**Introduction to Database Management**

**Semester I, 2019-2020**

**Instructor: Assistant Lecturer (Mustafa Jamal Jalal)**

**Third Class**

1. **Database Overview**

A database is an organized collection of data. The data is typically organized to model relevant aspects of reality, in a way that supports processes requiring this information. Database management systems are specially designed applications that interact with the user, other applications, and the database itself to capture and analyze data. A general-purpose database management system is a software system designed to allow the definition, creation, querying, update, and administration of databases. Well-known DBMSs include MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Microsoft Access, Oracle, SAP, dBASE, FoxPro, IBM DB2 and FilemakerPro. A database is not generally portable across different DBMS, but different DBMSs can inter-operate by using standards such as SQL and ODBC or JDBC to allow a single application to work with more than one database.

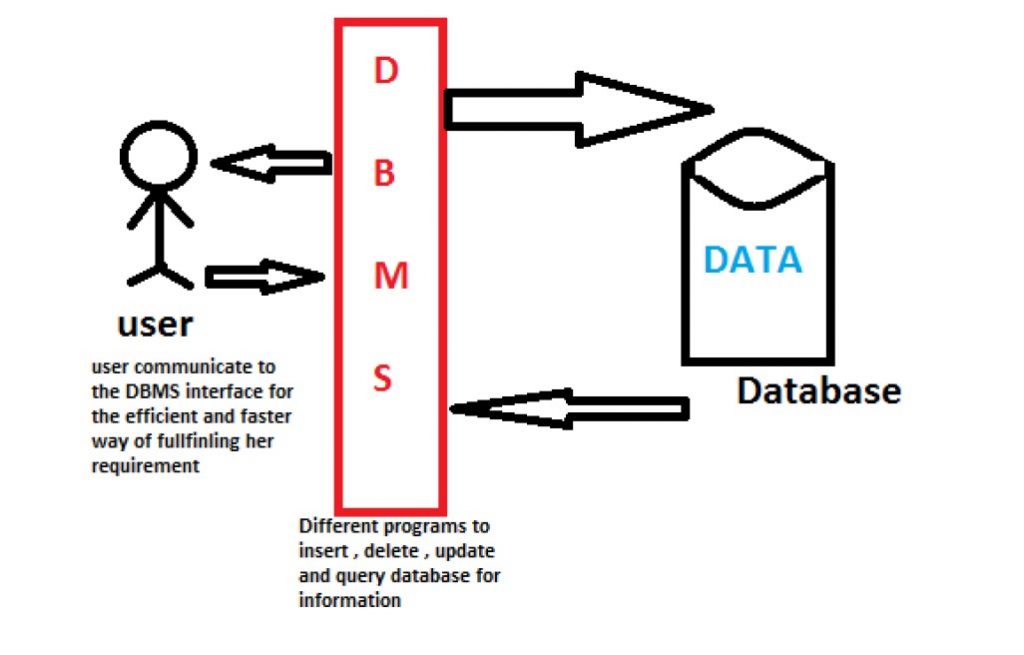


Figure 1.1: Database management

1. **How Are Databases Used in the Real World?**

Databases are obviously used in business applications and financial transactions; however, databases are not just used for business applications. Your grocery store, bank, video rental store and favorite clothing store all use databases to keep track of customer, inventory, employee and accounting information. Databases allow for data to be stored quickly and easily and are used in many aspects of your daily life. This article was stored in a database and its content was retrieved and displayed in your browser.

**2.1 History**

Several types of databases have been around since the early 1960s; however, the most commonly used type of database was not created until the early 1970s. Relational databases are the most commonly used type of database. Created by E.F. Codd, relational databases have given rise to a digital organizational tool used by countless companies and individuals. Computer systems replaced outdated forms of paper communication and paper file storage. Computer databases were used as a way to store and manage large amounts of information digitally. Companies began to use databases for a means of inventory tracking, customer management and accounting purposes.

**2.2 Significance**

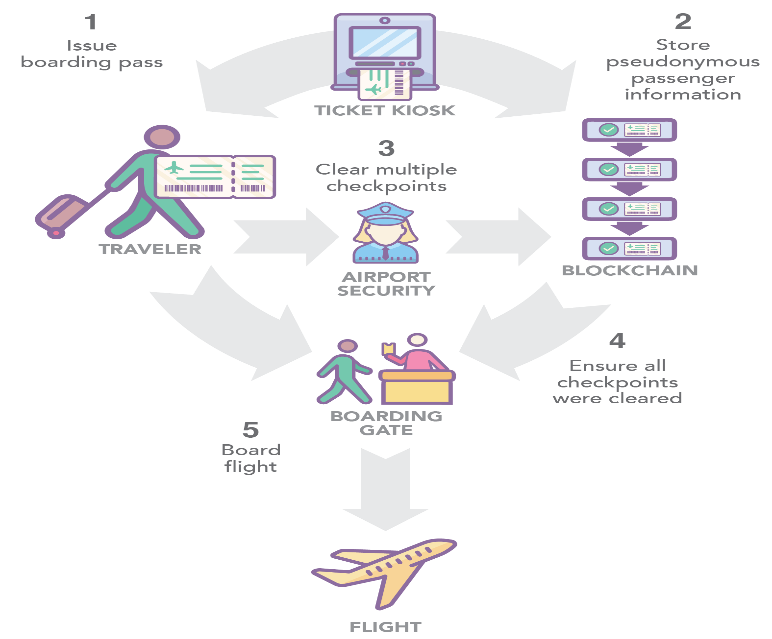
The move from paper to computer databases was a huge leap in information management and storage. Databases are much more efficient than paper storage in that they take up less space, are easily accessed by multiple users at once and can be transferred long distances with virtually no delay. The use of databases allowed for the rise of corporate infrastructure, credit card processing, email and the Internet. Databases allow for data to be shared across the world instead of being housed in one location on a physical piece of paper. 

Figure 2.2

Using Database in Airports

**3- types of database software**

There are many types of database software which use to manage the data .. such as

1. ADABAS.
2. IBM DB2.
3. Microsoft Access.
4. Microsoft Excel.
5. Microsoft SQL Server.
6. MySQL.
7. Oracle RDBMS.
8. Quick Base.

We will study Microsoft Excel as example

**3.1 Microsoft Excel**

What is Microsoft Excel?

Microsoft Excel is a program that provides worksheets comprised of rows and columns. Data can be stored in the worksheet, also called a spreadsheet, similarly to a Microsoft Word table, but the power of Excel is its ability to perform simple to complex mathematical calculations, and other functions. When you are ready to create some math formulas.

**Basic terms**

Workbook: The file in which you store your data. Can contain more than one worksheet – by default there are 3 in a workbook

Worksheet: The area you can see on screen, where you input data and perform calculations etc.

Sheet: tabs Appear at the bottom of the screen, allowing you to jump from one worksheet to another.

Cell: Each box on the screen is known as a “cell”. It is usually referred to by its address – e.g. A1 is the cell at the top left of the worksheet.

Active cell : The cell that has a dark border around it is “active” – that is where anything you type will appear.

Row: All the cells in one horizontal line, marked by the numbers on the left of the screen.

Column: All the cells in one vertical line, marked by the letters at the top of the screen.

**The Excel Worksheet (Spreadsheet) and Workbook**

An Excel worksheet, or spreadsheet, is a two-dimensional grid with columns and rows. Look at the spreadsheet below. The column names are letters of the alphabet starting with A, and the rows are numbered chronologically starting with the number one. The cells in the first roware A1, B1, C1, and so on. And the cells in the first column are A1, A2, A3, and so on. These are called cell names or cell references.

We use cell references when creating math formulas or functions. For example, the formula to add the contents of cells B2 and B3 together is: =B2+B3.

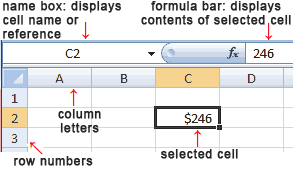


Figure (3.1.1) Structure of a Microsoft Excel Worksheet (Spreadsheet)

The Name Box is located in the area above Column A, and displays the cell reference of the selected cell - the cell where the cursor is resting. In our spreadsheet above, the selected cell is C2. Notice that the column letter (C) and the row number (2) change color.

The beginning of the Formula Bar can be seen in the area above Column D on our worksheet. The Formula Bar displays the contents of the selected cell.

A workbook is a collection of worksheets or spreadsheets. When the Excel program is opened, a workbook opens with three blank worksheets. The names of the worksheets are displayed on tabs at the bottom of the Excel window.

**3.1.2** **EXCEL THREE DATA TYPES AND THEIR DEFAULT FORMATS**

In Excel, the worksheet consists of a grid of columns and rows that form cells. You enter three types of data in cells: labels, values, and formulas.

**Labels (text)** are descriptive pieces of information, such as names, months, or other identifying statistics, and they usually include alphabetic characters.

**Values (numbers)** are generally raw numbers or dates.

**Formulas** are instructions for Excel to perform calculations.

To enter data in a cell in Excel 2010, you select the cell, type the data, and press Enter. Excel moves the cell cursor down one cell. You also can click the Enter button (the check mark) in the Formula bar to enter data. The check mark appears only when you are in the process of entering (or editing) data.

**How Excel aligns the data (by default) depends on what type of data it is:**

**Label**: Excel aligns text to the left side of the cell. If the text is too wide to fit, Excel extends that data past the cell width if the next cell is blank. If the next cell is not blank, Excel displays only enough text to fit the display width. Widening the column displays additional text.

**Whole value**: If the data is a whole value, such as 34 or 5763, Excel aligns the data to the right side of the cell.

**Value with a decimal:** If the data is a decimal value, Excel aligns the data to the right side of the cell, including the decimal point, with the exception of a trailing 0. For example, if you enter 246.75, then 246.75 is displayed; if you enter 246.70, however, 246.7 appears.

**Date**: If you enter a date, such 12/16, Dec 16, or 16 Dec, Excel automatically returns the value in your default date format (16-Dec if you haven’t changed it) in the cell, but the Formula bar displays 12/16/2010.

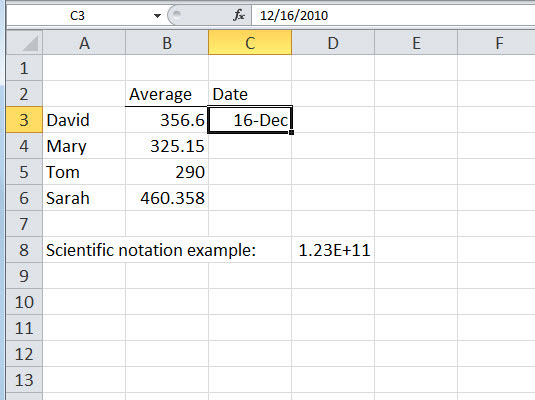


Figure (3.1.2) Showing The type of data in Microsoft Excel

**3.1.3 What does mean "Cell"**

A cell is the most basic storage unit available in a spreadsheet program such as Excel or Google Spreadsheet. Data entered into spreadsheet program is always stored in a cell.

How to Select Cells

There are a variety of ways to select cells in an Excel spreadsheet:

• To select one cell, click in the cell.

• To select one or more rows of cells, click on the row number(s).

• To select one or more columns of cells, click on the column letter(s).

• To select a group of contiguous cells, click in a corner cell and, with the left mouse button depressed, drag the cursor horizontally and/or vertically until all of the cells you want selected are outlined in black.

• To select multiple cells that are not contiguous, press and hold the Ctrl key while clicking in the desired cells.

• To select every cell in the worksheet, click in the upper right corner of the worksheet to the left of "A."

**3.1.4 OPENING EXCEL**

There are two ways to open Excel. You can:

Click the icon on your desktop called “Microsoft Excel” or something similar (i.e.: “Excel,” “MS Excel,” “Microsoft Excel”).

From the start menu go to Programs

Microsoft Office

Microsoft Office Excel

When you open Excel, you will see the following elements:

A bar at the top. This is called the title bar and shows you what program you are in. You also see the words Book1 in the title bar. This is the default title of your spreadsheet. Each spreadsheet is referred to as a Book

The ribbon tabs. The different ribbons take the place of the toolbars and menus of previous versions of Excel.

The Ribbons offer access to all the different features in Excel.

The work area. Excel always opens with cell A1 selected. A1 or any combination of letter and number is called the cell address.

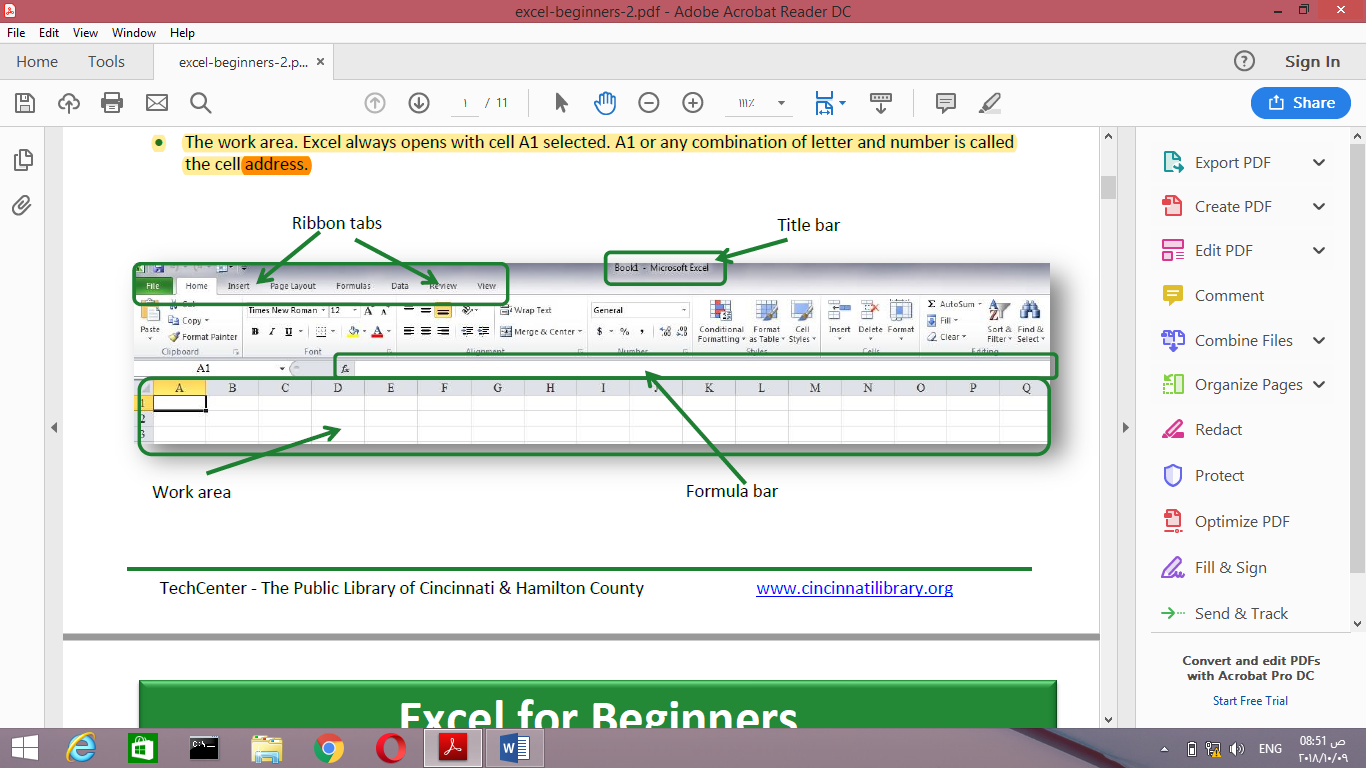


Figure (3.1.4) Microsoft Excel ribbon tabs (Spreadsheet)

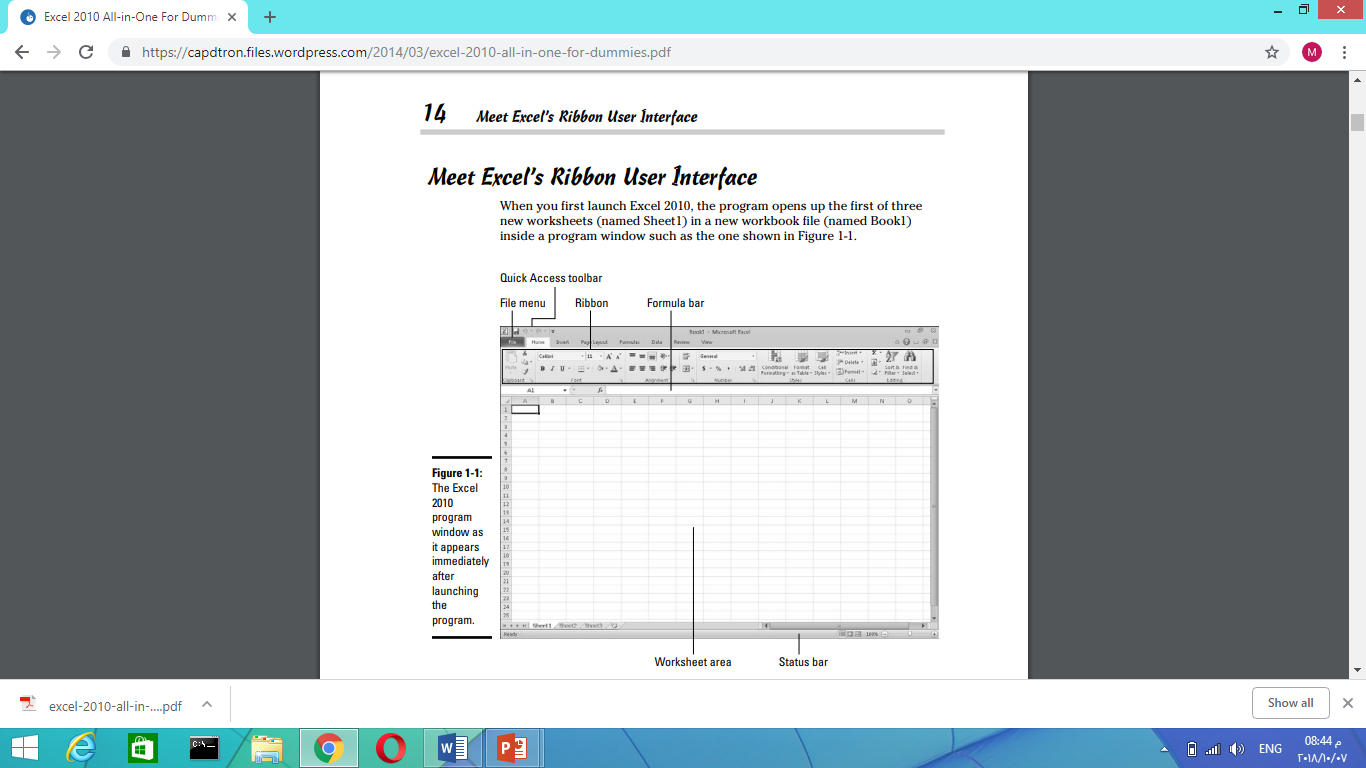


Figure (3.1.5) Microsoft Excel Worksheet (Spreadsheet)

**Save workbook**

Wherever you want to save your workbook (on your computer or the web, for example), you do all your saving on the File tab.

While you’ll use Save or press Ctrl+S to save an existing workbook in its current location, you need to use Save As to save your workbook for the first time, in a different location, or to create a copy of your workbook in the same or another location.

Click File > Save As.

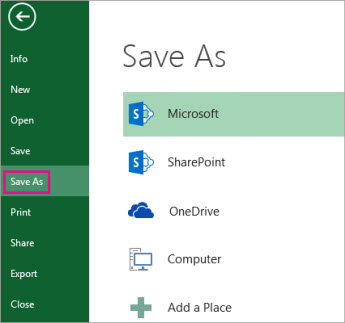


Figure (3.1.6) Saving Microsoft Excel Worksheet (Spreadsheet)

Under Save As, pick the place where you want to save your workbook. For example, to save to your desktop or in a folder on your computer, click Computer.



Figure (3.1.7) Save your workbook Microsoft Excel

Click Browse to find the location you want in Documents folder.

To pick another location on your computer, click Desktop, and then pick the exact place where you want to save your workbook.

In the File name box, enter a name for a new workbook.

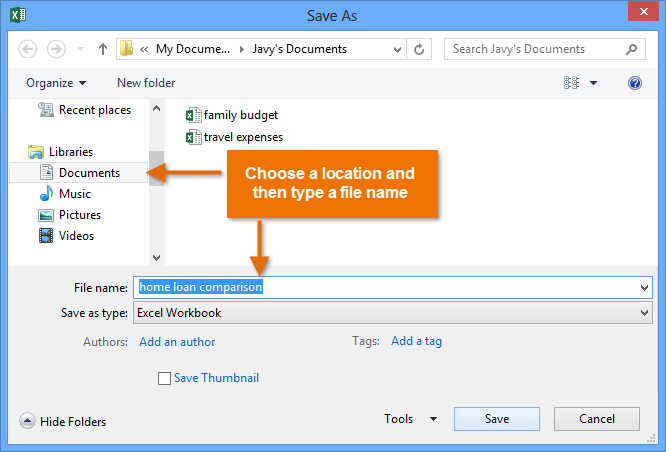


Figure (3.1.8) Saving Microsoft Excel Worksheet (Spreadsheet)

**How to enter data in Excel**

In this lesson, we'll look at the most basic way to enter data in an Excel worksheet - by typing. In future lessons, we'll look at a number of shortcuts for entering data faster.

To enter data in Excel, just select a cell and begin typing. You'll see the text appear both in the cell and in the formula bar above.

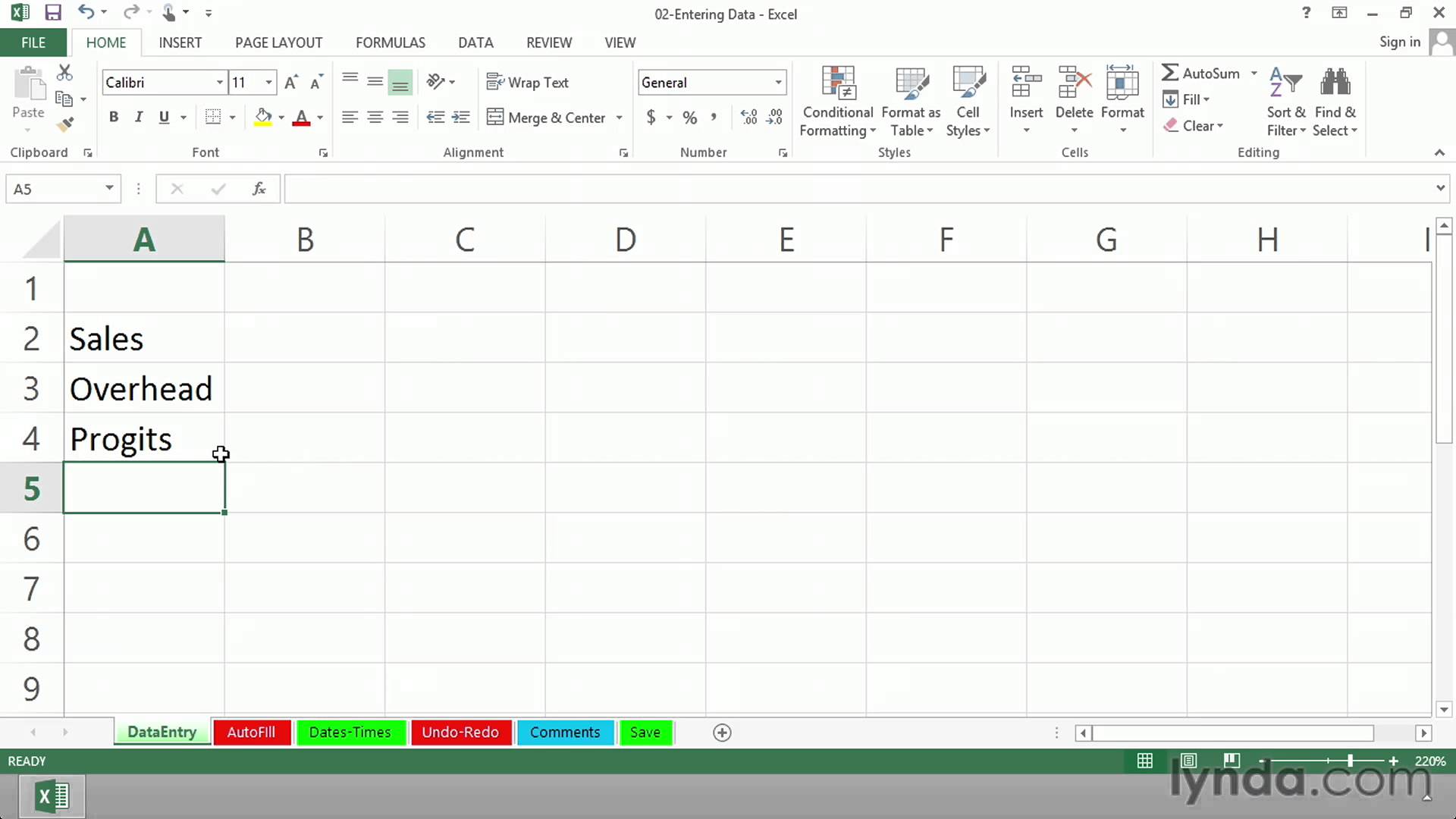
To tell Excel to accept the data you've typed, press enter. The information will be entered immediately, and the cursor will move down one cell.

You can also press the tab key instead of the enter key. If you press tab, the cursor will move one cell to the right once the information has been entered.

When Excel sees that you are typing into a list, pressing enter at the end of the row will move the cursor down one row and back to the first column.

At any time while you are typing you can press the escape key to cancel. This brings Excel back to the state it was in before you started typing.

When you want to delete information that has already been entered, just select the cells, and press the delete key.



Use a simple formula to add, subtract, multiply, or divide numeric values

To create a simple formula, you enter values and math operators into a cell, or the formula bar, to receive a result. Instead of entering values directly into the formula, you can also refer to the cells that contain the values that you want to calculate. Using a cell reference in a formula ensures that the result is updated if the values change.

Type a couple of values into cells.

For example, in cell A1, type 5, and in cell B1, type 20.

Click any blank cell, and then type an equal sign (=) to start a formula.

After the equal sign (=), you can type two numbers and a math operator to create a simple formula.

For example, you could simply type =5+20, or =5\*20. But to create a formula that you would not have to change, even if you change one of the values, type the cell reference and a math operator. For example, A1 + B1.

After you have tried the formula with a plus sign (+), type a minus sign (-) to subtract values, an asterisk (\*) to multiply values, and a forward slash (/) to divide values.

If you use the example numbers, the results are 25, -15, 100, and 0.25

**Sort data in a range or table**

Sorting data is an integral part of data analysis. You might want to arrange a list of names in alphabetical order, compile a list of product inventory levels from highest to lowest, or order rows by colors or icons. Sorting data helps you quickly visualize and understand your data better, organize and find the data that you want, and ultimately make more effective decisions.

You can sort data by text (A to Z or Z to A), numbers (smallest to largest or largest to smallest), and dates and times (oldest to newest and newest to oldest) in one or more columns. You can also sort by a custom list you create (such as Large, Medium, and Small) or by format, including cell color, font color, or icon set.

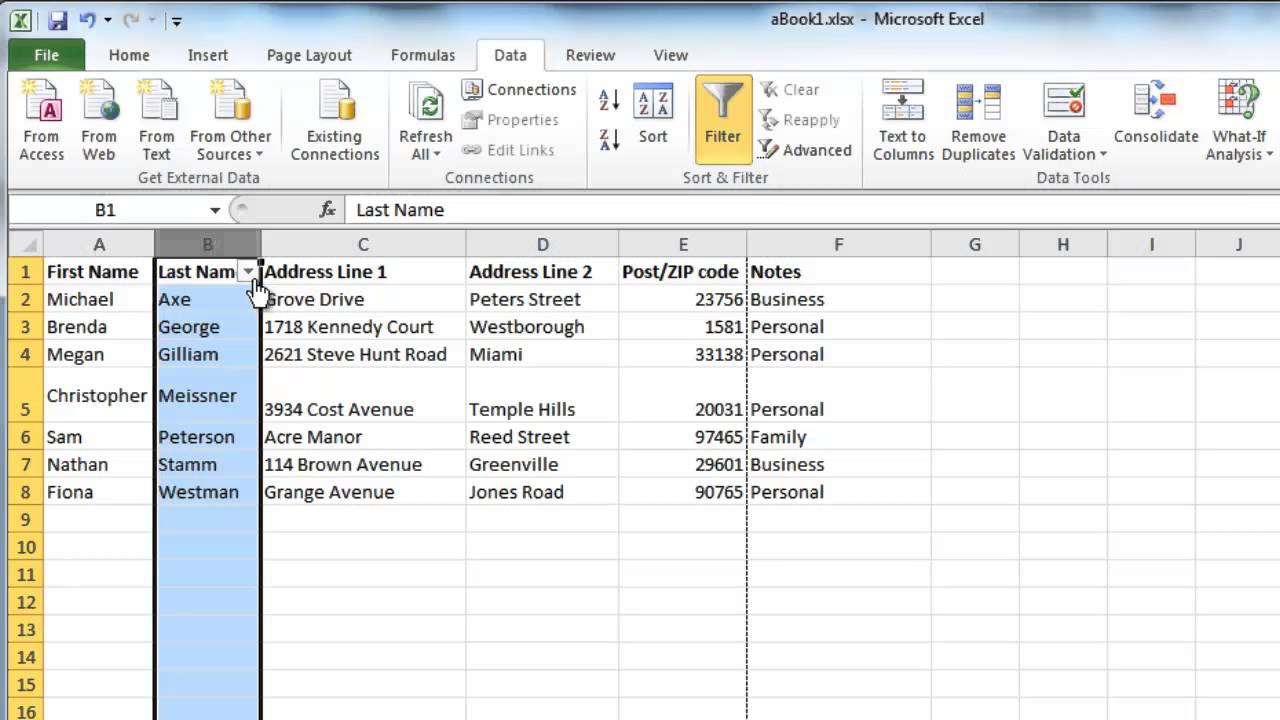


Figure (3.1.9) Microsoft Excel Sorting

**Quick Sort With Sort Buttons**

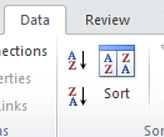
In Excel, you can quickly sort your data by using the A-Z and Z-A Sort buttons on the Ribbon's Data tab. But, be careful, or one column may be sorted, while others are not.

Only use this technique if there are no blank rows or columns within the data.

Select one cell in the column you want to sort.

On the Excel Ribbon, click the Data tab.

Click Sort A to Z (smallest to largest) or Sort Z to A (largest to smallest)



Before you do anything else, check the data, to ensure that the rows have sorted correctly. If things look wrong, immediately click the Undo button on the toolbar.

**Problems with Sorting Excel Data**

When you quickly sort data with the A-Z or Z-A button, things can go horribly wrong. If there is a blank row or blank columns within the data, part of the data might be sorted, while other data is ignored. Imagine the mess you'll have, if names and phone number no longer match, or if orders go to the wrong customers!

Follow these steps to help prevent problems when sorting Excel data:

Select one cell in the column you want to sort.

Press Ctrl + A, to select the entire region.

Check the selected area, to make sure that all the data is included. For example, in the screen shot below, hidden column E is blank, so columns at the left are not selected.



Figure (3.1.10) Microsoft Excel Sorting

If all the data was not selected, fix any blank columns or rows, and try again. Or, use the Sort Dialog box, as described in the next section.

If all the data is selected, click Sort A to Z (smallest to largest) or Sort Z to A (largest to smallest)

Before you do anything else, check the data, to ensure that the rows have sorted correctly. If things look wrong, click the Undo button on the toolbar.

**Functions in Excel**

The Function SQRT In Excel

The Excel SQRT function returns the Square root of a list of supplied numbers.

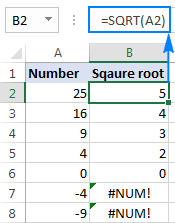
How to square root in Excel using SQRT function?

The easiest way to do square root in Excel is by using the function specially designed for this:

SQRT(number)

Where number is the number or reference to the cell containing the number for which you want to find the square root.

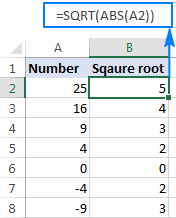
For example, to get a square root of 225, you use this formula: =SQRT (225)

To calculate square root of a number in A2, use this one: =SQRT(A2)  


If a number is negative, like in rows 7 and 8 in the screenshot above, the Excel SQRT function returns the #NUM! error. It happens because the square root of a negative number does not exist among the set of real numbers. Why's that? Since there is no way to square a number and get a negative result.

(The Excel ABS function returns the Absolute value of a list of supplied numbers.)

In case you wish to take a square root of a negative number as if it were a positive number, wrap the source number in the ABS function, which returns the absolute value of a number ignoring the sign:

=SQRT(ABS(A2)) 

How to do square root in Excel using a calculation

When calculating by hand, you write square root by using the radical symbol (√). Though, it's not possible to type that traditional square root symbol in Excel, there is a way to find square root without any function. For this, you use the caret character (^), which is located above the number 6 on most keyboards.

In Microsoft Excel, the caret symbol (^) acts as the exponent, or power, operator. For example, to square the number 5, i.e. raise 5 to the power of 2, you type =5^2 in a cell, which is equivalent to 52.

To get a square root, use the caret with (1/2) or 0.5 as the exponent:

number^(1/2)

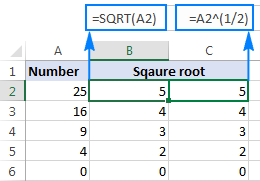
or

number^0.5

For example, to get the square root of 25, you type =25^(1/2) or =25^0.5 in a cell.

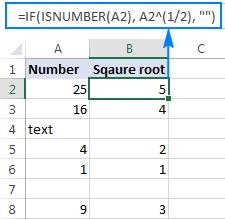
To find square root of a number in A2, you type: =A2^(1/2) or =A2^0.5

As shown in the screenshot below, the Excel SQRT function and the exponent formula yield identical results:



This square root expression can also be used as part of bigger formulas. For instance, the following IF statement tells Excel to calculate a square root on condition: get a square root if A2 contains a number, but return an empty string (blank cell) if A2 is a text value or blank:

=IF(ISNUMBER(A2), A2^(1/2), "")



Why is an exponent of 1/2 the same as square root?

For starters, what do we call a square root? It is nothing else but a number that, when multiplied by itself, gives the original number. For example, the square root of 25 is 5 because 5x5=25. That is crystal clear, isn't it?

Well, multiplying 251/2 by itself also gives 25:

25½ x 25½ = 25(½+½) = 25(1) = 25

Said another way:

√25 x √25 = 25

And:

25½ x 25½ = 25

So, 25½ is equivalent to √25.

The Function AVERAGE In Excel

The Excel AVERAGE function returns the arithmetic mean of a list of supplied numbers.

The syntax of the function is:

AVERAGE (number1, [number2], ...)

where the number arguments are a set of one or more numeric values, or arrays of numeric values, for which you want to calculate the average.

In current versions of Excel (Excel 2007 and later), you can provide up to 255 number arguments to the Average function, but in Excel 2003, the function can only accept up to 30 arguments. However, each argument can consist of an array of values or a range of cells, each of which can contain multiple values. Excel Average Function Examples

Cells B1-B4 of the spreadsheets below show the Excel Average function used to calculate the arithmetic mean of the numbers 8, 7, 9, 6 & 10. In each case, the values are supplied to the function in different ways.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Formulas:   |  |  |  | | --- | --- | --- | |  | A | B | | 1 | 8 | =AVERAGE( 8, 7, 9, 6, 10 ) | | 2 | 7 | =AVERAGE( {8,7,9}, 6, 10 ) | | 3 | 9 | =AVERAGE( A1, A2, A3, A4, A5 ) | | 4 | 6 | =AVERAGE( A1:A5 ) | | 5 | 10 |  | | Results:   |  |  |  | | --- | --- | --- | |  | A | B | | 1 | 8 | 8 | | 2 | 7 | 8 | | 3 | 9 | 8 | | 4 | 6 | 8 | | 5 | 10 |  | |

The examples in the above spreadsheets show that each argument to the Average function can be supplied as a single value or cell, or as an array of values or cells (note that in cell B2, the argument {8,7,9} is an array of numbers).

The Function MIN In Excel

The Excel MIN function returns the smallest value from a supplied set of numeric values.

The syntax of the function is:

MIN( number1, [number2], ... )

where the number arguments are one or more numeric values (or arrays of numeric values), that you want to return the smallest value of.

In current versions of Excel (Excel 2007 & later), you can provide up to 255 number arguments to the Min function, but in Excel 2003, the function can only accept up to 30 number arguments.

If an argument is supplied to the function as a reference to a cell, or an array of cells, the Min function will ignore blank cells, and text or logical values contained within the supplied cell range. However, logical values and text representations of numbers that are supplied directly to the function will be included in the calculation.

Excel Min Function Example

Cell B1 of the following spreadsheet shows the Excel Min function, used to retrieve the smallest value from the set of values in cells A1-A5.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Formulas:   |  |  |  | | --- | --- | --- | |  | A | B | | 1 | 4 | =MIN( A1:A5 ) | | 2 | 3 |  | | 3 | 1 |  | | 4 | 5 |  | | 5 | 2 |  | | Results:   |  |  |  | | --- | --- | --- | |  | A | B | | 1 | 4 | 1 | | 2 | 3 |  | | 3 | 1 |  | | 4 | 5 |  | | 5 | 2 |  | |

The Function MAX In Excel

The Excel MAX function returns the largest value from a supplied set of numeric values.

The syntax of the function is:

MAX( number1, [number2], ... )

where the number arguments are one or more numeric values (or arrays of numeric values), that you want to return the largest value of.

In current versions of Excel (Excel 2007 & later), you can provide up to 255 number arguments to the Max function, but in Excel 2003, you can only provide up to 30 number arguments.

If an argument is supplied to the function as a reference to a cell, or an array of cells, the Max function will ignore blank cells, and text or logical values contained within the supplied cell range. However, logical values and text representations of numbers that are supplied directly to the function are included in the calculation.

Excel Max Function Example

Cell B1 of the following spreadsheet shows the Excel Max function, used to retrieve the largest value from the set of values in cells A1-A5.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Formulas:   |  |  |  | | --- | --- | --- | |  | A | B | | 1 | 4 | =MAX( A1:A5 ) | | 2 | 3 |  | | 3 | 1 |  | | 4 | 5 |  | | 5 | 2 |  | | Results:   |  |  |  | | --- | --- | --- | |  | A | B | | 1 | 4 | 5 | | 2 | 3 |  | | 3 | 1 |  | | 4 | 5 |  | | 5 | 2 |  | |

The Function SUM In Excel

Function Description

The Excel SUM function adds together a supplied set of numbers and returns the sum of these values.

The syntax of the function is:

SUM( number1, [number2], ... )

where the number arguments are a set of numbers (or arrays of numbers) that you want to find the sum of.

In current versions of Excel (Excel 2007 and later), you can enter up to 255 number arguments to the Excel Sum function, but in Excel 2003, the function can only handle up to 30 arguments. However, each argument can consist of an array of values or a range of cells, each of which can contain many values.

Excel Sum Function Examples

The following spreadsheet shows three simple examples of the Excel Sum function:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Formulas:   |  |  |  | | --- | --- | --- | |  | A | B | | 1 | 5 | =SUM( 15, 29 ) | | 2 | 7 | =SUM( A1, A2 ) | | 3 | 9 | =SUM( A1:A3 ) | | Results:   |  |  |  | | --- | --- | --- | |  | A | B | | 1 | 5 | 44 | | 2 | 7 | 12 | | 3 | 9 | 21 | |

The Function COUNT In Excel

Function Description

The Excel Count function returns the count of numeric values in a supplied set of cells or values. This count includes both numbers and dates.

The syntax of the function is:

COUNT( value1, [value2], ... )

Where the arguments, value1, [value2], etc., can be any values, arrays of values, or references to cell ranges.

In recent versions of Excel (2007 and later), you can enter up to 255 value arguments to the Excel Count function, each of which may be single values or arrays of cells or values. However, in Excel 2003, the function can only handle up to 30 value arguments.

Which Values are Counted as Numeric Values?

Numbers and dates are always counted as numeric values by the Excel Count function. However, text representations and logical values are counted differently, depending on whether they are supplied as a value in a range of cells, or if they are supplied directly to the function.

The table below summarises which values are and which are not treated as numeric values by the Excel Count function:

|  |  |  |
| --- | --- | --- |
|  | Value Within a Range of Cells | Value Supplied Directly to Function |
| Numbers | ARE counted | ARE counted |
| Dates | ARE counted | ARE counted |
| Logical Values | NOT counted | ARE counted |
| Text Representations of Numbers & Dates | NOT counted | ARE counted |
| Other Text | NOT counted | NOT counted |
| Errors | NOT counted | NOT counted |

Count Function Examples

Example 1 - Values Supplied from a Range of Worksheet Cells

Column C of the following spreadsheet shows three examples of the Count function, used to return the number of numeric values in one or more supplied ranges of cells.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Formulas:   |  |  |  |  | | --- | --- | --- | --- | |  | A | B | C | | 1 | 5 | 0 | =COUNT( A1:A5 ) | | 2 | text |  | =COUNT( A1:A5, B1 ) | | 3 | FALSE |  | =COUNT( A1:B5 ) | | 4 | 01/01/2015 |  |  | | 5 | #N/A | 10 |  | | Results:   |  |  |  |  | | --- | --- | --- | --- | |  | A | B | C | | 1 | 5 | 0 | 2 | | 2 | text |  | 3 | | 3 | FALSE |  | 4 | | 4 | 01/01/2015 |  |  | | 5 | #N/A | 10 |  | |

Note that, in the above example:

The numbers and the date 01/01/2015 are counted by the function.

The text value "text", the logical value FALSE, and the error value #N/A are not counted by the function.

The empty cells are not counted by the function.

Example 2 - Values Supplied Directly to the Excel Count Function

In the following spreadsheet, the Excel Count function is used to count the number of numeric values in sets of values supplied directly to the function.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Formulas:   |  |  | | --- | --- | |  | A | | 1 | =COUNT( 100, DATE(2015,1,1) ) | | 2 | =COUNT( "100", "01/01/2015", FALSE ) | | 3 | =COUNT( "text", #N/A ) | | Results:   |  |  | | --- | --- | |  | A | | 1 | 2 | | 2 | 3 | | 3 | 0 | |

Note, in the above example:

The number 100 and the date 01/01/2015 are counted by the function.

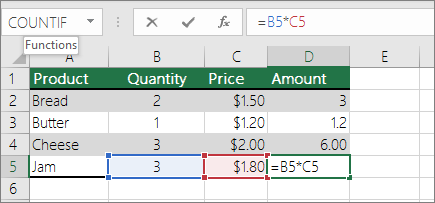
The text representations of the number "100" & the date, "01/01/2015", and the logical value FALSE, are counted by the function.

The text string "text" and the error #N/A are not counted by the function.

Switch between relative, absolute, and mixed references

By default, a cell reference is a relative reference, which means that the reference is relative to the location of the cell. If, for example, you refer to cell A2 from cell C2, you are actually referring to a cell that is two columns to the left (C minus A)—in the same row (2). When you copy a formula that contains a relative cell reference, that reference in the formula will change.

As an example, if you copy the formula =B4\*C4 from cell D4 to D5, the formula in D5 adjusts to the right by one column and becomes =B5\*C5. If you want to maintain the original cell reference in this example when you copy it, you make the cell reference absolute by preceding the columns (B and C) and row (2) with a dollar sign ($). Then, when you copy the formula =$B$4\*$C$4 from D4 to D5, the formula stays exactly the same.



Less often, you may want to mixed absolute and relative cell references by preceding either the column or the row value with a dollar sign—which fixes either the column or the row (for example, $B4 or C$4).

To change the type of cell reference:

Select the cell that contains the formula.

In the formula bar Button image , select the reference that you want to change.

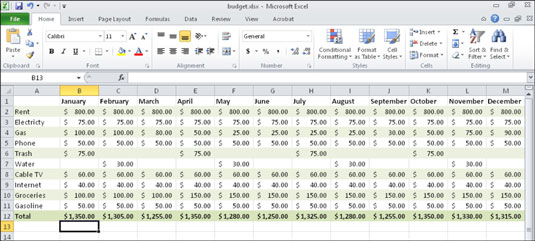
Press F4 to switch between the reference types.

The table below summarizes how a reference type updates if a formula containing the reference is copied two cells down and two cells to the right.

| For a formula being copied: | If the reference is: | It changes to: |
| --- | --- | --- |
| Formula being copied from A1, to two cells down and to the right | $A$1 (absolute column and absolute row) | $A$1 (the reference is absolute) |
|  | A$1 (relative column and absolute row) | C$1 (the reference is mixed) |
|  | $A1 (absolute column and relative row) | $A3 (the reference is mixed) |
|  | A1 (relative column and relative row) | C3 (the reference is relative) |

HOW TO CREATE A BASIC EXCEL CHART?

When you have a lot of numeric data on a Microsoft Excel worksheet, using a chart can help make more sense out of the numbers. Excel offers various chart types, each suited for a different type of data analysis.



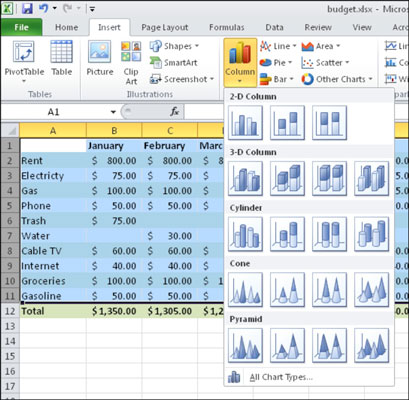
To create a chart, follow these steps:

Select the data to include on the chart. Include any cells that contain text labels that should be in the chart, too.

You may need to select a range of cells that don’t touch each other (noncontiguous) for Step 1. If so, hold down the Ctrl key while you select the cells you want.

On the Insert tab, click a chart type. (Use the buttons in the Charts group.)

A menu opens showing the chart subtypes.



Click the subtype you want.

A new chart is created and placed on the current sheet as a floating object.

**Microsoft PowerPoint**

**What is Microsoft PowerPoint**

Microsoft Office product that provides users with an interface to design multimedia slides to be displayed on a projection system or personal computer. The software incorporates images, sounds, videos, text, and charts to create an interactive presentation. Microsoft PowerPoint interacts with other Office products such as Microsoft Word and Excel, and is included with most Microsoft Office packages.

Benefits of PowerPoint

PowerPoint provides multiple benefits to users, including the following.

• Most popular presentation software, meaning it is more likely that other people will have PowerPoint software or a way to view PowerPoint files.

• Many features available for slide creation, including slide transitions, animations, layouts, templates, and more.

• Professional looking slides, great for educational and business presentations.

• Option of saving slides as more than just a PowerPoint file, including GIF and JPGimages, MPEG-4 video, PDF, RTF (rich text format), WMV (Windows Media Video), and PowerPoint XML.

PowerPoint example

PowerPoint slides can be plain with only text, or they can include pictures and even animation, including moving text and images. Text can be formatted in the same way as text can be formatted in Microsoft Word, including color, size, and font type.

While the look and feel of PowerPoint has changed over the years, the functionality has remained mostly the same.

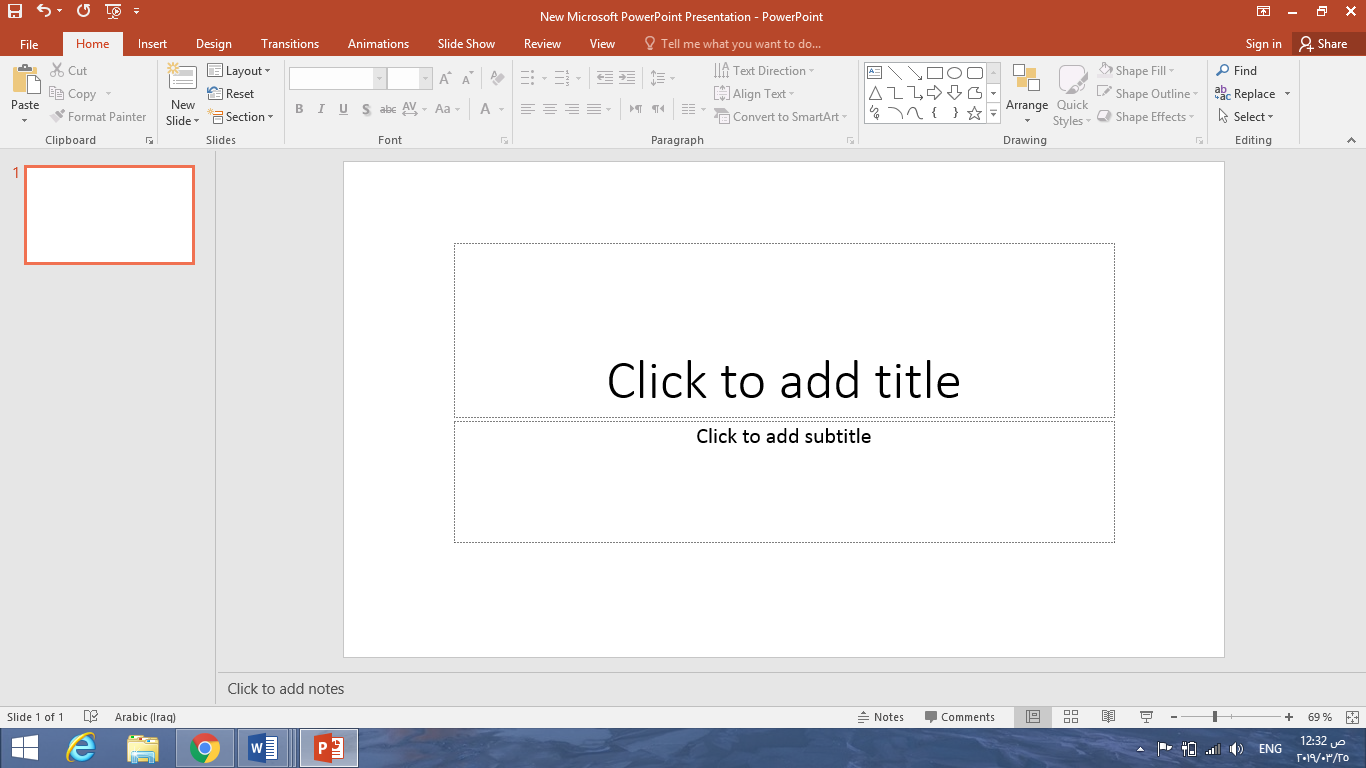


Figure (1.1) example of new PowerPoint presentation

How to create or add a slide in Microsoft PowerPoint

In Microsoft PowerPoint, a presentation is made up of multiple slides. There are several ways to create or add a slide in a PowerPoint presentation.

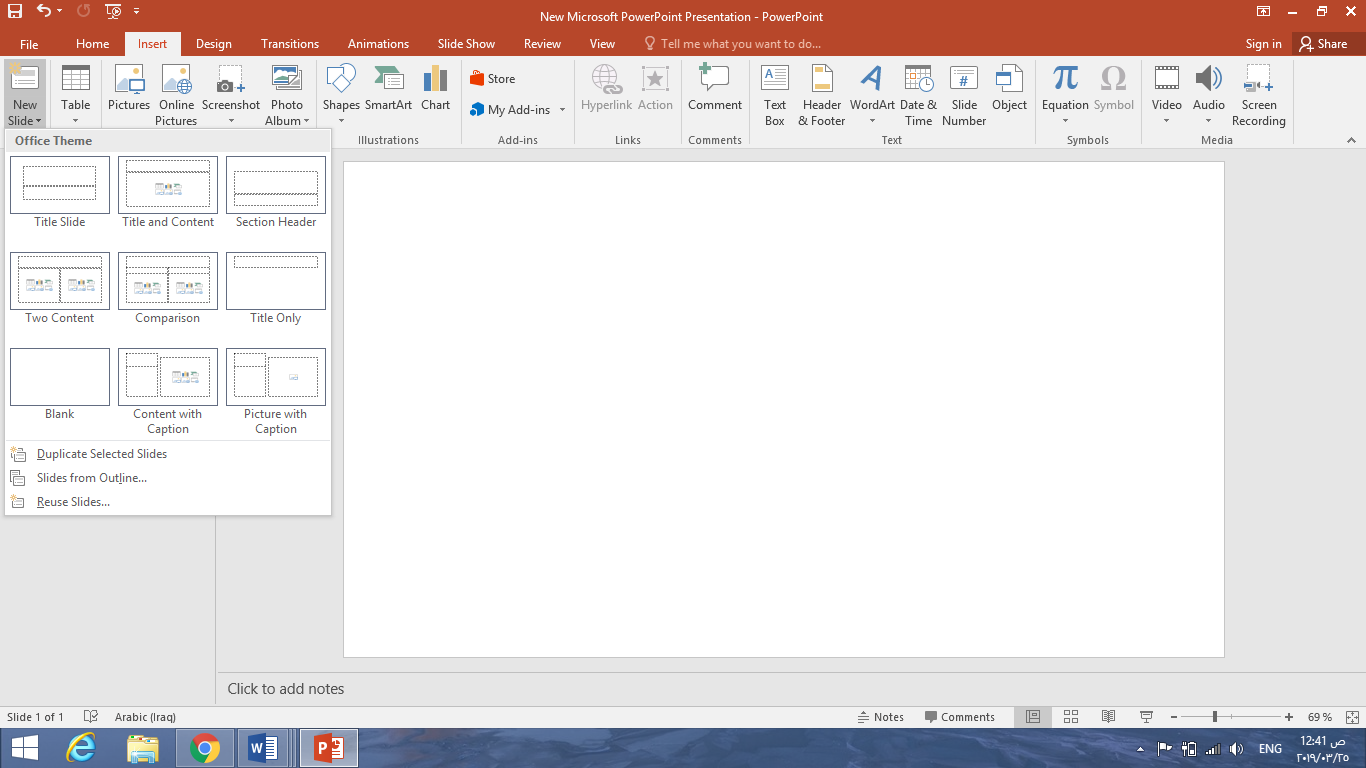
Insert new slide

To insert a new slide into a presentation:

1. In the slide preview pane on the left, left-click with your mouse in-between two slides where you want to insert a slide.

2. In the PowerPoint ribbon, on the Home or Insert tab, click the New Slide option.

3. In the drop-down menu that opens, select the type of slide to insert. The new slide will be inserted into the presentation where you clicked in step 1 above.

Figure (1.2) insert a new PowerPoint slide

Work in the PowerPoint user interface The PowerPoint user interface provides intuitive access to all the tools you need to develop a sophisticated presentation tailored to the needs of your audience. You can use PowerPoint 2016 to do the following (and much more):

■ Create, import, format, and edit slide content, including text, pictures, tables, charts, shapes, symbols, equations, SmartArt business diagrams, audio record- ings, and video recordings.

■ Capture screenshots, screen recordings, and audio recordings.

■ Organize and manage slides in sections.

■ Animate slide content and the transitions between slides; managing the form, timing, and sound associated with animations.

■ Document speaker notes for each slide.

■ Control the layout of content by creating custom masters; precisely align slide elements by using gridlines and Smart Guides.

■ Create, rehearse, present, and record custom slide shows.

■ Save, export, and send presentations in a wide variety of formats.

■ Create notes in a OneNote notebook that link to specific slide content.

The elements of PowerPoint interface

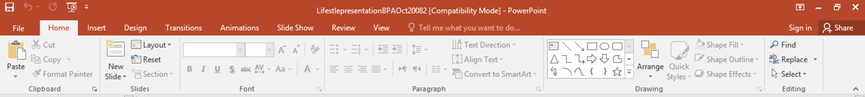
The PowerPoint app window contains the elements described in this section. Com- mands for tasks you perform often are readily available, and even those you might use infrequently are easy to find.

**Title bar**

At the top of the app window, this bar displays the name of the active file, identifies the app, and provides tools for managing the app window, ribbon, and content.

The title bar elements are always on the left end, in the center, and on the right end of the title bar

The Quick Access Toolbar at the left end of the title bar can be customized to include any commands that you want to have easily available. The default Quick Access Tool- bar in the PowerPoint app window displays the Save, Undo, Redo/Repeat, and Start From Beginning buttons. On a touchscreen device, the default Quick Access Toolbar also includes the Touch/Mouse Mode button.



Ribbon

The ribbon is located below the title bar. The commands you’ll use when working with a presentation are gathered together in this central location for efficiency

Applying theme in PowerPoint

To apply a standard theme to a presentation

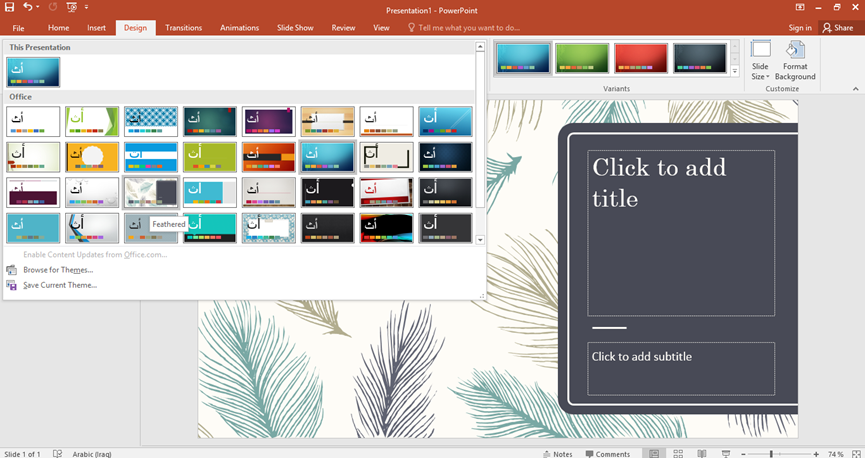
1. Display the presentation in Normal view.

2. On the Design tab, in the Themes group, click the More button (below the scroll arrows) to display the menu that includes the Office theme gallery and any custom templates on your computer.

3. Point to thumbnails in the gallery to display the theme names in tooltips and preview the effect of applying the themes to your presentation.

Choose a theme that enhances the content of your presentation

4. Click a theme thumbnail to apply that theme to the entire presentation.

 Figure (1.4) design of presentation

Animate text and pictures on slides

In the context of PowerPoint, animation refers to the movement of an element on a slide. When used appropriately, animated slide elements can both capture the audi- ence’s attention and effectively convey information. You can animate any individual objects on a slide, including text containers, pictures, and shapes. (You can’t animate objects that are part of the slide background or slide master, other than as part of the transition between slides.)

# **SPSS**

# Section 1: Overview

## 1.1 Introduction to SPSS Tutorials

This document is the first of a series of four modules intended for beginning SPSS users, providing an overview of SPSS for Windows. This first module introduces readers to the SPSS for Windows environment, and discusses how to create or import a dataset, transform variables, manipulate data, and perform descriptive statistics. The second module describes some commonly used inferential statistics, the third module discusses graphical display of output, and the fourth module covers other advanced topics. All modules can be found at <https://stat.utexas.edu/training/software-tutorials>. Throughout these modules, a single dataset, Employee data.sav, is used for all examples. This example dataset is provided with recent versions of SPSS. Thus, you will have access to the dataset and will be able to use SPSS to test your knowledge by replicating the examples contained in this document. Although the present documentation assumes SPSS Version 14, it will still be useful to users of SPSS on the Macintosh platform as well as many earlier, similar versions of SPSS. If you are a University of Texas affiliate and do not have access to SPSS or would like the software for your personal computer, visit the Software Distribution Services Web page at <http://www.utexas.edu/its/sds/>to get more information about obtaining the latest version of SPSS.

## 1.2 Introduction to SPSS

SPSS is a software package used for conducting statistical analyses, manipulating data, and generating tables and graphs that summarize data. Statistical analyses range from basic descriptive statistics, such as averages and frequencies, to advanced inferential statistics, such as regression models, analysis of variance, and factor analysis. SPSS also contains several tools for manipulating data, including functions for recoding data and computing new variables, as well as for merging and aggregating datasets. SPSS also has a number of ways to summarize and display data in the form of tables and graphs.

## 1.3 Overview of SPSS for Windows

SPSS for Windows consists of five different windows, each of which is associated with a particular SPSS file type. This document discusses the two windows most frequently used in analyzing data in SPSS, the Data Editor and the Output Viewer windows. In addition, the Syntax Editor and the use of SPSS command syntax is discussed briefly. The Data Editor is the window that is open at start-up and is used to enter and store data in a spreadsheet format. The Output Viewer opens automatically when you execute an analysis or create a graph using a dialog box or command syntax to execute a procedure. The Output Viewer contains the results of all statistical analyses and graphical displays of data. The Syntax Editor is a text editor where you compose SPSS commands and submit them to the SPSS processor. All output from these commands will appear in the Output Viewer. This document focuses on the methods necessary for inputting, defining, and organizing data in SPSS.

# Section 2: Entering Data in SPSS

## 2.1 Starting SPSS

To start SPSS, go to the Start icon on your Windows computer. You should find an SPSS icon under the Programs menu item. You can also start SPSS by double-clicking on an SPSS file.

When the program opens, it will present you with a “What would you like to do?” dialog box. For now, hit “Cancel” to dismiss the box.

## 2.2 The Data Editor

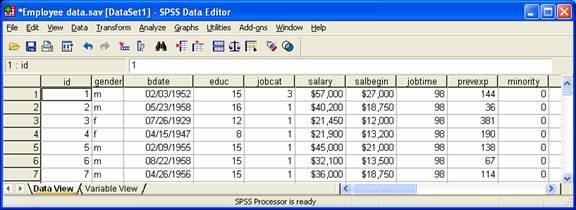
The Data Editor window displays the contents of the working dataset. It is arranged in a spreadsheet format that contains variables in columns and cases in rows. There are two sheets in the window. The *Data View* is the sheet that is visible when you first open the Data Editor; this sheet contains the data. You can access the second sheet by clicking on the tab labeled *Variable View.* While the second sheet is similar in appearance to the first, it does not actually contain data. Instead, this second sheet contains information about the variables in the dataset. Beginning with version 14, you can have multiple datasets open at one time in the Data Editor (however, this can be confusing, and we recommend keeping only one dataset open at a time while you are first getting familiar with the program.) Datasets that are currently open are called *working datasets;* all data manipulations, statistical functions, and other SPSS procedures operate on these datasets. Data can be directly entered in SPSS, or a file containing data can be opened in the Data Editor. From the menu in the Data Editor window, choose the following menu options:

**File**

**Open**

**Data...**

The *Open File* dialog box should automatically open to the SPSS directory of example files. Choose *Employee data.sav* from the list and click **Open.** Your Data Editor should now look like this:

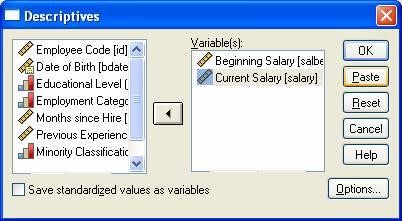


If the file you want to open is not an SPSS data file, you can often use the *Open* menu item to import that file directly into the Data Editor. If a data file is not in a format that SPSS recognizes, then try using the software package in which the file was originally created to translate it into a format that can be imported into SPSS (e.g., Excel).

## 2.3 The Syntax Editor

Another important window in the SPSS environment is the Syntax Editor. In earlier versions of SPSS, all of the procedures performed by SPSS were submitted through the use of syntax, which instructed SPSS on how to process your data. More recent versions contain pull-down menus with dialog boxes that allow you to submit commands to SPSS without writing syntax. This SPSS for Windows tutorial focuses on the use of dialog boxes to execute procedures; however, there are at least two reasons why you should be aware of SPSS syntax, even if you plan to primarily use the dialog boxes. First, not all procedures are available through the dialog boxes. Therefore, you may occasionally have to submit commands from the Syntax Editor. Second, the Syntax Editor is a useful way to save a log of what you have done, and to re-run what you have done at a later date. The dialog boxes available through the pull-down menus have a button labeled **Paste**, which will print the syntax for the procedure you are running in the dialog box environment to the Syntax Editor. Thus, you can easily generate SPSS syntax without typing in the Syntax Editor. This process is illustrated below.

The following dialog box is used to generate descriptive statistics. (You can get this dialog box by choosing **Analyze**, then **Descriptive Statistics**, then **Descriptives**, then clicking over the two variables using the arrow button.)



By clicking on the **Paste** button, the procedure that the above dialog box is prepared to run will be written in the form of SPSS syntax to the Syntax Editor. Thus, clicking the **Paste** button in the above example would produce the following syntax:

**DESCRIPTIVES**

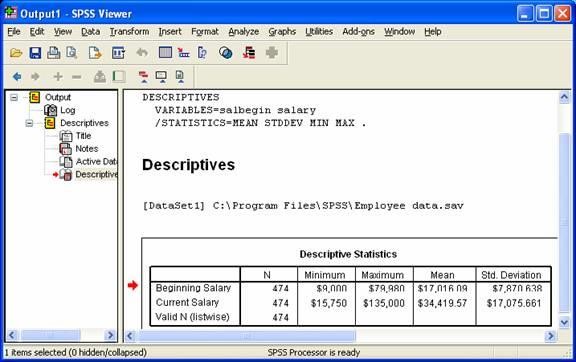
**VARIABLES=salbegin salary**

**/STATISTICS=MEAN STDDEV MIN MAX .**

This syntax will produce exactly the same output as would be generated by clicking the **OK** button in the above dialog box. The syntax that is printed to the Syntax Editor can then be saved and run at a later time, as long as the same dataset (or at least a dataset containing the variables with the same names) is active in the Data Editor window. Saving syntax is useful if you think you may want to rerun your analysis after you add more data, or if you want to run the same analysis on another dataset that contains the same variables.

## 2.4 The Output Viewer

When you execute a command for a statistical analysis, regardless of whether you used syntax or dialog boxes, the output will be printed in the Output Viewer. An example of the output viewer is shown below:



The left frame of the Output Viewer contains an outline of the objects contained in the window.

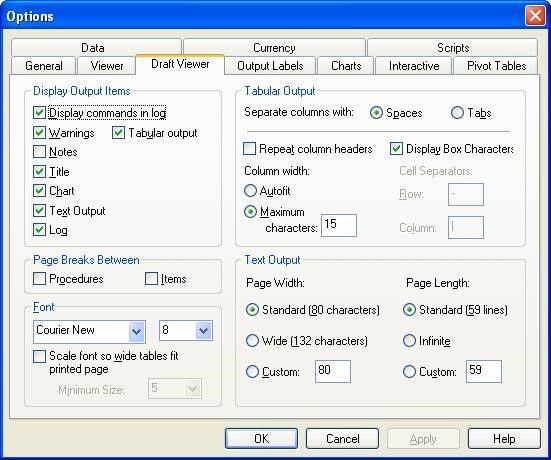
For example, the icon labeled *Log* represents the command syntax shown at the top of the figure. Everything under *Descriptives* in the outline refers to objects associated with the descriptive statistics. The *Title* object refers to the bold title *Descriptives* in the output. The *Active Dataset* object refers to the line in the output that designates which dataset was used to run the analysis. The highlighted icon labeled *Descriptive Statistics* refers to the table containing descriptive statistics. The *Notes* icon has no referent in the above example, but it would refer to any notes that appeared between the title and the table. This outline can be useful for navigating in your Output Viewer when you have large amounts of output. By clicking on an icon, you can move to the location of the output represented by that icon in the Output Viewer. You can also copy, paste, or delete objects by first highlighting them in the outline and then performing the operation you want.

You can control what is displayed in your output by using the *Options* menu item on the *Edit* menu:

**Edit**

**Options...**

Selecting this option will produce the following dialog box:



This figure shows the *Options* dialog box with the *Draft Viewer* tab selected, to choose which options you want to appear in the Output Viewer. Most commands are selected by default. Here, the *Display commands in log* option, normally unselected, was selected so that the command syntax will be written to the log in the Output Viewer. This can be useful for keeping track of which procedures you have executed.

## 2.5 Importing Data from Excel Files

Data can be imported into SPSS from Microsoft Excel and several other applications with relative ease. This document describes a method for importing an Excel spreadsheet into SPSS. If you are working with a spreadsheet in another software package, you may want to save your data as an Excel file, then import it into SPSS. If you have a spreadsheet that is arranged in a database format (e.g., you have several tables in your Workbook that are related through identification fields), there is another method for importing Excel file that you might consider that will merge tables within your database as part of the import procedure. It is described in our online tutorial of [SPSS Data Manipulation and Advanced Topics.](http://ssc.utexas.edu/software/software-tutorials#SPSS)

In order to easily import Excel data into SPSS, make sure your Excel spreadsheet is formatted as follows: (1) the spreadsheet should have a single row of variable names across the top of the file, and each variable name should begin with ordinary letters, rather than with any special characters, (2) the data should begin in the first column and second row of the Excel file, and (3) any graphs, labels, or extra text that is not part of the dataset should be deleted. To open an Excel file, select the following options from the menu in the Data Editor window in SPSS:

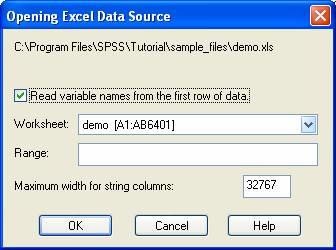
**File**

**Open**

**Data...**

First, select the desired location on disk using the *Look in* option. Next, select Excel from the

*Files of type* drop-down menu. The file you want should now appear in the main box in the *Open File* dialog box. You can open it by double-clicking on it. You will be presented with one more dialog box:



Your Excel spreadsheet should have all variable names on the top row, so leave the *Read variable names* option checked. Since Excel files can consist of multiple worksheets, the *Worksheet* drop-down menu allows you to choose which worksheet you wish to open. You may ignore the remaining options and choose **OK**. You should now see data in the Data Editor window. Check to make sure that all variables and cases were read correctly. Next, save your dataset in SPSS format by choosing the **Save** option in the **File** menu. 2.6 Importing data from ASCII files

Data are often stored in an ASCII file format, alternatively known as a text or flat file format. Typically, columns of data in an ASCII file are separated by a space, tab, comma, or some other character. To open these files, choose:

**File**

**Read Text Data**

The Text Import Wizard will first prompt you to select a file to import. After you have selected a file, you will go through a series of dialog boxes that will provide you with several options for importing data. Once you have imported your data and checked it for accuracy, be sure to save a copy of the dataset in SPSS format by selecting the *Save* or *Save As* options from the *File* menu:

**File**

**Save**

**Save As...**

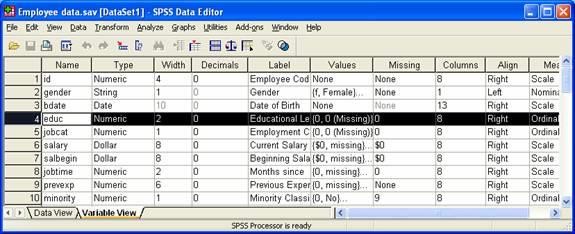
# Section 3: Creating and Modifying Data in SPSS

## 3.1 Creating and Defining Variables

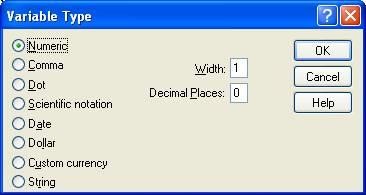
After data are in the Data Editor window, there are several things that you may want to do to describe your data. Before describing the process for defining variables, an important distinction should be made between two terms that are often confused: *variable* and *value*. A variable is a measure or classification scheme that can have several values. Values are the numbers or categorical classification representing individual instances of the variable being measured. For example, a variable could be created for employment classification status. Each individual in the dataset would be assigned a value representing their job classification. For instance, we could assign clerks the value 1, custodial workers the value 2, and managers the value 3.

One reason to define information about your variables is to help you interpret the output. For example, if your employment categories are coded as either 1, 2, or 3, it may be unwieldy to read the output if you are constantly trying to remember which number represents which categories. One advantage of defining variables is that these values can be assigned labels that will appear in your output, thus making it much easier to interpret. Another aspect of defining variable information is to provide SPSS with information about the type of data in your dataset, which is often critical for SPSS to correctly process analyses.

You can define information about your variables by clicking the *Variable View* tab. Doing so will bring the *Variable Information* sheet to the foreground. You can also access this sheet by double-clicking one of the variable names at the top of the columns in the Data Editor. The advantage of the second method is that it takes you to the row for the variable whose column head you clicked. Using the *Employee data.sav* dataset, you will see a spreadsheet organized as the one below:

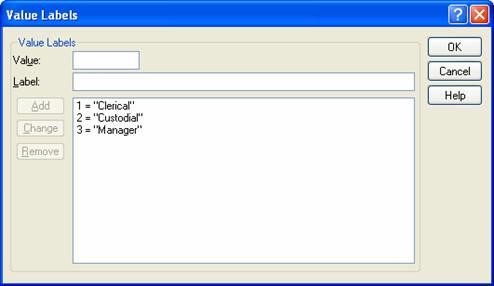


Many of the cells in the spreadsheet contain hidden dialog boxes that can be activated by clicking on a cell. If you see a gray box appear on the right side of the cell when you first click on the cell, this indicates that there is a hidden dialog box which can be accessed by clicking on that box. For example, clicking on the box in the cell for the *Type* column for the variable *jobcat* produces the following dialog box:



This box allows you to define the type of data for variables. For example, you will be presented with *Numeric*, *String*, and *Date* options. Thus, if you wanted to define *jobcat* as a string variable rather than the default numeric variable type, you would choose the *String* option.

Looking back at the Variable View, the *Missing Values* column allows you to define which values of a variable should be treated as missing data. The *Label* column is used to define labels for variables. The *Values* column is used to assign labels to the particular values of a variable. For example, the following dialog box shows that the *jobcat* variable that has been assigned the values 1, 2, and 3 for the labels *Clerical*, *Custodial*, and *Manager*.



To define variables as shown above, you should first enter the value (e.g., 1) in the box labeled *Value*, then enter the label associated with that value (e.g., *Clerical*), and click on the **Add** button. Repeat this process for each value you want to label.

## 3.2 Inserting and Deleting Cases and Variables

You may want to add new variables or cases to an existing dataset. For example, you may want to add data about participants' ages to an existing dataset. To insert a new variable, go to the Data View and right-click on the variable name next to the place that you would like to insert a new variable, then choose *Insert Variables*. To insert a case, right-click the row number below the place that you would like to insert the new row, and choose *Insert Case.*

You may also want to delete cases or variables from a dataset. To do that, select a row or column, and use the **Delete** key on your keyboard to delete the highlighted area. Alternatively, you can use the *Clear* option in the *Edit* menu.

## 3.3 Computing New Variables

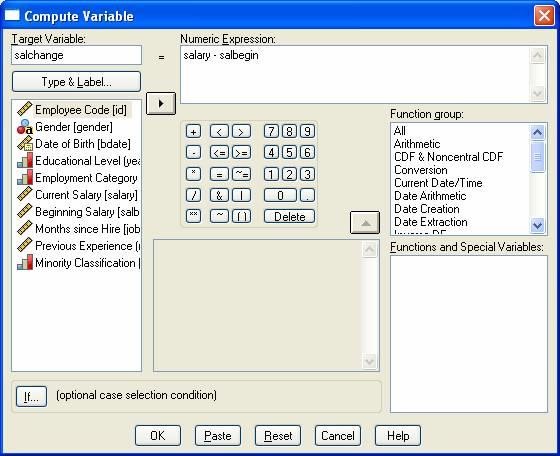
You may want to create new variables in your datasets. For example, *Employee data.sav* contains employees' salaries in terms of their beginning and current salaries, but you may also want the difference between starting salary and present salary; this new variable could be computed by subtracting the starting salary from the present salary. Alternatively, you may want to transform an existing variable. For example, the variable *jobtime* represents months of experience on the job, but you may wish to analyze data in terms of years on the job; in this case, you could compute the new variable by dividing *jobtime* by 12.

In both situations, the new variable can be created using the *Compute* option available from the menu in the Data Editor:

**Transform**

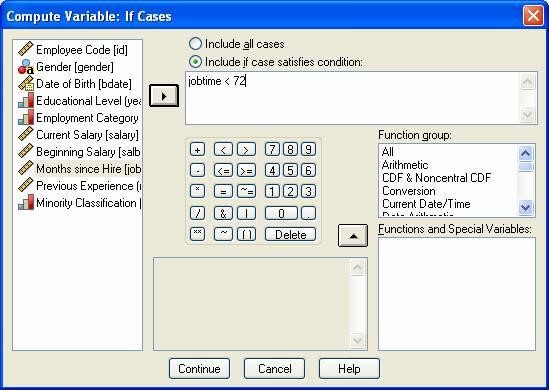
**Compute...**

To create a new variable, type its name in the box labeled *Target Variable*. The expression defining the variable being computed will appear in the box labeled *Numeric Expression*. This expression can either be typed into the box directly, or you can use the buttons located below the *Numeric Expression* box to input values or operators.



The example shown above demonstrates the computation of a new variable. This new variable, *salchange*, will be the difference between an employee's current salary and the employee's beginning salary. The new variable will appear in the rightmost column of the working dataset.

Variables can also be computed conditionally. For instance, if in the above example, you were only interested in the change in salaries for people who began working for the company within the last six years, you could create a condition that would compute a new variable only if an employee had begun employment within the last 72 months. To do this, first click on the button labeled **If**, which will produce the following dialog box:



First, click on the button labeled *Include if case satisfies condition* to activate the grayed-out areas of the dialog box. Then, specify the condition for computing a new variable in the box at the top right. You can either type in the condition or click on variables in the list on the left side of the dialog box and use the buttons on the bottom middle of the dialog box. Variables can be moved to the conditional box by clicking on the variable's name, then clicking the arrow button between the two boxes. Clicking on the buttons on the bottom left of the dialog box will cause the character on the button to be displayed at the location of the cursor in the input box.

The above example illustrates the definition of a condition that requires cases to have less than 72 months’ experience in order to be included in the computation of the new variable. The variable *jobtime* represents the number of months since an employee has been hired. Here, only employees with *jobtime < 72*, or who have been working at the company for less than 72 months, will be included. Click the **Continue** button to return to the previous dialog box, then click **OK**. The new variable should appear in the rightmost column of your dataset. The first several rows for this variable will be blank because these people have more than 72 months of experience; scroll down about 2/3 of the way down the dataset, and you will see the values for those with less than 72 months of experience.

## 3.4 Recoding Variables

You can also modify the values of existing variables in your dataset. For example, the variable *jobcat* codes an employee's status in three categories, but for a particular analysis you may want to combine two of these classifications into a single category. To do this, use the *Recode* option from the menu in the Data Editor:

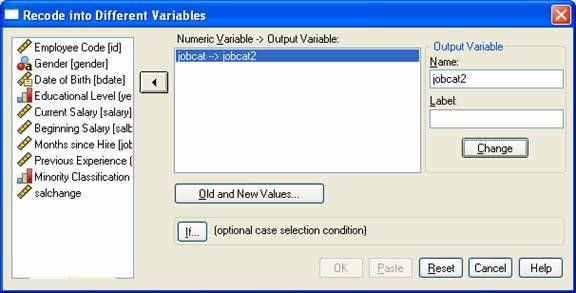
**Transform**

**Recode**

You will be offered two options for recoding variables in the *Recode* submenu. The *Into Same*

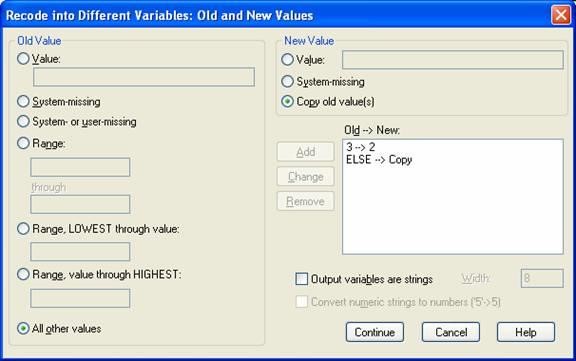
*Variables* option changes the values of the existing variables, whereas the *Into Different Variables* option is used to create a new variable with the recoded values. Both options are essentially the same, except that recoding into a different variable requires you to supply a new variable name. We typically recommend using the *Into Different Variables* option, so that you do not overwrite your original data.

The following example illustrates the use of the *Recode Into Different Variables* option to recode *jobcat* into a new two-category variable *jobcat2*.



First, a variable from the existing dataset should be selected by clicking on that variable, then clicking the arrow button in the middle of the dialog box. This will result in the selected variable being displayed in the box labeled *Numeric Variable -> Output Variable*. Next, you must supply the name of the new variable, and optionally you can supply a label for the new variable. Then click **Change***.* After a new variable name has been supplied, click on the button labeled **Old and New Values***.*

The original value of the variable being recoded is entered in the box labeled *Old Value*, and the new value is entered in the box labeled *New Value*. After values are entered in these boxes, click on the button labeled **Add** to complete the recode process.



Here, the old variable *jobcat* has 3 values, and we wish to recode it into only two values. In the example dataset, *jobcat* has three values: 1, 2, and 3. If the goal were to combine cases with the values 2 and 3, this could be accomplished by recoding cases with the value 3 into 2's. Enter 3 in the box labeled *Old Value* and enter 2 in the box labeled *New Value,* then click **Add**.This can be repeated for as many of the values as necessary. In this case, we want to simply copy all the other values; under *Old Value*, click on *All other values*, and under *New Value*, click *Copy old values*, then choose **Add.** Now that all the old values have been accounted for, you may click **Continue** and then **OK**.

Values can also be recoded conditionally. The process for recoding values on the basis of a condition is essentially identical to the process for conditionally computing new variables discussed in the previous section: when you click on the **If** button in the main *Recode* dialog box, the same dialog box that was obtained from clicking **If** in the *Compute* dialog box will appear with the same options.

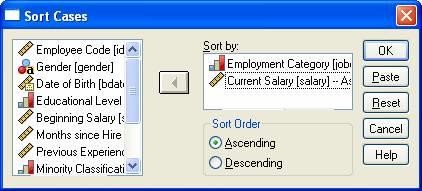
## 3.5 Sorting Cases

Sorting cases allows you to organize rows of data in ascending or descending order on the basis of one or more variable. For example, the data could be sorted by job category, so that all of the cases coded as job category 1 appear first in the dataset, followed by all of the cases that are labeled 2 and 3 respectively. The data could also be sorted by more than one variable. For example, within job category, cases could be listed in order of their salary. The *Sort Cases* option is available under the *Data* menu item in the Data Editor:

**Data**

**Sort Cases...**

The dialog box that results from selecting *Sort Cases* presents only a few options:



To choose whether the data are sorted in ascending or descending order, select the appropriate button. You must also specify on which variables the data are to be sorted. The hierarchy of such a sorting is determined by the order in which variables are entered in the *Sort by* box. Variables are sorted by the first variable entered, then the next variable is sorted within that first variable. Here, *jobcat* was the first variable entered, followed by *salary*; accordingly, the data would first be sorted by *jobcat*, then, within each of the job categories, data would be sorted by *salary*.

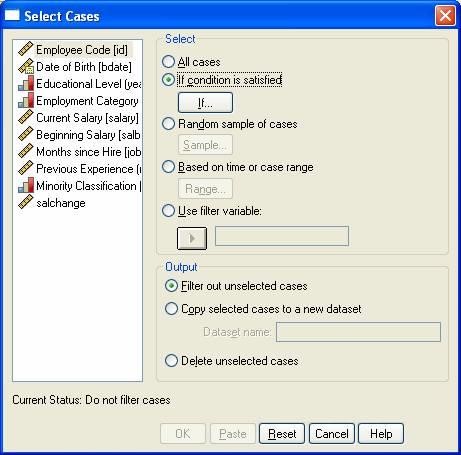
## 3.6 Selecting Cases

You can analyze a specific subset of your data by selecting only certain cases in which you are interested. For example, you may want to do a particular analysis on employees only if the employees have been with the company for greater than six years. This can be done by using the *Select Cases* menu option, which will either temporarily or permanently remove cases you don't want from the dataset. The *Select Cases* option is available under the *Data* menu item:

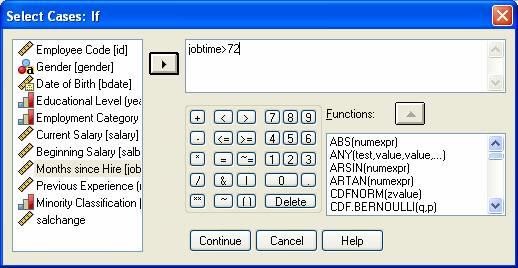
**Data**

**Select Cases...**

Selecting this menu item will produce the following dialog box. This box contains a list of the variables in the active data file on the left and several options for selecting cases on the right.



Selecting one of these options will produce a second dialog box that prompts you for the particular specifications in which you are interested. For example, selecting the *If condition is satisfied* option and clicking on the **If** button (as was done in the example) results in a second dialog box, as shown below. The portion of the dialog box labeled *Output* gives you the option of temporarily or permanently removing data from the dataset. The *Filter* option will remove data from subsequent analyses until the *All Cases* option is reset, at which time all cases will again be active and used in further analyses. The *Copy* option will save the selected cases to a new dataset. The *Delete* option will remove unselected cases from the working dataset; be very careful with this option, because if the dataset is subsequently saved, these cases will be permanently deleted. Here, we have chosen to use the *Filter* option.



Clicking on the **If** button opens the *Select Cases: If* dialog box. Here, we select all of the cases in the dataset that meet a specific criterion: employees that have worked at the company for greater than six years (72 months) are selected. After this selection has been made, subsequent analyses will use only this subset of the data. Because we selected the *Filter* option in the previous dialog box, SPSS will indicate the inactive cases in the Data Editor by placing a slash over the row number. To select the entire dataset again, return to the *Select Cases* dialog box and select the *All Cases* option.

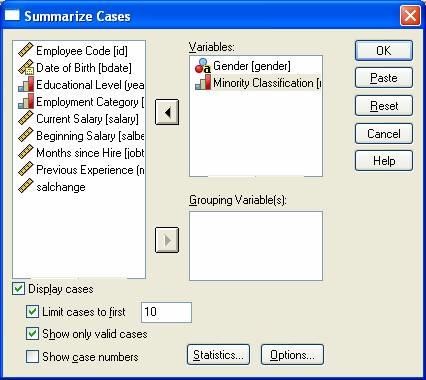
## 3.7 Listing Cases

You may sometimes want to print a list of your cases and the values of variables associated with each case, or perhaps a list of only some of the cases and variables. For example, if you want to visually examine the gender and minority status of each person in your dataset, you can generate a list of only these variables in the Output Viewer. This can done by using the *Summarize Cases* menu option, available under the *Analyze* menu item:

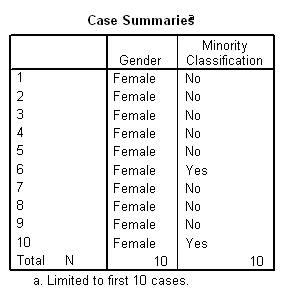
**Analyze**

**Reports**

**Case Summaries . . .**



The above example requests a listing of each person’s gender and minority status. The option *Limit cases to first* has been checked, which allows you to request a listing of only the first 10 cases. If you sorted the dataset earlier in this tutorial, then the Output Viewer would show this table:



# 