Formulas and Functions with Excel

Kiruthika Ragupathi
Assistant Director
(Educational Technology)

Centre for Development of Teaching and Learning (CDTL)
National University of Singapore
email: kiruthika@nus.edu.sg

TABLE OF CONTENTS

Synopsis

Need to perform complex calculations and data analysis? The usage of formulas and functions is what gives an Excel spreadsheet much of its power. This workshop explores the formula and functions that are useful for computing and managing data. This course begins with mathematical principles and moves on to basic formula construction. Participants will learn about mathematical functions, statistical functions, conditional calculations, and lookup tables. In addition, the use of Analysis ToolPak will be discussed aiding in the generation of descriptive statistics and regression.

Objectives

By the end of this workshop, participants should be able to:

- Set up and verify data
- Understand the mathematical order of operations used by Excel
- Use the appropriate type of cell reference
- Construct basic formulas
- Use statistical functions
- Perform conditional calculations
- Use Analysis Toolpak

Kiruthika Ragupathi, email: kiruthika@nus.edu.sg
Centre for Development of Teaching and Learning

Introduction

Microsoft Excel has many capabilities that make it suitable for use as a data management tool. It provides multiple features for organising and managing data, so you can ensure that data is entered correctly and calculations and formulas are valid.

Data organisation features enable you to:

- sort and filter data,
- summarise and group data, and
- outline data so that you can focus on the key parts of your data.

Validation features are very important for maintaining accurate records and to ensure that

- the data is correct,
- it is entered in the proper format, and
- formulas are working correctly,

Excel makes use of formulas (mathematical expressions that you create) and functions (mathematical expressions that are already available in Excel) to dynamically calculate results from the data available in your worksheets.

Exploring the Excel Workspace

Let us start with an overview of the most important elements of the command area.

1. Use the **File menu** to open new or saved workbooks, save, print or close workbooks or manage Excel options.
2. You can set up a custom **Quick Access toolbar** on the upper left corner of the screen. You can place the most commonly used commands here.
3. The **Ribbon** is split into tabs where the commands are organised into logical groups. Each tab corresponds to the various toolbars used in the previous versions of Excel.
4. An **Excel-help function** as a question-mark icon on the outer right of the Ribbon.

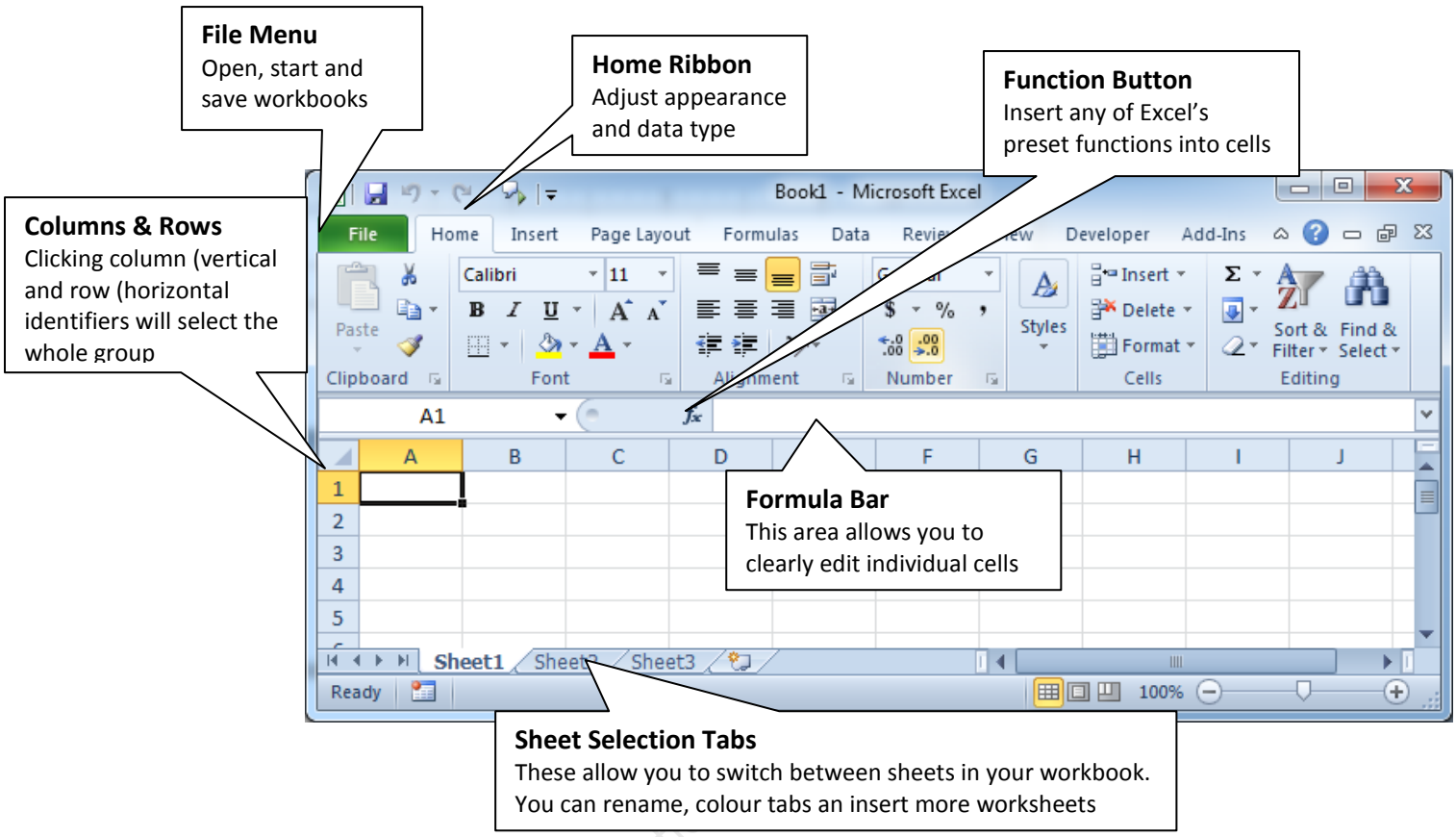
Navigating the Workbook

This section explains how to navigate a spreadsheet to edit and format cells. The concepts can easily be applied to early versions of Excel; menu locations may differ but commands are the same. The core document of Excel is a workbook. An Excel workbook can hold any number of sheets. Several worksheets can be saved together to form a workbook. The main types of sheets are:

- Work sheets
- Chart sheets

At any one time, only one sheet is active in a workbook. Each spreadsheet is gridded into columns and rows. Columns are denoted using letters, and rows are denoted by numbers. Cells are referenced by their column and row. For example, the cell in the upper left hand corner of the spreadsheet is called “A1,” for it is in column A and row 1.

Below is a picture of a blank workbook and descriptions of the more frequently used items.



Creating a new workbook

Start a new workbook by clicking the **File menu** and selecting **New**. In the **New Workbook** window, click **Blank** workbook. When you start Excel, a blank worksheet opens.

Selecting cells

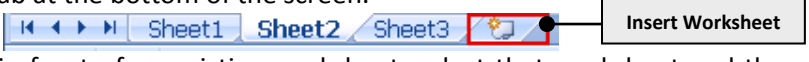
Edit individual cells by clicking on them. Click a cell and drag the mouse pointer to select a range of cells. You can also select entire rows and columns by clicking on the number or letter heading of that alignment. To select non-adjacent cells, hold down the **CTRL** key, and then click the cells that you want.

Adding data manually

To enter data manually into the spreadsheet, double click on the cell you want to edit. Notice that whatever you type appears in both the cell and the cell editor at the top. When you are finished typing, press Enter.

Adding a new worksheet

Click the **Insert Worksheet** tab at the bottom of the screen.



To insert a new worksheet in front of an existing worksheet, select that worksheet and then, on the **Home** tab, in the **Cells** group, click **Insert**, and then click **Insert Sheet**. You can also right-click the tab of an existing worksheet, and then click **Insert**. On the **General** tab, click **Worksheet**, and then click **OK**.

Renaming the worksheet

Double-click the **Name** tab; when the default name is selected, type the new name. You can also right click on the Name Tab and choose **Rename** to type in the new name. Each worksheet can be colour-coded, as setting different colors is helpful when you have a large number of worksheets in a workbook. To set the color of the Worksheet tabs, right click on the tab and select a new color.

Navigating

To change the active cell, use the arrow keys, Page Up or Page Down keys, or use the mouse to click a new cell or drag the scroll bars.

Embedding charts and pictures

To create a new chart in Excel, on the **Insert** menu, from the **Charts** group, choose the chart type you prefer. To insert another file, such as a clip art, picture or a scanned image, from the **Insert** tab, under the **Illustrations** group, choose **Picture** or **Clipart**.

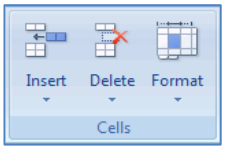
Setting up Data

To use a spreadsheet efficiently, it helps to organise the data so that it is easy to read. When grading, columns are typically used for each assessment item, whereas rows are used for the individuals. The first row will be used for the column titles and the first column will be used for the serial number. You can use the second row to indicate the maximum possible score for each of the assessment item.

Inserting a new column or row

The process for inserting rows and columns is similar. To insert a column, click the header where you want to add a new column or row. On the header, right-click and in the drop-down context menu, select **Insert**.

You may also use the **Insert** button to add new columns found in the **Home** tab, **Cells** group. When you insert a column or row, content found in columns will be moved to the right, and content in rows will be moved down.

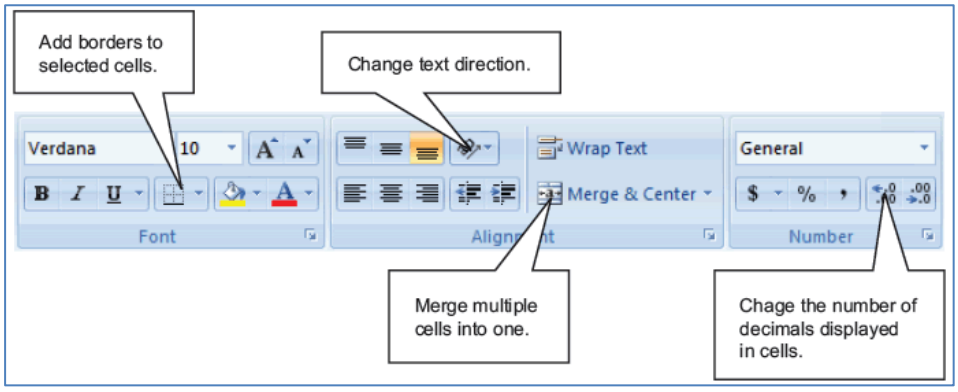


Formatting the worksheet

The formatting of a cell refers to both the way it is styled (stylistic formatting) and the way it functions (numeric formatting). Formatting includes display characteristics such as font, size, alignment, style, color, as well as the type of data that the cell contains. For instance, a cell can be formatted to treat any data entered as a monetary amount and display only whole dollar amounts. Stylistic formatting options make your worksheet more attractive and easier to read.

Appearance

Formatting the appearance of cells, rows, and columns can be done with buttons in the **Home** ribbon or by right-clicking and choosing options from **Format** menu. To format cells, select the **Home** ribbon then select the column, row, group or the cells that you want to format. Then, select the formatting options you desire.



Kiruthika Ragupathi, email: kiruthika@nus.edu.sg
Centre for Development of Teaching and Learning

A new formatting concept called the ***document themes*** has been introduced in Excel 2007 and can be found in the **Page Layout** Tab. These allow you to set many formatting options at once.

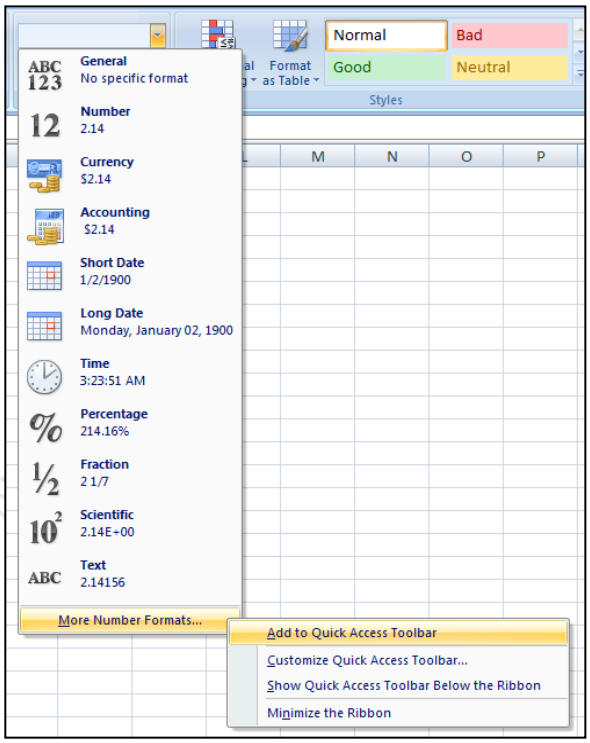
Data Types

The **data type** manages how Excel will display and interpret data in the cells. For instance, you may choose percentages as a fraction, decimal or whole number. It is important that excel interprets your data correctly since Excel's functions depend on the type of data being manipulated. After you type numbers in a cell, you can change the format in which they are displayed.

The most commonly used data types for student data management in Excel are:

Number: cells that contain only numerals, commas, and decimal points that can be used in numerical calculations.

- Click the cell that contains the numbers that you want to format.
- On the **Home** tab, in the **Number** group, point to **General**, and then click the format that you want.
- When you click on **More Number Formats** on the category list, the **Format cells** dialog box opens (similar to that in the older versions).
- To add the format cells icon to your Quick Access toolbar, right-click on the **More Number formats** and choose the **Add to Quick Access Toolbar**.



Percentage: multiplies the cell value by 100 and displays the value with a **%**.

Text: cells that contain letters, numbers, spaces, or any other keyboard character.

Freeze Panes

This is used to keep titles in sight when you scroll down a page. To freeze panes, first "split" the worksheet so that you can look at two areas independently.

On the **View** tab, in the **Window** group, select **Split**. A thick grey bar should appear on the sheet. To make a split, drag the grey bar to a position next to or below the title row. When you have positioned the bar where you would like it, select **Freeze Panes** from the **Window** group of the **View** tab. The thick grey bar will be replaced by a thin black line and the row/column will stay in place.

Verifying Data

When it is necessary to enter large amounts of data, it is good practice to have the data checked for accuracy. This can be done immediately by the person entering the data, or entries can be verified by a different person. Wouldn't it be convenient to have a tool that could provide immediate feedback of the data entered?

Text to Speech feature

Excel provides such a tool that enables a quick and easy double-check of entries before you continue working with the data. It is the Text to Speech feature.

Through this **Speak Cells** functionality, Excel can read back to you what you have typed while you check that audio against your original data. For this to work, some commands have to be added to the quick access toolbar.

- Using the **Excel Options**, and in the **Customize** section, choose **All Commands.**
- Then scroll down to select the **Speak Cells** command
- Click **Add** and click **OK**. And this option is now available for use in the Quick Access toolbar.
- Each time, you want to use a text-to-speech command, choose the icon  , from the Quick Access Toolbar.

Formulas and Functions

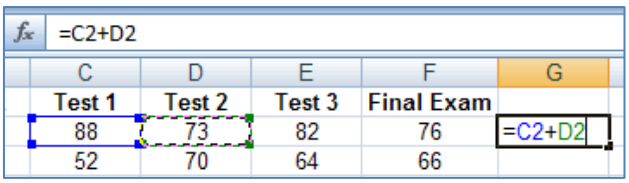
Formulas are the real power of an Excel spreadsheet. A formula uses standard mathematical symbols to operate on cell addresses and/or numbers. A formula can be a combination of values (numbers or cell references) and mathematical operators into an algebraic expression. Mathematical operators include the following symbols:

+	for addition
-	for subtraction
*	for multiplication
/	for division
%	for percentage
^	for exponentiation (power)

In addition you can use parentheses and decimal points, where needed.

Constructing a formula

- Excel requires that EVERY formula begin with an equal sign (=). If you just type without this symbol, Excel treats the entry as text.
- To start entering a formula in a cell, click in that cell and then type the formula. Type Enter or Tab to move to the next cell when you have finished entering the formula.
- Formulas containing numbers will produce results that will not change. e.g., *the formula* ***= 200*7*** *always produces the result 1400.*
- However, a formula that contains cell references produces a result that may change depending on the data in the cell. e.g., *the formula* = C2+D2 *will produce a result based on the data in cells C2 and D2.*



- To view the formulas that are contained in your worksheet, press CTRL and ` (the apostrophe key) together. Repeat to hide the formulas and show the values.
- All formulas follow the standard mathematical order of operations when calculating the results. e.g., *the formula* = A2 + B2 * C2 *will add data in cell A2 with the multiplied product of B2 and C2.*
- If a part of the formula is in parenthesis, then that part will be calculated first.
- After expressions in the parentheses, Excel will calculate your formula using the Math operators in the following order: **Multiplication, Division, Addition** and **Subtraction.**

Copy formulas into additional cells

You can copy formulas into additional cells by selecting the formula, and then paste into the desired cell. Usually, cells referenced in a formula are based on their relative location to the cell with the formula (known as **relative referencing**). The cell references will change depending on where you copy it. If cell **G2** contains the formula =**C2+D2,** and you copy this to **G3**, the resulting formula will be =**C3+D3**.

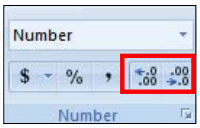
But sometimes, your formula will always have to point to exactly the same cell or cells, no matter where you copy and paste your formula. This is known as **absolute referencing** and have an added feature of a dollar sign $ placed in front of the row and column references. For example, if you wanted to add cells A1 and B1 using an absolute reference, your formula would be **= $A$1 + $B$1**.

Repeat the calculation for every cell in a column

To do this select the cell with the formula and when you move the pointer to the lower right hand corner of the cell, the cursor changes to a bold plus (**+**) sign. Click and hold down the mouse while you drag to highlight cells down the column to the last row of data. Every highlighted cell now contains the formula. Note: Display of #NA indicates problem(s) with the formula or the result is larger than the width of cell.

Round values to remove decimal places

To reduce decimal places click on the **Decrease Decimal** button. Each click of this button shrinks the result one decimal place. Be cautious with this tool though as numbers will be rounded. So, for instance, an 85.55 will be rounded to 86.



Numeric Precision

Excel maintains an internal numeric precision of 15 digits. Excel stores 15 digits internally, but rounds the value for the screen display according to the format of the cell. When you apply a formula, Excel looks at the actual value held in the cell, not the number displayed. This means that if you have formatted a number to no decimal places and a whole number is displayed, the total will not necessarily be the same as the sum of the resulting whole numbers as all the fractional parts will have been added as well.

Sample Formulas

Formulas can be a combination of built-in functions and your own formulas. Let us take a look at some sample formulas.

1. This formula multiplies 200 x 7, returning 1400 and only uses literal values:

   `= 200*7`

2. This adds the values in cells B2 and C2:

   `= B2+C2`

3. This subtracts the value in the cell B3 from the value in cell B2 and then multiplies the result by the value in cell B4

   `= (B2-B3)*B4`

4. This formula uses the SUM function to ass the values in the range C1:C12

   `= SUM(C1:C12)`

5. This one compares the value in cell A1 and Value in cell B1. If the values in the two cells are identical, then the formula returns TRUE otherwise, it returns FALSE.

   `= A2=B1`

6. For example, to find the average of the values in the cells A2 to A10 you could do this several ways:

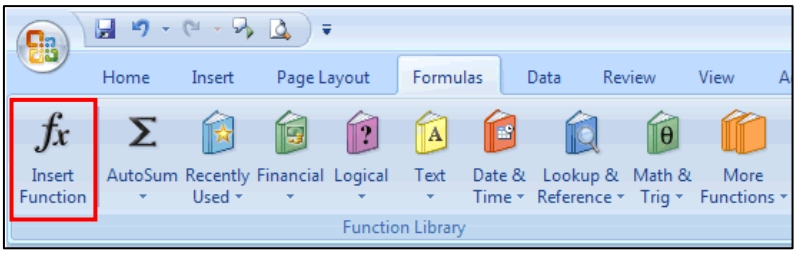
   ```
   = (A2 + A3 + A4 + A5)/4
   = SUM (A2,A3,A4,A5)/4
   = SUM (A2:A5)/4
   = AVERAGE (A2,A3,A4,A5)
   = AVERAGE (A2:A5)
   ```

   These formulas would all give the same result if all the cells contained values but the easiest one to type is probably the bottom one. It is also safer to use the AVERAGE function as that calculates how many values there are in the range you select, using =SUM(A2:A5)/4 would lead to an error if one of those cells was intentionally a blank row as you may only have 3 actual values in that range.

Inserting a function

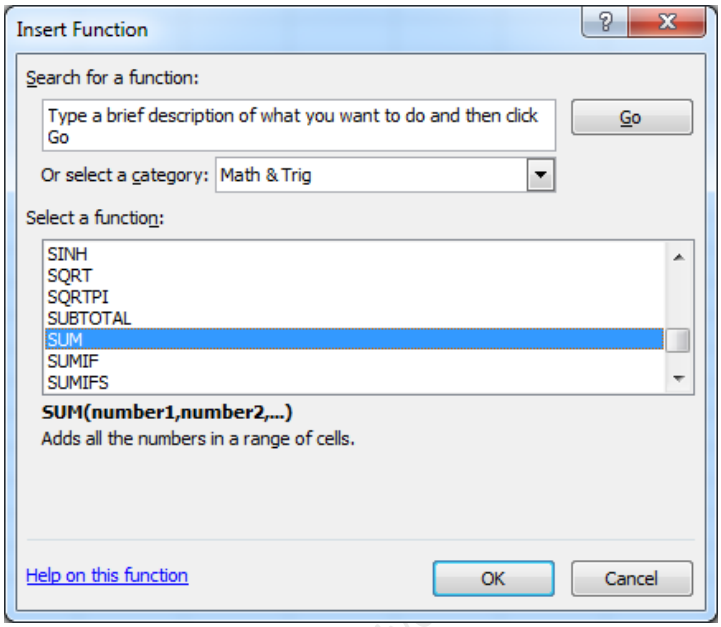
A function is a predefined/prewritten formula that takes a value or values, performs an operation on a range of cells you select, and returns a value or values. Functions are used to simplify and shorten formulas on a worksheet, especially those that perform lengthy or complex calculations. Excel refers to each range of cells in the function as an argument. For complex calculations, it might be necessary to use as many as 3 to 4 arguments.

1. Select the cell where you would like the average score to appear.
2. From the **Formulas tab**, choose **Insert Function**.

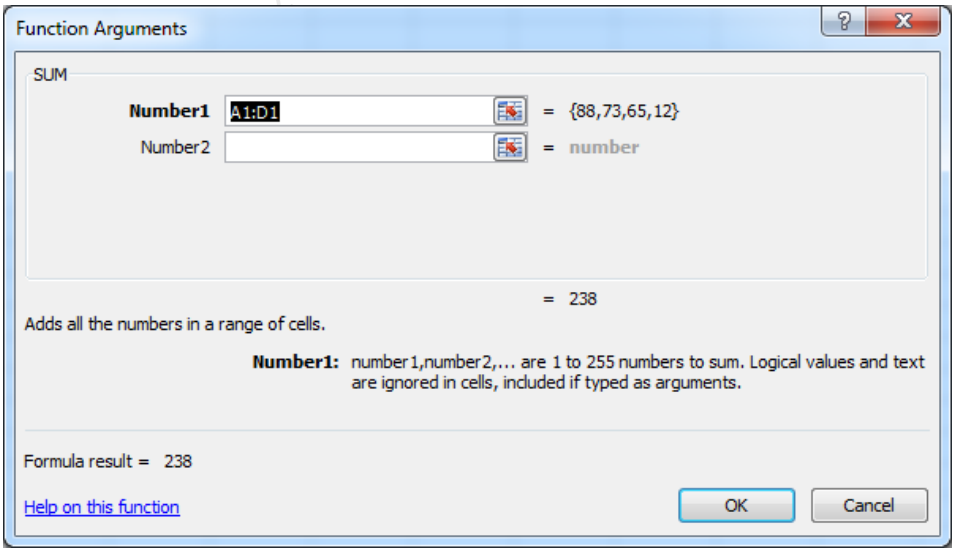


Kiruthika Ragupathi, email: kiruthika@nus.edu.sg
Centre for Development of Teaching and Learning

3. Or, you could also click on the Function button,  found just before the formula bar to use any of Excel's preset functions.
4. The **Insert Function** dialog box appears.



5. Under **Select a function**, choose from the range of functions available.
6. For example, choose the **SUM** function and click OK.
7. The **Function Arguments** dialog box as shown below appears.
8. For example let us consider the **SUM** function that adds all the numbers in a range of cells.

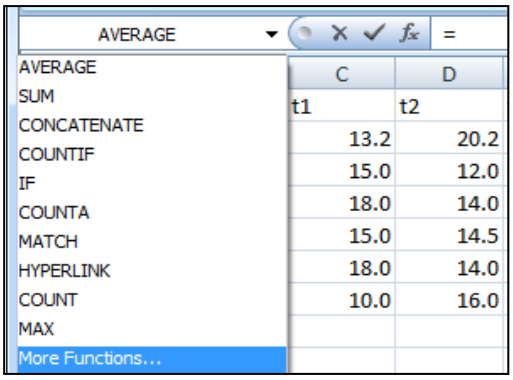


9. Under the **Number1** type the range of cells that you want to use or you can also drag the mouse to select the range of cells to be included as the function's arguments and click **Ok**.
10. The formula result is displayed on the window for you as a preview.

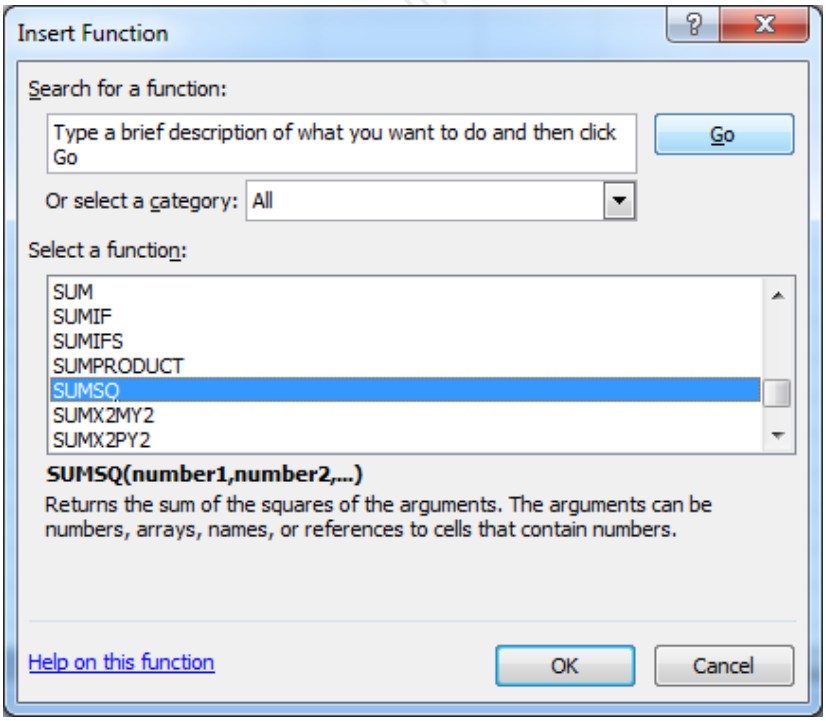
Built-in Functions

Excel has a variety of built-in functions that can be accessed using the Formula Wizard.

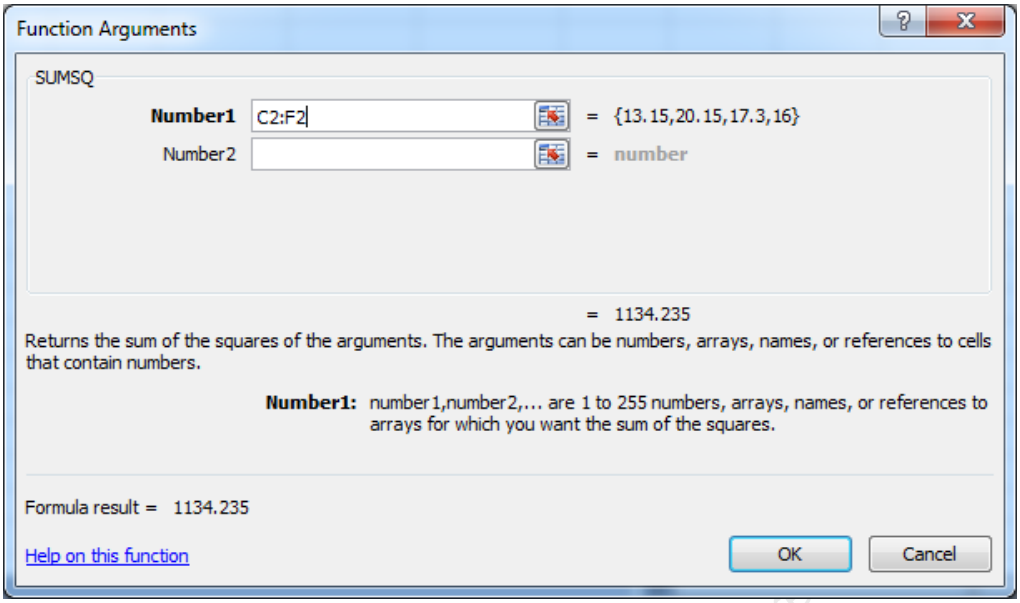
1. Click in the cell where you want the result of the formula to be placed. Now click on the **=** sign in the formula bar.



2. Click on the drop-down arrow to the left of the formula bar to select the function you wish to use.
3. Click on the **More Functions...** option at the bottom of the list to display a window showing all the available functions.
4. When you have selected the function, the **Insert Function** dialog box opens to help you complete the arguments after the function so that Excel calculates the right result.

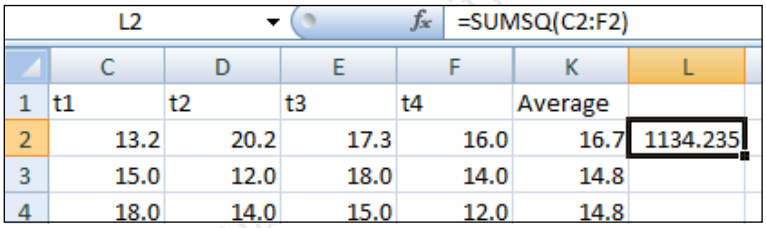


5. When each function is selected a short description of the function and the type of arguments to be used is displayed in the dialog box.
6. For a function like the **SUMSQ** function you need to select a range of cells (C2:F2 in this case) to add together the sums of the squares of the arguments chosen.



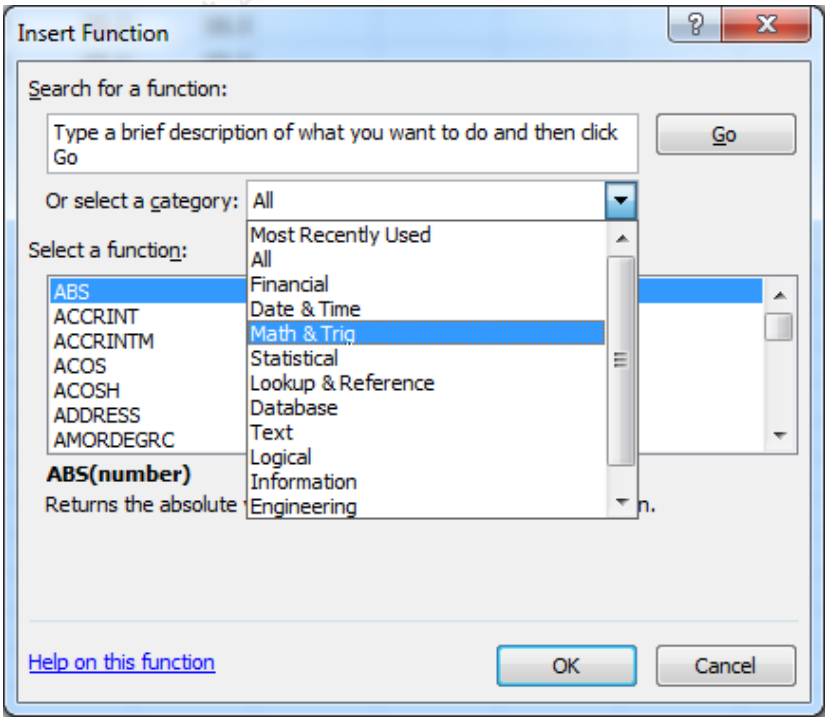
7. An example of the result of by computing the formula is shown at the bottom of the dialog box. Then, click the OK button when you are satisfied with the output.

8. The result will now be displayed on the selected cell of your spreadsheet. If you click on the cell containing the result, the formula used will be displayed in the formula bar.



	C	D	E	F	K	L
1	t1	t2	t3	t4	Average	
2	13.2	20.2	17.3	16.0	16.7	1134.235
3	15.0	12.0	18.0	14.0	14.8	
4	18.0	14.0	15.0	12.0	14.8	

In the **Insert Function** dialog box, you can choose **All** from the list or choose a specific category of functions you are interested in that may contain the function and look for it in there.



Built-in functions can be categorised as follows:

Function Category	What it does
Mathematical/Trigonometric	Take a number as data, transform it, and produce a numerical result. For example, sine and absolute value are math functions.
Logical	Consider a condition and return True or False
Text	Manipulate or create strings
Lookup/Reference	Manipulate or examine areas of a worksheet
Statistical	Return statistical values to sets of numbers. For example, average, count, min, max, etc.
Database	Same as statistical, however datasets are taken from a database
Date/Time	Perform calculations on dates, times, and combinations of dates and times
Engineering	Perform commonly used engineering calculations, many of which relate to Bessel Functions, Complex Numbers or converting between different bases
Financial	Helpful when calculating such things as interest, monthly payments, and assist in what-if scenarios
Information	Used to get information about the contents of a cell

Mathematical Calculations

Function	What it Does
SUM	Adds its arguments
SUMIF	Adds the cells specified by one or many given criteria
SUMPRODUCT	Returns the sums of the products of two arrays
SUBTOTAL	Returns a subtotal of a filtered list or database
TRUNC	Truncates a number to an integer
ROUND	Rounds a number to a specified number of digits
ROUNDUP	Rounds a number up, away from zero
INT	Rounds a number down to the nearest integer
ABS	Returns the absolute value of a number
MOD	Returns the remainder from division
SQRT	Returns a positive square root
POWER	Returns the result of a number raised to a power

SUM

=SUM(A1,B6,G6)	will return the sum of the values in cells A1, B6 and G6
=SUM(A1:A23)	will return the sum of the values in cells A1 to A23
=SUM(A1:A23,F3:F34)	will return the sum of the values in cells A1 to A23 plus the sum of the values in cells F3 to F34

For example, if you would require a yearly summary report, you want to sum the values in cells B2 of each of the monthly sheets. You have named your sheets "Jan" , "Feb", etc. Then to calculate, key in the following formula:

=Jan!B2+Feb!B2+Mar!B2...+Dec!B2

You can also write this as follows:

=SUM(Jan:Dec!B2)

SUMIF

The general format for the SUMIF formula is as follows:
=SUMIF(range where the criteria is to be applied, criteria, sum what).

	A	B	C	D
1	Date	Country	Name	Amount
2	05/01/2011	IN	Lim	2
3	06/01/2011	SG	Peter	6
4	07/01/2011	US	Lim	8
5	05/01/2011	AU	Peter	3
6	06/01/2011	DE	Lim	4
7	07/01/2011	CN	Peter	6

Example 1: Sum the values of the cells D2 to D7, only if they are greater than 5

=SUMIF(D2:D7,">5",D2:D7); Result should be 20.

Example 2: Sum the values of the cells D2 to D7 if the value in cells C2 to C7 is “Lim” or equivalent to the value in cell C2.

=SUMIF(C2:C7,"Lim" ,D2:D7) ; Result should be 14.

or

=SUMIF(C2:C7,C2 ,D2:D7)

SUMPRODUCT

Let's say that you have a series of quantities in cells A1 to A5 and a series of unit prices in B1 to B5. With SUMPRODUCT you can calculate total sales with this formula:

=SUMPRODUCT(A1:A5,B1:B5)

Basically SUMPRODUCT sums A1 multiplied by B1 plus A2 multiplied by B2 and so on. But you could also apply specific conditions to it.

=SUMPRODUCT(A1:A5,B1:B5, (C1:C5="Lim")*1)

SUBTOTAL

When the Filter Data is used, then the SUBTOTAL function is a very useful and interesting function. The function SUBTOTAL allows to count, to sum or to calculate the average of filtered elements of a database.

The function requires two arguments:

- the first is a number between **1** and **11** that specifies the operation to be executed: 1(for AVERAGE); 2(for COUNT); 3(for COUNTA); 4(for MAX); 5(for MIN); 6(for PRODUCT); 7(for STDEV); 8(for STDEVP); 9(for SUM); 10(for VAR); 11(for VARP).
- the second is the range covered by the functionand

Hence, if you want the sum of B2:45, then use the following:

=SUBTOTAL(9,B2:B45)

Rounding up the decimals

When you specify in the format of a cell that you want only 2 decimals Excel shows only 2 decimals (rounding up) BUT you should note that still uses all the decimals. For example, if in cell A1 you entered 2.1456 and use a 2 decimal format, you will see 2.15. Now if in cell B1 you write the formula =A1 and make the format "General" you will see still see 2.1456. Hence, functions like **INT, TRUNC, ROUND, ROUNDUP** and **ROUNDDOWN** will enable you to use a specific number of decimals in your calculations.

TRUNC

The **TRUNC** function does the same as the INT or ROUNDDOWN functions. The TRUNC function removes decimals without rounding. If you have 2.2 or 2.7 in cell A1 **=TRUNC (A1,0)** will return 2. Interestingly enough if you have 12345 in B1 using a minus sign in the second argument of TRUNC **=TRUNC(B1,-3)** will return (12,000). This is very useful when you don't want to show the hundreds, the tens and units in a report.

ROUND

This function removes decimals rounding up the last decimal if the next one is 5 or over. So if you have 4.126 in cell A1 and use the formula **=ROUND(A1,2)** the result will be 4.13 if the value in A1 is 4.123 the result will be 4.12.

ABS

=ABS(A1) will return 5 if in cell A1 you have -5 or 5. This function removes the sign.

MOD

The modulo is the remainder left after a division.

=MOD(32,6)

is 2 because you have 5 times 6 in 30 and the remainder is 2.

SQRT and POWER

To extract the square root of a number you will use a formula like:

=SQRT(16) that will result in 4 because 4 multiplied by 4 is 16 or

=SQRT(A1) that will also result in 4 if the value in cell A1 is 16.

There is no specific Excel function to extract the cubic root or any other root, but can be achieved by using POWER function. You can raise a number to a power (multiplying it by itself a certain number of times with this function. Hence:

=POWER(4,2) will result in 16 (4 times 4) or

=POWER(A1,2) will also result in 16 if the value in cell A1 is 4.

=POWER(A1,1/2) will give the square root of the value in cell A1

=POWER(A1,1/3) will give the cubic root of the value in cell A1

Conditional Calculations

Excel has several logical functions that allow you to test cells and perform different operations depending on their contents.

IF( ) function

The **IF()** function enables you to specify two different calculations based on a certain condition:

=IF(condition, calculation if condition is true, calculation if condition is false)

If the condition specified in the first argument is true, Excel performs the calculation specified in the second argument, otherwise Excel calculates the third argument.

In the figure below, suppose you have an experimental test results that allows you to take a additional value 4. Specifically, if the experimental value 3 is higher the additional value taken, then the adjusted value will be the average of value 3 and the additional value 3. If the additional value is not higher than the value 3, then their adjusted value score is their actual Test value 3. The **IF( )** function below performs this calculation to adjust the score for the cells in column F.

`= IF(E2>D2, (D2+E2)/2, D2)` or **`= IF(D2>E2, D2, (D2+E2)/2)`**

	A	B	C	D	E	F	G
1		**Value 1**	**Value 2**	**Value 3**	**Additional Value 3**	**Adjusted Value**	**Final Value**
2	Iteration 1	88	73	82	80	82	81
3	Iteration 2	52	70	64	70	67	63
4	Iteration 3	87	73	82	80	82	81
5	Iteration 4	76	74	78	80	79	76

The first argument, E2>D2, tests whether the value 3 is better than the additional value 3. If the condition is true, the second calculation is performed (the average of the two values, (E2+D2)/2). If the condition is false, the adjusted value is simply the actual value 3, D2 in this case.

AND( ) and OR( ) functions

To specify multiple conditions within an IF( ) function, use Excel's AND( ) and OR( ) functions.

= AND(condition 1, condition 2, ...condition n)

= OR(condition 1, condition 2, ...condition n)

AND( ) returns the value of TRUE if all its conditions are true, and returns FALSE otherwise. OR( ) returns TRUE if at least one of the specified conditions is true.

Note: ISBLANK( ) function tests whether a certain cell is blank. This function returns TRUE if the cell is blank and FALSE if it's not.

The modified formula would then be:

`=IF(AND(D2>E2,NOT(ISBLANK(D2))),D2 , (D2+E2)/2)`

The first argument of the IF( ) function is a AND( ) function that tests both conditions, Improvement Test 3 is greater than Test 3 and Test 3 is not blank. If both conditions are true, then the second calculation is performed, else the calculation in the third argument is performed.

Cutting up & Piecing together Text Strings using Text functions

Function	What it Does
CONCATENATE	Joins together two or more text strings **=CONCATENATE(A2,B2)**
LEFT	Returns a specified number of characters from the start of a supplied text string **=LEFT(C2,8)**
MID	Returns a specified number of characters from the middle of a supplied text string **=MID(A2,2,2)**
RIGHT	Returns a specified number of characters from the end of a supplied text string **=RIGHT(A2,3)**
REPT	Returns a string consisting of a supplied text string, repeated a specified number of times **=REPT(A2,3)**
LEN	Returns the length of a supplied text string **=LEN(A2)**
EXACT	Tests if two supplied text strings are exactly the same and if so, returns TRUE; Otherwise, returns FALSE. **=EXACT(A2,F2)**
MATCH	Returns the relative position of an item in an array that matches the specified value in a specified order. **=MATCH(D2,$D$2:D2,0)**

Using Look up tables / References

Excel's VLOOKUP function, which stands for vertical lookup, is used to find specific information that has been stored in a spreadsheet table. VLOOKUP is a function that is used in a worksheet to return a value from a table (either in the same sheet, another sheet or another workbook) that is related to the value you give it. Say you have a parts list in a table on one sheet containing thousands of parts numbers and their related information. The table shows the part number in the first column, the part description in the second column, the parts supplier in the third column, the parts price in the fourth column and so on. On a separate sheet you have an invoice with columns for the Quantity, the Part number, the part Description, the Price and the Total.

When you enter the part number in the Part number column of the Invoice, the VLOOKUP function is used to get the details for Description column and the unit price column for the part number. Although the IF( ) function can be very useful in these circumstances, it is limited to either TRUE or FALSE outcomes. But in these types of examples with large inventories or contact list, we need a function that can handle multiple outcomes. The **VLOOKUP( )** function is ideally suited for this sort of calculation. Here's the syntax of a VLOOKUP Function:

=VLOOKUP(lookup value, lookup table range, value column)

With the **VLOOKUP( )** function (short for **vertical lookup**) you can specify lookup values for different outcomes. For example, if you have a list of parts in a worksheet, you can create a formula that assigns the part description or remarks based on the part number. To use a VLOOKUP( ) function, you must first create a lookup table with a range of values. A sample lookup table to handle parts and their description/remarks is shown below.

Part Number	Description	Suppliers	Price
D001	Nut	A1 traders Pte ltd	$2.00
D002	Bearing	XYZ Dealers	$4.00
D003	Bolt	Universal Parts Distributors	$1.50
D004	Motor Pulley	Sam Motors and suppliers	$15.00
D005	Gear	Amro Parts co. Ltd	$70.00

In this sample, the formula is

```
=VLOOKUP($D19,PartsList!$A$1:$D$11,2,0)
```

The first argument, **the lookup value**, is the value you wish to look for in the lookup table which is the part number in this example.

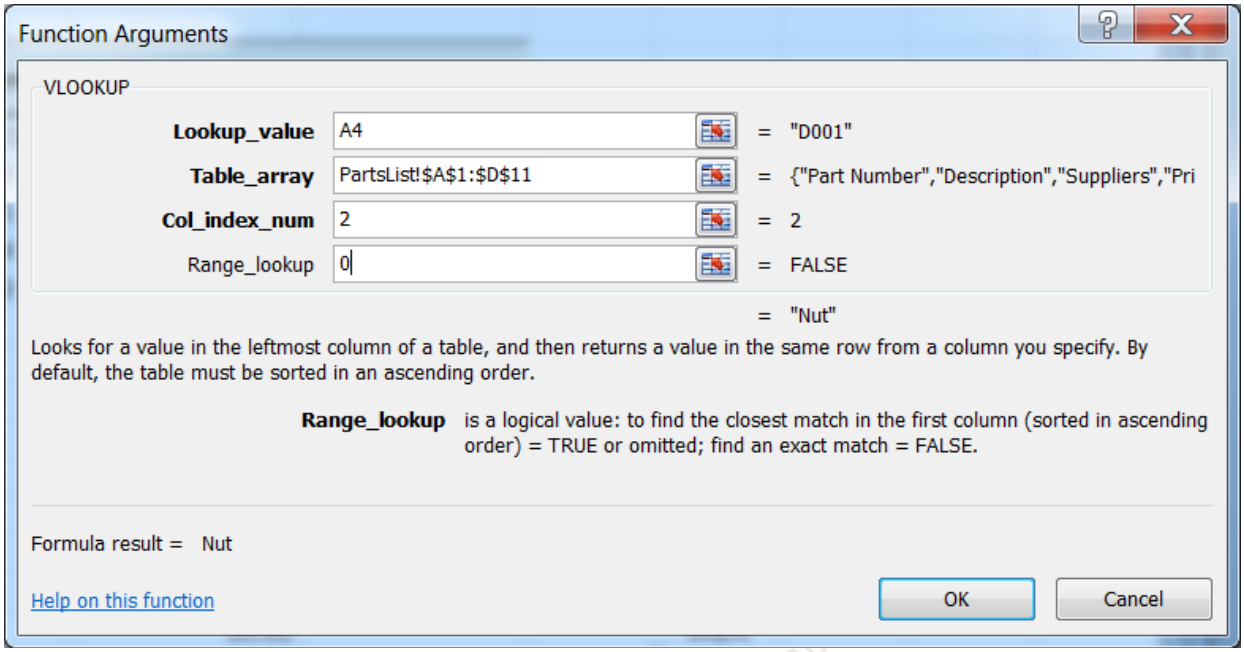
The **lookup table range** is the range on the worksheet that contains the lookup table. You can use a reference to a range or a range name. The values in the first column of the table are the values searched by lookup_value. These values can be text, numbers, or logical values.

Finally, **the value column** in argument 3 is to indicate which column of the table is to be used for the actual result. Specify the value column by indicating what numeric column of the lookup table to use. In the Parts-Description lookup table above, use column 2 if you want part description.

Range_lookup specifies whether you want VLOOKUP to find an exact match or an approximate match.

Note: The first column of a vertical lookup table must be in sorted order.

1. In the **Lookup_value** (first) argument box, select cell **A4** (the value for which the calculation is made).
2. Select the **Table_array** (second) argument box; use the separate table that has been created earlier. In our example, the lookup table is in **'PartsList'!A2:D11**. Look up table has been created in the worksheet named “PartsList” and is in the columns A2 to D11. Now adjust the formula such that the table array becomes an absolute reference, **'PartsList'!$A$2:$D$11**.
3. In the **Col_Index_num** (third) argument box, type in the number of the column in the **data** table. A **Col_index_num** of 1 returns the value in the 1st column in table_array; a **Col_index_num** of 2 returns the value in the 2nd column in table_array and so on. In this example, we are using the 2nd column, hence the value is 2
4. **Range_lookup** can be either a 0 or 1, else you could use TRUE or FALSE. By default, it is set as TRUE. However, if an exact match is needed, choose 0 or FALSE.



5. Hence, the lookup function to get details of a particular part number will be:
   =VLOOKUP(A4,Lookup!$A$2:$D$11,2,FALSE)

6. Check if the result is as expected, and then click OK.

Pivot Tables

The Pivot tables in Excel are very useful and powerful feature and can be used to summarize, analyse, explore and present your data with ease.

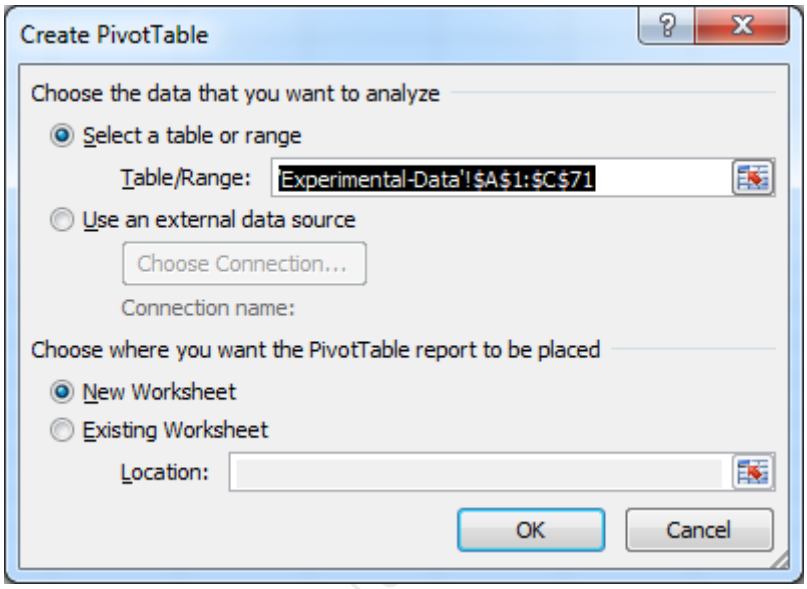
PivotTable report is especially designed for:

- Querying large amounts of data
- Sub-totalling and aggregating numeric data, i.e., summarising data by categories and subcategories, and creating custom calculations and formulas.
- Expanding and collapsing levels of data to focus your results, and drilling down to details from the summary data for areas of interest to you.
- Transposing data – moving rows to columns or columns to rows (or "pivoting") to see different summaries of the source data.
- Filtering, sorting, grouping, and conditionally formatting the most useful and interesting subset of data to enable you to focus on the information that you want, without having to write any formulas
- Presenting concise, attractive, and annotated online or printed reports.

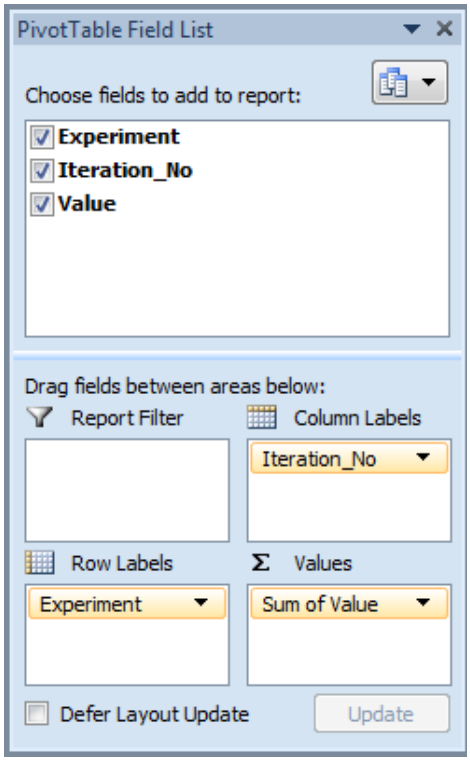
Create a PivotTable from worksheet data

When you create a PivotTable report from worksheet data, that data becomes the source data for the PivotTable report.

1. Select the range of cells that contains the data along with column headings.
2. On the **Insert** tab, in the **Tables** group, click **PivotTable**.
3. The **Create PivotTable** dialog box is open.



4. Under **Choose the data that you want to analyze**, make sure that **Select a table or range** is selected, and then in the **Table/Range** box, make sure the range of cells that you want to use is listed.
5. Under **Choose where you want the PivotTable report to be placed**, choose either the **New Worksheet** or **Existing Worksheet** and click **OK**.
6. An empty PivotTable report is added on the specified worksheet along with the PivotTable Field List from which you can select the fields that you would like to add to create a layout and customize the PivotTable report.
7. To place any field in the default area of the layout section, **select the check box** next to the field name in the field section. By default, nonnumeric fields are added to the Row Labels area, numeric fields are added to the Values area, while date/time hierarchies are added to the Column Labels area.
8. To place a field in a specific area of the layout section, you can also right-click the field name in the field section, and then select **Add to Report Filter**, **Add to Column Label**, **Add to Row Label**, or **Add to Values.**
9. You could also drag a field to the area that you want by clicking and holding the field name in the field section, and then dragging it to an area in the layout section.



Statistical functions

Excel has a wide variety of built-in statistics functions that give, for instance, the slope and y-intercept of a line, the standard deviation of a data sample, and the mean, median and mode of a set of values. Here, we will cover a few of the more useful and popular statistics functions from which you can easily obtain summary statistics directly; else you can use the Analysis Tool available from the Tools menu.

Basic Statistical built-in functions

Function	What it Does
AVERAGE(range)	Returns the average of its arguments
AVERAGEA(range)	Returns the average of its arguments, including numbers, text, and logical values
MEDIAN(range)	Returns the number in the middle of a range of data
MODE(range)	Returns the most frequently occurring or repetitive value in a range of data
COUNT(range)	Counts how many numbers are in a range of data
COUNTA(range)	Counts how many values are in a range of data
MAX(range)	Returns the maximum value of a range
MIN(range)	Returns the minimum value of a range
LARGE(range, n)	Returns the k-th largest value in a data set
SMALL(range, n)	Returns the k-th smallest value in a data set

We will use the familiar example of a class grades to illustrate the use of some of the more basic Excel functions, like AVERAGE( ), MODE( ) AND MAX( ). Assume a class's grade distribution is as follows: 3, 0, 4, 4, 4, 2, 4, 1, 4, 0, 3, 3, 1, 1, 3, 4, 2, 4, 0, 3, 3, 1, 3. These grades are based on a 4-point scale with 4=A and 0=F. Using the AVERAGE( ) function, we find the class's average (or arithmetic mean) grade is a disappointing 2.48, or a mid-C. The syntax for this common function is =AVERAGE(number1, number2, ...). However, we don't get a clear picture of the class's performance by simply looking at its average. We can further analyze the data using the MEDIAN( ) function. The median gives the middle number in a set of numbers and its syntax is =MEDIAN(number1, number2,...). When the median grade is calculated, it is 3.0, meaning that half of the grades are higher than 3.0, and half are lower. Therefore, despite the low class average, more students scored 3's and 4's than 2's, 1's and 0's.

	A	B	C	D	E	F	G
1	First Name	Last Name	Class Grades				
2	Caroline	Tan	3				
3	Don	Poole	0		Average	2.48	AVERAGE (C2:C26)
4	Sam	Thangvelu	4		Median	3	MEDIAN(C2:C26)
5	Carmel	Lim	4		Mode	3	MODE(C2:C26)
6	Daphne	Liu	4				
7	Victor	Tang	2		Maximum Grade	4	MAX(C2:C26)
8	Dylan	Tiong	4		Minimum Grade	0	MIN(C2:C26)
9	Shirley	Williams	1				
10	Glen	Wright	4		Number of students	25	COUNT(C2:C26)
11	Linda	Tan	0				

Additionally, we can also analyze the grade distribution by using the MODE( ) function. The mode gives the most frequently occurring value of a set of numbers and its syntax is =MODE(number1,

number2,...). Excel's built-in functions to determine the number of grades entered, and the maximum and minimum grades of the distribution.

AVERAGE and AVERAGEA

The difference between AVERAGE and AVERAGEA becomes evident when one of the cells contains a text OR A SPACE and don't forget the SPACE. A cell containing a space is NOT empty.

COUNT and COUNTA

If you want to count the number of cells that are not blank COUNT and COUNTA will return a different result if in one of the cells there is **a text** or **a space**

=COUNT(B2:B7) will return 6 is only numbers are present in cells B2 to B7 and 5 if there is a letter, an empty cell OR A SPACE in one of the cells.

=COUNTA(B2:B7) will return 6 unless one of the cells is empty. If all the cells contain numbers, letters OR SPACES the result will be 6.

LARGE and SMALL

The MAX and MIN functions would give the largest and smallest value from a list of values. But what if you want the second or third largest value or the second smallest value, use LARGE and SMALL functions as follows:

=LARGE(A1:A5,2), =LARGE(A1:A5,3), =SMALL(A1:A5,2)

As a matter of factsyou can also get the MIN and MAX values using these functions.

=LARGE(A1:A5,1), =SMALL(A1:A5,1)

Linear regression functions

Excel has some built-in functions that allow a method for determining the slope, y-intercept, correlation coefficient, and R-squared values of a set of data. The functions are SLOPE(), INTERCEPT(), and CORREL(). These functions are easier and faster to compute than plotting the data. However, a visual graph shows trends in the data better than any other tool.

Function	What it Does
SLOPE	Returns the slope of the regression line through the given data points =SLOPE(y cell range, x cell range) **=SLOPE(C2:C6,A2:A6)**
INTERCEPT	Calculates the point at which a line will intersect the y-axis using a best-fit regression line plotted through the known x values and y values =INTERCEPT(y cell range, x cell range) **=INTERCEPT(C2:C6,A2:A6)**
CORREL	Return the correlation coefficient between two data sets. =CORREL(y cell range, x cell range)

If we consider the example of a car in motion that is coming to a stop and are required to determine its acceleration and initial velocity. Then, we can determine the car's acceleration and its initial velocity with the help of the **SLOPE( )** and **INTERCEPT( )** functions. Hence, the y-axis values represent the square of the car's velocity and the x-axis values represent the car's position or distance travelled. In order to find the acceleration, we divide the slope by 2 and to find the initial velocity, we take the square root of the y-intercept.

Distance travelled (in m)	Velocity (in m/s)	Square of Velocity in m^2/s^2
2.00	6.90	47.61
4.00	6.00	36.00
6.00	4.90	24.01
8.00	3.40	11.56
10.00	0.00	0.00

The CORREL( ) function can then be used to ensure that the data displays a linear trend. It is always a good idea to plot the data as well as use these statistics functions because sometimes trends are not obvious. Additionally, a plot of the data allows us to visualize the data and gross blunders and errant data points are easily detected.

STDEV

For error analysis, we use the STDEV function. When we carry out a number of repetitive measurements of one quantity, we find the average value. This does not however tell us anything of the precision of our measurement. The standard deviation of the measured values will give a measure of the precision. To quickly determine the standard deviation of any measurement, use Excel's built-in STDEV( ) function.

=STDEV(A2:A6)

Statistical Analysis

Microsoft Excel has numerous Add-in features that support statistical analysis. Statistical data analysis in Excel is not recommended for analysing datasets with a large sample size or a large number of variables, performing advanced statistical analyses, or for projects in which a number of procedures need to be performed. Excel is a useful tool for answering basic analysis. The primary reason to use Excel for statistical data analysis is because it is so widely available.

There are a number of disadvantages worth considering before using Excel for statistical analysis:

- Missing values are handled inconsistently, and sometimes incorrectly.
- Data has to be organised differently according to the analysis you wish to perform.
- Most analyses can only be done on one column at a time. This makes it inconvenient to do the same analysis on many columns.
- There is no log or record of how an analysis was accomplished.
- It also lacks many important features for advanced analyses.

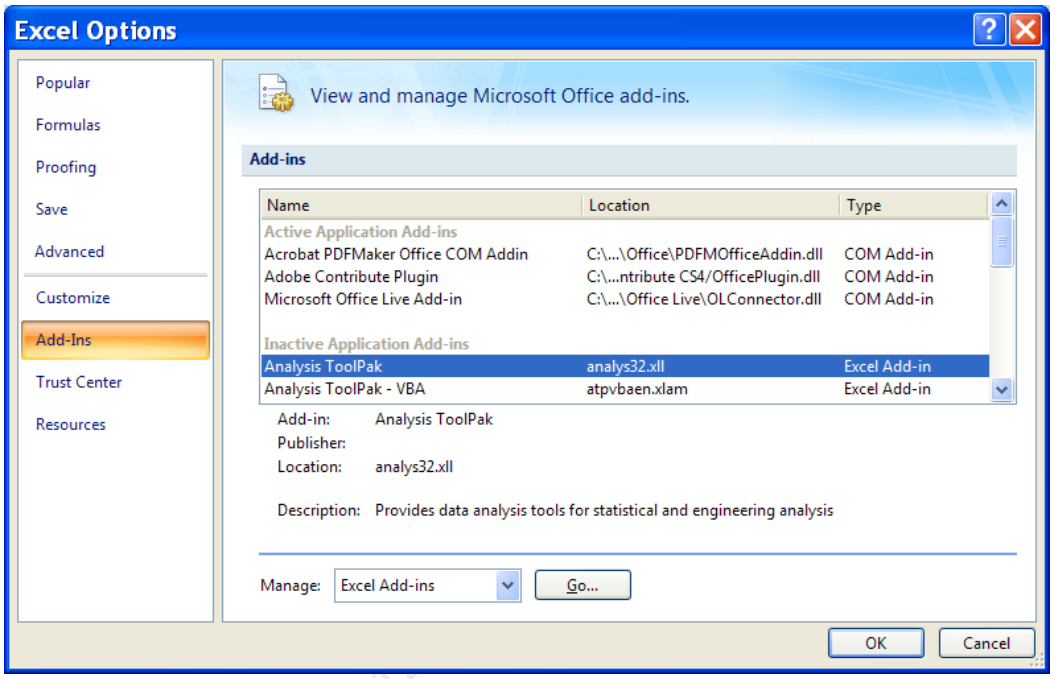
Using the Analysis ToolPak

Two specific tools are useful in generating descriptive statistics and histograms of grade distributions. To access these features, the Analysis ToolPak must be loaded.

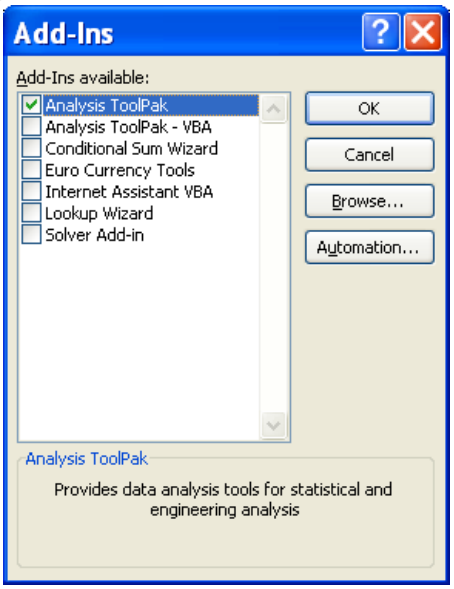
Check to see if the **Data Analysis** command is available in the **Analysis** group on the **Data** tab. If the Data Analysis menu is available, then it indicates that the Analysis ToolPak is already loaded.

If the menu is not visible, then

1. Choose **Add-Ins** from the Excel Options found through the Microsoft Office Button.



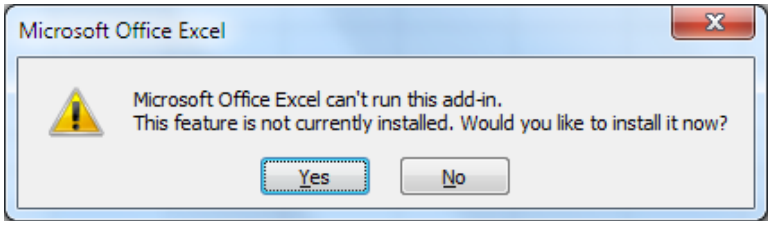
2. In the **Manage** drop-down menu, select **Excel Add-ins**, and click **Go**.



3. In the **Add-Ins available** box, select the **Analysis ToolPak** check box, and then click **OK**.

Kiruthika Ragupathi, email: kiruthika@nus.edu.sg
Centre for Development of Teaching and Learning

4. If you get prompted that the Analysis ToolPak is not currently installed on your computer, click **Yes** to install it.



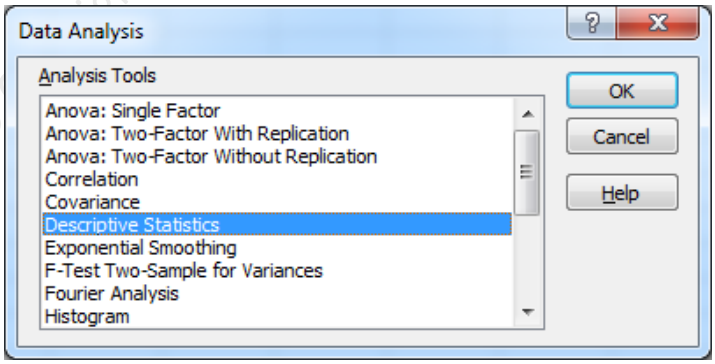
5. A configuration progress screen for Microsoft Office appears and once completed, the **Data Analysis** menu should appear on the **Data** menu in the **Analysis** group.

Descriptive Statistics

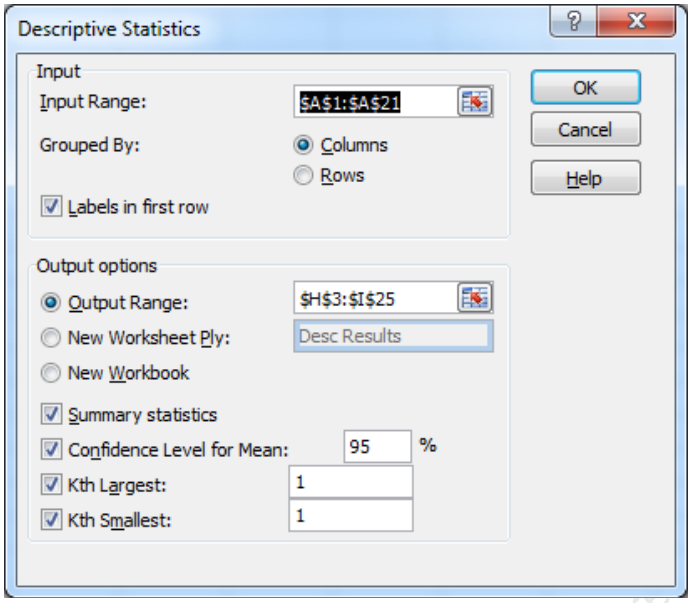
The quickest way to get mean and standard deviation for an entire group is by using the Descriptive Statistics in the Data Analysis tools. This generates simple descriptive statistics like average, median, and standard deviation for a collection of data. You can choose several adjacent columns for the Input Range and each column is analysed separately. The labels in the first row are used to label the output, and the empty cells are ignored. If there are more non-adjacent columns to be analysed, then one has to repeat the process for each group of adjacent columns. The procedure is straightforward, can manage many columns reasonably efficiently, and empty cells are treated properly. To generate these statistics,

1. From the **Data menu** in the **Analysis** group, select **Data Analysis**.

2. The **Data Analysis** dialog box opens.



3. In the **Data Analysis dialog box**, choose **Descriptive Statistics**.
4. In the **Descriptive Statistics** dialog box (see below), specify the cells that contain your data in the Input Range box. Select the **Summary Statistics** checkbox in the lower left corner.



5. By default, the statistics is generated on a new worksheet. If you want the statistics to appear on the same worksheet, click the **Output Range** button and specify a destination cell for the statistics. You can also specify a name for the new worksheet.

Sample output from the Descriptive Statistics tool appears as shown in the table below.

Measured Values	
Mean	343.491
Standard Error	0.666719
Median	344
Mode	344
Standard Deviation	2.981658
Sample Variance	8.890283
Kurtosis	1.775491
Skewness	-1.22725
Range	12.17
Minimum	335.53
Maximum	347.7
Sum	6869.82
Count	20
Largest(1)	347.7
Smallest(1)	335.53
Confidence Level (95.0%)	1.395459

Note:
Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point. Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution.

More Information

Excel Help

Excel has a Help system built into the software. You can click on at any time and browse through the topics for your specific topic or can search for it using the search option. To get help with a specific procedure, click on the Help on this function in the Dialog box for the procedure you would like to run.

Office online

Microsoft also offers a very comprehensive online help at the Microsoft office - Help for Excel 2007, http://office.microsoft.com/en-us/excel/FX100646951033.aspx?CTT=96&Origin=CL100570551033