Data Manipulation with Excel

Seminar for the MBA students at the



August 2014



Table of Contents

Section	Page
Important Settings	2
Data Manipulation	
Basic Tools	6
Functions	7
Auto Filter	8
Subtotal Function	10
Advanced Filter	11
VLookup Function	14
Absolute References	15
Conditional Formatting	16
Text Functions	17
Paste Special	18
Grouping Rows and Columns	19
Database Functions	20



Important Settings



Setting Up Your System

 The Options menu in Excel contains many important settings that control the operation of the spreadsheets

```
        Excel 2003
        Excel 2010

        ➤ Tools ➤ Options
        ➤ File [F] ➤ Options [T]
```

One important setting is as follows:

Editing Options

- Turn off "Allow Editing Directly in Cell"
- When this selection is "off" and you edit a cell (F2 key), you will be prompted to edit the formula in the formula bar at the top of the screen. This is optimal.
- Alternatively, when this selection is "on" and you edit a cell, you will be prompted to edit the formula directly in the cell. This can be problematic because the formula may cover up numerous other cells on the screen that you want to see while editing the formula.
- The other benefit from keeping this selection "off" is that if you double click on a cell that contains a formula, the cursor will jump to the precedent cells (it will only jump to the first cell reference in the formula if the first cell in the formula is on a different worksheet).



Setting Up Your System (Cont'd)

Some other important settings are as follows:

Calculation Options

- Check the "Automatic" box in the Calculation section.
- This setting ensures that all formulas calculate automatically.
- If this option is set to "Manual", nothing will happen when you enter a formula until you press the F9 key to manually recalculate the spreadsheet.
- You may want to set this option to "Manual" if you are working on an extremely large file because it could take Excel a few seconds to recalculate the spreadsheet every time you make a change.
- Turn off the "Iteration" check box (the "Iteration" check box should only be turned on when you are working on a spreadsheet that contains an intentional circular reference).
- The settings on the "Excel Options" menu affect all spreadsheets that are currently open. Therefore, if a setting is changed, the change will take place on all the Excel files that are open.



Data Manipulation



Basic Tools

 The following tools are fairly simple but can be very helpful when manipulating data.

Freezing Panes

Excel 2003 ➤ Window ➤ Freeze Panes Excel 2010 ➤ View [W] ➤ Freeze Panes [F]

- This tool allows you to freeze rows and/or columns in a spreadsheet.
- The key to freezing panes is to properly locate the cursor prior to running this tool. All cells on *top* and to the *left* of the cursor will be frozen.

Sorting Data

- The Sorting Tool is an important, albeit overused tool.
- Many people sort their data numerous times when there are often more effective ways to extract the desired output.

```
Excel 2003

> Data > Sort

Data > Sort [SS]

> Home [H] > Sort & Filter [S]
```

- Select any cell within your data and then run the Sort tool.
- The sort tool can also be found on the Home tab



Functions

• The following functions are particularly helpful for use in manipulating data:

Function Table

Function	Description
= Average(A1:A10)	Calculates the average of a set of data
= Count(A1:A10)	Counts the number of cells in a range that contain numbers
= CountA(A1:A10)	Counts the number of cells in a range that are not empty
= Countif(A1:A10,">10")	Counts the number of cells that meet a certain condition
= Large(A1:A10,n)	Returns the nth largest value in a data set
= Max(A1:A10)	Finds the maximum value in a set of data
= Median(A1:A10)	Finds the median value in a set of data
= Min(A1:A10)	Finds the minimum value in a set of data
= Mround(A1,5)	Returns a number rounded to the desired multiple
= Product(A1:A10)	Multiplies a range of data
= Rand()	Returns a random number between 0 and 1
= Randbetween(bottom,top)	Returns a random number between the numbers specified
= Round(A1,2)	Rounds a selected cell to a specified number of digits
= Small(A1:A10,n)	Returns the nth smallest value in a data set
= Sum(A1:A10)	Sums a range of data
= Sumproduct(A1:A10,B1:B10)	Sums the Products of pairs of data
= Sumif(A1:A10,">10")	Sums a set of data that meet selected criteria



Auto Filter

- Excel has two filter tools that allow a user to extract a subset of information from a range of data.
- The simpler of these two filter tools is the AutoFilter.

Excel 2003

➤ Data ➤ Filter ➤ AutoFilter

Excel 2010

➤ Data [A] ➤ Filter [T]

- The AutoFilter allows a user to filter data based on selected criteria
- The raw data needs to include column headings because the AutoFilter inserts drop down menus beside each heading
- Use the Alt + ♥ keys to open the drop down menus with the keyboard

Excel 2003

Choose the "Custom" option within the filter pull-down menu to customize the criteria

Excel 2010

Within the filter drop-down menu, you will have an option to customize the criteria

- When a Filter is on, the triangle on the drop down button will turn blue
- The row numbers on the left side of the screen will also turn blue when a filter is on
- To reset the filters:

Excel 2003

▶ Data ➤ Filter ➤ Show All

Excel 2010

➤Data [A] ➤ Clear [C]



Auto Filter (Cont'd)

• The AutoFilter is a very simple tool. In fact, it is often too simple to create the required subset of information.

Common Pitfalls

- There are two common pitfalls with the AutoFilter tool:
 - 1. The Data needs to be filtered in place. Every time the drop down menus are changed, the filtered subset will change. When using an AutoFilter, if you want to create a subset of data and then keep the subset, you will need to manually copy the subset elsewhere.
 - 2. The AutoFilter cannot run sophisticated criteria.
- If either of the two pitfalls above becomes an impediment to using the AutoFilter tool, the solution is to use Excel's Advanced Filter (see subsequent pages).



Subtotal Function

 When using an AutoFilter, the only way to perform mathematical calculations on the subset of data is to use Excel's Subtotal function.

= SUBTOTAL(Function_Num,Reference)

- The Subtotal function is necessary when using an AutoFilter because it only performs operations on the visible cells
- The functions in the table below are the only options available. If a different calculation is required, use the Advanced Filter to filter the data.

Subtotal Function Table

Function_Num	Function	Description
1	Average	Calculates the average of the subset of data
2	Count	Counts the number of cells in the subset that contain numbers
3	CountA	Counts the number of cells in the subset that contain text OR numbers
4	Max	Finds and returns the maximum value in the subset of data
5	Min	Finds and returns the minimum value in the subset of data
6	Product	Multiplies the cells in the subset of data
7	Stdev	Estimates the standard deviation based on a sample
8	StdevP	Calculates the standard deviation based on the entire population
9	Sum	Sums the subset of data
10	Var	Estimates the variance based on a sample
11	VarP	Calculates the variance based on the entire population



Advanced Filter

- If you need a subset of data and you discover that the AutoFilter is too restrictive, you will need to use Excel's Advanced Filter (see discussion of pitfalls on AutoFilter pages).
- The Advanced Filter is a very important tool because it allows you to:
 - filter data based on more elaborate criteria than the AutoFilter; and
 - place the filtered subset in a new location to keep the original data intact.
- The steps to set up an Advanced Filter are as follows:
 - 1. Ensure that the raw data has column headings;
 - 2. Add a few blank rows on top the raw data;
 - 3. Copy the column headings to a blank row a few rows above the raw data. This will be the criteria section for the Advanced Filter;
 - 4. Within the Criteria section, place criteria on top of one another to use the *OR* logical operator. Place criteria beside one another to use the *AND* logical operator. Each row within the Criteria section is a separate request to extract rows from the raw data (see example on subsequent pages).



Advanced Filter (Cont'd)

The steps to run an Advanced Filter are as follows:

Excel 2003 ➤ Data ➤ Filter ➤ Advanced Filter Excel 2010 ➤ Data [A] ➤ Advanced Filter [Q]

In the menu box that appears, fill in the following information:

- Check Box: Choose whether you want to filter the data in place or copy the subset to another location.
- List Range: Select the entire range of raw data, *including* the column headings.
- Criteria Range: Select the few rows at the top of the spreadsheet that contain the criteria. Again, be sure to include the column headings.
- Copy to: If you previously selected the box that says, "Copy to another location", you need to indicate where Excel should put the new subset (the new subset must be placed on the same worksheet as the raw data).





Advanced Filter (Cont'd)

The following tables provide an example:

Table 1

	Α	В	С
1	Name	University	Age
2	Michelle	Western	24
3	Paul	McGill	32
4	Ken	Queens	21
5	Cindy	Western	26
6	Susie	McGill	29
7	Sophie	Queens	33
8			
9			
10			
11			

Table 1 is a list of raw data that contains the name, university and age of various students.

Table 3

	Α	В	С
1	Name	University	Age
2		Western	> 24
3		Queens	> 20
4			
5	Name	University	Age
6	Michelle	Western	24
7	Paul	McGill	32
8	Ken	Queens	21
9	Cindy	Western	26
10	Susie	McGill	29
11	Sophie	Queens	33

In Table 3, the user has made two separate requests. The user would like to see a list of all students who are either 1. at Western and >24; **OR** 2. at Queens and >20. The filtered subset will contain Ken, Cindy and Sophie.

Table 2

	Α	В	С
1	Name	University	Age
2			
3			
4			
5	Name	University	Age
6	Michelle	Western	24
7	Paul	McGill	32
8	Ken	Queens	21
9	Cindy	Western	26
10	Susie	McGill	29
11	Sophie	Queens	33

In Table 2, a few blank rows were added above the data and the column headings were repeated.

Table 4

	Α	В	С
1	Name	Age	Age
2		> 24	<= 32
3			
4			
5	Name	University	Age
6	Michelle	Western	24
7	Paul	McGill	32
8	Ken	Queens	21
9	Cindy	Western	26
10	Susie	McGill	29
11	Sophie	Queens	33

In Table 4, the user has requested a list of all students who are >24 **AND** <=32. To filter *in between* a range, repeat the column heading in the criteria section twice. The filtered subset will contain Paul, Cindy and Susie.



VLookup Function

- Lookup Functions are very useful for extracting information from large tables
 - = VLOOKUP(Lookup_value, Table_array, Col_index_num, Range_lookup)
 - VLOOKUP (vertical lookup) searches for a value in the leftmost column of a table, and then returns a value in the same row from a column you specify

The VLookup function has 4 arguments:

• Lookup_value: This is the item you are searching for. This item must be

found in the leftmost column of the Table_array.

• **Table_array**: The entire table in which you are searching for the data.

• Col_index_num: When Excel finds the item you are searching for (the

Lookup-value) in the leftmost column of the table (the Table_array), it will move a certain number of columns to the right to extract the data that you are looking to retrieve.

• Range_lookup: To find a close match, type TRUE (or omit).

To find an exact match, type FALSE.

Example Table

	Α	В	С
1	Name	University	Age
2	Michelle	Western	24
3	Paul	McGill	32
4	Ken	Queens	21
5	Cindy	Western	26

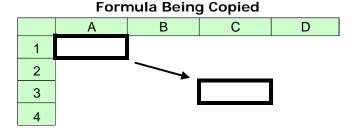
=VLOOKUP("Ken",A1:C5,3,False)

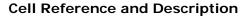
- This function will return 21, because it will search for the word "Ken" in column A and then return the corresponding age in the third column



Absolute References

- The ability to create formulas with absolute references is extremely important when manipulating data and building financial models
 - By default, Excel uses relative referencing
 - Absolute references allow you to "freeze" a reference to a particular cell, row or column by inserting dollar signs in front of the row reference, the column reference or both
 - As seen in the following example, if a formula references cell A1 and you copy that formula down two rows and over two columns, the following references will result





Reference	Description	Changes to
=\$A\$1	(absolute column and absolute row)	=\$A\$1
=A\$1	(relative column and absolute row)	=C\$1
=\$A1	(absolute column and relative row)	=\$A3
=A1	(relative column and relative row)	=C3

 To insert the dollar signs more quickly while editing a formula, use the F4 key to toggle between the various absolute referencing options



Conditional Formatting

Excel 2003

> Format > Conditional Formatting

Excel 2010

➤ Home [H] ➤ Conditional Formatting [L]

- This function formats cells that meet pre-specified criteria
- If you have more than one rule, the first rule will take priority over the second rule, so it's important to enter the rules in the appropriate order
- The first step is to highlight all data that you want the formatting applied to (when highlighting a column, you must start at the top and highlight down)
- There are two ways to conditionally format cells:
 - 1. Based on the cell's values
 - 2. Based on a formula applied to the cells
- Using method 1, if "Cell Value Is" greater than 25 and less than 100, all values between 25 and 100 will be formatted
- Using method 2, enter the following formula =D2>average(\$D\$2:\$D\$558)
- In the formula above, any values larger than the average will be shaded (think of the = sign at the beginning of the formula as the word "if")
 - For the "Formula Is" option, the formula has to return a value of TRUE or FALSE
 - If the formula returns TRUE, the selected formatting will be applied
 - If the formula returns FALSE, the selected formatting will not be applied



Text Functions

- Text functions are very important when editing and manipulating data
- The following are some important text functions:

Text Function Table

Function	Description
= Left(text,num_chars)	Returns x number of characters from a cell, starting from the left
= Right(text,num_chars)	Returns x number of characters from a cell, starting from the right
= Mid(text,start_num,num_chars)	Returns a specific number of characters from a text string starting at the position you specify
= Find(find_text,within_text,start_num)	Finds one text string within another text string and returns the number of the starting position of the found string (case sensitive)
= Len(text)	Returns the number of characters in a text string
= Upper(text)	Converts a text string to all uppercase letters
= Lower(text)	Converts a text string to all lowercase letters
= Proper(text)	Converts a text string to proper case so that the first letter in each word is uppercase and all other letters are lowercase
= Text(value,format_text)	Converts a value to text in a specific number format
= Trim(text)	Removes all spaces from a text string except for single spaces between words



Paste Special Command

Paste Special

Excel 2003

➤ Edit ➤ Paste Special

Excel 2010

➤ Home [H] ➤ Paste [V] ➤ Paste Special [S]

- Allows you to paste specific elements from one cell to another
- Can be used to paste Formats, Values, or Mathematical Operations
- Can also be used to transpose text or data
- If you want to multiply a bunch of numbers by 1,000:
 - Enter 1,000 into a blank cell
 - Put the cursor on the cell containing the 1,000 and select Copy (Ctrl + C)
 - Highlight all of the cells you want to multiply by 1,000 and then select:

Excel 2003

➤ <u>E</u>dit ➤ Paste <u>Special</u> ➤ <u>M</u>ultiply

Excel 2010

➤ Home [H] ➤ Paste [V] ➤ Paste Special [S] ➤ Multiply



Grouping Rows and Columns

- Occasionally, you may not want to show a particular row or column
- Under most circumstances, you should NOT use the HIDE function
 - When a row or column is hidden, it is very difficult for others using the file to know that it is hidden
- Instead, you should use the GROUPING function if you don't want certain rows or columns to be visible
 - The GROUPING function puts handles with buttons around each of the grouped rows or columns so that you can instantly realize that certain rows or columns are not visible

Excel 2003

➤ Data ➤ Group and Outline ➤ Group

Excel 2010

➤ Data [A] ➤ Group [G] ➤ Group [G]

Keyboard Shortcut: Shift + Alt + left or right arrow



Database Functions



Database Functions

- Excel's Database functions are some of the most powerful and underutilized functions in Excel.
 - Whenever there is a large spreadsheet of data (i.e. a list of employees, deals, transactions, etc.), people often want to know the sum or average of a subset of the data.
 - For example, if there was a spreadsheet that contained a listing of all the employees at a given company, it might be nice to know the following:
 - 1. What was the total compensation to all employees in the finance department?
 - 2. What is the average age of employees in the marketing department?
 - Most Excel users solve these types of questions by first sorting or filtering the data and then calculating a sum or average on the subset.
 - If there are a lot of queries, it can take a long time to first sort or filter the data before answering each question.
 - The Database functions calculate the answers to these types of questions without requiring that the data be sorted or filtered.



• The following table provides a listing of all the Database functions in Excel:

Database Functions:

Function	Description
= DAVERAGE	Returns the average of selected database entries
= DCOUNT	Counts the cells that contain numbers in a database
= DCOUNTA	Counts nonblank cells in a database
= DGET	Extracts from a database a single record that matches the specified criteria
= DMAX	Returns the maximum value from selected database entries
= DMIN	Returns the minimum value from selected database entries
= DPRODUCT	Multiplies the values in a particular field of records that match the criteria
= DSTDEV	Estimates the standard deviation based on a sample of selected database entries
= DSTDEVP	Calculates the standard deviation based on the entire population of entries
= DSUM	Adds the numbers in the field column in the database that match the criteria
= DVAR	Estimates variance based on a sample from selected database entries
= DVARP	Calculates variance based on the entire population of selected database entries



 The Database functions all use the same syntax, as described in the following DSUM Function:

=DSUM(Database,Field,Criteria)

Database: This is the entire range of data. This range must have column headings.

Field: This is the column heading of the data that you would like to

perform the mathematical operation on (i.e. sum, count, average). You can either click on the cell that contains the column heading, or retype the column heading in the function

surrounded by quotation marks.

Criteria: This is another set of cells that tells the database function

which criteria to use in order to calculate the correct answer. The criteria section should be set up exactly the same way as the criteria section for an Advanced Filter (see next 2 pages).



- The steps to set up a Database function are as follows:
 - 1. Ensure that the raw data has column headings.
 - 2. Add a few blank rows on top of the raw data.
 - 3. Copy the column headings to a blank row a few rows above the raw data. This will be the criteria section for the database functions.
 - 4. Within the criteria section, place criteria on top of one another to use the **OR** logical operator. Place criteria beside one another to use the **AND** logical operator.



The following tables provide some examples:

Table 1

	А	В	С	
1	Name	University	Age	
2	John	Rotman	22	
3	Peter	McMaster	25	
4	Miranda	Ivey	27	
5	David	Rotman	21	
6	Katie	McMaster	23	
7	Sarah	Ivey	24	
8				
9				
10				

Table 1 is a list of raw data that contains the name, university and age of various students.

Table 3

	Α	В	С
1	Name	University	Age
2		Rotman	
3			
4	Name	University	Age
5	John	Rotman	22
6	Peter	McMaster	25
7	Miranda	Ivey	27
8	David	Rotman	21
9	Katie	McMaster	23
10	Sarah	Ivey	24

=DAVERAGE(A4:C10,C4,B1:B2) will calculate the average age of Rotman students and return a value of 21.5.

Table 2

	А	В	С
1	Name	University	Age
2			
3			
4	Name	University	Age
5	John	Rotman	22
6	Peter	McMaster	25
7	Miranda	Ivey	27
8	David	Rotman	21
9	Katie	McMaster	23
10	Sarah	Ivey	24

In Table 2, a few blank rows were added above the data and the column headings were repeated.

Table 4

	Α	В	С	
1	Name	University	Age	
2		Ivey	>24	
3				
4	Name	University	Age	
5	John	Rotman	22	
6	Peter	McMaster	25	
7	Miranda	Ivey	27	
8	David	Rotman	21	
9	Katie	McMaster	23	
10	Sarah	Ivey	24	

=DCOUNT(A4:C10,C4,B1:C2) will count the number of Ivey students who are older than 24 and return a value of 1

