Formulas & Functions in Microsoft Excel

Theresa A Scott, MS
Biostatistician III
Department of Biostatistics
Vanderbilt University

theresa.scott@vanderbilt.edu

Table of Contents

Preface

All of the lecture notes and supplementary sample data files are located at `http://biostat.mc.vanderbilt.edu/TheresaScott` under *Current Teaching Material*.

If you have any questions, feel free to contact me at theresa.scott@vanderbilt.edu or (615) 343-1713, or drop by my office D-2217 MCN.

References for this lecture

The following references were used to compile this lecture:

- The Excel 2003 Module information available from Carnegie Mellon University's Computer Skills Workshop Class (`www.cmu.edu/computing/csw`).

- The BayCon Group Microsoft Excel Online Tutorial (`www.baycongroup.com/el0.htm`).

- The Florida Gulf Coast University Excel 2000 Tutorial (`www.fgcu.edu/support/office2000/excel`).

Section 1

Introduction

The distinguishing feature of a spreadsheet program such as Excel is that it allows you to create mathematical *formulas* and execute *functions*. Otherwise, it is not much more than a large table displaying text.

- Recall, your spreadsheet environment will become a *dynamic* and *responsive* work environment when you use formulas and functions; they automatically updating when you change your data.

The following should be considered when creating a spreadsheet:

- Which terms are data/numbers that you will type.

- What terms are data/numbers that you will calculate from the data.

- Which numbers are constant and are to be used in a variety of calculations.

- What arrangement of the columns and rows will make it easiest for you and your intended audience to work with the spreadsheet.

  - Feel free to check out an additional lecture called "*Guidelines to Data Collection and Data Entry*" that can also be found on my website.
  - You can also find examples of a '*Spreadsheet from Heaven*' and a '*Spreadsheet from Hell*'.

1.1 Using Excel for Analysis

Excel is a great tool to use for data collection and entry, and even to use for some derivation of other columns. However, Excel **IS NOT** the best tool to use to conduct advanced analyses, especially statistical analyses.

Missing values can be very dangerous in Excel.

- In formulas and functions, missing values (i.e. blank cells) are sometimes taken as zeros, when they should represent data that are truly missing.

- Recommend consulting with an experienced statistician when wanting to conduct advanced and/or statistical analysis, or use packages like SPSS, STATA, R, or SAS.

- *Tip*: Be aware of Excel's *faults* and, if possible, set up a simple example to test the function's handling of blank cells.

Section 2

Formulas and Functions

As mentioned, the ability to perform calculations is one of the purposes of using a spreadsheet application. Some examples of the types of calculations that can be done are:

- totals

- subtotals

- average

- standard deviation

In Excel, the calculation can be specified using either a *formula* or a *function*.

- **Formulas** are *self-defined* instructions for performing calculations.

- In contrast, **functions** are *pre-defined* formulas that come with Excel.

In either case, all formulas and functions are entered in a cell and must begin with an equal sign '='.

2.1   Entering Formulas

After the equal sign, a formula includes the addresses of the cells whose values will be manipulated with appropriate *operands* placed in between. The operands are the *standard arithmetic operators*:

Operator	Meaning	Example
+	Addition	=A7+A9
-	Subtraction	=A7-A9
*	Multiplication	=A7*A9
/	Division	=A7/A9
^	Exponents	=A7^A9

→ *Practice Exercise:* Enter the following information into a blank worksheet (ignore any formatting) in columns A, B, and C, and in rows 1 through 6. Then calculate the Total Cost for the the Fall semester using a *formula* to add up the individual Costs (Tuition, Housing, etc.).

Costs	Fall	Spring
Tuition	10000	10000
Housing	5000	5000
Books	1000	700
Spending	1500	1000
Total Cost		

- The most logical solution would be to type the formula '=B2+B3+B4+B5' into cell B6.

You can also enter formulas by using the *point mode*, where you either click on a cell with your left mouse button or you use the arrow keys.

- To enter '=B2+B3+B4+B5' into cell B6 using the point:
  - Left click on cell B6 to make it active.
  - Type '='.
  - Use the up arrow key to move to cell B5, or left click on cell B5.
  - Type '+'.
  - Use the up arrow or mouse to add cells B4, B3, and B2 in the same fashion.
  - Press Enter when you are finished entering the formula.

You'll notice that the calculation executes immediately after the formula is typed into the cell and entered by pressing either the Enter or Tab key or by clicking the check mark in the formula bar.

- The *result* of the calculation is displayed in the cell (i.e., in B6).

- The formula itself is now visible in the formula bar.

What happens if you enter 'B2+B3+B4+B5' without the leading equal sign '=' into cell B6?

- *REMEMBER:* All formulas and functions must begin with a '='.

- The text of the formula or function will be displayed in the cell if you do not use an '=' and the calculation will not be executed.

IMPORTANT: Why use cell references (i.e., cell addresses; e.g., B2 or C5) in formulas instead of the actual values of the cells (e.g., 10000 or 700)? The answer: *automatic calculation.* Let's illustrate the concept of automatic calculation with a practice exercise:

$\rightarrow$ *Practice Exercise:* Let's make a change to the Costs for the Fall and Spring semester spreadsheet and note how Excel automatically recalculates the formulas and functions we have already entered.

- Change the amount entered in cell C2 from '10000' to '15000'.

- Notice how all the calculations referencing cell C2 automatically changed.

To reiterate the use of cell addresses in formulas and functions, imagine we had constructed our formulas and functions by typing the actual numbers contained in the cells instead of the cell addresses. That is, to calculate the Total Cost for the Fall semester we had entered the formula '=10000+5000+1000+1500'. What would happen if we changed the amount entered in cell C2 from '10000' to '15000' as suggested above? Obviously, our calculation defined by our formula would not automatically change, and we would have to edit the formula by hand. Imagine if the cell we modified was referenced by ten separate formulas or functions. That would be a lot of editing.

In summary, *automatic calculation* is activated, Excel automatically recalculates the result of any formulas as cell entries change. Automatic calculation also applies to *functions*, which we'll be covering in the next section.

Side note: You can check to make sure automatic calculation is turned on:

- Select Tools from the Options drop menu.

- Click on the Calculation tab.

- Select Automatic if it is not already selected.

- Click OK.

2.2 Entering Functions

Functions differ from regular formulas in that, after the equal sign, you supply the cell addresses but not the arithmetic operators. Functions perform calculations by using specific values, called *arguments*, in a particular order called *syntax*. When using a function, remember the following:

- Use an equal sign to begin the function.

- Specify the function name.

- Enclose all of the function's arguments within parentheses.

- Use a comma to separate the function's individual arguments.

→ *Practice Exercise:* In the Costs for the Fall and Spring semester spreadsheet, calculate the Average Cost of Tuition for the Fall and Spring semesters using the AVERAGE *function.*

- The most logical solution would be to type '=AVERAGE(B2, C2)' into cell D2.

  - The equal sign begins the function.
  - AVERAGE is the name of the function.
  - B2, and C2 are the arguments.
  - Parentheses enclose all of the arguments.
  - Commas separate each of the arguments.

2.2.1 Reference Operators

Reference operators refer to a cell or a group of cells. There are two main types of reference operators:

1. **Range:**

   - Refers to all of the cells between and including the reference.
   - Consists of two cell addresses separated by a colon.
   - *EXAMPLE:* 'A1:A3' includes cells A1, A2, and A3.
   - *EXAMPLE:* 'A1:C3' includes cells A1, A2, A3, B1, B2, B3, C1, C2, and C3.

2. **Union:**

   - Includes two or more references.
   - Consists of two or more cell addresses separated by a comma.
   - *EXAMPLE:* 'A7, B8, C9' refers to cells A7, B8, and C9.
   - *EXAMPLE:* 'A7, B8:D9, E4' refers to cells A7, B8, B9, C8, C9, D8, D9 and E4.

In the function practice exercise, we used a union reference when we typed '=AVERAGE(B2, C2)' in cell D2, but we could have used a range reference instead: '=AVERAGE(B2:C2)'.

2.2.2 Autosum

You can use the Autosum icon on the standard toolbar, which automatically adds the contents of a cluster of adjacent cells.

- Select the cell that the sum will appear in that is outside the cluster of cells whose values will be added.

- Click the Autosum button (Greek letter sigma, Σ).

- Highlight the group of cells that will be summed.

- Press the Enter key on the keyboard or click the green check mark on the formula bar.

$\rightarrow$ *Practice Exercise:* In the Costs for the Fall and Spring semester spreadsheet, calculate the Total Cost for the the Spring semester using the Autosum icon.

- Click on cell C6 to activate it.

- Click the Autosum button.

- Highlight cells C2 through C5.

- Press Enter.

2.2.3 Function Wizard

You can access all of the available functions in Excel using the Function Wizard.

- Select the cell where the function will be placed and click the Function Wizard button on the standard toolbar.

- Other ways of starting the Function Wizard are:

  - Select Function from the Insert drop menu.
  - Click on the drop down arrow next to the Autosum icon button.
    * You will first see the commonly used functions in Excel, and at the bottom of the menu, the More Functions option.
    * Clicking on More Functions will give you an alphabetical and categorical listing of all available functions in Excel.

From the Paste Function dialog box, browse through the functions by clicking in the Function category menu on the left and select from the Function name choices on the right.

- As each function name is highlighted a description and example of use is provided below the two boxes.

Clicking OK to select the function, opens the Function Arguments dialog box, which allows you to choose the cells that will be included in the function.

- As in the last previous function, we can type the cell addresses in the necessary argument boxes.

- We can also enter the cells using the point mode (i.e., the left mouse button).

  - Click and drag across a group of cells to enter a range of cells addresses.
  - Use the Ctrl key and mouse to enter a union of cells addresses.

- Excel may automatically select cells for you, but you can delete those selected cells from the argument box and enter the desired cell addresses.

Click OK when all the cells for the function have been entered into the necessary argument boxes.

→ *Practice Exercise:* In the Costs for the Fall and Spring semester spreadsheet, calculate the average Housing cost for the Fall and Spring semesters using the Function Wizard and the AVERAGE function.

- Click on cell D3 to activate it.

- Activate the Function Wizard.

- Choose the AVERAGE function from the 'Most Recently Used' Function category.

- Using the mouse, highlight cells B3 and C3, or type 'B3:C3' in the Number1 argument box.

- Click OK to execute the calculation.

All other functions can be entered and executed in a similar fashion.

2.3 A Function instead of a Formula

Functions can be a more efficient way of performing mathematical operations than formulas. Specifically, in many cases, a function will simplify formulas that you can type in manually, such as average or sum.

- *EXAMPLE:* If you wanted to add the values of cells D1 through D10, you could type the formula '=D1+D2+D3+D4+D5+D6+D7+D8+D9+D10', or a shorter way would be to use the SUM function and simply type '=SUM(D1:D10)'.

The following table contains the definitions and examples of several other available functions:

Function	Example	Description
SUM	=SUM(A1:A100)	Finds the sum of cells A1 through A100.
AVERAGE	=AVERAGE(B1:B10)	Finds the average of the cells B1 through B10.
MAX	=MAX(C1:C100)	Returns the highest number from cells C1 through C100.
MIN	=MIN(D1:D100)	Returns the lowest number from cells D1 through D100.
SQRT	=SQRT(D10)	Finds the square root of the value in D10.
TODAY	=TODAY()	Returns the current date (leave the parentheses empty).

Section 3

Copying Formulas/Functions and Cell References

As we've seen, the best way to construct a formula or a function is to use *cell references* (i.e., use the cell addresses) instead of typing actual numbers. In turn, this enables Excel to automatically update the results of the formulas and functions when you change the values in the cells referenced.

Using cell references in formulas and functions is also important when you *copy* a formula or function. When you copy the formula or function, the cell addresses will be copied in a particular way depending on how the cells are referenced. Specifically, Excel records cell addresses in formulas and functions in three different ways:

- Relative cell reference.

- Absolute cell reference.

- Mixed cell reference.

As we saw in our first lecture, several different ways exist to copy and paste a formula or a function:

- Use Copy and Paste from the Edit drop menu.

- Use the Ctrl+C and Ctrl+P keyboard shortcuts.

- Use the Autofill feature – activate the cell containing the formula or function by selecting it and drag the Autofill handle down the column or across the row to fill in the remaining cells.

Copying formulas and functions instead of typing them into a new cell obviously avoids typing errors and will save you time.

3.1   Relative Cell References

Calling cells by just their column and row labels (such as 'A1') is called a *relative cell reference.* When formulas or functions contain relative cell referencing and they are copied from one cell to another, Excel does not create an exact copy of the formula or function. Instead, Excel changes the cell addresses relative to the row and column they are moved to.

- *EXAMPLE:* If a simple addition formula of '=A1+B1' in cell C1 is copied to cell C2, the formula would change to '=A2+B2' to reflect the new row.

- *EXAMPLE*: If a simple average function of '=AVERAGE(B2:B5)' in cell B6 is copied to cell C6, the function would change to '=AVERAGE(C2:C5)' to reflect the new column.

→ *Practice Exercise:* Let's work with a new data example. Enter the following information into a blank worksheet (ignore any formatting) in columns A, B, and C, and in rows 1 through 5. Then define a formula in cell D2 to calculate the Cost of the group of Biology Textbooks, and copy the formula into cells D3 through D5.

Textbook	Quantity	Price
Biology	4	99.99
Chemistry	2	79.95
Calculus	7	65.99
English	12	49.99

- The most logical solution would be to type the formula '=B2*C2' into cell D2. We may also want to add the column label of Cost into cell D1. We can then use one of the methods mentioned above to copy and paste the formula into the desired cells.

Compare the formulas in cells D2 through D5 by looking in the Formula bar. You'll notice that the formulas are the same except that the row numbers were automatically changed (from 2 through 5) because of the relative cell referencing.

→ *Practice Exercise:* In the Costs for the Fall and Spring semester spreadsheet, copy the AVERAGE function in cell D3, which calculates the average Housing cost for the Fall and Spring semesters, to cell D4 through D6. Compare the copied functions by looking in the Formula bar.

3.2   Absolute Cell References

An *absolute cell reference* refers to the same cell, no matter where the formula or function is copied. In other words, when a formula or function containing an absolute cell reference is copied to a new location, the cell reference is not adjusted.

- To create an absolute cell reference, you will need to add dollar signs ('$') in front of both the column and row identifiers for the cell referenced – this fixes the row *AND* column.
  - You can add the dollar signs ('$') automatically by using the F4 keyboard shortcut – press the F4 key once after entering the cell address into the formula or function by typing or by using the point mode.
- *EXAMPLE*: When the formula '=B7*C1' is copied from cell C7 to cell C8, the relative cell reference changes 'B7' to 'B8', but the 'C1' absolute cell reference remains unchanged.

An absolute cell reference is most often used when you want to use a *constant* in a formula or function.

- *Advantage of placing the constant in its own cell and using an absolute cell reference*: Referencing the cell containing the constant in all the desired formulas and functions with an absolute cell reference allows you to change the value of the constant in the one cell and all the formulas and functions absolutely referencing the cell will be automatically updated. This is much better than typing the actual value of the constant into the formulas or functions.

→ *Practice Exercise:* The following exercise demonstrates what happens when you don't use a absolute cell reference when you should.

1. Create the following spreadsheet in columns A through F and rows 1 through 4 (ignore any formatting).

Subject	Trial 1	Trial 2	Total	Variance	CONSTANT
A	500	650			0.03
B	50	567			
C	670	780			

2. In cell D2, calculate the Total of Trial 1 and Trial 2 for subject A using the SUM function.

3. Use the Autofill feature to copy the function for subjects B and C.

4. In cell E2 enter the formulas '=D2*F2' either by typing or by using the point mode.

5. Use the Autofill feature to copy the formula for subjects B and C.

   - *NOTE:* The results of the formula for subjects B and C are incorrect. This is because the formula referenced the CONSTANT using a *relative* cell reference.

6. Edit the formula in cell E2 to reference the CONSTANT using an *absolute* cell reference: '=D2*F2'.

7. Use the Autofill feature to copy the *correct* formula for subjects B and C.

   - *NOTE:* The results of the formula for subjects B and C are now correct. The Total value adjusts because of the relative cell reference, but the CONSTANT value does not because of the absolute cell reference.

3.3 Mixed Cell References

You use a *mixed cell reference* to reference a cell that is part absolute and part relative. In other words, a mixed cell reference can be used when you wish to fix only the row *OR* the column.

- *EXAMPLE:* In the formula '=A$1 + $B2', the row of cell A1 is fixed and the column of cell B2 is fixed.

As with absolute cell referencing, you need to add a dollar sign ('$'), but only in front of the column *OR* row identifier for the cell referenced.

As before, you can add the dollar sign ('$') automatically by using the F4 keyboard shortcut:

- In the active cell, type '='.

- Enter a cell address either by typing or using the point mode.

- Press F4 once – two dollar signs ('$') are entered (both the column and row identifiers are absolute).

- Press F4 again – the column identified is now relative and the row identifier is now absolute.

- Press F4 again – the column identified is now absolute and the row identifier is now relative.

- Press the Escape key to deactivate the F4 key.

Section 4

Other Topics to Consider

4.1 Order of Operations

As you can imagine you can define a complex formula by hand or by combining multiple functions into a complex formula using the mentioned standard arithmetic operators.

When you perform calculations in Excel, be careful of precedence. Excel follows the standard *order of operations* when it performs all calculations:

1. *Parentheses*: Computations enclosed in parentheses are performed first.

2. *Exponents*: Computations involving exponents are performed next.

3. *Multiplication and Division*: Performed in the order which they occur (from left to right).

4. *Addition and Subtraction*: Performed in the order which they occur (left to right).

In general, calculations are performed from left to right, with multiplication and divisions performed before addition and subtraction.

→ *Practice Exercise:* Let's enter a dummy formula into cell A8 in the Costs for the Fall and Spring semester spreadsheet.

- Type '=3+3+12/2*4' into cell A8.

  – Excel divides 12 by 2, multiplies the answer by 4, adds 3, and then adds another 3, which results in 30.

- Edit the formula in cell A8 to '=(3+3+12)/2*4'.

  – Excel adds 3 plus 3 plus 12, divides the answer by 2 and then multiplies the result by 4, which results in 36.

4.2 Using Named Ranges Formulas and Functions

When you are working with a large spreadsheet you can name a range of cells and use the name in formulas or functions.

- To name a range, select the cells to be included in the range, and then type a name in the Name Box located above the worksheet.

- Recall, the Name Box normally contains the address if the current cell.

→ *Practice Exercise:* In the Costs for the Fall and Spring semester spreadsheet:

1. Delete the formula calculating the Total Cost for the Fall semester in cell B6.

2. Highlight the cells in the Fall semester column.

3. Type 'Fall' in the Name box and press Enter – the range of numbers is now identified by the name Fall.

4. Type '=SUM(Fall)' in cell B6.

4.3 Linking Worksheets

You may want to use the value from a cell in another worksheet within the same workbook in a formula.

- Use the format 'sheetname!celladdress'.

- *EXAMPLE*: The value of cell A1 in the current worksheet and cell A2 in the second worksheet can be added using the formula '=A1+Sheet2!A2'.