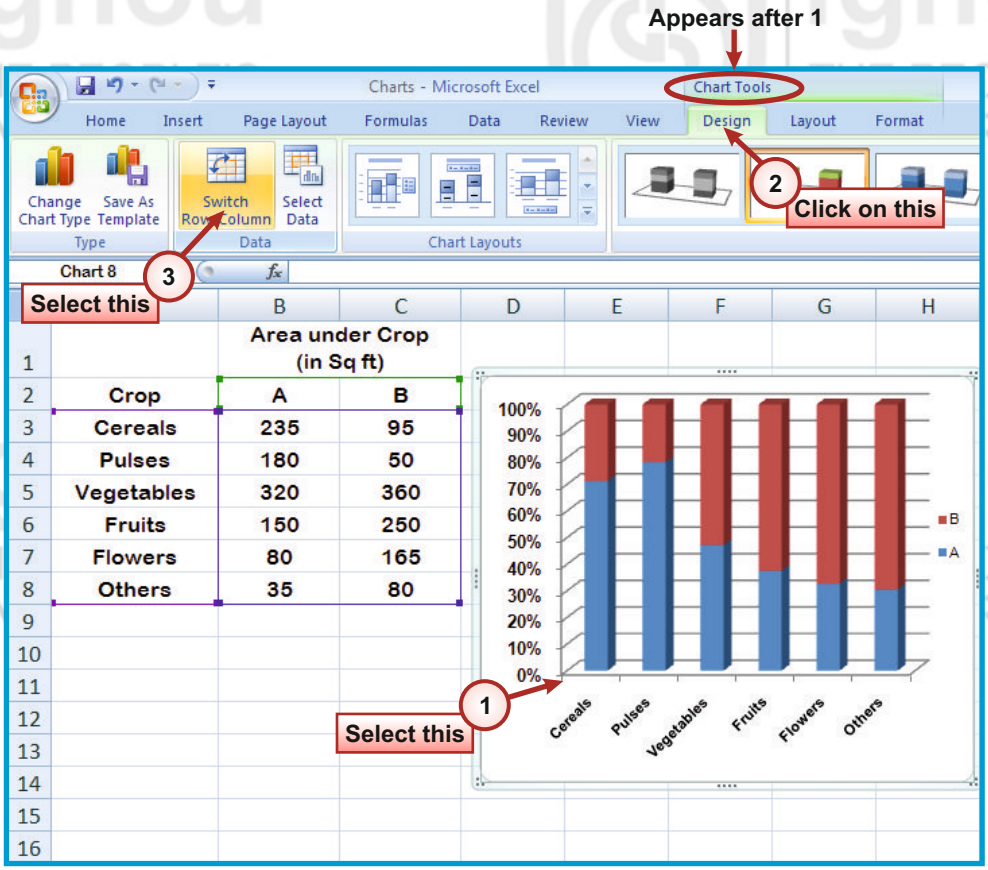


Fig. 3.20

Step 5: After switching the rows and columns, we get the percentage bar chart shown in Fig. 3.21.

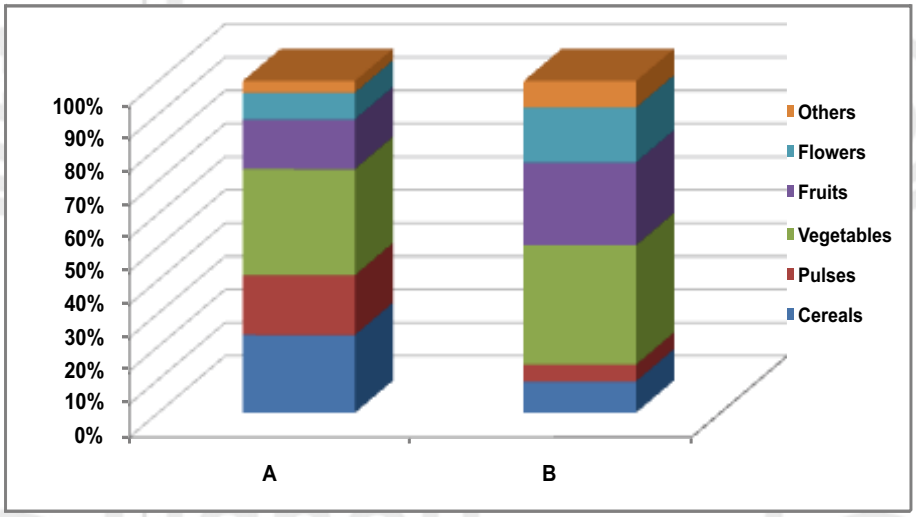


Fig. 3.21

Step 6: We format the chart as explained in Steps 4 to 9 of Sec. 3.3. The resulting percentage bar chart is shown in Fig. 3.22.

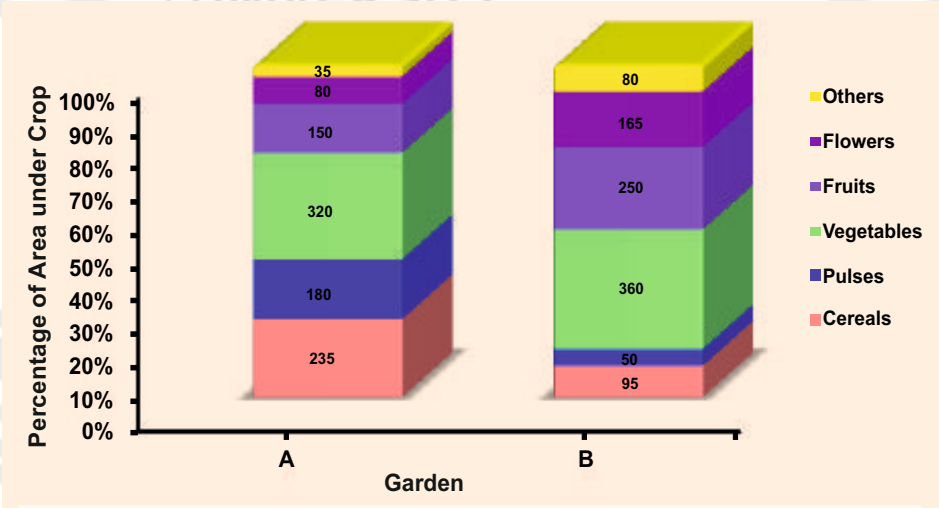


Fig. 3.22: Percentage bar diagram.

3.7 PIE DIAGRAM

You have learnt in Unit 14 of MST-001 that the pie diagram is the most popular diagram used to represent the breakdown of an aggregate into its components or sub-divisions. A pie diagram shows the components in terms of percentages and the area of each sector is proportional to the angles subtended at the centre of the circle. We briefly mention the main steps as follows:

Step 1: In the pie diagram, we convert each component in terms of degrees in a circle considering aggregate angle as 360^0.

Step 2: We draw a circle and divide it into various sectors with areas equal to the corresponding components.

Step 3: We use different colours, patterns, etc., to distinguish each sector of the circle representing different components of the variable.

Steps in Excel

In Problem 3, we have two fields (A and B) and each field has 6 categories of the crops. We can also represent the data of Table 3 to compare the breakup of

each crop in the two fields using a pie diagram. You have learnt about the manual plotting of pie chart in Unit 14 of MST-001. Here we describe the procedure of plotting the pie diagram using MS Excel 2007 for Problem 3.

Step 1: We need to plot separate pie diagram for the fields A and B in Excel 2007. To obtain the pie diagram for A, we follow the steps shown in Fig. 3.23 and

1. select the data given in Cells A2:B8,
2. click on the *Insert* tab,
3. click on the *Pie* option in the *Charts* group, and
4. select a chart subtype *Pie in 3-D* that we wish to use.

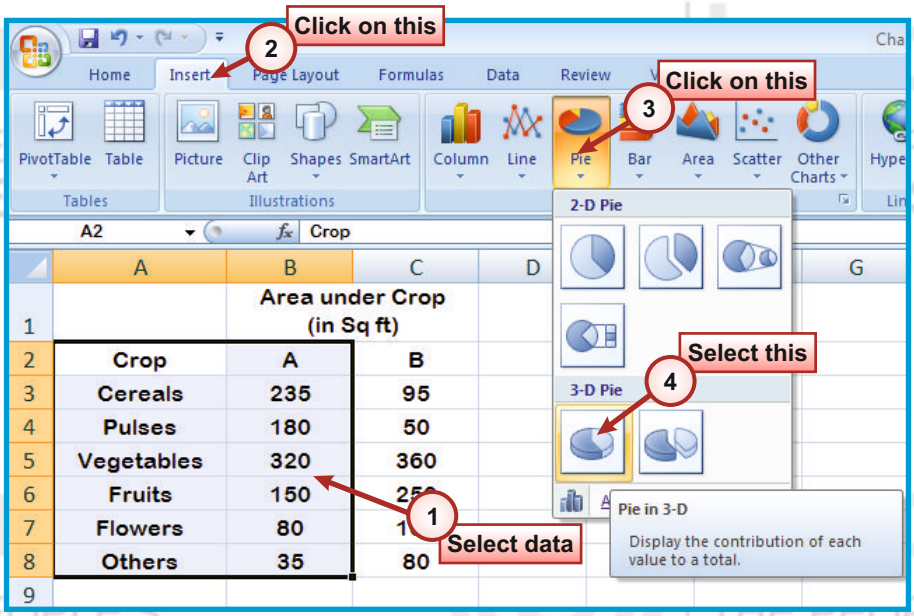


Fig. 3.23

Step 2: We format the chart as explained in Steps 4 to 7 of Sec. 3.3. The resulting pie diagram is shown in Fig. 3.24.

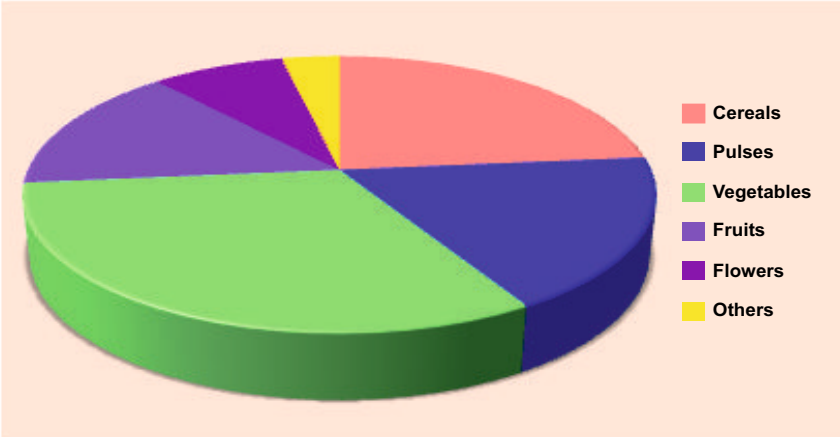


Fig. 3.24

Step 3: We can also add the corresponding values and percentages on each sectors of the pie diagram. For this, we

1. click on the chart shown in Fig. 3.24,
2. click on the *Layout* tab under the *Chart Tools*, and
3. choose the *More Data Label Options* under the *Data Labels* as shown in Fig. 3.25.

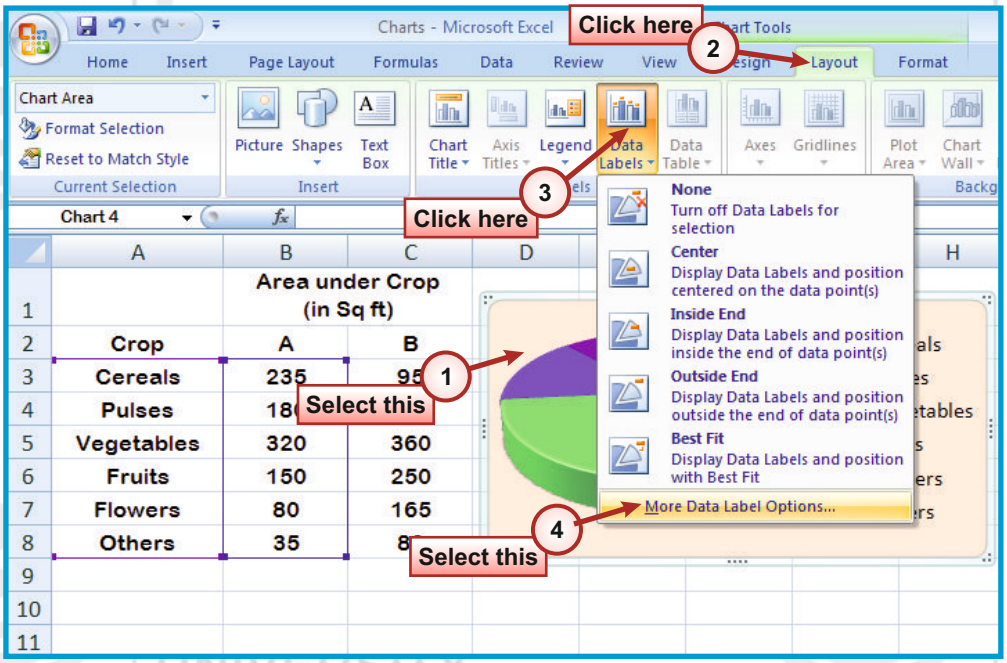
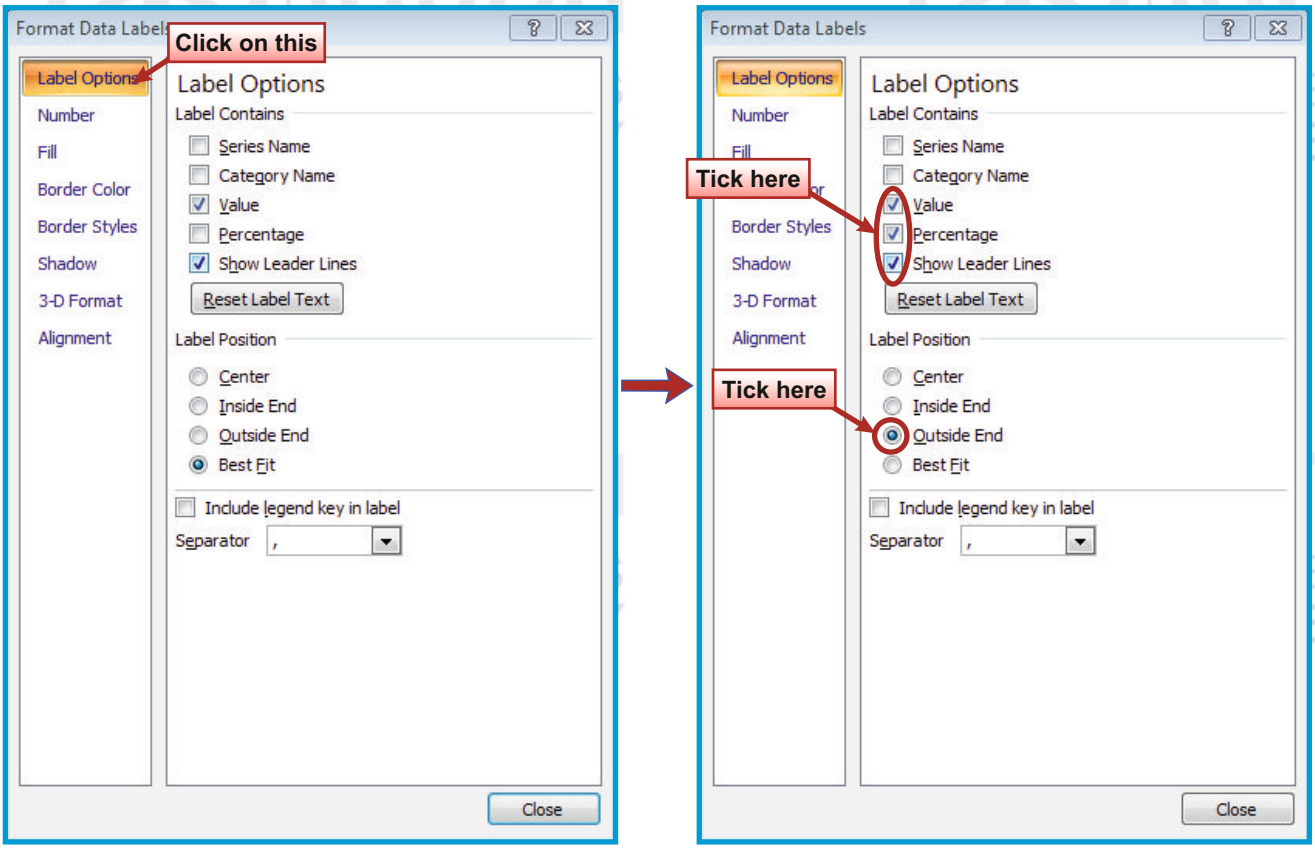


Fig. 3.25

Step 4: The dialog box shown in Fig. 3.26a will appear after Step 3.
In the dialog box, we

1. tick on the *Value*, *Percentage* and *Show Leader Lines* options under *Label Contains*,
2. tick on *Outside End* under *Label Position*, and
3. click on *Close*. These steps are shown in Fig 3.26b.



(a) (b)

Fig. 3.26

Step 5: When we click on *Close* in Step 4, we get the pie diagram with data labels as shown in Fig. 3.27.

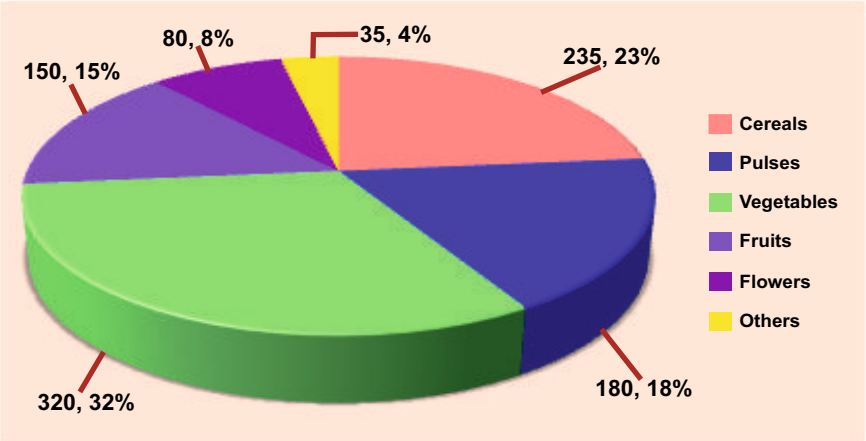


Fig. 3.27

Step 6: To obtain the pie diagram for field B, we follow the steps shown in Fig. 3.28 and

1. select the data given in Cells A2:A8 and C2:C8 by holding *Ctrl* key,
2. click on the *Insert* tab,
3. click on the option *Pie* in the *Charts* group, and
4. select a chart subtype *Pie in 3-D* that we wish to use.

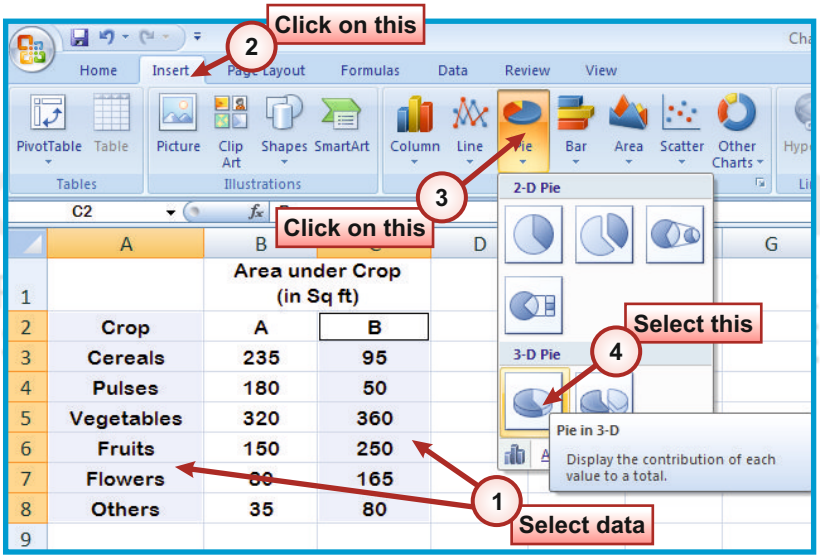


Fig. 3.28

Step 7: We format the chart as explained in Steps 4 to 7 of Sec. 3.3 and add the labels as explained in Steps 3 and 4 in this section. The resulting pie diagram is shown in Fig. 3.29.

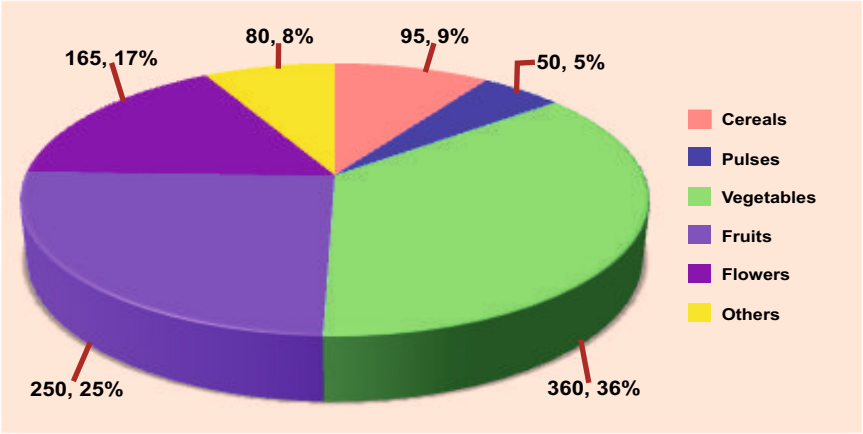


Fig. 3.29

You should now apply these methods to other problems for practice.

Activity

Construct suitable diagrams with the help of MS Excel 2007 and interpret the results for

A1) Examples 1, 2, 3, 4 and 10 given in Unit 14 of MST-001.

A2) Exercises E1 to E5, E9 and E10 given in Unit 14 of MST-001.

Match the results with the manual plotting of the charts done in Unit 14 of MST-001.

Continuous Assessment 3

1. A survey was conducted on 100 families for their preference of cooking oil. The data collected in the survey are recorded in Table 4.

Table 4: Preferences of families for cooking oils

Cooking Oil	Number of Families
Sunflower	19
Soybean	15
Cottonseed	18
Olive	11
Mustard	18
Ghee	12
Others	7

Draw a bar diagram for the given data.

2. The cooking oil preferences of 100 families in three regions I, II and III are recorded in Table 5.

Table 5: Preferences of families for cooking oils in three regions

Cooking Oil	Number of Families		
	I	II	III
Sunflower	19	17	28
Soybean	15	13	10
Cottonseed	18	15	16
Olive	11	14	14
Mustard	18	5	8
Ghee	12	13	10
Others	7	23	14

Draw the multiple and sub-divided bar diagrams to compare the cooking oil preferences of families in the three regions.

3. The average amount of time that Mr. A and Mr. B spend each day on various activities is recorded in Table 6. What is the percentage of the pie chart representing the time Mr. A spends playing sport and Mr. B spends on social media?

Table 6: Amount of time spent each day

Work	Mr. A	Mr. B
Office	8	10
Eating	2	1
Sports	1	0
Social Media	3	5
Sleeping	8	7
Others	2	1

Draw the percentage bar diagram and pie diagram to compare their work patterns.

Home Work: Do It Yourself

1) Follow the steps explained in Secs. 3.3 to, 3.7 to present the data of Tables 1, 2 and 3 diagrammatically. Use a different format for the charts. Take the final screenshots and keep them in your record book.
2) Develop the spreadsheets for the exercises of "Continuous Assessment 3" 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.

Graphical Presentation of Data

SESSION 4

Structure

4.1 Introduction
Objectives
4.2 Problem Description
4.3 Frequency Bar Graph
4.4 Histogram
4.5 Histogram using Data Analysis ToolPak
4.6 Frequency Polygon
4.7 Frequency Curve
4.8 Cumulative Frequency Curves (Ogives)

4.1 INTRODUCTION

In Lab Session 3, you have learnt how to present data in the form of different types of bar diagrams and pie diagram. In Unit 15 of MST-001 (Foundation in Mathematics and Statistics), you have learnt that we can use graphs for plotting frequency distributions. We can represent the discrete frequency distribution using a frequency bar graph and the continuous frequency distribution using a histogram. We can use cumulative frequency curves (ogives) to represent the cumulative frequency distributions. In Lab Session 2, we have explained the construction of frequency distributions using Excel 2007.

In this lab session, you will learn how to draw the frequency bar graph to represent the discrete frequency distribution graphically. You will also learn how to draw the most commonly used graphs, e.g., histogram, frequency polygon, frequency curve for the continuous frequency distributions and cumulative frequency curves (or ogives) for the cumulative frequency distributions. We shall explain how to draw graphs for time series data in Lab Session 5.

Prerequisite

- Lab Sessions 2 and 3 of MSTL-001 (Basic Statistics Lab).
- Unit 15 of MST-001 (Foundation in Mathematics and Statistics).

Objectives

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

- prepare the spreadsheet in MS Excel 2007;
- plot the frequency bar graph for discrete frequency distributions;
- draw the histogram, frequency polygon, frequency curve for continuous frequency distributions; and
- construct the cumulative frequency curves for cumulative frequency distributions.

4.2 PROBLEM DESCRIPTION

For this lab session, we consider

1. The data of Problem 2 given in Lab Session 2 to represent the number of mobile phones graphically in a frequency bar graph.
2. The data of Problem 3 given in Lab Session 2 to represent the life span of the bulbs graphically using histogram, frequency polygon, frequency curve and both the ogives.

4.3 FREQUENCY BAR GRAPH

You have learnt in Unit 15 of MST-001 that we use the frequency bar graph to represent the discrete frequency distribution graphically. It is the same as simple bar diagram explained in Sec. 3.3 of Lab Session 3. In the frequency bar graph, we use bars of the same width and these bars are separated from each other by equal intervals. We briefly mention the main steps as follows:

Step 1: We draw the horizontal (X) axis and the vertical (Y) axis.

Step 2: We take different values of the variable on the horizontal axis and the frequencies on the vertical axis.

Step 3: We draw a perpendicular bar, for each value of the variable on the horizontal axis. The height of the bar represents the frequency of the corresponding value of that variable.

Steps in Excel

You have learnt about the manual plotting of the frequency bar graph to represent the discrete frequency distribution in Unit 15 of MST-001. In order to plot the frequency bar graph for the data of Problem 1 in MS Excel 2007, we follow the steps given below:

Step 1: We directly enter the discrete frequency distribution of the data, which was formed in Sec. 2.4 of Lab Session 2 in Excel 2007 spreadsheet as shown in Fig. 4.1.

	A	B	C
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	
10			

Fig. 4.1: Partial screenshot of the spreadsheet for the given data.

Step 2: We select Cells B2:B9 and follow Steps 2 to 9 explained in Sec. 3.3 of Lab Session 3 to obtain the formatted frequency bar graph. The resulting graph is shown in Fig. 4.2.

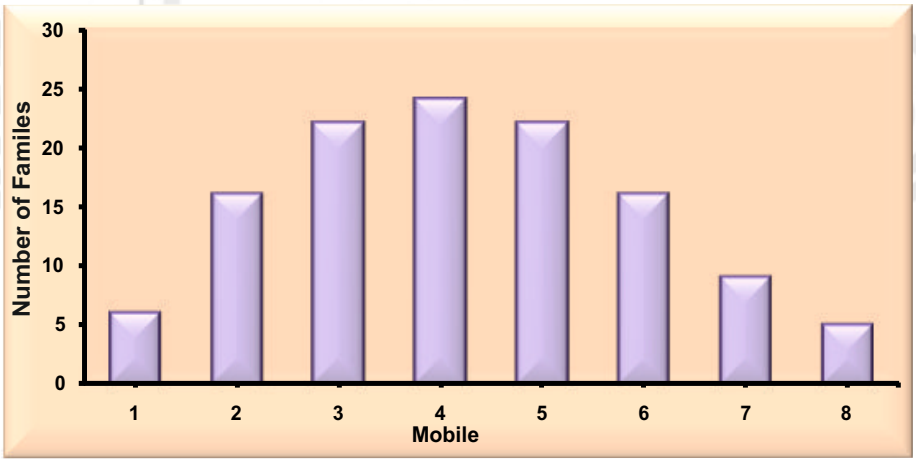


Fig. 4.2

Note that here the values of the variable (number of mobile phones) are in sequence from 1 to 8. If the values of the variable are not in a sequence starting from 1, we change the horizontal axis labels with respect to the values of the variable as explained in Steps 7 to 11 of Sec. 4.4.

4.4 HISTOGRAM

Histogram is used to represent the continuous frequency distribution. It is similar to a bar diagram and represents a frequency distribution with continuous classes. The main difference between a bar diagram and a histogram is that in the histogram, there is no gap between the bars and it has a continuous appearance. We briefly mention the main steps as follows:

Step 1: We take the class intervals on the horizontal axis (X-axis) and frequencies on the vertical axis (Y-axis).

Step 2: For constructing a histogram, we draw adjacent rectangles (or bars) over the class intervals such that the height of the rectangles is proportional to the corresponding class frequency and the width of all bars is equal to the class interval.

We can construct the histogram using Excel 2007 in two different ways:

i) Using Bar Diagram

ii) Using Data Analysis ToolPak

We shall explain how to draw a histogram using bar diagram in this section. In the next section, we shall use the *Data Analysis ToolPak* to draw the histogram.

Steps in Excel

You have learnt how to construct a histogram manually in Unit 15 of MST-001. In this lab session, we explain the method of constructing the histogram for Problem 2 in Excel 2007 using the bar diagram.

Step 1: For plotting the histogram using the bar diagram as explained in Sec. 3.3 of Lab Session 3, we directly use the continuous frequency distribution of the data which was formed in Sec. 2.5 of Lab Session 2 Here we enter only the continuous frequency distribution in Excel 2007 spreadsheet as shown in Fig. 4.3.

	A	B	C
1	Class Interval	No. of Bulbs (Frequency)	
2	500-600	11	
3	600-700	20	
4	700-800	25	
5	800-900	32	
6	900-1000	27	
7	1000-1100	19	
8	1100-1200	11	
9	1200-1300	5	
10			

Fig. 4.3: Partial screenshot of the spreadsheet for the given data.

Step 2: We now draw a bar diagram by following the steps shown in Fig. 4.4. For this, we

1. select Cells B2:B9,
2. click on the *Insert* tab,
3. click on the *Column* option under the *Chart* group, and
4. choose a *Clustered Column* subtype under *2-D Column*.

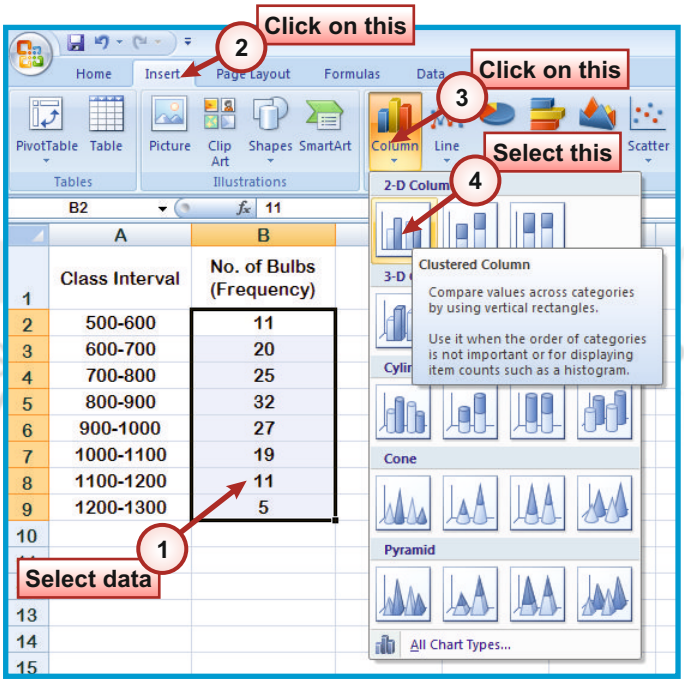


Fig. 4.4

Step 3: The resulting chart is shown in Fig. 4.5.

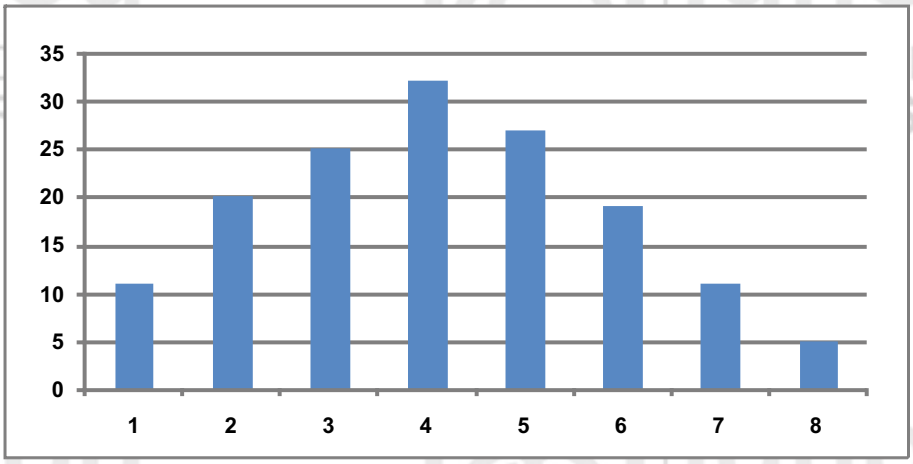


Fig. 4.5

Step 4: To remove the gap between the bars, we

1. select bars given on the chart,
2. click on the *Format* tab under *Chart Tools*, and
3. click on the *Format Selection* as shown in Fig. 4.6.

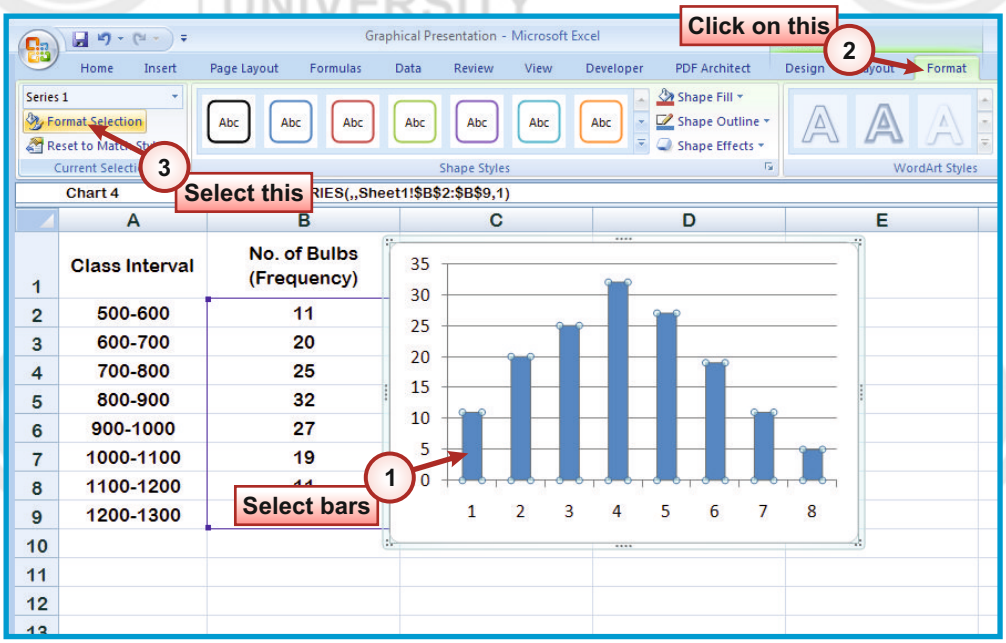
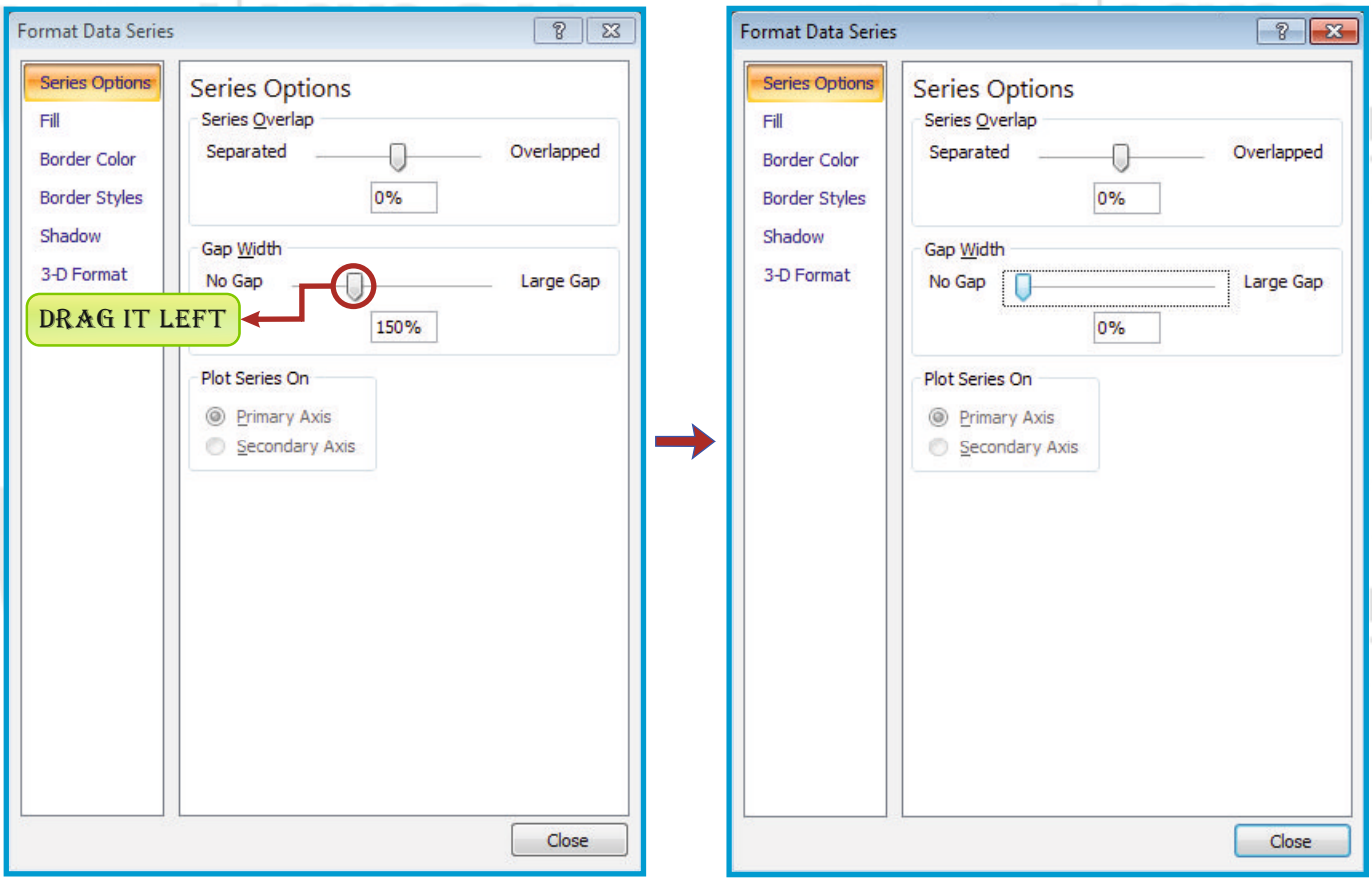


Fig. 4.6

Step 5: A new dialog box as shown in Fig. 4.7a appears. We now drag the *Gap Width* up to *No Gap*, i.e., we take the gap width as 0% (Fig. 4.7b).



(a) **(b)**

Fig. 4.7

Step 6: Thus, the gap between the bars is eliminated and the resulting graph is shown in Fig. 4.8. It is known as the histogram.

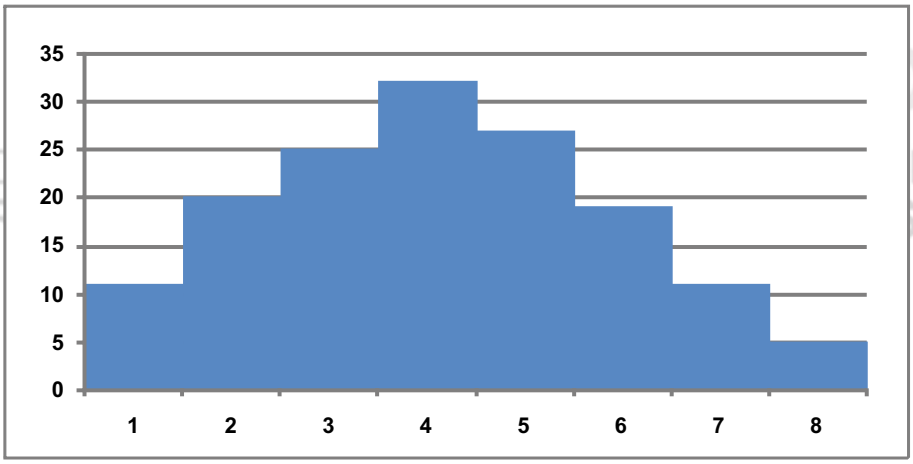


Fig. 4.8

Step 7: We now change the horizontal axis labels. For this, we

1. select the horizontal axis labels of the chart shown in Fig. 4.9,
2. click on *Design* tab under *Chart Tools*, and
3. click on *Select Data*.

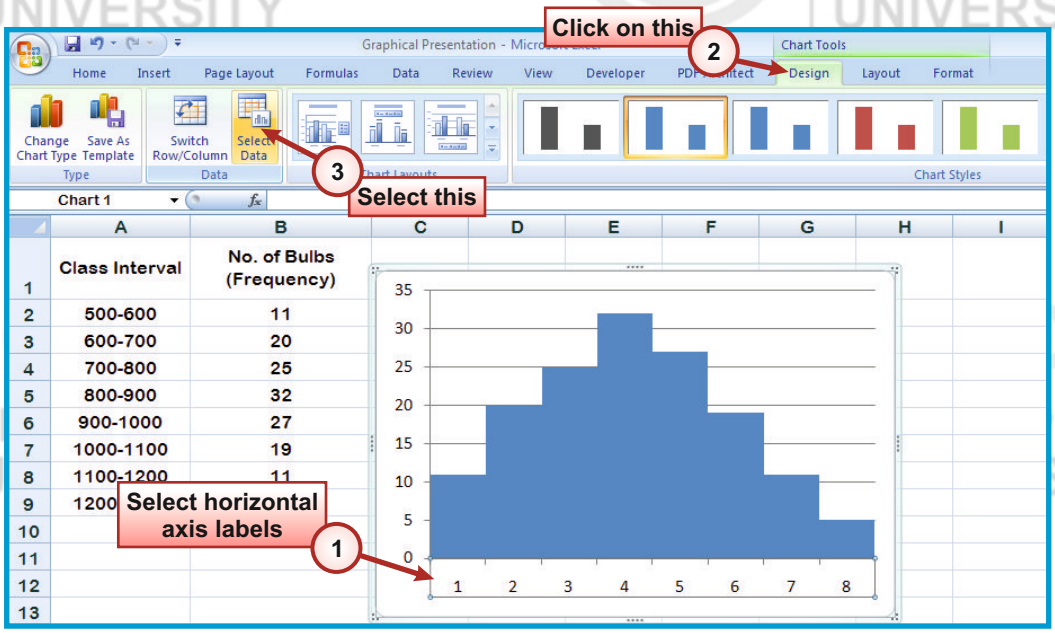


Fig. 4.9

Step 8: A new dialog box as shown in Fig. 4.10 appears. We click on *Edit* under *Horizontal (Category) Axis Labels* as shown in Fig. 4.10.

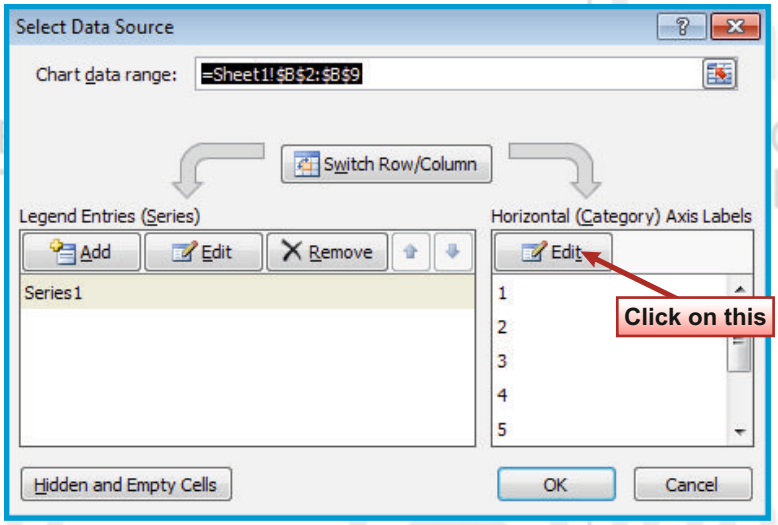
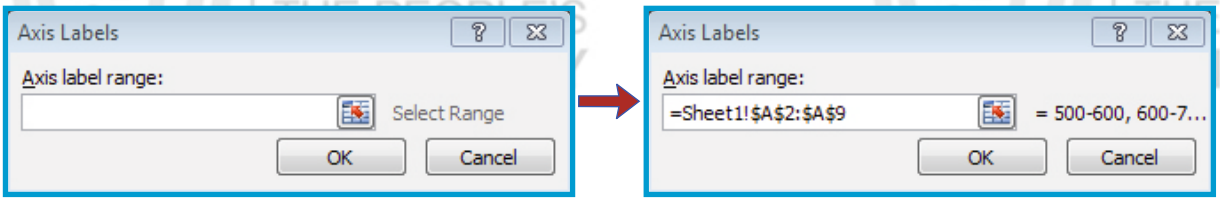


Fig. 4.10

Step 9: Again a new dialog box shown in Fig. 4.11a appears. We select Cells A2:A9 as *Axis label range* as shown in Fig. 4.11b and click on *OK*.



(a) (b)

Fig. 4.11

Step 10: Now the dialog box shown in Fig. 4.10 changes to the one shown in Fig 4.12 and we click on *OK*.

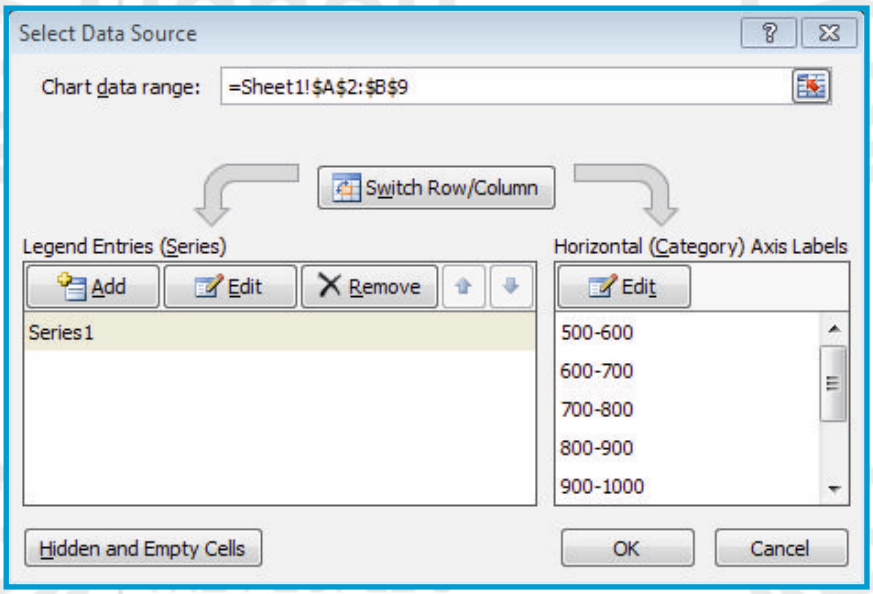


Fig. 4.12

Step 11: Thus, we obtain the histogram with class intervals on the horizontal axis as shown in Fig. 4.13.

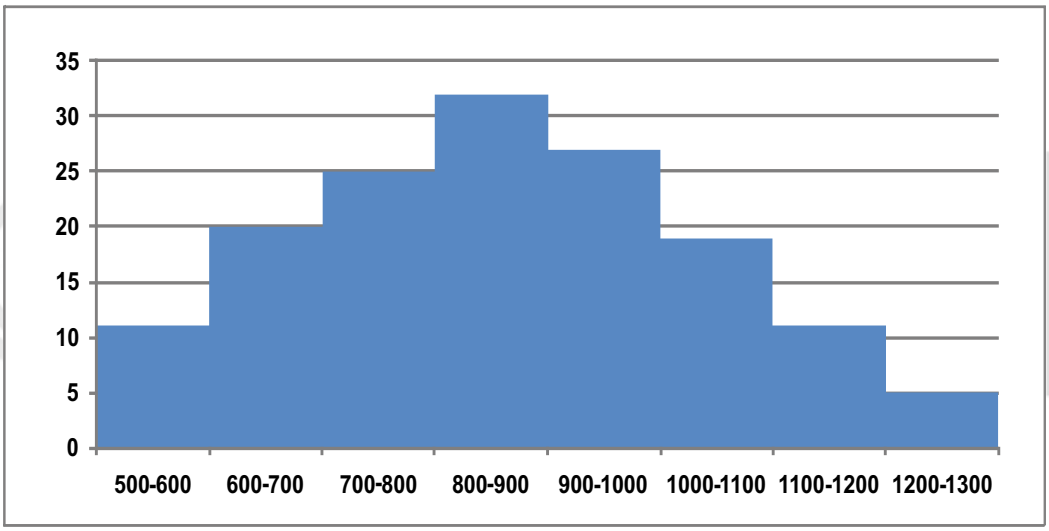


Fig. 4.13

Step 12: We format the histogram as explained in Steps 4 to 9 of Sec. 3.3 under the heading 'Steps in Excel' of Lab Session 3. The resulting histogram is shown in Fig. 4.14.

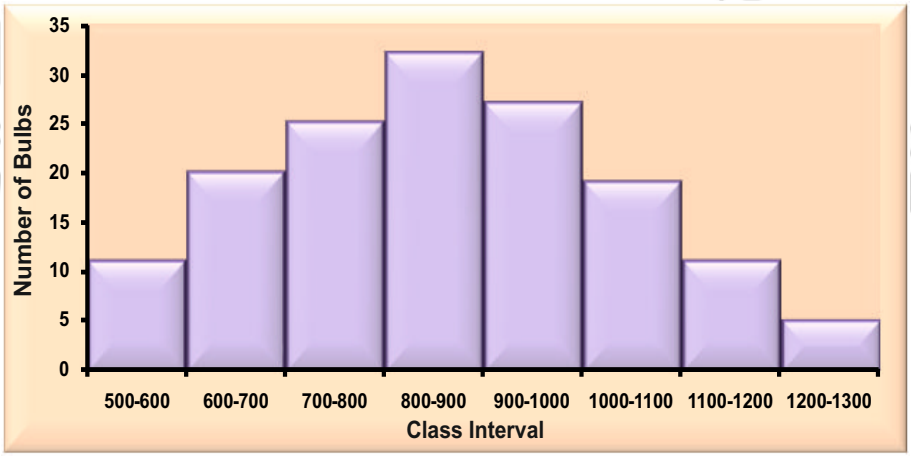


Fig. 4.14

We now explain how to construct a histogram using the *Data Analysis ToolPak*.

4.5 HISTOGRAM USING DATA ANALYSIS TOOLPAK

Step 1: To construct a histogram using *Data Analysis ToolPak*, we consider the raw data given in Table 3 of Lab Session 2 (Problem 2 of this session). We enter the data in Excel sheet as shown in Fig. 4.15.

	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	
12	11	957	
13	12	745	

Fig. 4.15: Partial screenshot of the spreadsheet for the given data.

Step 2: Here we need the values of **Bin** computed in Step 7 of Sec. 2.5 of Lab Session 2 under the heading 'Steps in Excel'. We type the values of **Bin** in Cells C2:C9 as shown in Fig. 4.16.

	A	B	C	D
1	S.No.	Life of Bulb (in hours)	Bin	
2	1	1087	599	
3	2	1289	699	
4	3	876	799	
5	4	725	899	
6	5	900	999	
7	6	1080	1099	
8	7	952	1199	
9	8	741	1299	
10	9	1000		

Fig. 4.16

Step 3: We now plot the histogram using the *Data Analysis ToolPak*. We click on *Data* tab and then click on *Data Analysis* as shown in Fig. 4.17.

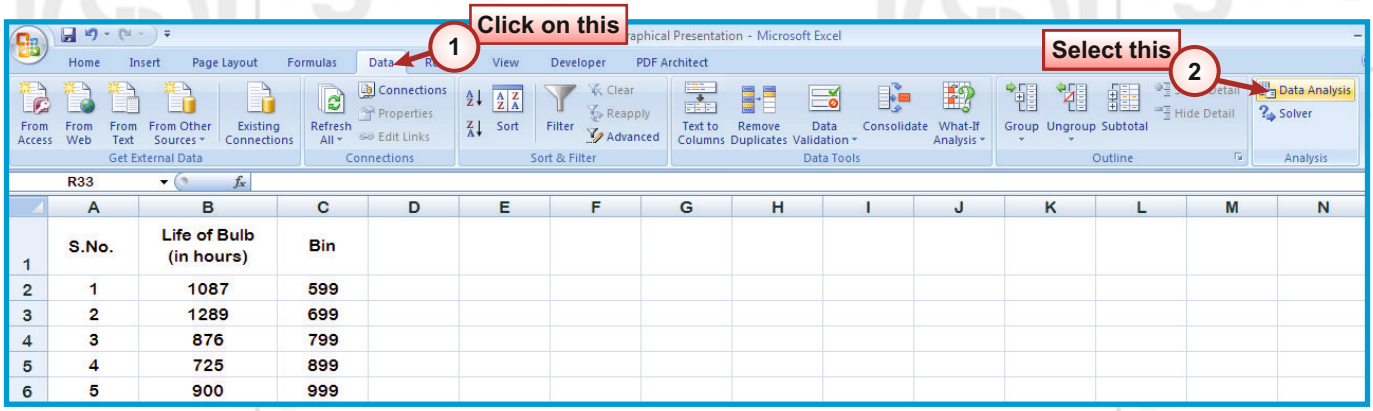


Fig. 4.17

Step 4: The dialog box shown in 4.18 appears. We select the option *Histogram* and click on *OK*.

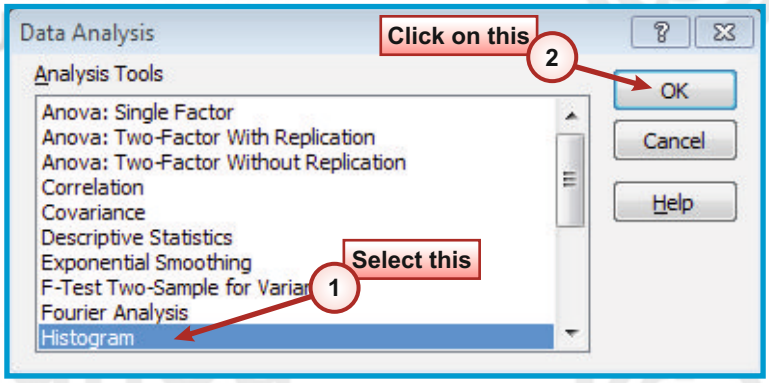
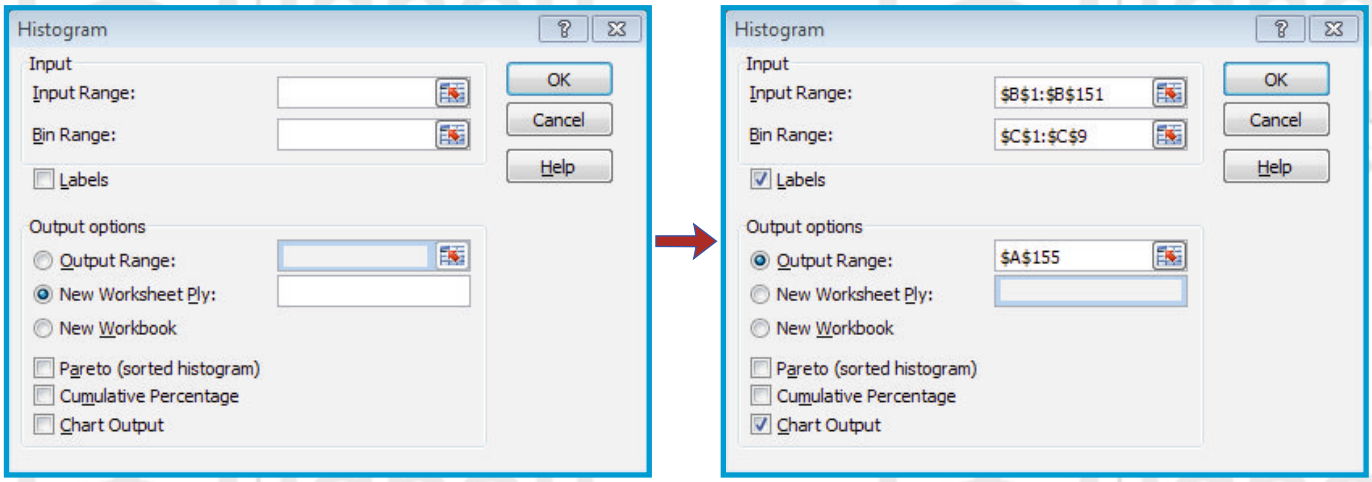


Fig. 4.18

Step 5: Again a new dialog box as shown in Fig. 4.19a appears. We fill the desired ranges in this dialog box as shown in Fig. 4.19b. We

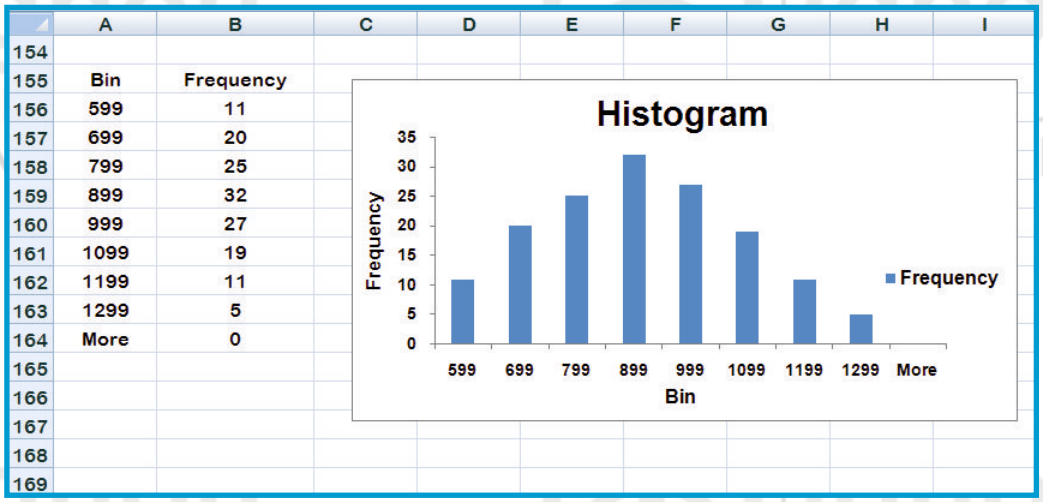
1. select *Input Range* as Cells B1:B151,
2. select *Bin Range* as Cells C1:C9,
3. tick on the *Labels*,
4. provide *Output Range* as Cell A155 (You can also use *New Worksheet* or *New Workbook* to get the output), and
5. tick on the *Chart Output* and click on *OK*.



(a) **(b)**

Fig. 4.19

Step 6: The output is shown in Fig. 4.20.



Bin	Frequency
599	11
699	20
799	25
899	32
999	27
1099	19
1199	11
1299	5
More	0

Fig. 4.20

We now follow Steps 5 to 12 of Sec. 4.4 under the heading 'Steps in Excel' and obtain the same histogram as shown in Fig. 4.14.

4.6 FREQUENCY POLYGON

You have learnt how to construct a frequency polygon for the continuous frequency distribution in Unit 15 of MST-001. Here we briefly mention the main steps as follows:

Step 1: To draw the frequency polygon, we compute the mid values of all class intervals as

$$\text{Mid value} = \frac{\text{lower limit} + \text{upper limit}}{2}$$

Step 2: We take the mid values on the horizontal axis (X-axis) and frequencies on the vertical axis (Y-axis). We plot the respective points on the graph using the mid values and frequencies.

Step 3: We join these points by line segments.

Step 4: We add two imaginary classes of the same width as the class intervals with zero frequencies, one just before the first class interval and the other just after the last class interval.

Step 5: We plot the points (mid values) of the imaginary classes with zero frequencies and join these points with the plotted graph. The graph thus obtained is known as the frequency polygon.

Step 6: We can also plot the frequency polygon along with a histogram. For this, we mark the points in the middle of the upper side of each rectangle (or bar) and join them as explained in Steps 3 to 5.

Steps in Excel

You have learnt how to construct the frequency polygon manually in Unit 15 of MST-001. In this lab session, we shall consider Problem 2 and explain the method of constructing the frequency polygon using Excel 2007.

Step 1: For plotting the frequency polygon, we consider the data used in Sec. 4.4 as shown in Fig. 4.21.

	A	B	C
1	Class Interval	No. of Bulbs (Frequency)	
2	500-600	11	
3	600-700	20	
4	700-800	25	
5	800-900	32	
6	900-1000	27	
7	1000-1100	19	
8	1100-1200	11	
9	1200-1300	5	
10			

Fig. 4.21

Step 2: We now insert a new row before Row 2 as shown in Fig. 4.22. For this, we

1. select Row 2 as shown in Fig. 4.22a,
2. click on the *Home* tab,
3. choose *Insert Sheet Rows* under *Insert* option as shown in Fig. 4.22a, and
4. obtain a new blank Row 2 as shown in Fig. 4.22b.

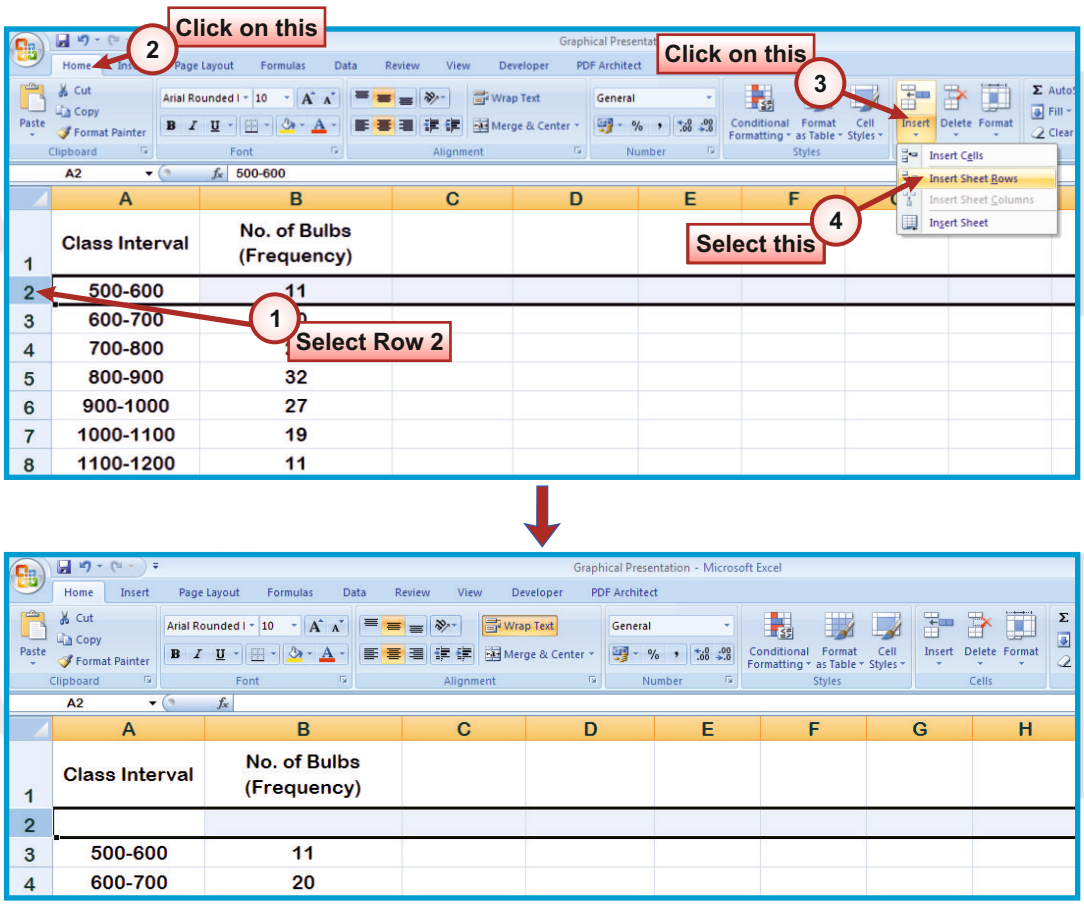


Fig. 4.22

Step 3: We add one imaginary class (400-500) with frequency zero, i.e., Cells A2:B2 before the first class (500-600) of equal width and one imaginary class (1300-1400) with zero frequency, i.e., Cells A11:B11 after the last class (1200-1300) as shown in Fig. 4.23.

	A	B
1	Class Interval	No. of Bulbs (Frequency)
2	400-500	0
3	500-600	11
4	600-700	20
5	700-800	25
6	800-900	32
7	900-1000	27
8	1000-1100	19
9	1100-1200	11
10	1200-1300	5
11	1300-1400	0

Fig. 4.23

Step 4: We type the corresponding lower and upper limits in Cells C2:C11 and Cells D2:D11, respectively, as shown in Fig. 4.24.

	A	B	C	D
1	Class Interval	No. of Bulbs (Frequency)	Lower Limit	Upper Limit
2	400-500	0	400	500
3	500-600	11	500	600
4	600-700	20	600	700
5	700-800	25	700	800
6	800-900	32	800	900
7	900-1000	27	900	1000
8	1000-1100	19	1000	1100
9	1100-1200	11	1100	1200
10	1200-1300	5	1200	1300
11	1300-1400	0	1300	1400

Fig. 4.24

Step 5: We compute the mid value for the first class interval by typing "=(C2+D2)/2"in Cell E2 as shown in Fig. 4.25.

E2 f_x =(C2+D2)/2

	C	D	E	F
1	Lower Limit	Upper Limit	Mid Value	
2	400	500	450	
3	500	600		
4	600	700		
5	700	800		
6	800	900		

DRAG IT DOWN

Fig. 4.25

Step 6: We drag down Cell E2 up to Cell E11 to compute the mid values for the remaining class intervals as shown in Fig.4.26.

	C	D	E	F
1	Lower Limit	Upper Limit	Mid Value	
2	400	500	450	
3	500	600	550	
4	600	700	650	
5	700	800	750	
6	800	900	850	
7	900	1000	950	
8	1000	1100	1050	
9	1100	1200	1150	
10	1200	1300	1250	
11	1300	1400	1350	

Fig. 4.26

We can plot a frequency polygon in two different ways:

i) With Histogram, and

ii) Without Histogram.

We first explain how to plot the frequency polygon with histogram.

Step 7: To plot the frequency polygon with histogram, we plot the histogram as explained in Sec. 4.4 by selecting the data given in Cells B2:B11 and choose the horizontal axis labels given in Cells A2:A11. The resulting histogram after formatting is shown in Fig. 4.27.

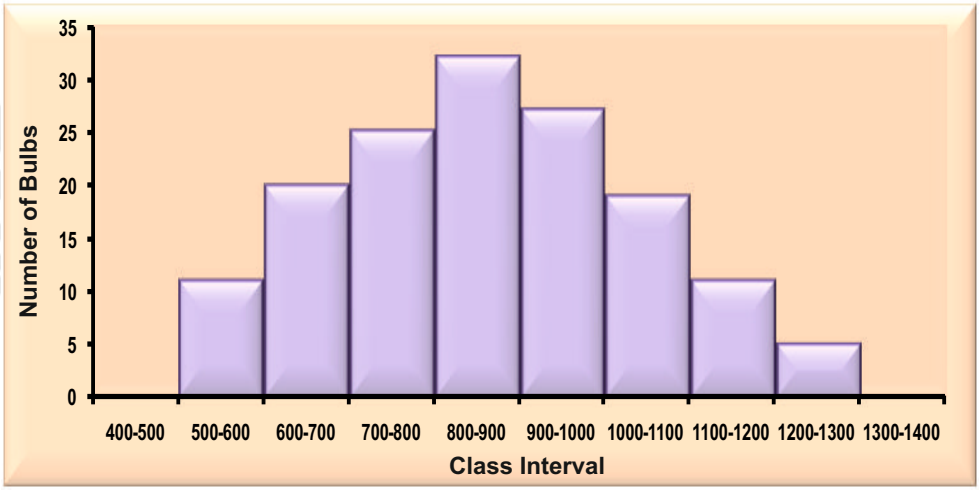


Fig. 4.27

Step 8: To add frequency polygon in the histogram shown in Fig. 4.27, we

1. click on the chart,
2. click on the *Design* tab, and
3. click on the *Select Data* as shown in Fig. 4.28.

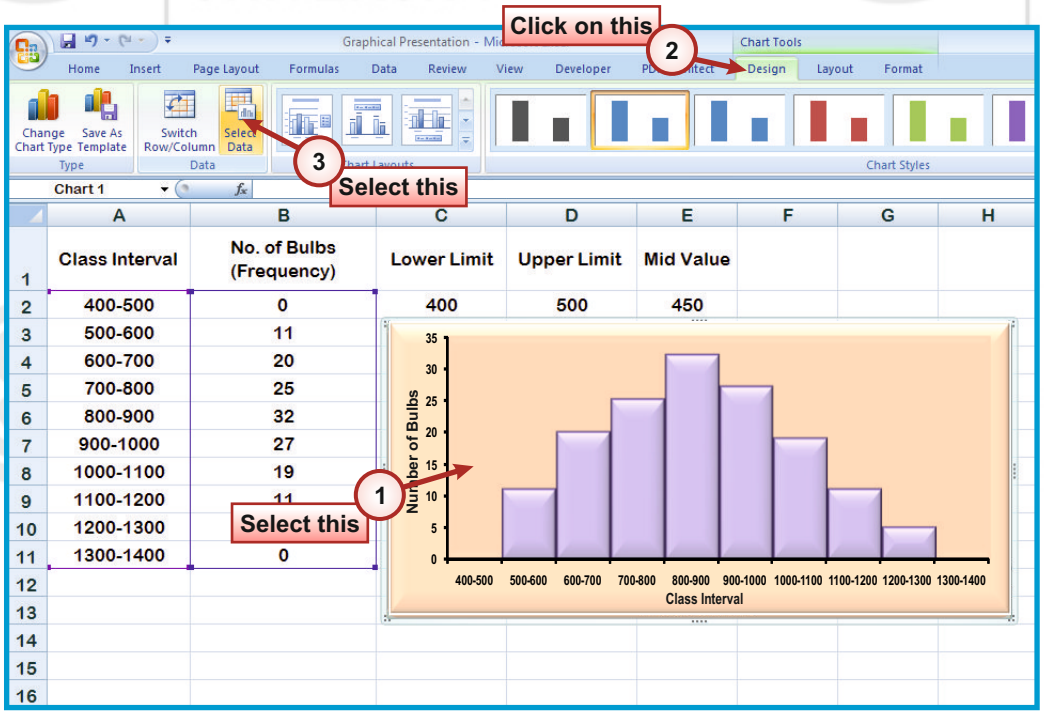


Fig. 4.28

Step 9: This opens a new dialog box as shown in Fig. 4.29. In this dialog box, we click on *Add*.

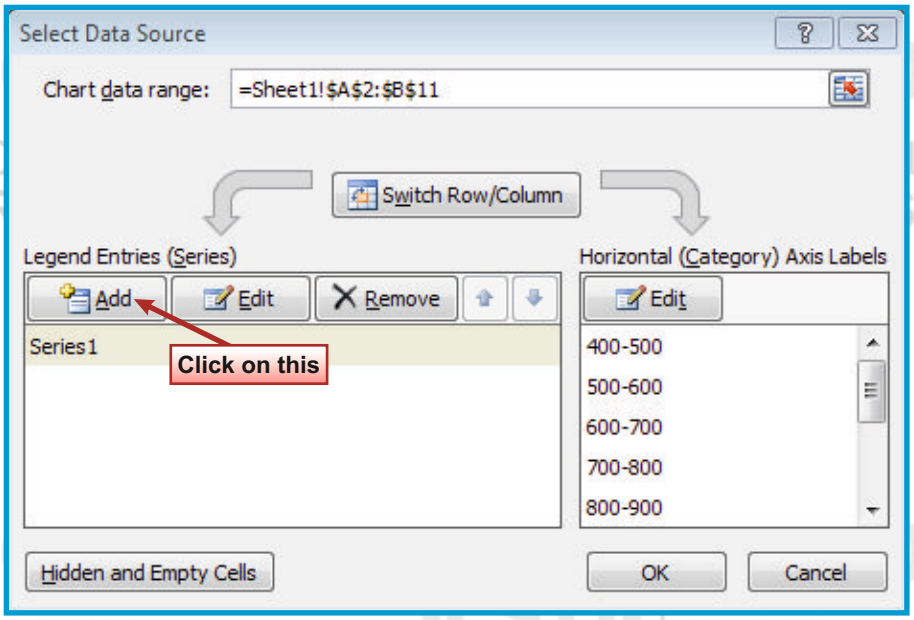
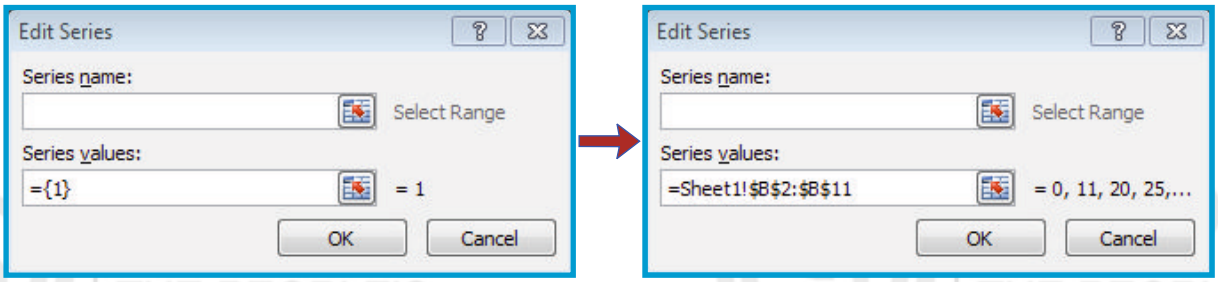


Fig. 4.29

Step 10: It opens another dialog box shown in Fig. 4.30a. We select ***Series Values*** by selecting Cells B2:B11 as shown in Fig. 4.30b and click on ***OK***.



(a) Fig. 4.30 (b)

Step 11: The dialog box shown in Fig. 4.29 will change to the one shown in Fig. 4.31. We click on ***OK***.

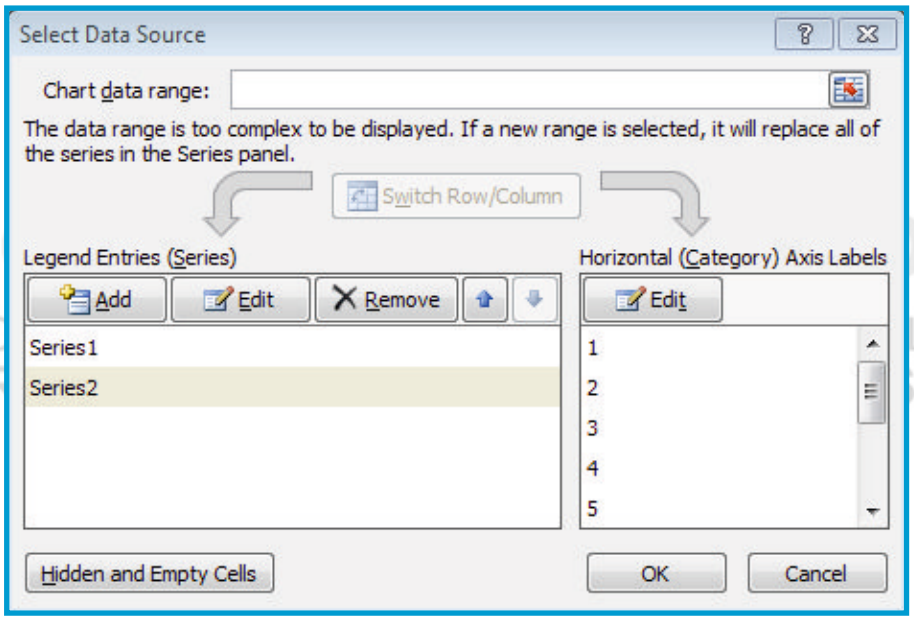


Fig. 4.31

Step 12: The resulting chart is shown in Fig. 4.32. The new series appears in red coloured bars.

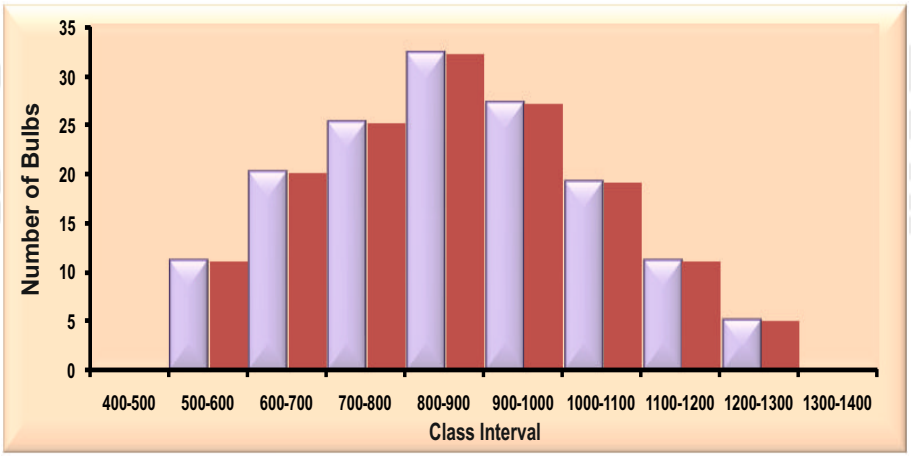


Fig. 4.32

Step 13: We now select the red coloured bars and click on *Change Chart Type* option in *Design* tab under *Chart Tools* as shown in Fig. 4.33.

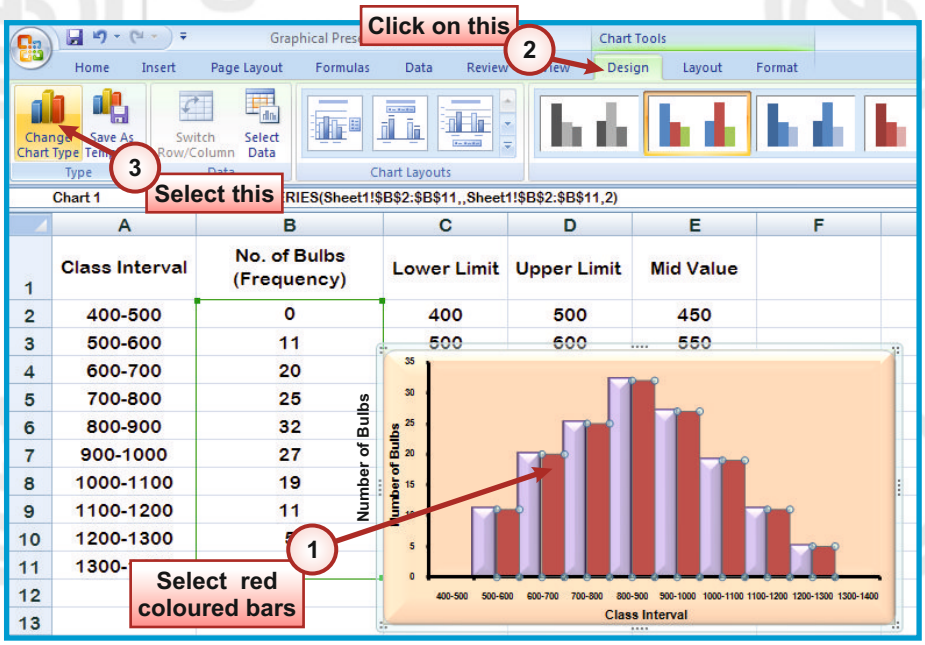


Fig. 4.33

Step 14: The dialog box shown in Fig 4.34 appears. We select the fourth chart subtype *Line with Markers* under *Line* option as shown in Fig. 4.34 and click on *OK*.

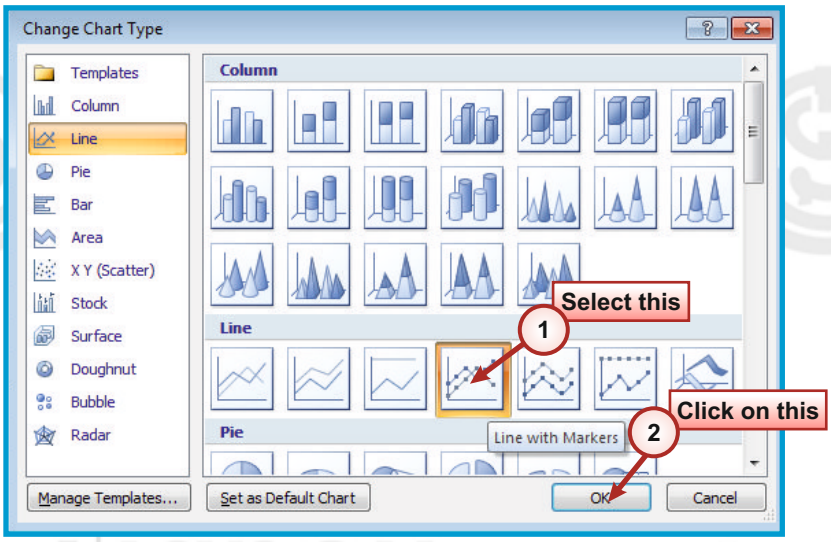


Fig. 4.34

Step 15: The resulting frequency polygon along with the histogram is shown in Fig. 4.35.

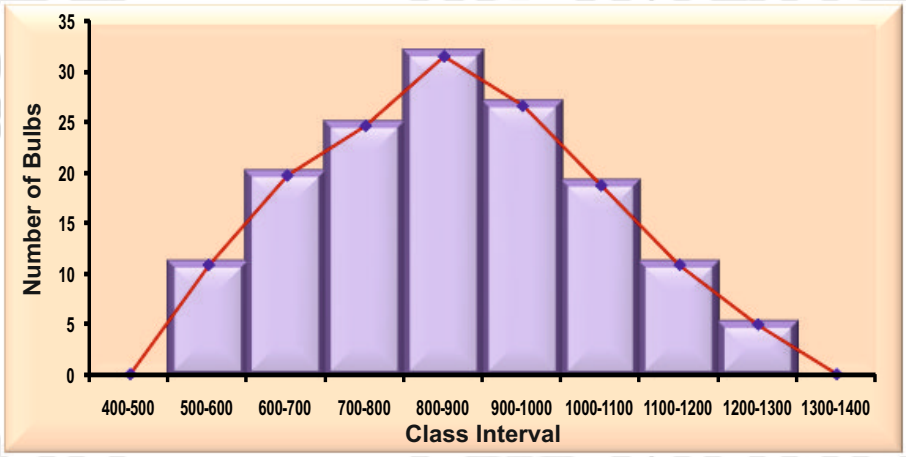


Fig. 4.35

We can plot frequency polygon directly without drawing the histogram as follows:

Step 16: We first click on *Scatter* under *Insert* tab and choose a *Scatter with Straight Lines and Markers* chart subtype as shown in Fig. 4.36. In this way, we get the blank chart shown in Fig. 4.37.

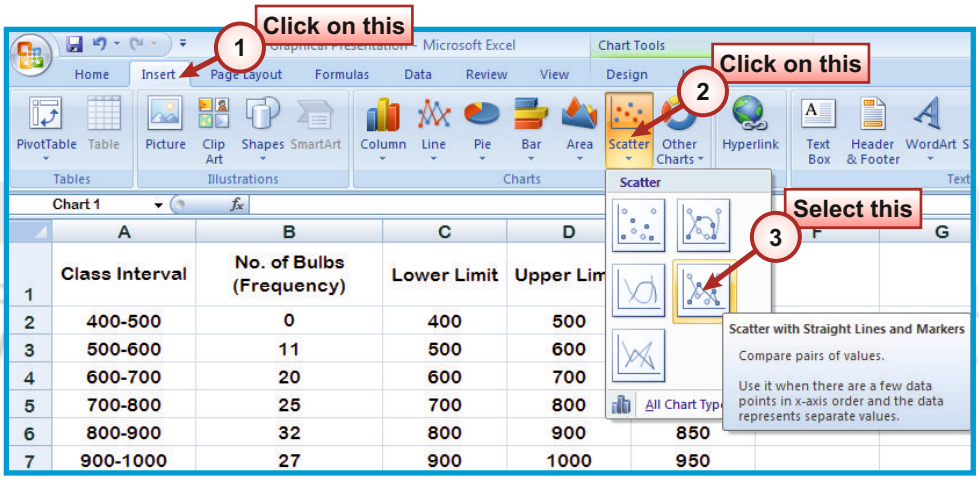


Fig. 4.36

Step 17: We click on the chart, the *Design* tab under *Chart Tools* and then click on the *Select Data* option as shown in Fig. 4.37.

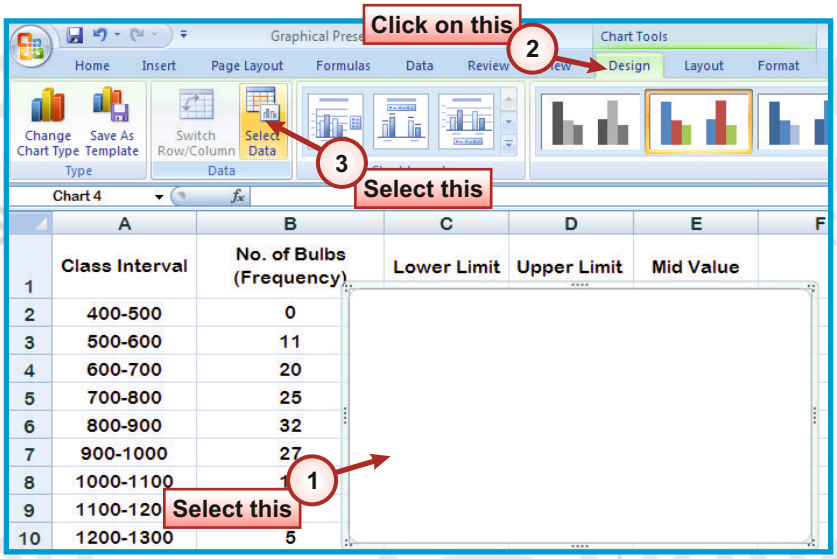


Fig. 4.37

Step 18: A new dialog box opens as shown in Fig. 4.38.

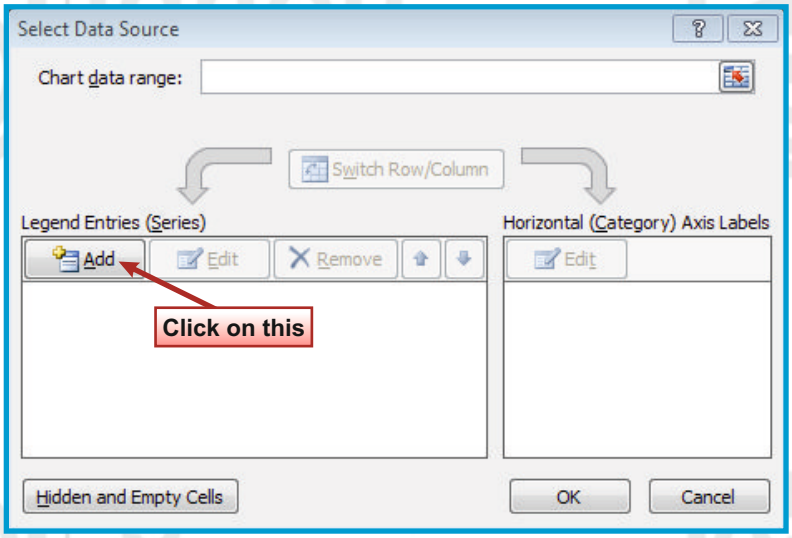
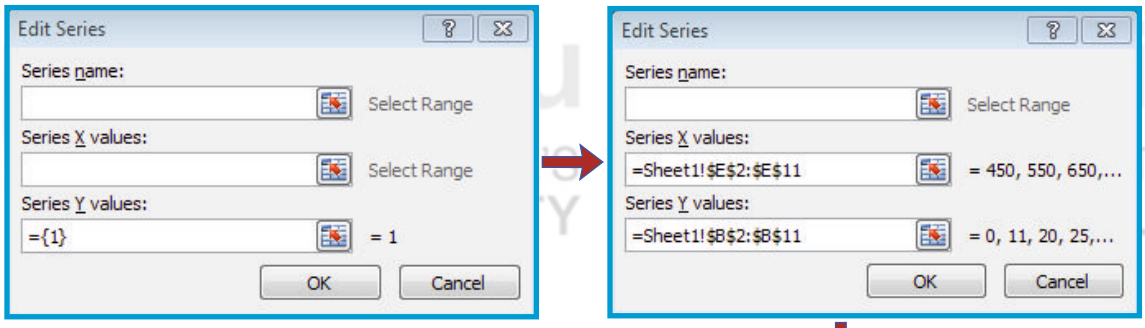


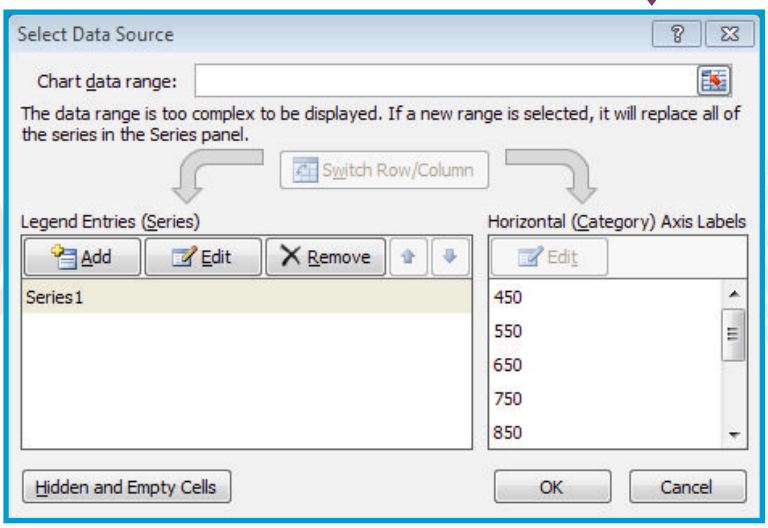
Fig. 4.38

Step 19: When we click on *Add*, the dialog box shown in Fig. 4.39a appears. Then we,

1. select Cells E2:E11 as *Series X values*,
2. select Cells B2:B11 as *Series Y values*,
3. click on *OK* as shown in Fig. 4.39b, and
4. click on *OK* as shown in Fig. 4.39c.



(a) (b)



(c)

Fig. 4.39

Step 20: The resulting graph shown in Fig. 4.40 is called the frequency polygon.

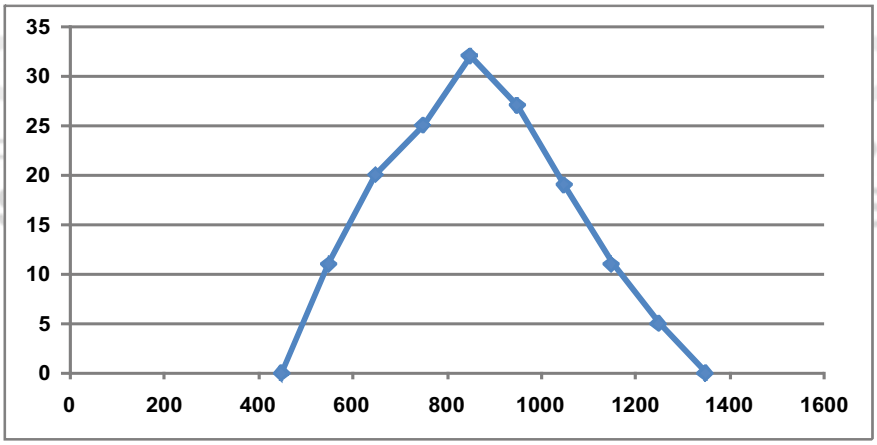


Fig. 4.40

Step 21: We now format this graph as explained in Sec. 3.3 of Lab Session 3. The resulting graph is shown in Fig. 4.41.

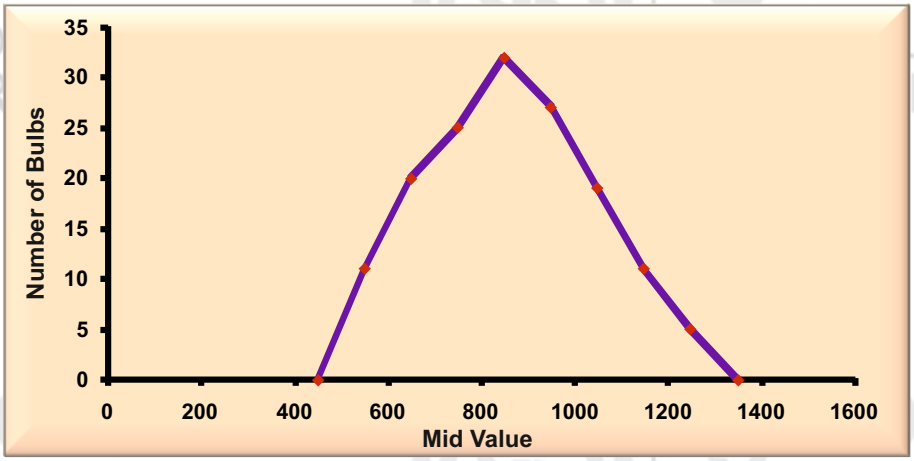


Fig. 4.41

Step 22: We can also change the minimum and maximum values of the horizontal axis. For this, we

1. select the horizontal axis,
2. click on the *Format* tab under *Chart Tools*, and
3. choose *Format Selection* option as shown in Fig. 4.42.

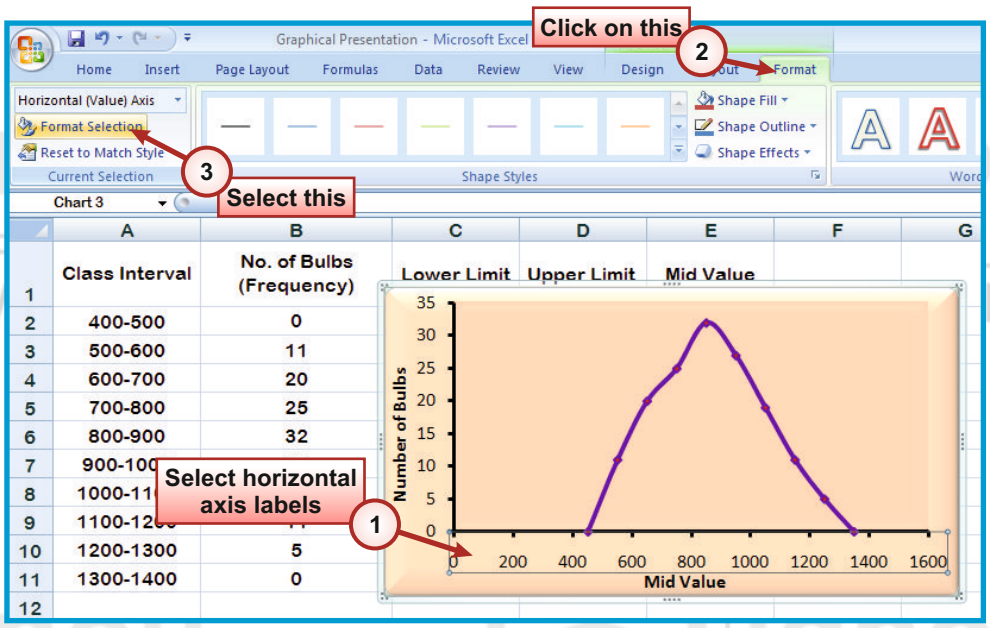


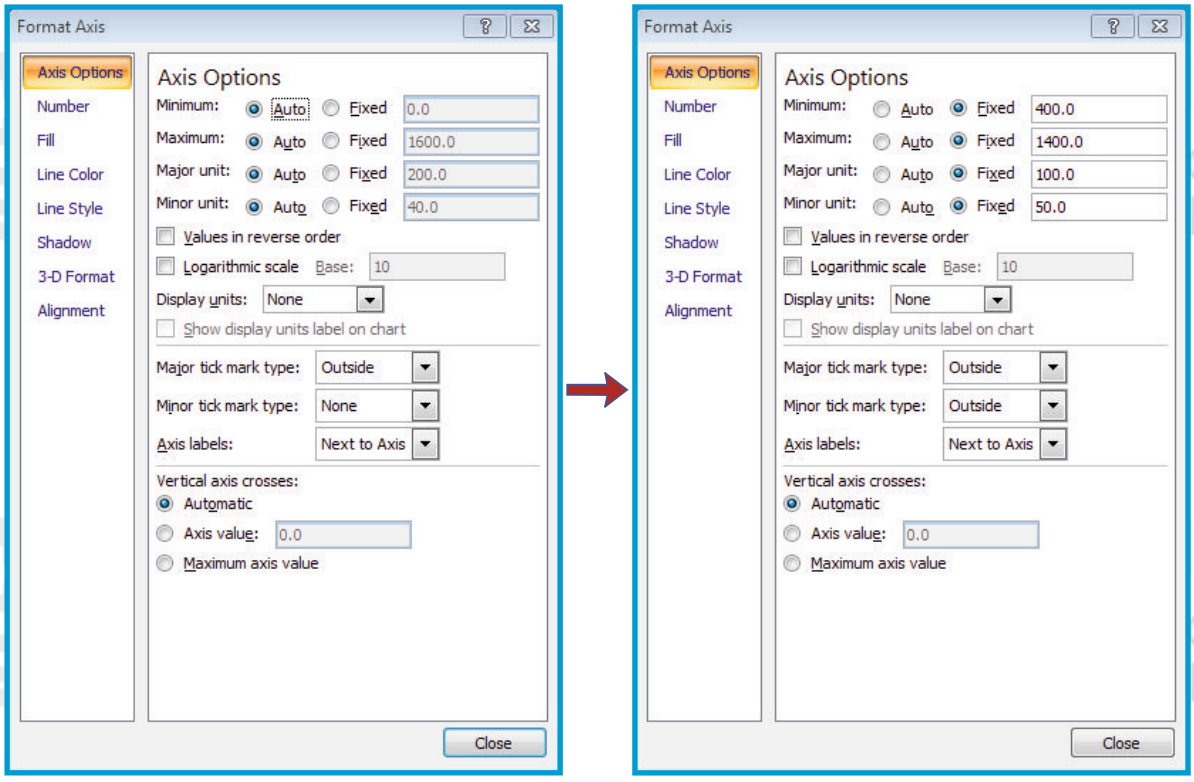
Fig. 4.42

Step 23: The dialog box shown in Fig. 4.43a appears. As shown in Fig. 4.43b, we change the

- ✓ *Minimum* label to 400.0,
- ✓ *Maximum* label to 1400.0,
- ✓ *Major Unit* label to 100.0,
- ✓ *Minor Unit* label to 50.0,

Then we

- ✓ Choose *Minor tick mark type* as ***Outside,*** and
- ✓ tick on *Close*.



(a) **(b)**

Fig. 4.43

Step 24: The resulting graph with the changed horizontal axis is shown in Fig. 4.44.

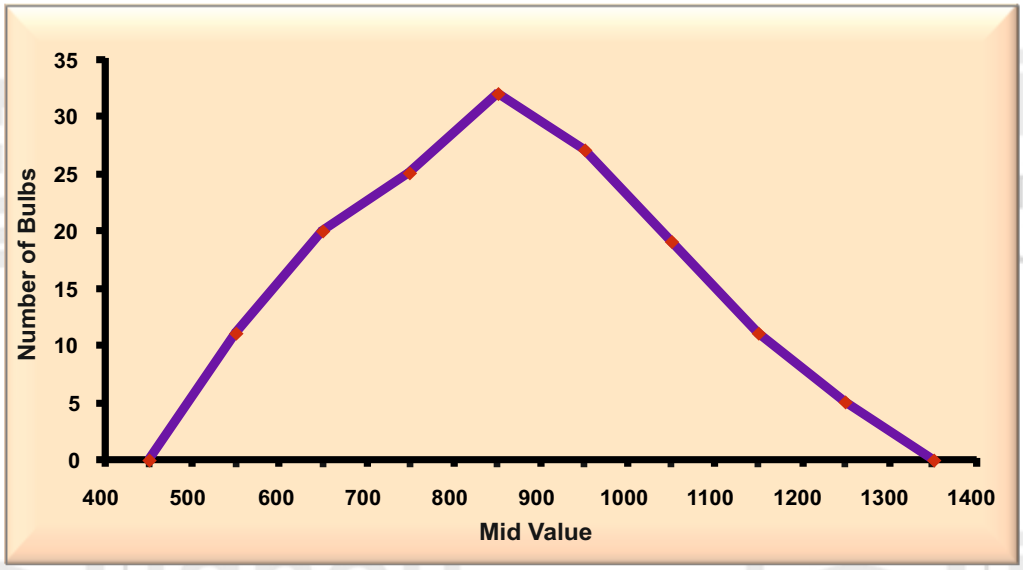


Fig. 4.44

4.7 FREQUENCY CURVE

In Unit 15 of MST-001, you have learnt how to construct the frequency curve for a continuous frequency distribution. The frequency curve is a smooth curve. We state the main steps as follows:

Steps 1 and 2 : The first two steps of plotting frequency curve are the same as explained in Sec. 4.6.

Step 3 : We join the points by a smooth curve instead of straight lines.

Steps 4 to 6 : The last three steps are the same as explained in Sec. 4.6.

Steps in Excel

You have learnt how to construct the frequency curve manually in Unit 15 of MST-001. In this lab session, we consider Problem 2 and explain the method of constructing the frequency curve using Excel 2007.

Step 1: The method of plotting the frequency curve in Excel is the same as explained in Steps 16 to 24 of Sec. 4.6. The only difference is that we select a chart subtype *Scatter with Smooth Lines and Markers* instead of a chart subtype *Scatter with Straight Lines and Markers* in Step 16 of Sec. 4.6. This is shown in Fig. 4.45.

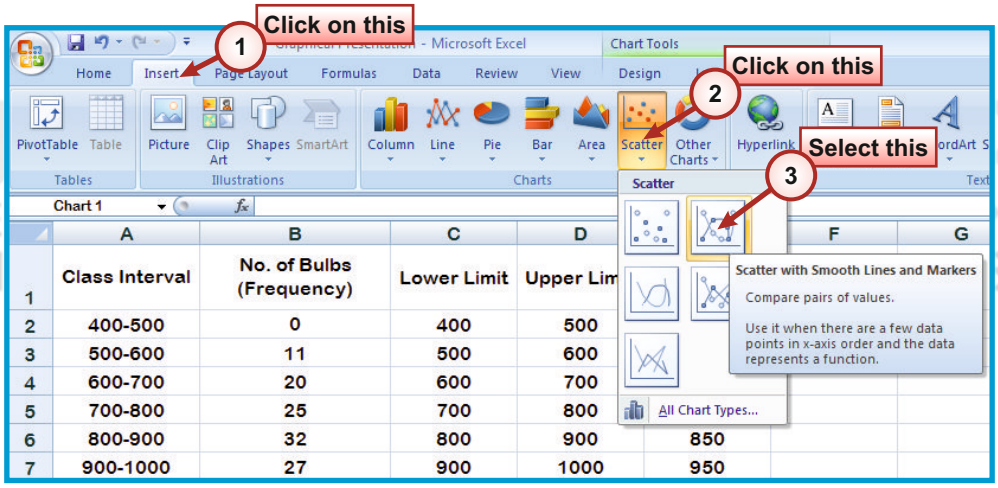


Fig. 4.45

Step 2: We now format this graph as explained in Sec. 3.3 of Lab Session 3. The resulting frequency curve is shown in Fig. 4.46.

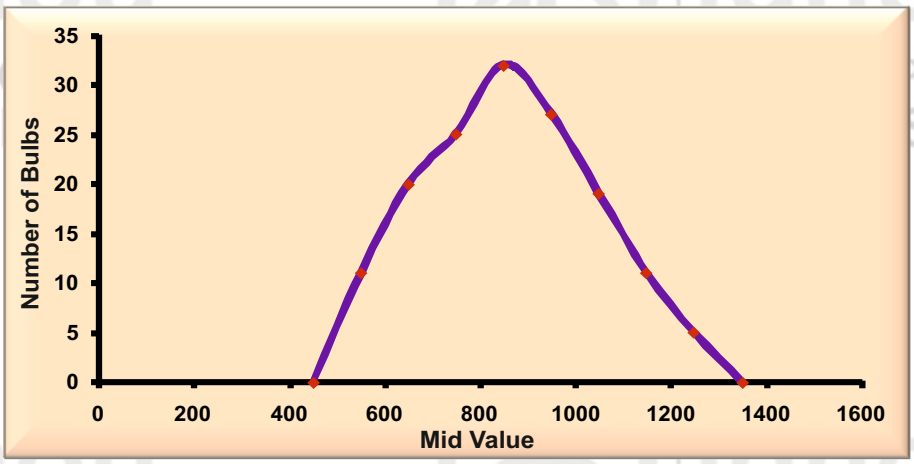


Fig. 4.46

4.8 CUMULATIVE FREQUENCY CURVES (OGIVES)

We plot two types of cumulative frequency curves, namely, less than frequency curve (or less than ogive) and more than frequency curve (or more than ogive) to study the graphical representation of less than and more than cumulative frequency distributions, respectively. Here we briefly mention the main steps as follows:

Less than ogive

Step 1: We take the upper limits of the class intervals on the X-axis and less than cumulative frequencies on the Y-axis and plot the respective points.

Step 2: We join the points by line segments. The curve so obtained is known as **less than cumulative frequency curve** or **less than ogive**.

More than ogive

Step 1: We take the lower limits of the class intervals on the X-axis and more than cumulative frequencies on the Y-axis and plot the respective points.

Step 2: We join the points by line segments. The curve so obtained is known as **more than cumulative frequency curve** or **more than ogive**.

Steps in Excel

In Unit 15 of MST-001, you have learnt that we use two forms of ogives depending on the types of cumulative frequencies. Here we consider Problem 2 and explain how to plot two types of ogives using MS Excel.

You can also continue with same Excel sheet to plot the ogives, which we used for histogram, frequency polygon and frequency curve.

Step 1: We enter the class intervals, frequencies, lower and upper limits of the class intervals in an Excel sheet in the same way as explained in Secs. 4.4 and 4.5 and name the sheet "**Ogive**" as shown in Fig. 4.47.

	A	B	C	D
1	Class Interval	No. of Bulbs (Frequency)	Lower Limit	Upper Limit
2	500-600	11	500	600
3	600-700	20	600	700
4	700-800	25	700	800
5	800-900	32	800	900
6	900-1000	27	900	1000
7	1000-1100	19	1000	1100
8	1100-1200	11	1100	1200
9	1200-1300	5	1200	1300

Ogive

Fig. 4.47

Step 2: We compute the less than and more than type cumulative frequencies as explained in Steps 16 to 19 of Sec. 2.5 under the heading 'Steps in Excel' of Lab Session 2 (Fig. 4.48).

	E	F
1	No. of Bulbs (Less Than Cumulative Frequency)	No. of Bulbs (More Than Cumulative Frequency)
2	11	150
3	31	139
4	56	119
5	88	94
6	115	62
7	134	35
8	145	16
9	150	5

Fig. 4.48

Step 3: To plot the more than type cumulative frequency curve (more than ogive), we

1. select Cells C2:C9 and F2:F9 by holding ***Ctrl*** key,
2. click on the ***Scatter*** under ***Insert*** tab, and
3. choose a ***Scatter with Straight Lines and Markers*** chart subtype as shown in Fig. 4.49.

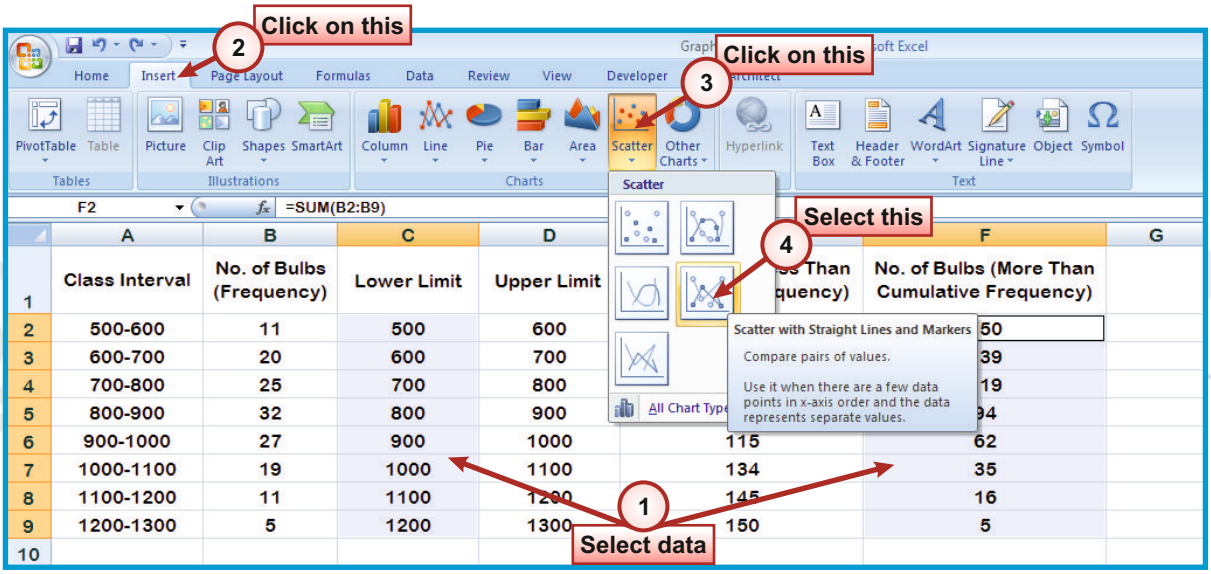


Fig. 4.49

Step 4: In this way, we get a more than type ogive as shown in Fig. 4.50.

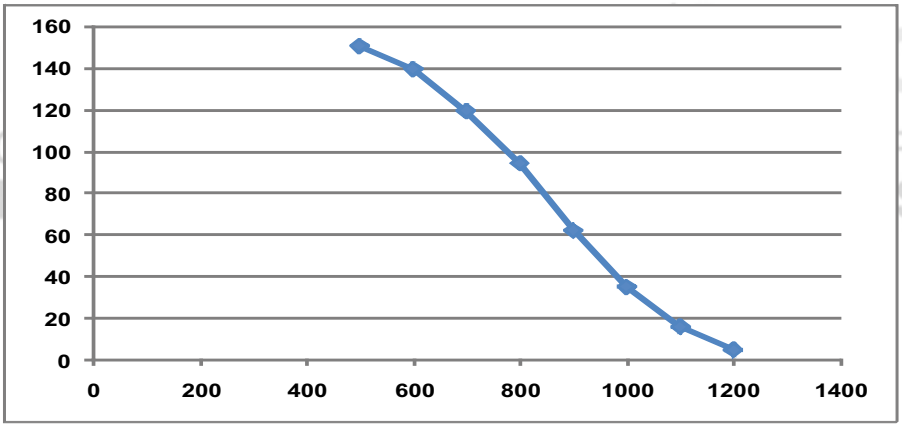


Fig. 4.50

Step 5: We now format this graph as explained in Sec. 3.3 of Lab Session 3. The resulting graph is shown in Fig. 4.51.

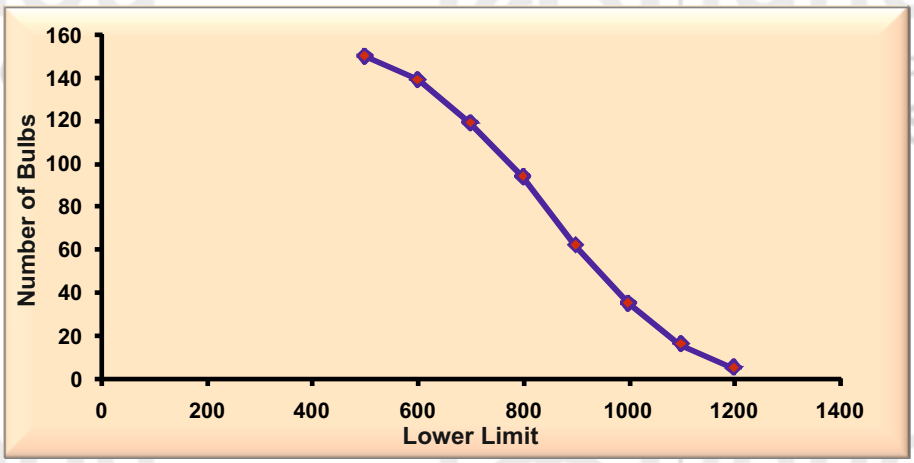


Fig. 4.51

Step 6: We can also change the minimum and maximum horizontal axis labels as shown in Fig. 4.52. For this, we

1. select the horizontal axis labels of the chart,
2. click on the *Format* tab under *Chart Tools*, and
3. choose *Format Selection* option.

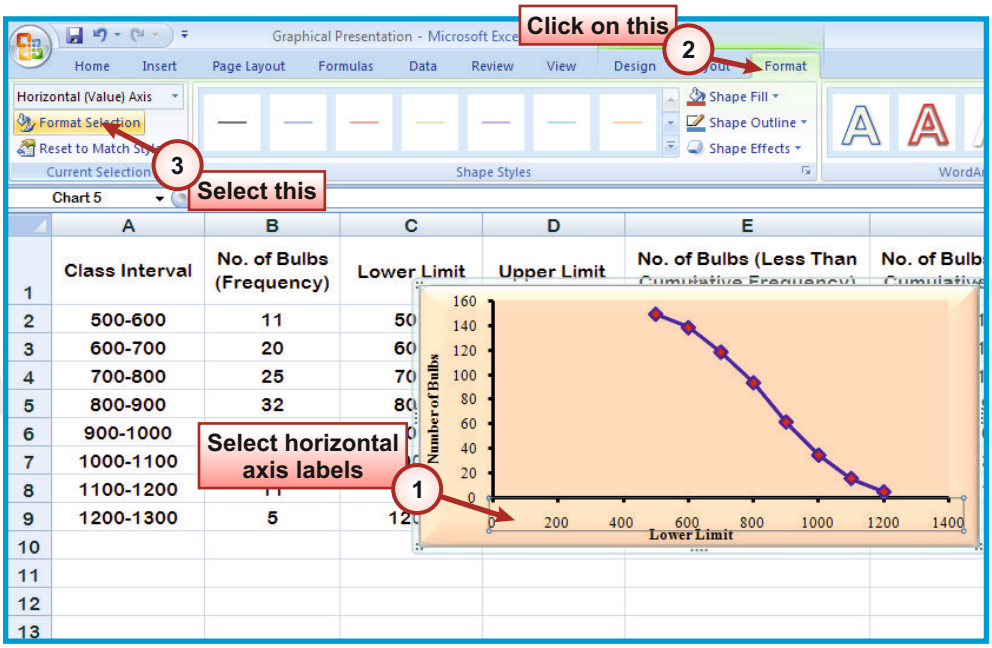


Fig. 4.52

Step 7: The dialog box shown in Fig. 4.53 appears. We change the

- ✓ *Minimum* label to 500.0,
- ✓ *Maximum* label to 1300.0, and
- ✓ *Close* the box.

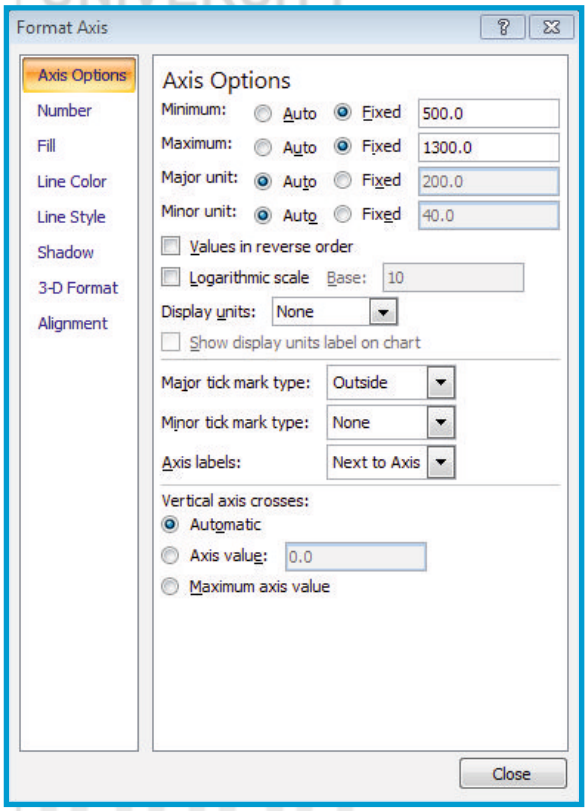


Fig. 4.53

Step 8: The more than ogive with the changed horizontal axis is shown in Fig. 4.54.

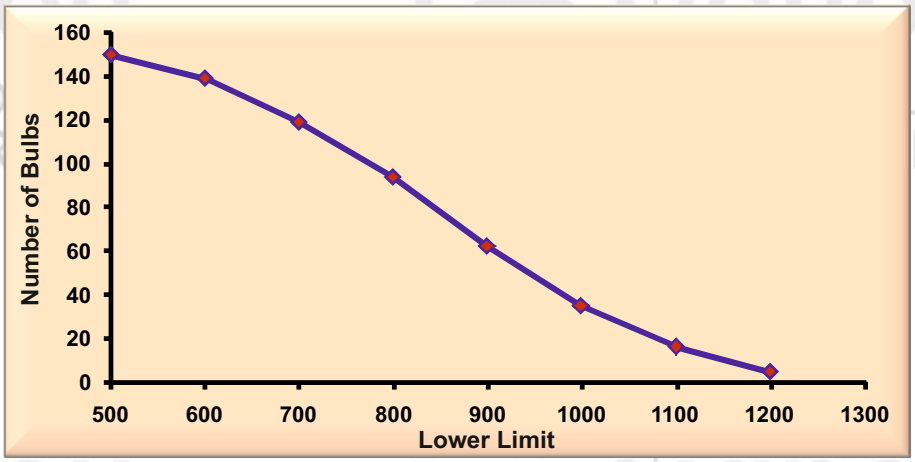


Fig. 4.54

Step 9: To plot the less than type cumulative frequency curve (less than ogive), we follow the steps shown in Fig. 4.55 and

1. select Cells D2:E9,
2. click on *Scatter* under *Insert* tab, and
3. choose the *Scatter with Straight Lines and Markers* chart subtype.

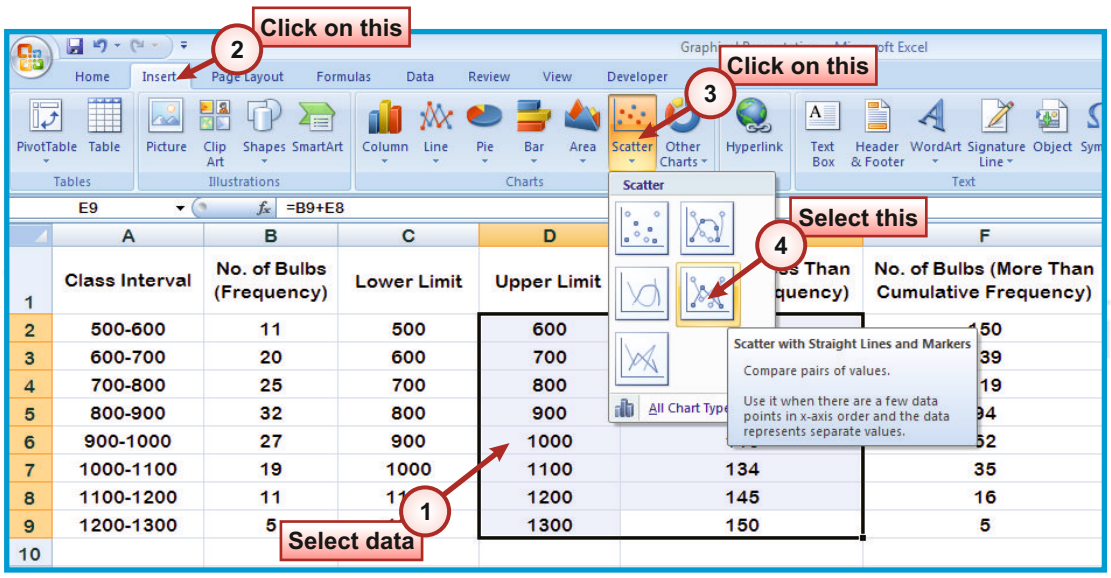


Fig. 4.55

Step 10: In this way, we get a less than type ogive as shown in Fig. 4.56.

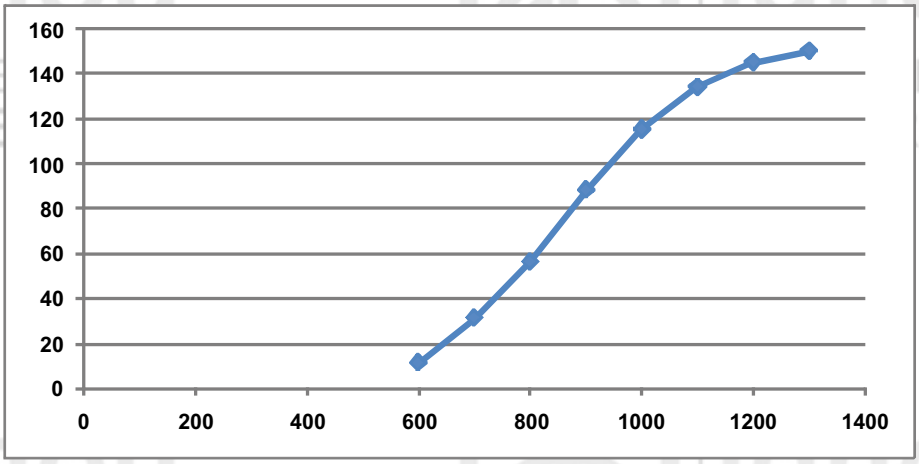


Fig. 4.56

Step 11: We now format this graph as explained in Sec. 3.3 of Lab Session 3. The resulting graph is shown in Fig. 4.57.

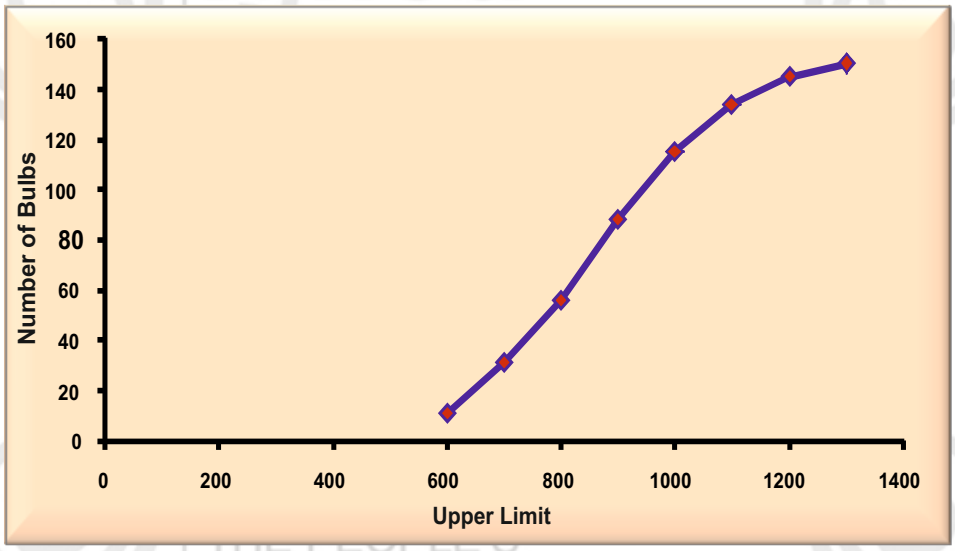


Fig. 4.57

Step 12: We can also change the minimum and maximum horizontal axis labels. For this, we repeat Step 6. The dialog box shown in Fig. 4.58 appears. We change the

- ✓ *Minimum* label to 500.0,
- ✓ *Maximum* label to 1400.0, and
- ✓ *Close* the box.

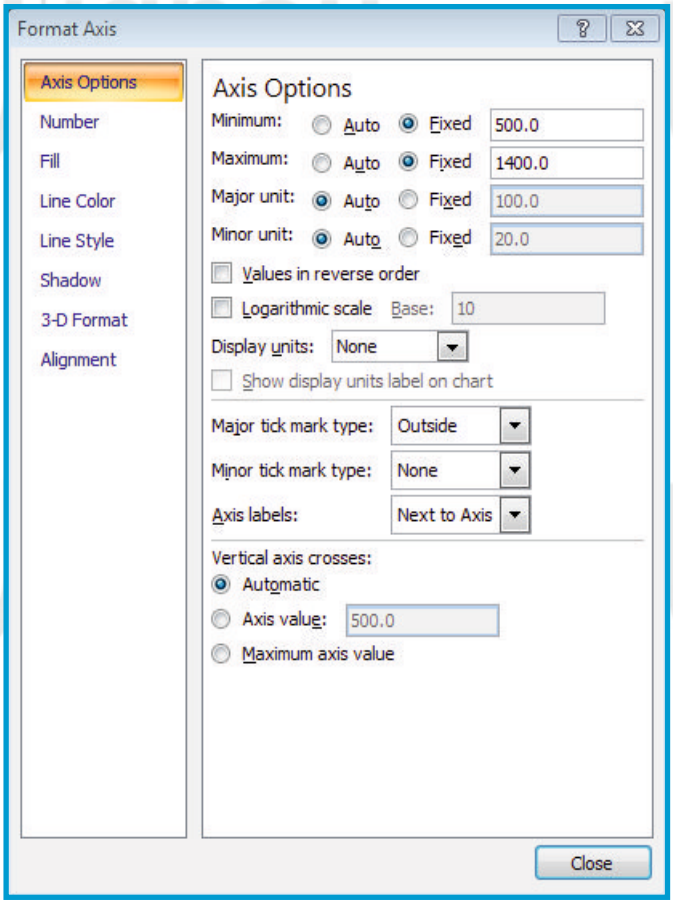


Fig. 4.58

Step 13: The less than ogive with the changed horizontal axis is shown in Fig. 4.59.

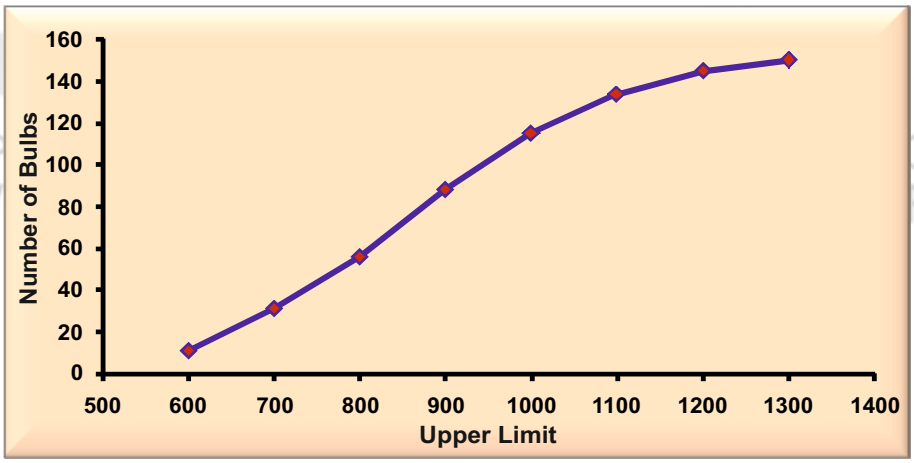


Fig. 4.59

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

Activity

Construct suitable graphs with the help of MS Excel 2007 for

A1) Examples 1 and 2 given in Unit 15 of MST-001.

A2) Exercises E1 to E7 given in Unit 15 of MST-001.

Match the results with the manual graphical representation of data carried out in Unit 15 of MST-001.

Continuous Assessment 4

1. Consider the data given in Table 5 of Lab Session 2. Represent the egg consumption in 120 families graphically using a frequency bar graph.
2. Consider the data given in Table 6 of Lab Session 2. Represent the rainfall patterns in 120 different areas of a state graphically using a histogram, frequency polygon, frequency curve and both ogives.

Home Work: Do It Yourself

1) Follow the steps explained in Secs. 4.3 to 4.8 to represent the data of Tables 2 and 3 of Lab Session 2 graphically. Use a different format for the graphs. Take the screenshots and keep them in your record book.
2) Develop the spreadsheets for the exercises of "Continuous Assessment 4" as explained in this lab session. Take the screenshots of the final spreadsheets and the graphs.
3) **Do not forget** to keep these screenshots in your record book as these will contribute to your continuous assessment in the Laboratory.

Graphical Presentation of Time Series Data

SESSION 5

Structure

5.1 Introduction
Objectives
5.2 Problem Description
5.3 Line Graph
5.4 Range Graph
5.5 Band Graph
5.6 Box Plot

5.1 INTRODUCTION

In Lab Session 4, you have learnt how to plot different types of graphs such as frequency bar graph, histogram, frequency polygon and frequency curve for representing frequency distributions. We have also discussed the less than and more than cumulative frequency curves, i.e., ogives for representing less than and more than cumulative frequency distributions, respectively. In Unit 16 of MST-001 (Foundation in Mathematics and Statistics), we have discussed the other graphical methods used for representing the time series data such as line, range and band graphs. You have also learnt how to construct the box plot in Unit 16.

Prerequisite

- Lab Sessions 1 and 3 of MSTL-001 (Basic Statistics Lab).
- Unit 16 of MST-001 (Foundation in Mathematics and Statistics).

In this lab session, we shall explain how to use Excel 2007 to draw three types of graphs, which we use to represent the time series data. These are the line, range and band graphs. We shall also explain how to construct the box plot used for identifying outliers and comparing distributions using Excel 2007.

Objectives

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

- prepare the spreadsheet in MS Excel 2007;
- plot the line, range and band graphs for time series data; and
- construct the box plot.

5.2 PROBLEM DESCRIPTION

1. The data of annual rainfall for the period from 1999 to 2013 are collected to study the rainfall patterns in a state and are given in Table 1.

Table 1: Annual rainfall data

Year	Annual Rainfall (in mm)	Year	Annual Rainfall (in mm)
1999	187.2	2007	173.1
2000	207.9	2008	192.7
2001	171.5	2009	157.3
2002	217.7	2010	244.4
2003	199.5	2011	182.2
2004	185.7	2012	241.5
2005	261.2	2013	232.4
2006	201.5		

Represent the time series data using a suitable graph.

2. The minimum and maximum temperatures during a day are collected for one week. The data are recorded in Table 2.

Table 2: Minimum and maximum temperature data

Days	Maximum Temperature (in ^{o}C)	Minimum Temperature (in ^{o}C)
Monday	40	18
Tuesday	41	20
Wednesday	38	16
Thursday	36	15
Friday	43	21
Saturday	42	20
Sunday	37	17

Represent the time series data using a suitable graph.

3. The production unit of a manufacturing company used five machines to produce glass bottles. The production of each machine for one year is recorded in Table 3.

Table 3: Number of glass bottles produced in one year

Month	Machine I	Machine II	Machine III	Machine IV	Machine V
January	155	135	180	195	165
February	164	144	189	204	174
March	159	139	184	199	169
April	168	148	193	208	178
May	154	134	179	194	164
June	163	143	188	203	173
July	158	138	183	198	168
August	167	147	192	207	177
September	153	133	178	193	163
October	162	142	187	202	172
November	157	137	182	197	167
December	166	146	191	206	176

Plot a suitable graph for this data.

4. A survey was conducted for 30 days to note the number of hours the television was watched in a family. A sample of four families was taken. The number of hours spent in watching TV by the families for 30 days are recorded in Table 4.

Table 4: Number of hours spent in watching TV

Days	Television Viewing Hours			
	Family A	Family B	Family C	Family D
1	5	2	6	3
2	1	2	7	5
3	5	6	6	7
4	2	4	3	3
5	2	3	3	6
6	4	3	2	2
7	2	1	2	5
8	3	2	6	7
9	4	1	5	4
10	5	2	3	6
11	2	3	5	4
12	2	4	4	4
13	5	5	6	7
14	3	6	3	2
15	3	1	4	4
16	3	4	7	6
17	2	3	5	6
18	2	3	2	2
19	5	6	6	8
20	2	3	2	2
21	4	5	5	3
22	4	3	3	5
23	2	4	2	2
24	2	4	5	8
25	2	5	5	3
26	3	4	2	2
27	3	5	5	5
28	5	6	4	7
29	2	5	6	2
30	4	2	2	7

Draw the box plot to compare the hours of TV watched by the families for the data of Table 4.

5.3 LINE GRAPH

In Unit 16 of MST-001, you have learnt how to present time series data using the line graph. We use line graph to represent the values of the variables for different time periods. Here we briefly mention the main steps as follows:

Step 1: We take the time (year, months, weeks, days, etc.) on the horizontal axis and the observed data on the vertical axis.

Step 2: We plot the points against the magnitudes of the observed data corresponding to each successive time period.

Step 3: We join these points by line segments.

Steps in Excel

You have learnt the manual plotting of the line graph in Unit 16 of MST-001. In Problem 1, the annual rainfall data are given for 15 years. We can represent the data of Table 1 by the line graph. Here we describe the procedure of plotting the line graph in MS Excel 2007 as follows:

Step 1: We enter the data given in Table 1 in the Excel 2007 spreadsheet as shown in Fig. 5.1.

	A	B
1	Year	Annual Rainfall (in mm)
2	1999	187.2
3	2000	207.9
4	2001	171.5
5	2002	217.7
6	2003	199.5
7	2004	185.7
8	2005	261.2
9	2006	201.5
10	2007	173.1

Fig. 5.1: Partial screenshot of the spreadsheet for the given data.

Step 2: To obtain the line graph, we refer to Fig. 5.2. We

1. select the data given in Cells B2:B16,
2. click on the *Insert* tab,
3. click on the *Line* option in the *Charts* group, and
4. select a chart subtype *Line with Markers* that we wish to use.

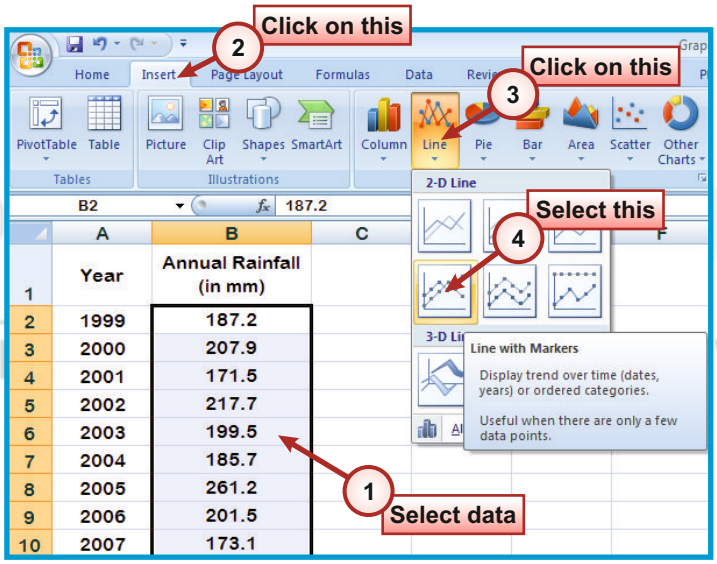


Fig. 5.2

Step 3: When we click on the chart subtype, we get the chart shown in Fig. 5.3. It is called the line graph.

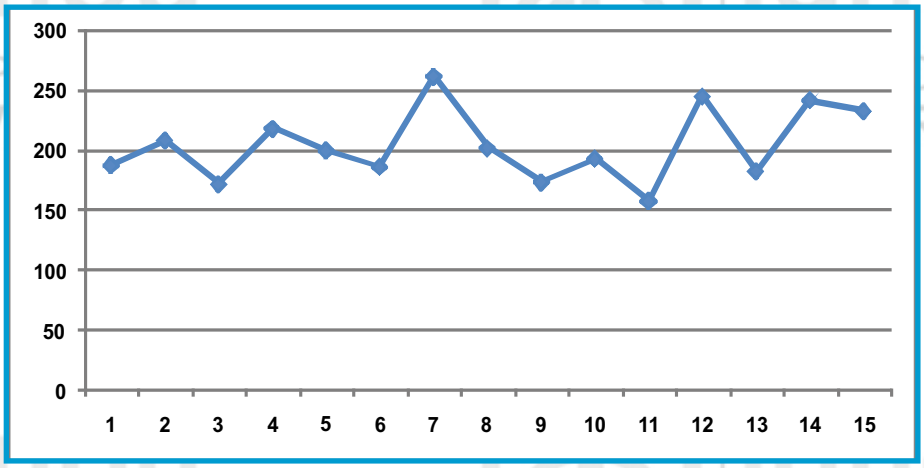


Fig. 5.3

Step 4: We format the chart as explained in Steps 4 to 9 of Sec. 3.3 of Lab Session 3. The resulting line graph is shown in Fig. 5.4.

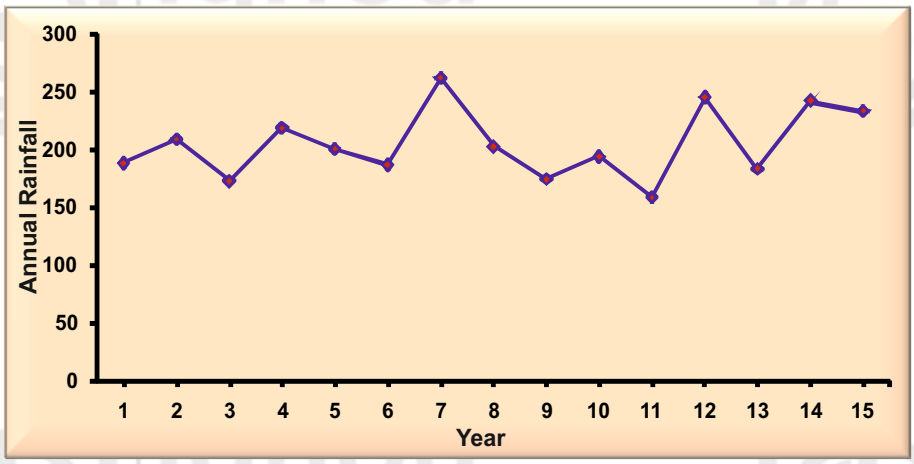


Fig. 5.4

Step 5: We now change the labels of the horizontal axis to years ranging from 1999 to 2013. For this, we follow the steps shown in Fig. 5.5 and

1. select the horizontal axis labels of the chart,
2. click on the *Design* tab under *Chart Tools*, and
3. click on *Select Data*.

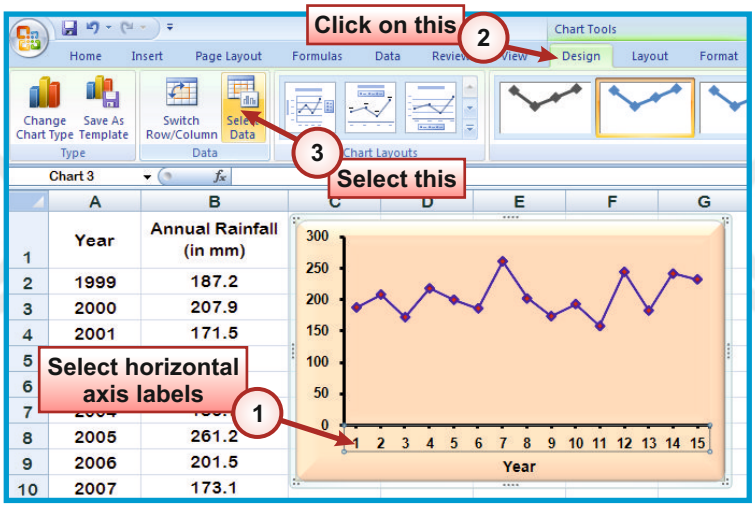


Fig. 5.5

Step 6: A new dialog box as shown in Fig. 5.6 appears. We click on *Edit* under *Horizontal (Category) Axis Labels* as shown in Fig 5.6.

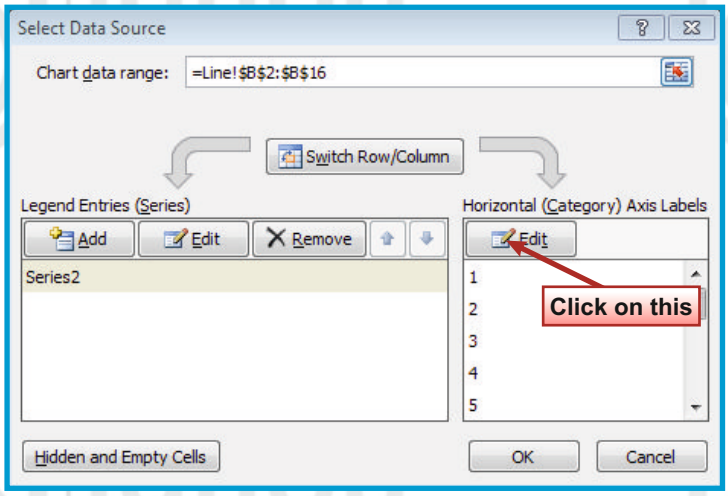
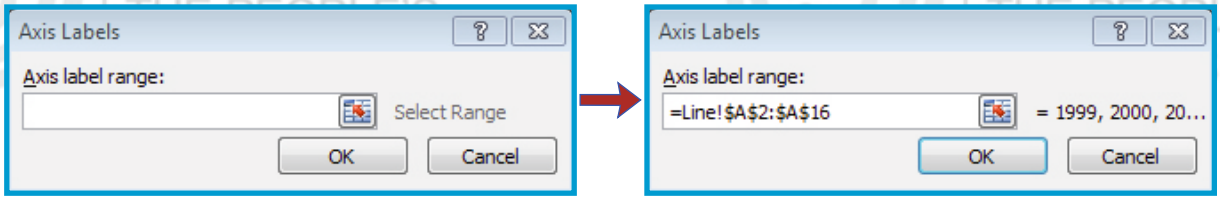


Fig. 5.6

Step 7: Again a new dialog box as shown in Fig. 5.7a appears. We select Cells A2:A16 as *Axis label range* as shown in Fig. 5.7b and click on *OK*.



(a) (b)

Fig. 5.7

Step 8: Now the dialog box shown in Fig. 5.6 changes to the one shown in Fig 5.8 and we click on *OK*.

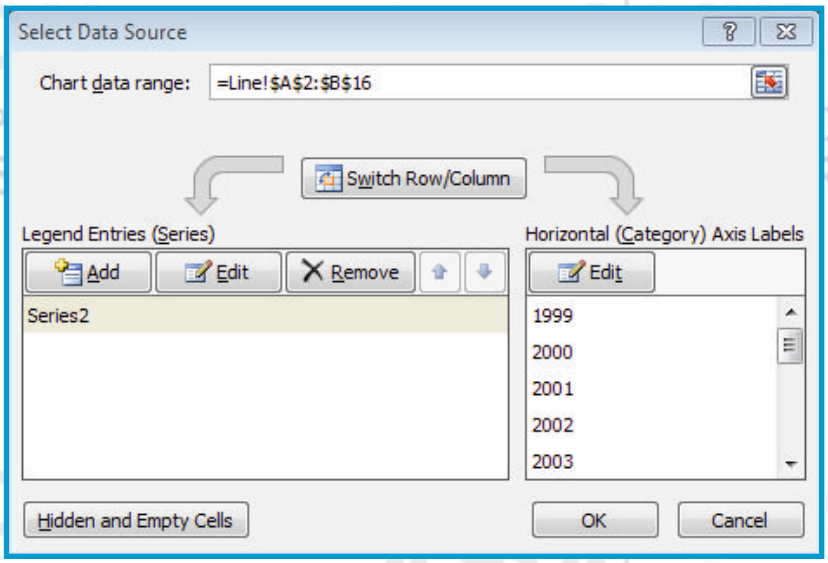


Fig. 5.8

Step 9: The graph with years on the horizontal axis is shown in Fig. 5.9.

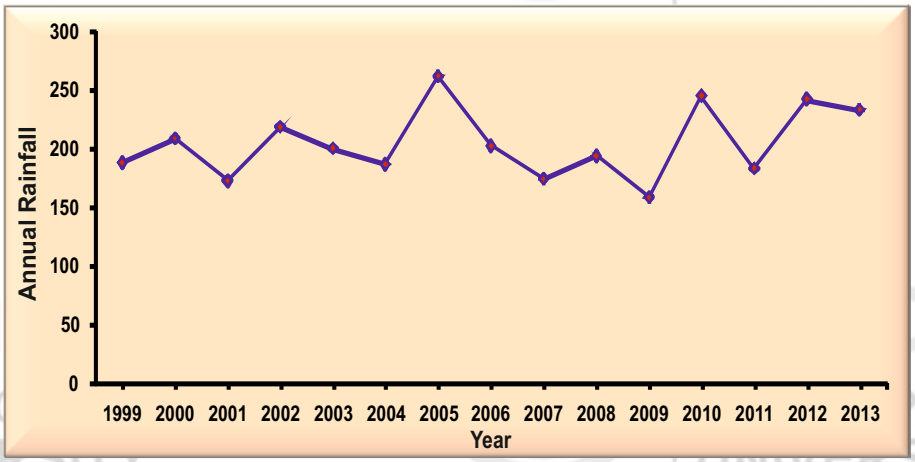


Fig. 5.9

5.4 RANGE GRAPH

In Unit 16 of MST-001, you have learnt how to present data using the range graph. We use the range graph to show the variation and fluctuation between the maximum and minimum values of a variable at the same point of time. For example, the minimum and maximum price of a product, marks of a student, temperature, etc. Here we briefly mention the main steps as follows:

Step 1: We take time on the horizontal axis and corresponding value of the given variable on the vertical axis.

Step 2: We draw two line graphs for the minimum and maximum values of data.

Step 3: The gap between both the curves represents the range of variation. We draw the vertical lines joining two line graphs.

Steps in Excel

You have learnt the manual plotting of range graph in Unit 16 of MST-001. In Problem 2, we can show the variation of the minimum and maximum temperatures using the range graph. Here we describe the procedure of plotting the range graph in MS Excel 2007 as follows:

Step 1: We enter the data given in Table 2 in Excel 2007 spreadsheet as shown in Fig. 5.10.

	A	B	C
1	Days	Minimum Temperature	Maximum Temperature
2	Monday	40	18
3	Tuesday	41	20
4	Wednesday	38	16
5	Thursday	36	15
6	Friday	43	21
7	Saturday	42	20
8	Sunday	37	17
9			

Fig. 5.10: Partial screenshot of the spreadsheet for the given data.

Step 2: To obtain the line graph, we refer to Fig. 5.11. We

1. select the data given in Cells B1:C8,
2. click on the *Insert* tab,
3. click on the *Line* option in the *Charts* group, and
4. select a chart subtype *Line with Markers* that we wish to use.

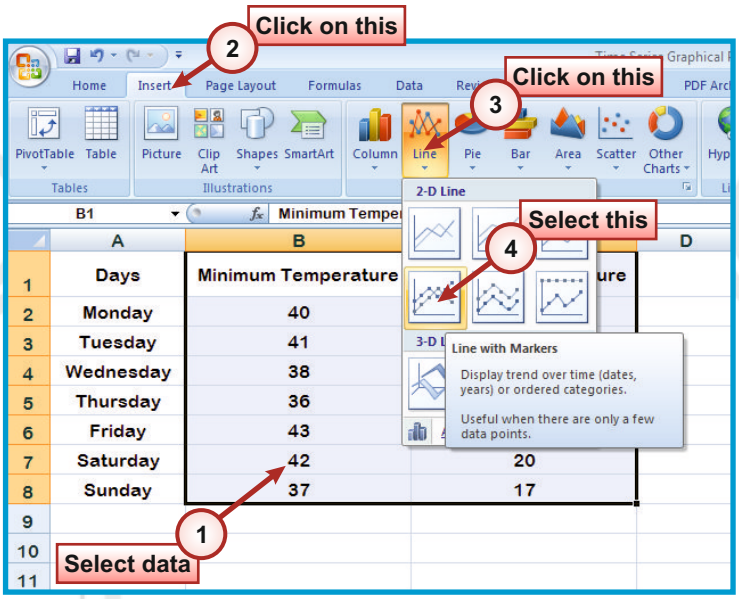


Fig. 5.11

Step 3: The resulting line graph is shown in Fig. 5.12.

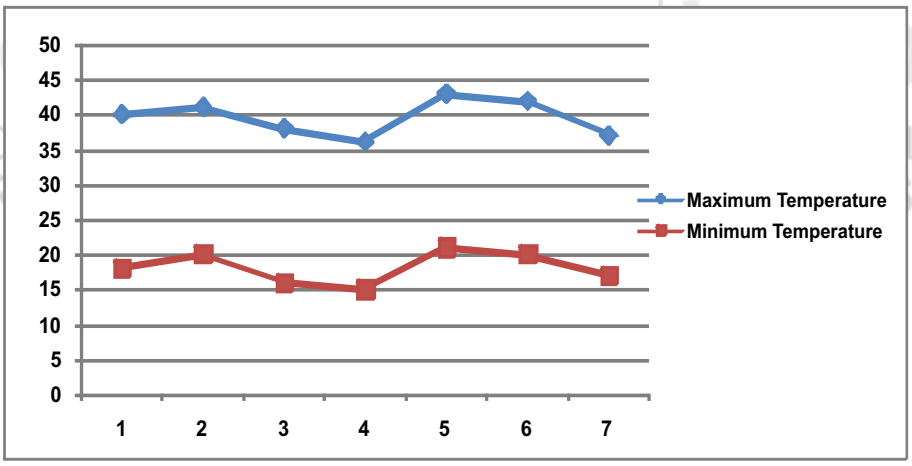


Fig. 5.12

Step 4: We format the graph as explained in Steps 4 to 9 of Sec. 3.3 of Lab Session 3. The resulting graph is shown in Fig. 5.13.

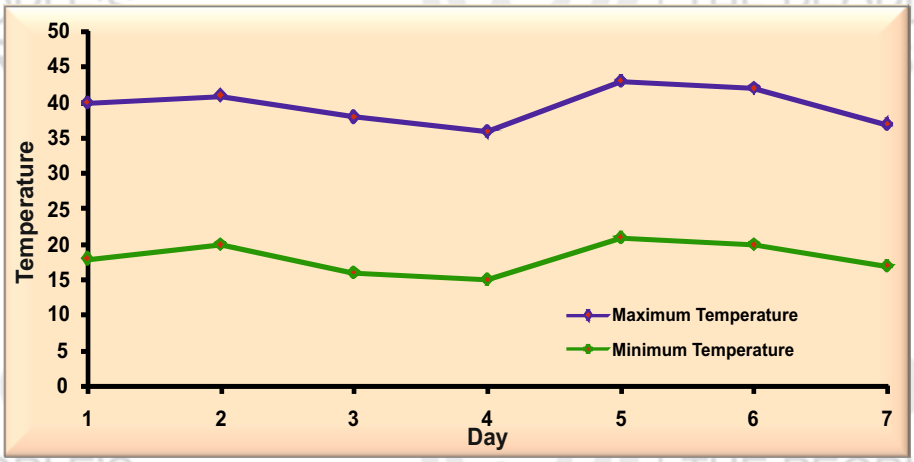


Fig. 5.13

Step 5: We now change the labels of the horizontal axis to days ranging from Monday to Sunday as explained in Steps 5 to 8 of Sec. 5.3. The resulting graph is shown in Fig. 5.14.

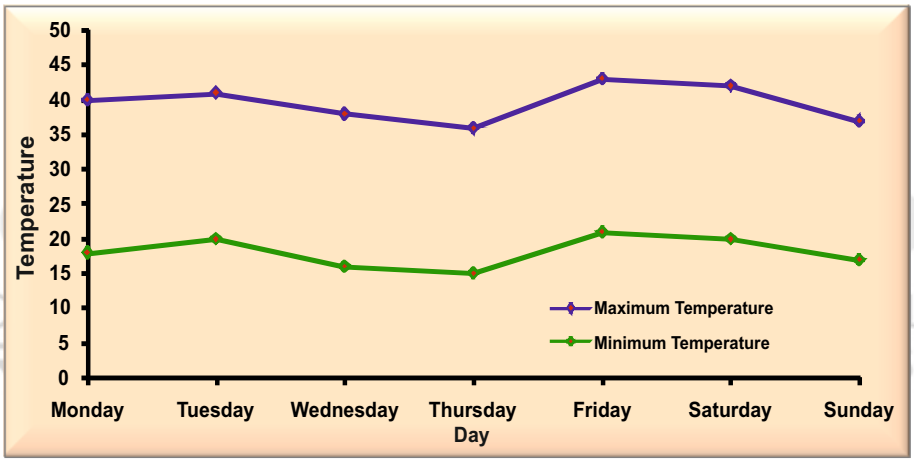


Fig. 5.14

Step 6: To change the look of the line graph shown in Fig. 5.14 as in the range graph, we follow the steps shown in Fig. 5.15 and

1. click on the chart,
2. click on the *Layout* tab under *Chart Tools*, and
3. choose *High-Low Lines* under *Lines* option.

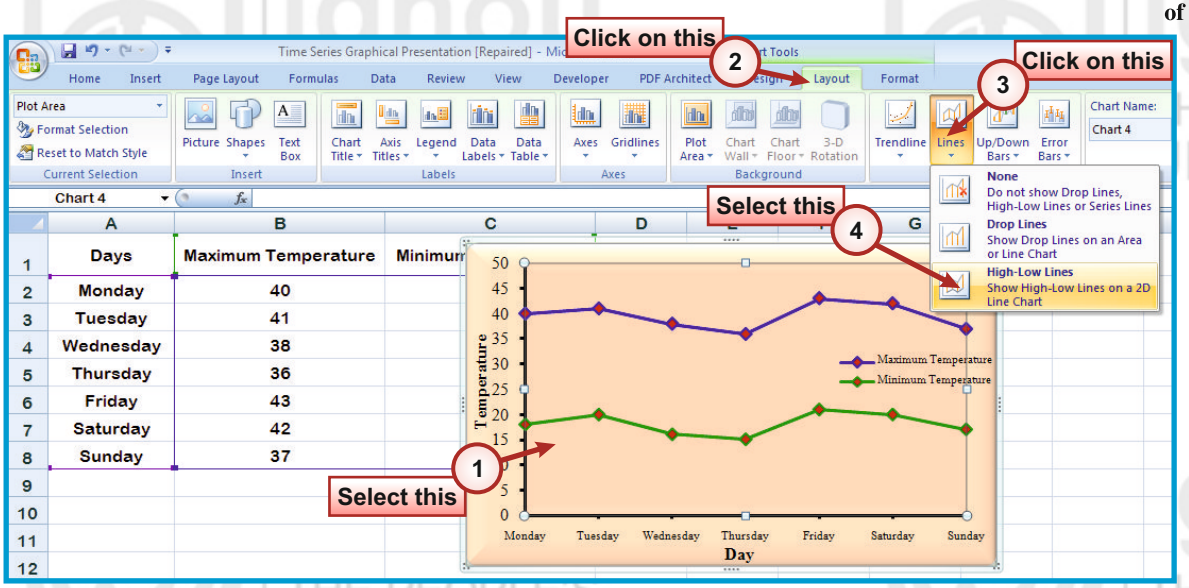


Fig. 5.15

Step 7: The resulting range graph is shown in Fig. 5.16.

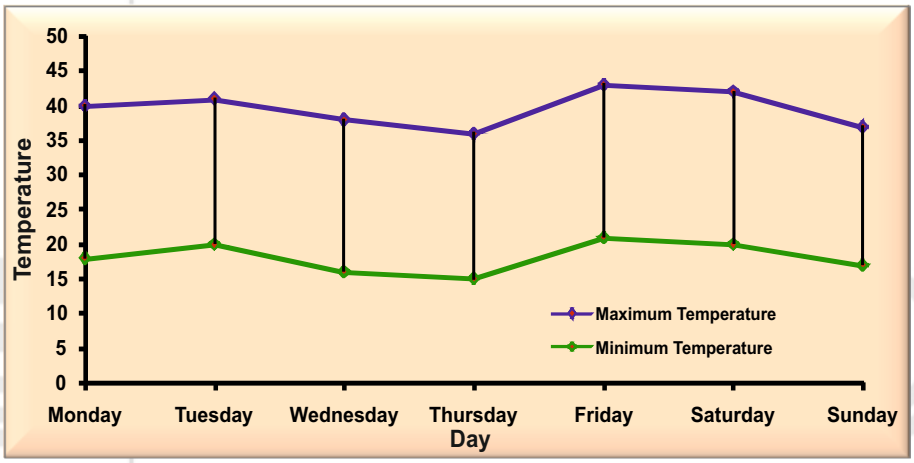


Fig. 5.16

5.5 BAND GRAPH

In Unit 16 of MST-001, you have learnt how to present time series data using the band graph. We use the band graph to show the total components, which may be broken up into sub divisions for each component of the total. We plot the different components of the total as line graphs, one over the other in the form of a series of bands. Here we briefly mention the main steps for drawing the band graph as follows:

Step 1: We take time on the horizontal (X) axis and the variable under consideration on the vertical (Y) axis.

Step 2: We plot the points for different time periods corresponding to the first component since here we have more than one component.

Step 3: We join the points by line segments.

Step 4: We now add the values of the first and second components. We plot these values on the graph and join them by line segments.

Step 5: For plotting the third component, we add the first, second and third components and plot them.

Step 6: We repeat this procedure for all components.

Step 7: We differentiate the gaps between the successive lines using different shades, colours, etc., so that it looks like a series of bands.

Steps in Excel

You have learnt the manual plotting of band graph in Unit 16 of MST-001. In Problem 3, we can show the number of glass bottles produced by five machines for different time periods using the band graph. Here we describe the procedure of plotting the band graph in MS Excel 2007 as follows:

Step 1: We enter the data given in Table 3 in Excel 2007 spreadsheet as shown in Fig. 5.17.

	A	B	C	D	E	F
1	Month	Machine I	Machine II	Machine III	Machine IV	Machine V
2	January	155	135	180	195	165
3	February	164	144	189	204	174
4	March	159	139	184	199	169
5	April	168	148	193	208	178
6	May	154	134	179	194	164
7	June	163	143	188	203	173
8	July	158	138	183	198	168
9	August	167	147	192	207	177
10	September	153	133	178	193	163
11	October	162	142	187	202	172
12	November	157	137	182	197	167
13	December	166	146	191	206	176

Fig. 5.17: Partial screenshot of the spreadsheet for the given data.

Step 2: To obtain the band graph, we follow the steps shown in Fig. 5.18. We

1. select the data given in Cells A1:F13,
2. click on the *Insert* tab,
3. click on the *Area* option in the *Charts* group, and
4. select a second chart subtype *Stacked Area* that we wish to use.

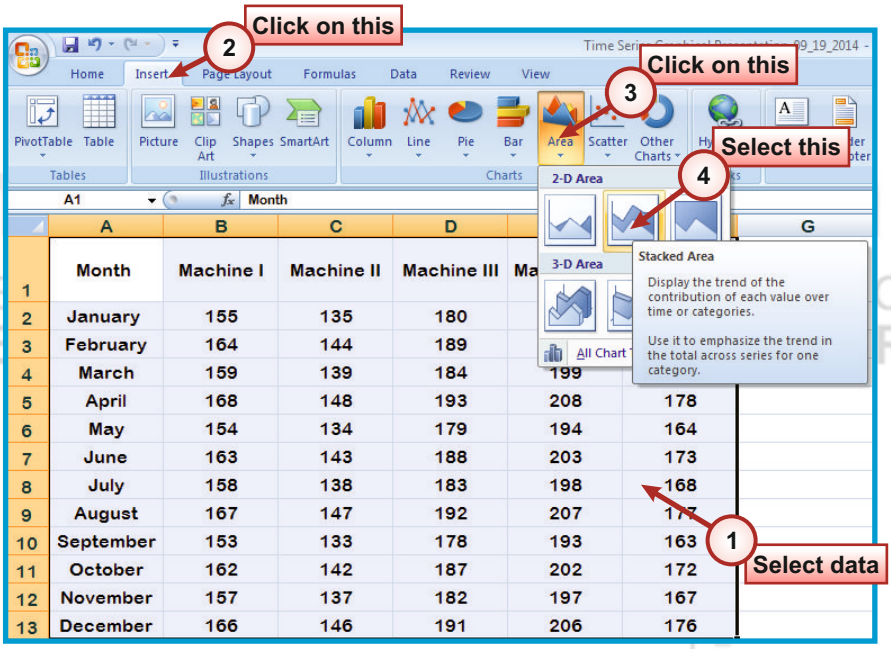


Fig. 5.18

Step 3: The resulting band graph is shown in Fig. 5.19.

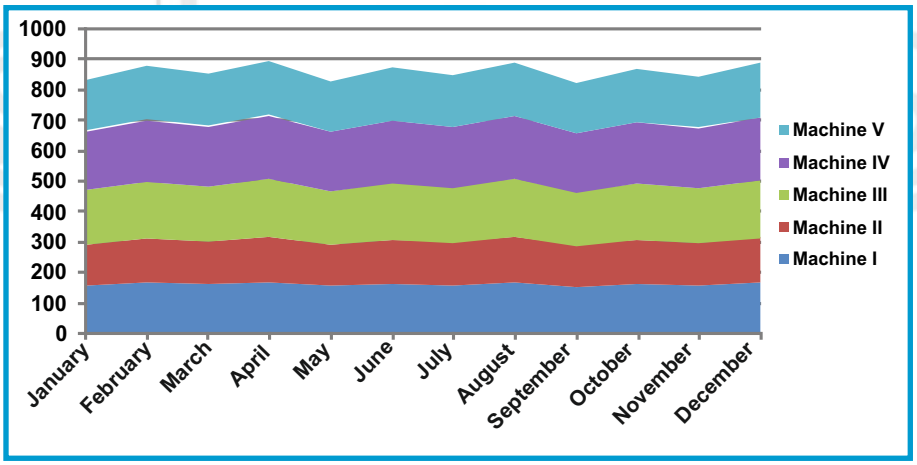


Fig. 5.19

Step 4: We format the graph as explained in Steps 4 to 9 of Sec. 3.3 of Lab Session 3. The resulting graph is shown in Fig. 5.20.

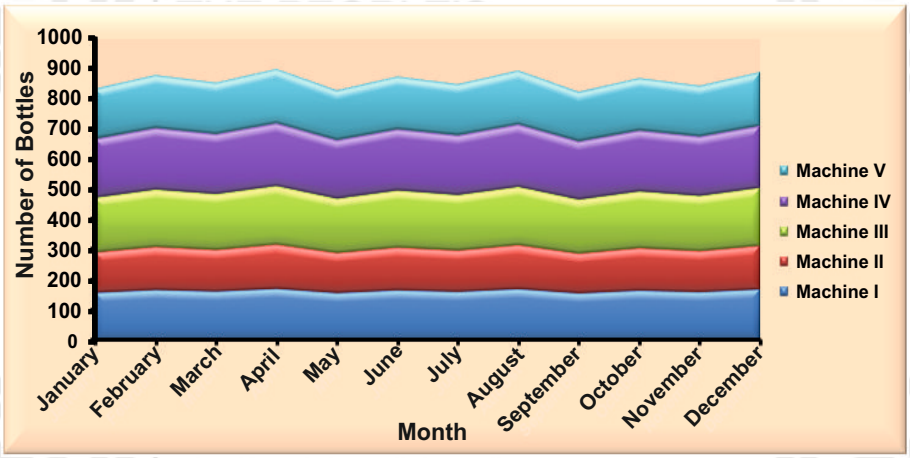


Fig. 5.20

5.6 BOX PLOT

In Unit 16 of MST-001, we have explained the box plot in detail. The box plot is also known as the box-and-whisker plot. We use the box plot to represent the distribution of the data visually. It helps in identifying outliers and comparing distributions. It is more useful when we have a large number of observations and also when we need to compare two or more data sets. The box plot provides a graphical representation of the data through its five number summary, i.e., minimum value, maximum value, median, and the first and third quartiles. Here we briefly mention the main steps as follows:

Step 1: We arrange the data in ascending order.

Step 2: We determine the minimum and maximum value of the data.

Step 3: We also compute the median, the first and third quartiles of the data.

Step 4: We plot the values computed in Steps 2 and 3 on the graph.

Step 5: We draw the box around the median taking the end points at the first and third quartiles.

Step 6: We draw the line from the first quartile to the minimum value and also from third quartile to the maximum value. These lines are called whiskers.

Steps in Excel

You have learnt the manual construction of box plot in Unit 16 of MST-001. Here we describe the procedure of constructing the box plot in MS Excel 2007 as follows:

Step 1: We enter the data given in Table 4 in Excel 2007 spreadsheet as shown in Fig. 5.21.

	A	B	C	D	E
1		Television Viewing Hours			
2	Days	Family A	Family B	Family C	Family D
3	1	5	2	6	3
4	2	1	2	7	5
5	3	5	6	6	7
6	4	2	4	3	3

Fig. 5.21: Partial screenshot of the spreadsheet for the given data.

Step 2: For constructing the box plot, we type "Third Quartile", "Maximum Value", "Minimum Value", "Median" and "First Quartile" in Cells A34:A38 without changing the sequence as shown in Fig. 5.22.

	A
33	
34	Third Quartile
35	Maximum Value
36	Minimum Value
37	Median
38	First Quartile

Fig. 5.22

Step 3: To compute the third quartile for the given data, we

1. select Cell B34,
2. click on ***More Functions*** under the ***Formulas*** tab, and
3. choose the ***Quartile*** function under ***Statistical*** as shown in Fig. 5.23.

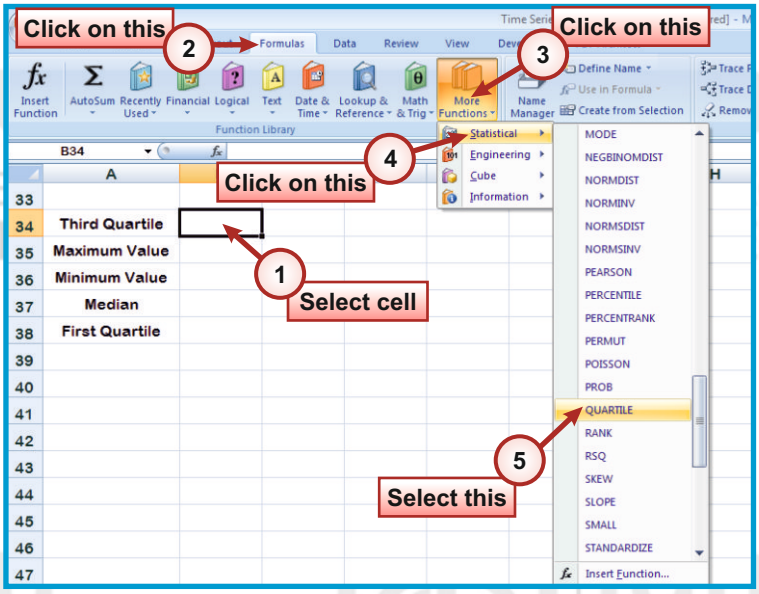


Fig. 5.23

Step 4: When we choose the ***Quartile*** function, the dialog box shown in Fig. 5.24 appears.

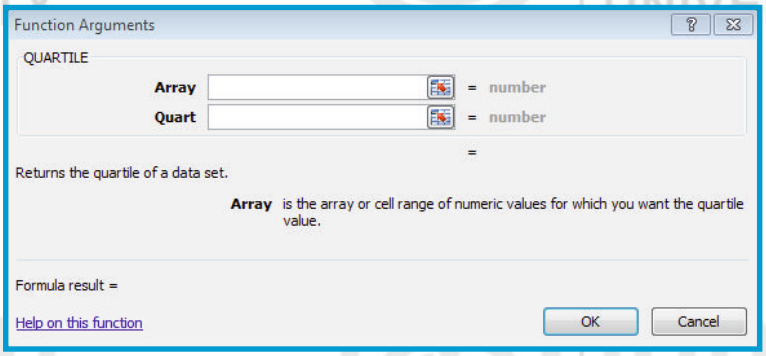


Fig. 5.24

Step 5: We select Cells B3:B32 as *Array* and type "3" in *Quart* as shown in Fig. 5.25. We click on *OK* to obtain the result.

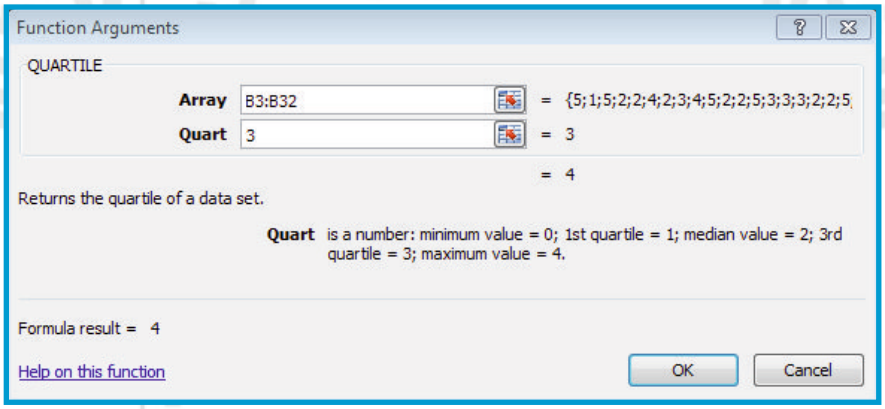


Fig. 5.25

Step 6: We obtain the value of the third quartile, i.e., 4, in Cell B34 as shown in Fig. 5.26.

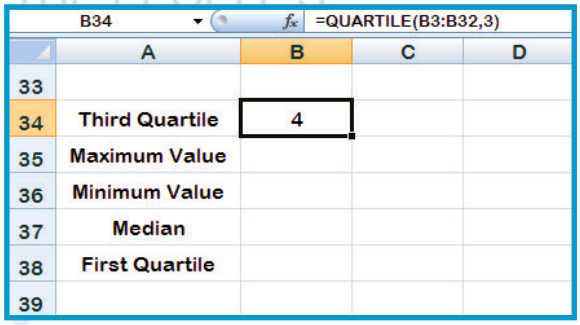


Fig. 5.26

Step 7: After selecting Cell B35, we repeat Step 3 to compute the maximum value. We select Cells B3:B32 as *Array*, type "4" in *Quart* and click on *OK* as shown in Fig. 5.27.

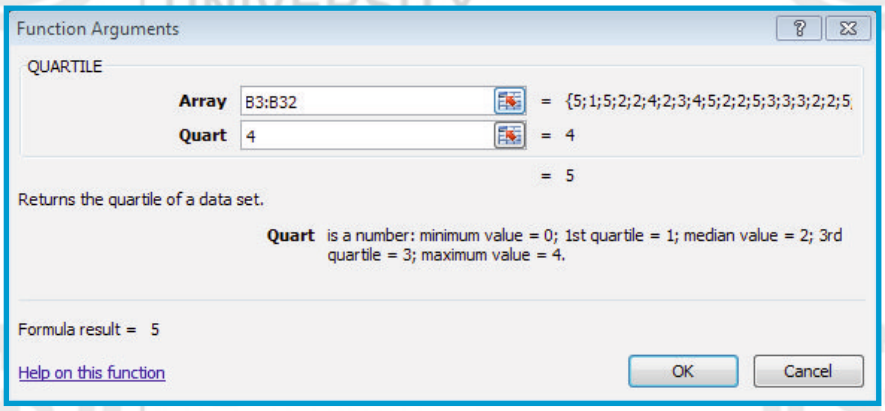


Fig. 5.27

Step 8: The maximum value of the given data is shown in Cell B35 (Fig. 5.28).

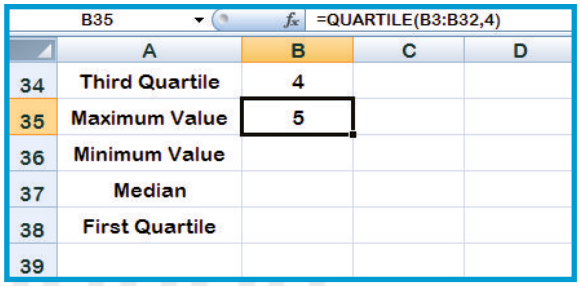


Fig. 5.28

We can also use "=Max(B3:B32)" function instead of "=Quartile (B3:B32,4)" function in Cell B35 to compute the maximum value of the data.

Step 9: After selecting Cell B36, we repeat Step 3 to compute the minimum value. We Select Cells B3:B32 as *Array*, type "0"in *Quart* and click on *OK* as shown in Fig. 5.29.

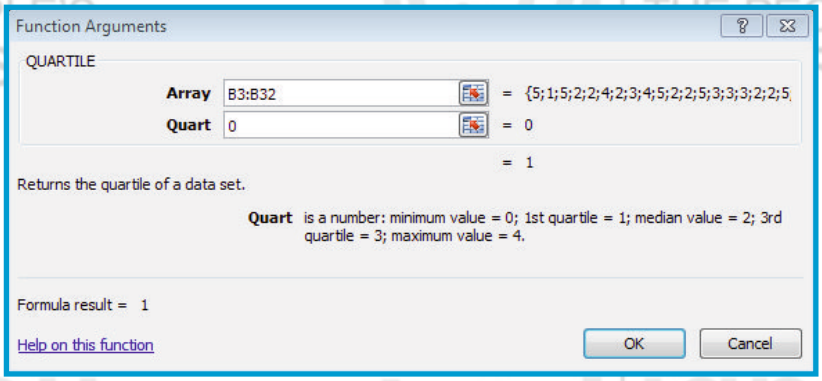


Fig. 5.29

Step 10: The minimum value of the given data is shown in Cell B36 (Fig. 5.30).

We can also use "=Min(B3:B32)" function instead of "=Quartile (B3:B32,0)" function in Cell B36 to compute the minimum value of the data.

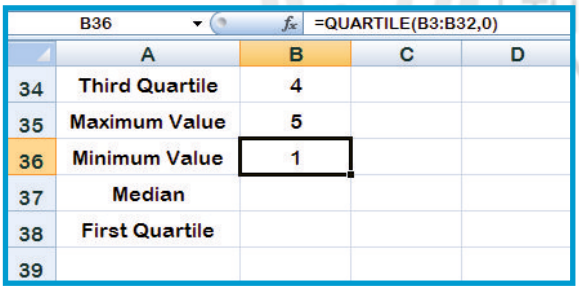


Fig. 5.30

Step 11: After selecting Cell B37, we repeat Step 3 to compute the median of the given data. We select Cells B3:B32 as *Array*, type "2" in *Quart* and click on *OK* as shown in Fig. 5.31.

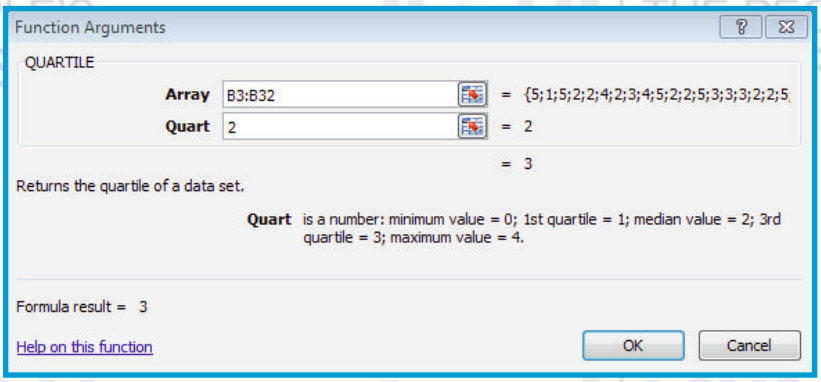


Fig. 5.31

Step 12: The median of the given data is shown in Cell B37 (Fig. 5.32).

We can also use "=Median(B3:B32)" function instead of "=Quartile(B3:B32,2)" function in Cell B7 to compute the median.

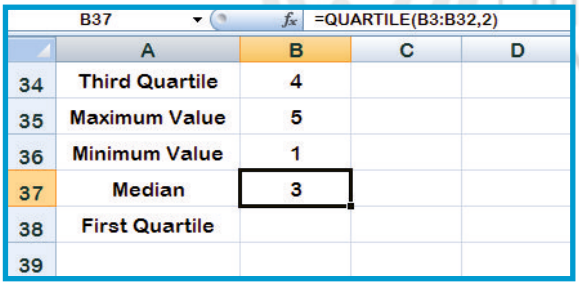


Fig. 5.32

Step 13: After selecting Cell B38, we repeat Step 3 to compute the first quartile. We select Cells B3:B32 as *Array*, type "1" in *Quart* and click on *OK* as shown in Fig. 5.33.

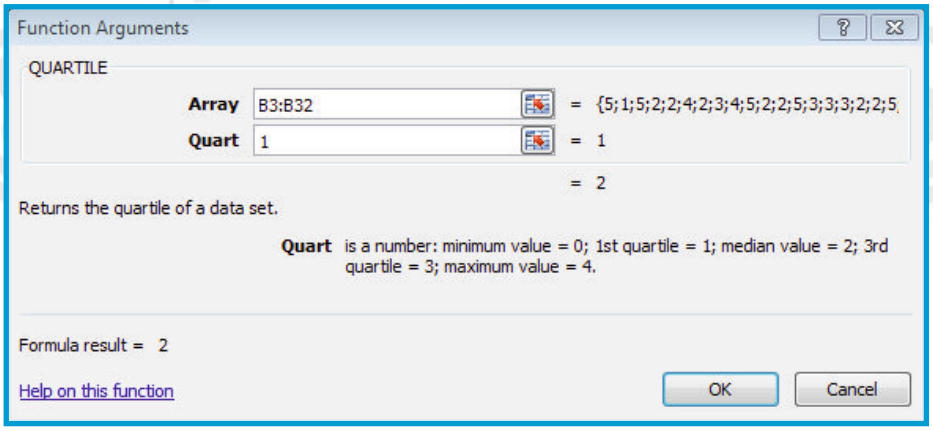
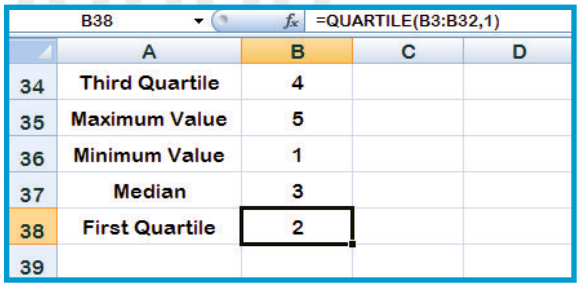


Fig. 5.33

Step 14: The first quartile of the given data is shown in Cell B38 (Fig. 5.34).

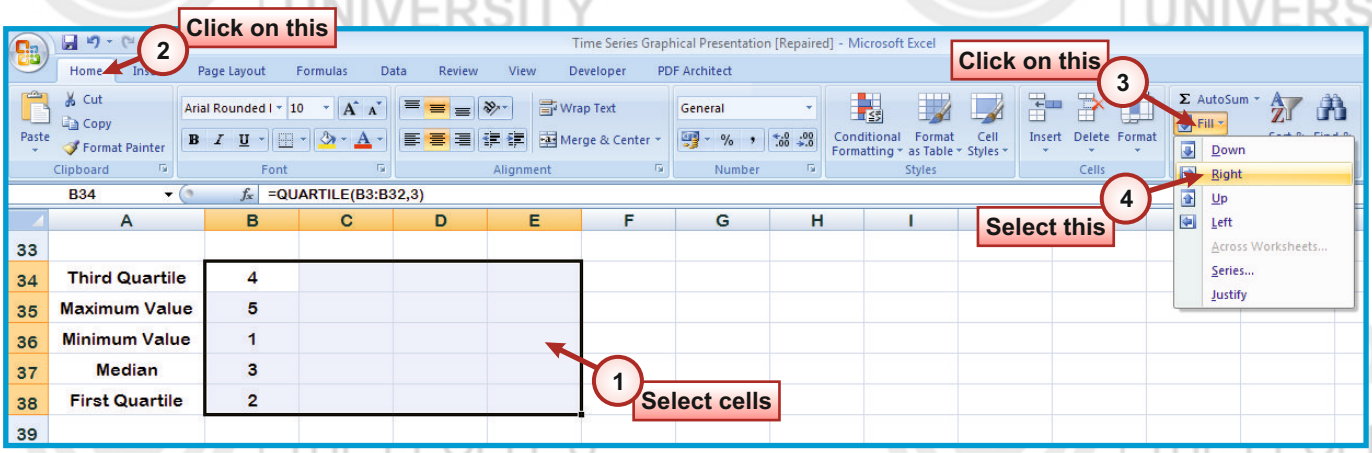


	A	B	C	D
34	Third Quartile	4		
35	Maximum Value	5		
36	Minimum Value	1		
37	Median	3		
38	First Quartile	2		
39				

Fig. 5.34

Step 15: To obtain these five values for the remaining families, we follow Fig. 5.35 and

1. select Cells B34:E38,
2. click on the *Home* tab, and
3. choose *Right* under the *Fill* option.



	A	B	C	D	E	F
33		Family A	Family B	Family C	Family D	
34	Third Quartile	4	5	5.75	6	
35	Maximum Value	5	6	7	8	
36	Minimum Value	1	1	2	2	
37	Median	3	3.5	4.5	4.5	
38	First Quartile	2	2.25	3	3	
39						

We can also drag Cells B34:B38 right up to column E using mouse.

Fig. 5.35

We do not have a built-in chart type in Excel to create box-and-whiskers plots. However, we create it by modifying the *Stock* chart given in Excel. It is a type of *Open-High-Low-Close* chart.

If you have less than 4 columns of data, you will not be able to use this *Stock* chart and you will get an error message. Here is a trick for you. If you have less than four columns (or variables) in the data, you should include additional blank columns in your selection of data while plotting the chart to make it four. The Excel will not plot the blank columns. Once the graph is created, you can use *Select Data* option and reduce the size of your series. It create problem only when we initiate to create the graph.

Step 16: For constructing the box plot, we follow Fig. 5.36 and

1. select Cells A33:E37,
2. click on *Other Charts* under *Insert* tab, and
3. choose second chart subtype *Open-High-Low-Close* under *Stock* option.

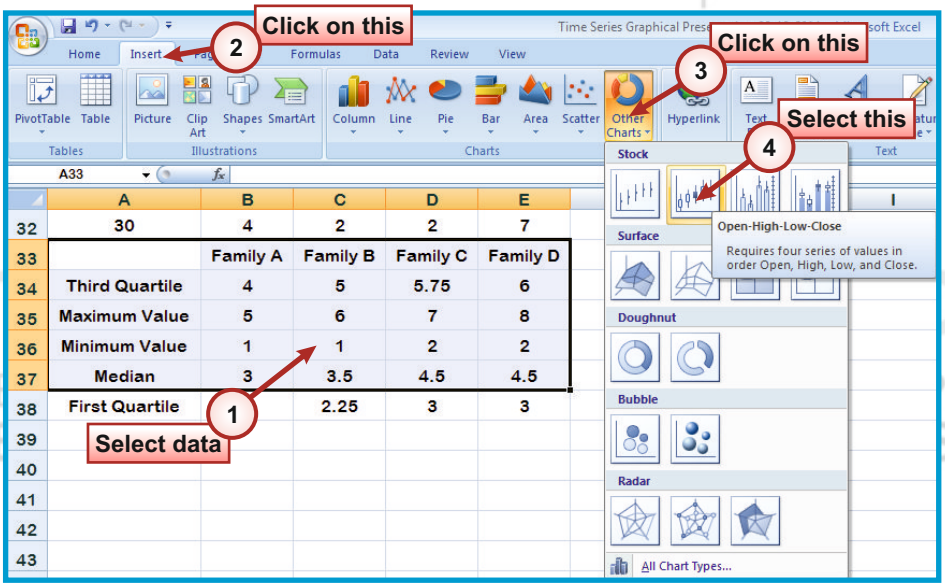


Fig. 5.36

Step 17: The resulting graph is shown in Fig. 5.37.

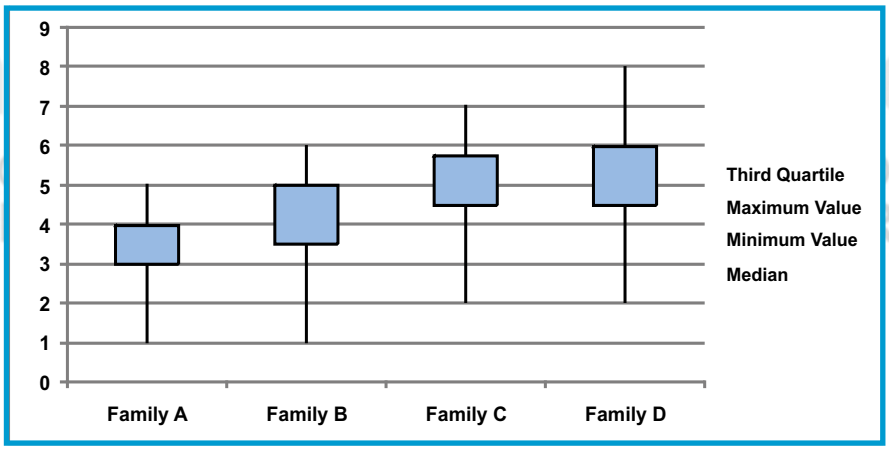


Fig. 5.37

Step 18: We format the graph as explained in Steps 4 to 9 of Sec. 3.3 of Lab Session 3. The resulting graph is shown in Fig. 5.38.

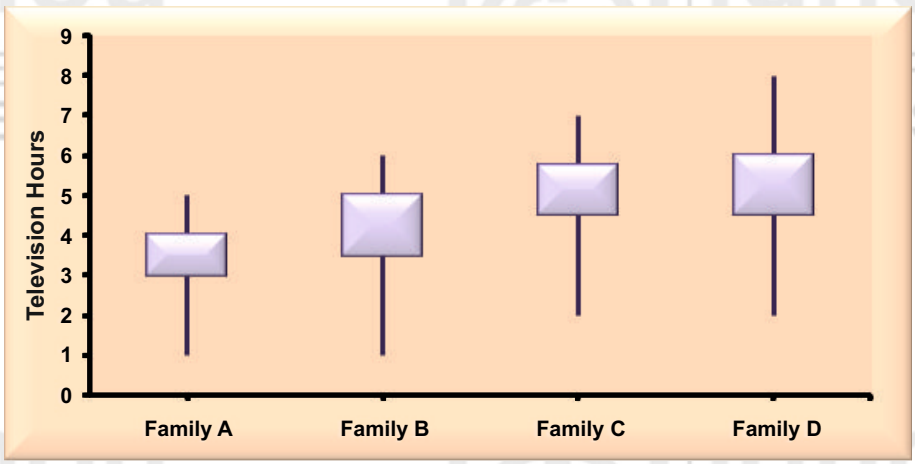


Fig. 5.38

Step 19: We now select the joint points of the lower lines (or whiskers) the boxes as shown in Fig. 5.39 and

1. click on the *Layout* tab under *Chart Tools*, and
2. choose the *Format Selection* option.

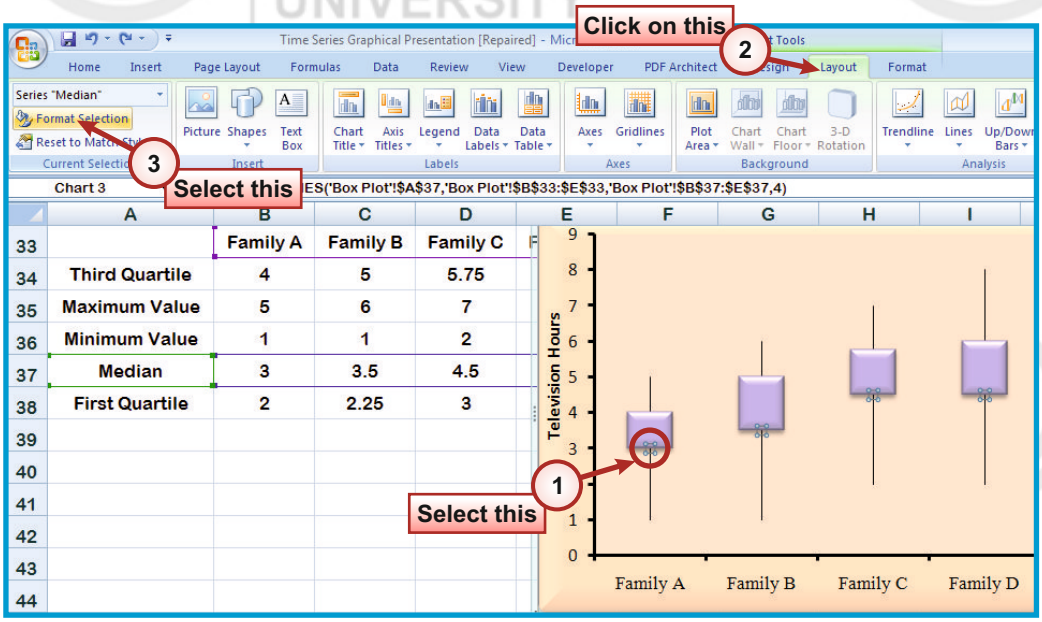


Fig. 5.39

Step 20: The dialog box shown in Fig. 5.40 appears. We now

1. click on *Marker Options*,
2. tick on the *Built-in* option,
3. choose the markers *Type* (–) as shown in Fig. 5.40, and
4. increase the *Size* as desired.

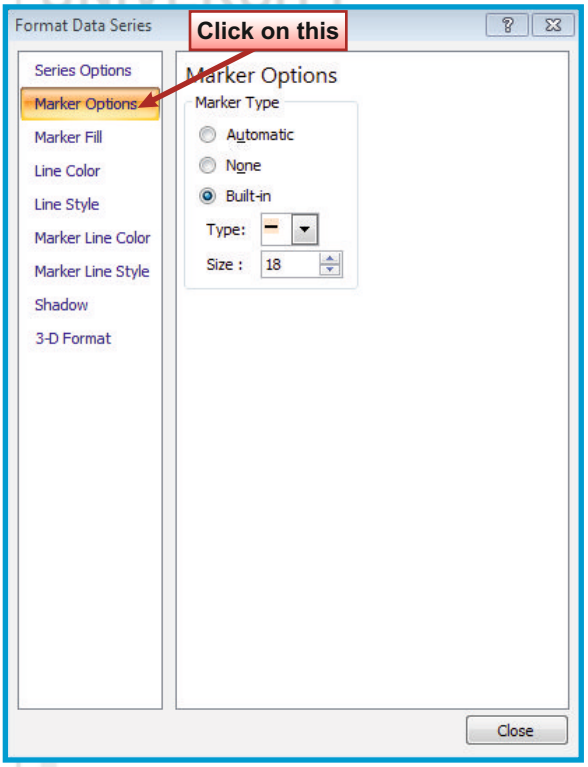


Fig. 5.40

Step 21: The resulting chart with line markers is shown in Fig. 5.41.

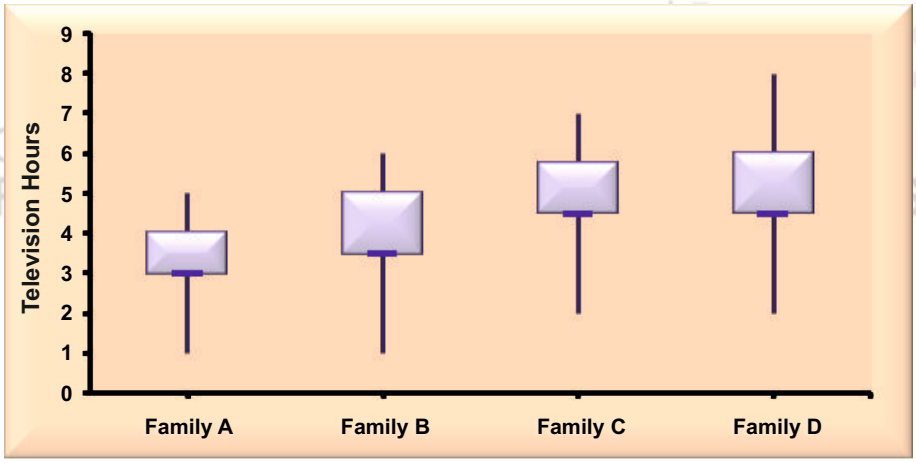


Fig. 5.41

Step 22: The box plot is still not complete. We now follow Fig. 5.42 and

1. click on the chart,
2. click on the *Design* tab under *Chart Tools*, and
3. choose the *Select Data* option.

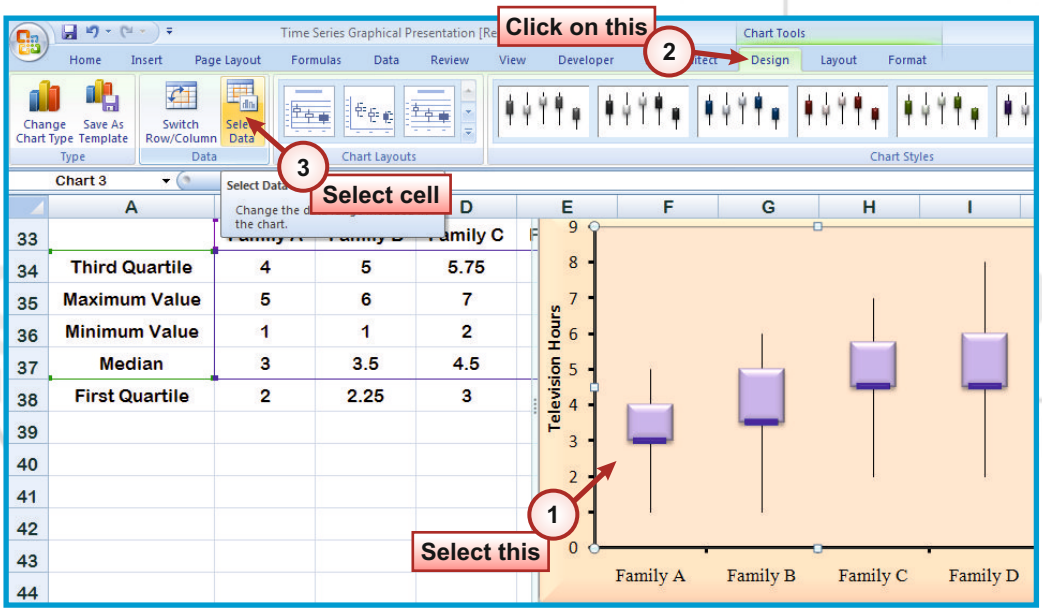


Fig. 5.42

Step 23: The dialog box shown in Fig. 5.43 appears. We click on *Add*.

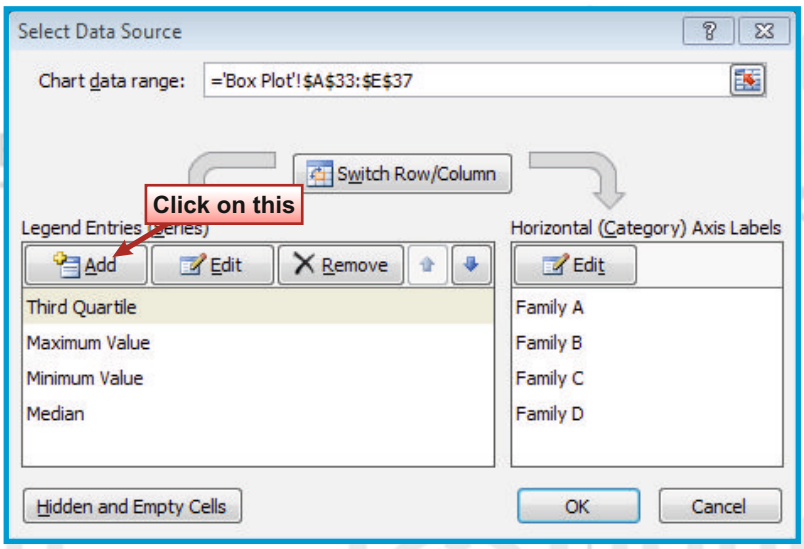
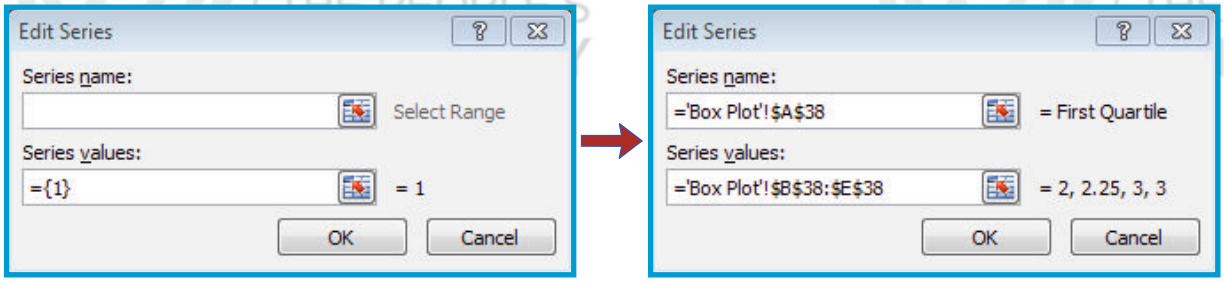


Fig. 5.43

Step 24: One more dialog box shown in Fig 5.44a appears. We select Cell A38 as *Series name* and Cells B38:E38 as *Series values* and click on *OK* as shown in Fig. 5.44b.



(a) (b)

Fig. 5.44

Step 25: The dialog box shown in Fig. 5.43 transforms to the dialog box shown in Fig. 5.45. We now click on *OK*.

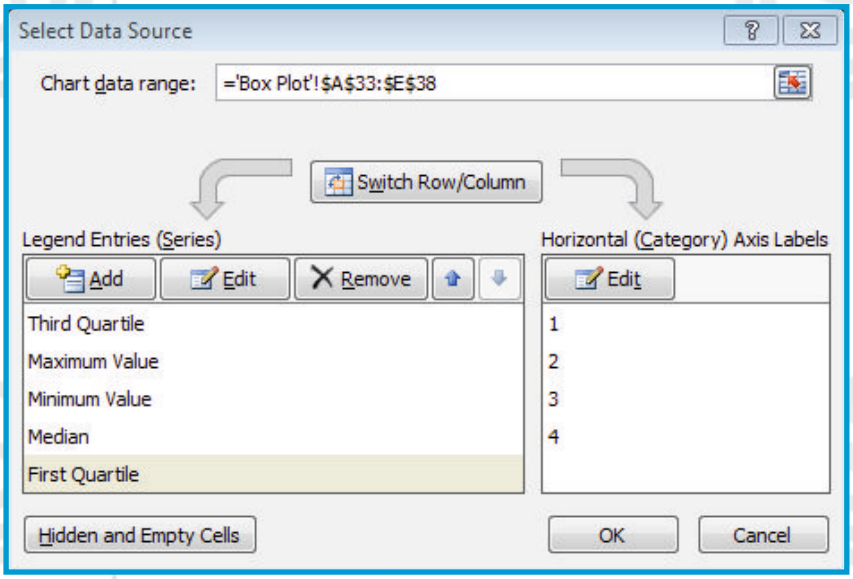


Fig. 5.45

Step 26: The graph shown in Fig. 5.46 is known as the box plot.

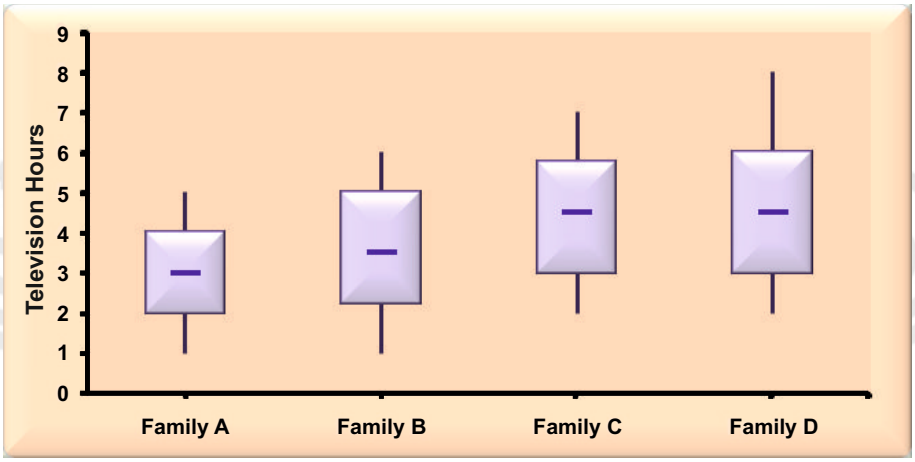


Fig. 5.46

The horizontal line inside the box represents the median. The vertical lines on both sides of the box extending to the minimum and maximum values of the given data are known as whiskers.

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

Activity

Construct suitable graphs with the help of MS Excel 2007 for

A1) Example 2 given in Unit 16 of MST-001.

A2) Exercises E1, E2, E6 and E7 given in Unit 16 of MST-001.

Match the results with the manual plotting of the graphs done in Unit 16 of MST-001.

Continuous Assessment 5

1. Suppose the number of cars on a particular road were recorded (in thousands) for 12 months as a part of a hypothetical study of a car manufacturing company in a city. The data are given in Table 5.

Table 5: Number of cars on road

Month	Number of Cars (in thousands)	Month	Number of Cars (in thousands)
1	278	7	418
2	286	8	378
3	398	9	435
4	421	10	498
5	367	11	425
6	387	12	447

Represent the time series data using a suitable graph.

2. The minimum and maximum prices of a particular variety of bag sold by a shopkeeper during a day are collected for fifteen days. The data are recorded in Table 6.

Table 6: Minimum and maximum price data

Days	Minimum Price (in ₹)	Maximum Price (in ₹)
1	800	680
2	820	650
3	775	700
4	720	575
5	860	700
6	840	750
7	740	600
8	780	650
9	800	700
10	750	575
11	780	705
12	800	608
13	775	550
14	750	650
15	840	680

Represent the time series data using a suitable graph.

3. To monitor the pattern of the agricultural production and study the changes in the annual cropping pattern of a field, the data of the yield of four varieties of crops, namely, wheat, rice, sugarcane and corn for the period of 1999 to 2013 were collected and are given in Table 7.

Table 7: Annual yield data for four crops

Year	Yield/Hectare (in tons)			
	Wheat	Rice	Sugarcane	Corn
1999	12.5	10.8	6.9	3.9
2000	14.5	11.5	5.9	3.1
2001	13.6	10.6	5.6	4.9
2002	14.5	11.5	6.5	4.5
2003	15.7	9.2	5.2	4.2
2004	13.5	10.5	5.5	3.9
2005	12.6	12.4	4.6	4.3
2006	13.5	10.5	5.5	3.5
2007	14.2	11.2	6.2	4.2
2008	12.5	9.5	5.8	3.7
2009	11.6	11.7	6.7	4.7
2010	12.5	9.5	4.5	2.9
2011	14.2	7.4	6.2	4.2
2012	13.5	10.5	5.5	3.5
2013	12.6	9.6	4.6	2.6

Plot a suitable graph for this data.

4. A survey was conducted to note the number of packed juices of four different brands sold each day. A sample of twenty-five days was taken. The numbers of juice cartons sold are recorded in Table 8.

Table 8: Data for sales of juice cartons

Days	Number of Sold Juice Cartons			
	Brand A	Brand B	Brand C	Brand D
1	25	10	30	15
2	5	10	35	25
3	25	30	30	35
4	10	20	15	15
5	10	15	15	30
6	20	15	10	10
7	10	5	10	25
8	15	10	30	35
9	20	5	25	20
10	25	10	15	30
11	10	15	25	20
12	10	20	20	20
13	25	25	30	35
14	15	30	15	10
15	15	5	20	20
16	15	20	35	30
17	10	15	25	30
18	10	15	10	10
19	25	30	30	40
20	10	15	10	10
21	20	25	25	15
22	20	15	15	25
23	25	30	20	35
24	10	25	30	10
25	20	10	10	35

Draw the box plot to compare the sales of the given brands of juice.

Home Work: Do It Yourself

1) Follow the steps explained in Secs. 5.3 to 5.6 to represent the data of Tables 1 to 4 graphically. Use a different format for the graphs. Take the screenshots and keep them in your record book.

2) Develop the spreadsheets for the exercises of "Continuous Assessment 5" as explained in this lab session. Take the screenshots of the final spreadsheets and graphs.

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