Unit 3

MS Excel *version 2010 for Windows PC* for Epidemiology

With many thanks to the following windows users:
Liz Procter-Gray, Yiwei Jiang, and Carrie Nobles

"Technical skills, like fire, can be an admirable servant and a dangerous master."

- A. Bradford Hill (1971)

Microsoft Excel, MS Excel, is the standard program for creating spreadsheets, maintaining them, and producing charts. Other programs are available, such as Quattro Pro or Lotus 1-2-3, but you are unlikely to encounter them.

This introduction to MS Excel focuses on its use for data set creation, manipulation (eg- sorting and selecting) and selected calculations.

While it can be done, we will not be using MS Excel to do statistical analyses and data visualizations. The focus of BIOSTATS 690C is, instead, on R and Stata.

Table of Contents

Learning Objectives

When you have finished this unit, you should be able to:

- Navigate in and out of Excel (launch, exit, enter data, format cells, arrange columns, freeze rows and columns for easy viewing, sort, and autofill)**;**
- Specify the format and layout of an excel spreadsheet for printing (portrait v landscape, headers, footers, etc);
- Create new fields (what we think of as variables) using functions and user-specified formulae; and
- Create a data set and document it.

1. Introduction to MS Excel

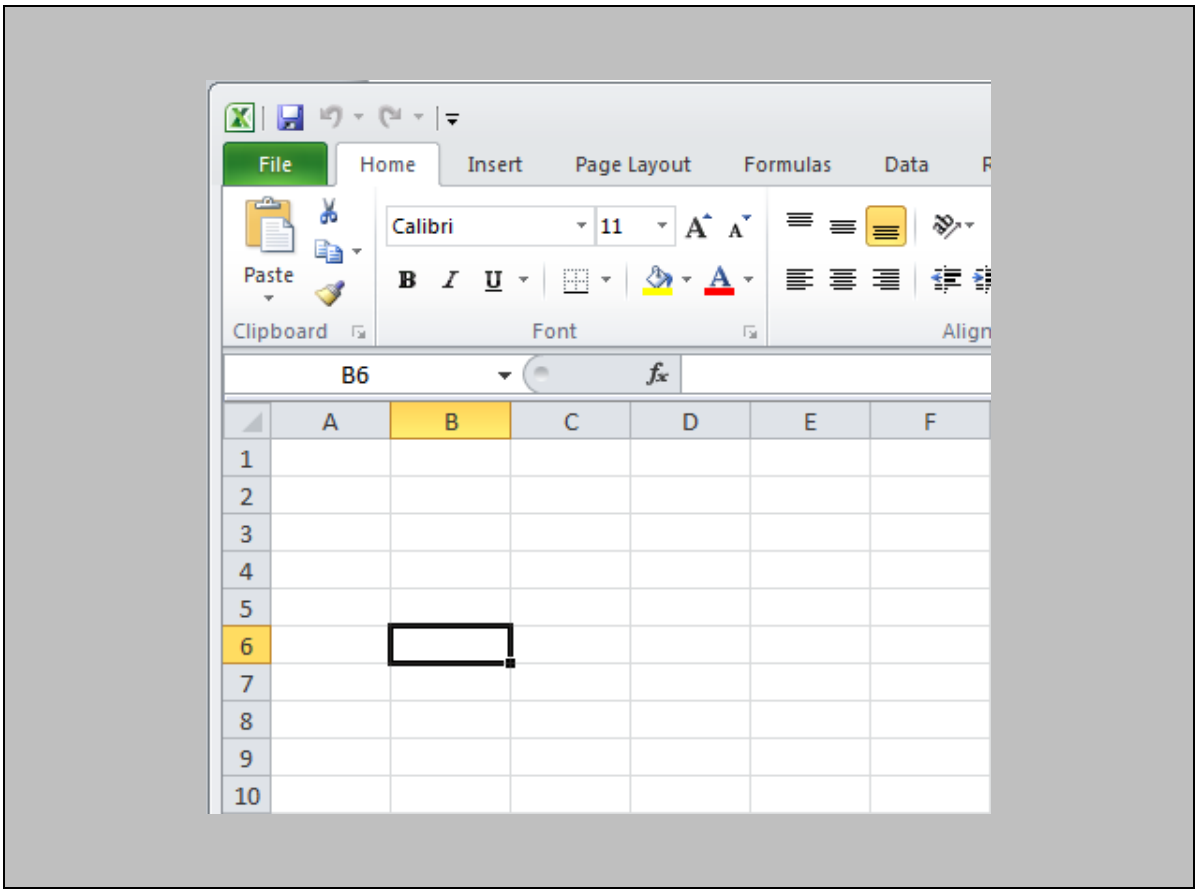
1.1 What is MS Excel?

MS Excel is a widely-used software package used for creating **spreadsheets**. In Excel spreadsheets are called **worksheets.**

What is a spreadsheet? *It is simply a grid of information that might include numbers or words or a mix. Its storage in a grid is handy way to do its organization.*

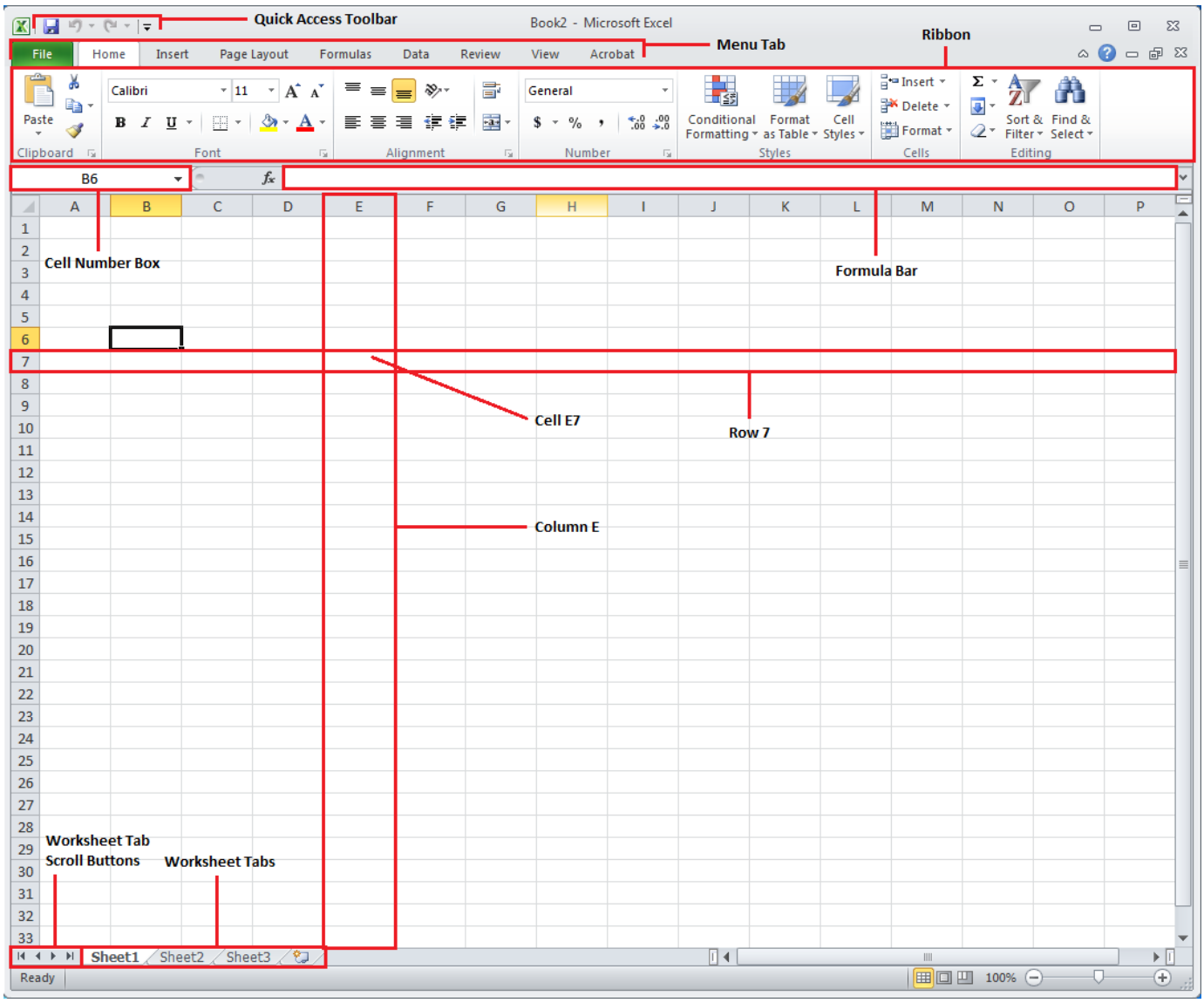
What might I like to do with a spreadsheet?
Lots of things, actually – lists, sorted lists, picture summaries. Also charts.

Each worksheet is a grid of **columns**, indicated by letters, and **rows**, indicated by numbers. Thus each space, or **cell**, on the worksheet is identified by a row-column designation, such as **B6**. Each cell can hold a number, text, or even the result of calculation of a mathematical formula.



Design ..... **Data Collection** ..... **Data Management** ..... **Data Summarization** ..... **Statistical Analysis** ..... **Reporting**

Microsoft Excel 2010 Screen



1.2 Advantages and disadvantages of MS EXCEL

Advantages

- It is very commonly used and very easily shared.
- Many statistical analysis packages permit the direct import of Excel data.
- Data can be sorted by any column while still retaining the integrity of each record.
- It is easy to create new variables that are mathematical functions of variables. For example, you can tell Excel to calculate the mean of the fields in columns A, B, and C and store the result as a new field, such as column D.
- Blocks of data can be copied and moved from one part of the worksheet to another or from worksheet to worksheet.
- Excel offers lots of formats for data display (eg – number of significant digits, display of dates as month-day-year) with no loss of information.

Disadvantages

- The entry of a negative number is awkward. Excel might interpret the negative sign as the beginning of a mathematical formula. Solution: enter the negative number with a leading **apostrophe ‘** in the cell, e.g., ‘-0.28.
- Excel can make mistakes in mathematical formulae if you inadvertently mix character and numeric fields. For example, if column A is character and Column B is numeric, the addition of entries in column A and column B may be incorrect.
- Sometimes, the charts produced by Excel are not correct.

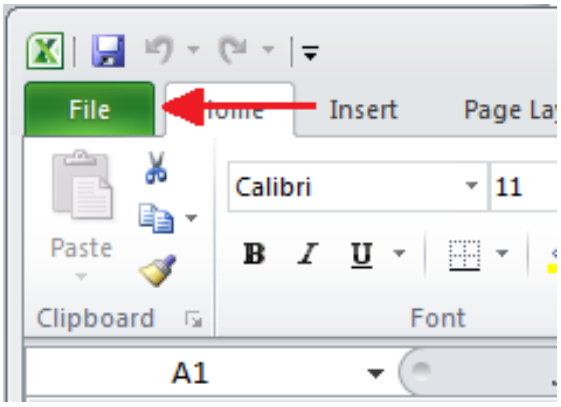
2. Getting Started - Spreadsheet Basics

2.1 The Toolbars and Moving Through Cells

Excel 2010 has many similar features to the previous versions and also many new features that you'll be able to utilize. Among them, there are three features that you're going to use most as you work with Excel 2010: the Microsoft Office Button, the Ribbon, and the Quick Access Toolbar.

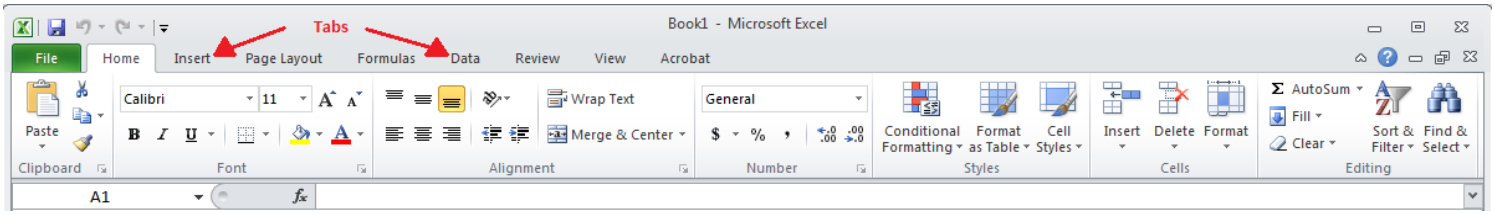
Microsoft Office Button

The Microsoft Office Button performs like the **File** menu in older versions. You can create a new workbook, open an existing workbook, save and save as, print, send, and close, etc. using the functions located in the drop down menu.



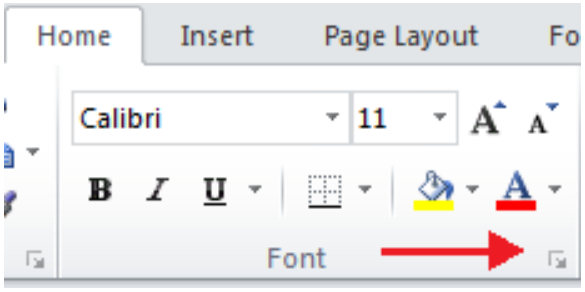
Ribbon

The Ribbon is a panel beside the Microsoft Office Button which has eight tabs: **File, Home**, **Insert**, **Page Layouts**, **Formulas**, **Data**, **Review**, and **View**. Each tab is divided into groups which are logical connections of features.



File: New, Open, Save, Print, Send, Close, etc.
Home: Clipboard, Fonts, Alignment, Number, Styles, Cells, Editing
Insert: Tables, Illustrations, Charts, Links, Text
Page Layouts: Themes, Page Setup, Scale to Fit, Sheet Options, Arrange
Formulas: Function Library, Defined Names, Formula Auditing, Calculation
Data: Get External Data, Connections, Sort & Filter, Data Tools, Outline
Review: Proofing, Comments, Changes
View: Workbook Views, Show/Hide, Zoom, Window, Macros

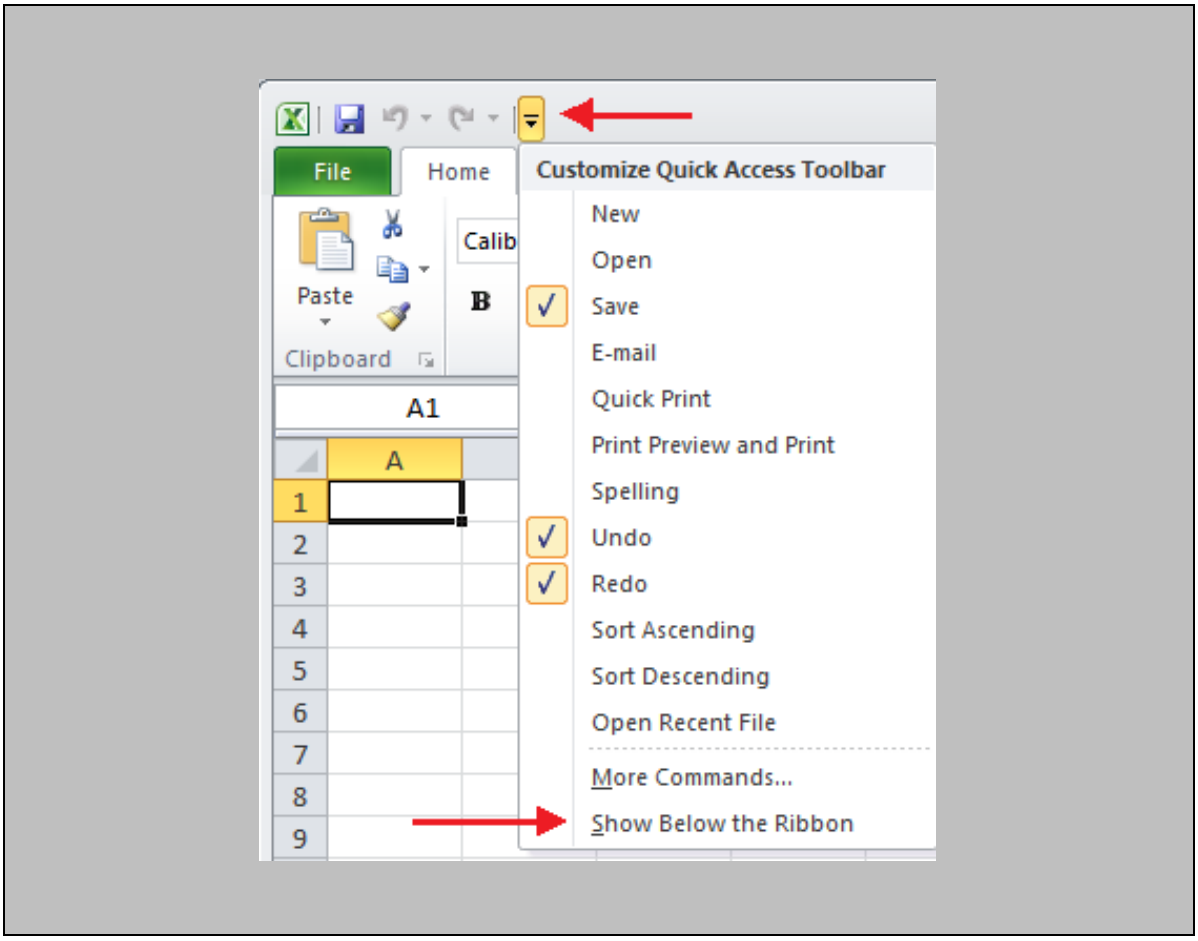
Besides commonly utilized features that are displayed on the Ribbon, you could see additional features within each group by clicking the arrow at the bottom right corner of each group.



Design ..... **Data Collection** ..... **Data Management** ..... **Data Summarization** ..... **Statistical Analysis** ..... **Reporting**

Quick Access Toolbar

The Quick Access Toolbar is a customizable toolbar that contains commands you may use. You could create buttons of commands to show in the toolbar by clicking them in the drop down menu and change the position of the toolbar.



Moving through cells

You already know how to move from one cell to another using the mouse. Here are some other ways to move within the worksheet.

Movement	**Key stroke**
One cell up	up arrow key
One cell down	down arrow key or **ENTER**
One cell left	left arrow key
One cell right	right arrow key or **TAB**
Top of the worksheet (cell A1)	**CTRL+HOME**
End of the worksheet (last cell containing data)	**CTRL+END**
End of the row	**CTRL**+ right arrow key
End of the column	**CTRL**+ down arrow key
Any cell	**Home|Editing|Find&Select|Go To** command

2.2 Modifying a Worksheet

As you add data to your worksheet, you may find you need to modify the layout in various ways:

- **Widen or shrink rows or columns**
  - To resize a row: Position your cursor over the boundary line between two rows at the far left of the worksheet. The appearance of your cursor will change from a little arrow to a cross. Left-click and drag to obtain the row size you want and then release.
  - To resize a column: Position your cursor over the boundary line between two columns at the top of the worksheet. The appearance of your cursor will change from a little arrow to a cross. Left click and drag to obtain t he column size you want. Release.

Note - Another way to resize a row or a column is using options in the drop down menu of the Format button in the Cells group of the Home tab.

- **Highlight a cell or cells**
  Excel offers some shortcuts for selecting cells:

Cells to select	Mouse action
One cell	click once in the cell
Entire row	click the row label (row number at far left)
Entire column	click the column label (column letter at top)
Entire worksheet	click the whole sheet button (cell just to the left of label "A")
Cluster of cells	drag mouse over the cells or hold down the **SHIFT** key while using the arrow keys

Source: Florida Gulf Coast University, Excel 2000 tutorial ***http://www.fgcu.edu/support/office2000/excel/index.html***

- **Insert a row above** -- (1) Highlight the row to be below the new row. (2) Then, from the **Cells** group of the **Home** tab (or right click the highlighted row), choose **Insert>Insert Sheet Rows** (or just **Insert**). Insert a column in a similar manner.

- **Move or copy cells** -- (1) Highlight the cells to be moved or copied using the **Cut** or **Copy** buttons from **Clipboard** group of the **Home** tab (or the CTRL-X or CTRL-C keys). (2) Then click the upper left-hand cell of the area you want to move them to and click the **Paste** button (or CTRL-V).

- **Freeze panes - This is a wonderful feature!** It allows you to retain for viewing some top rows (such as row 1 which might contain your variable names) and some important column (such as column 1 which might contain study id) (1) Position your cursor in the cell that is simultaneously below the rows you want to freeze and to the right of the columns you want to freeze (2) Then, from the **Window** group of the **View** tab, choose **Freeze Panes> Freeze Panes** from the menu. You can undo this operation by selecting **Freeze Panes >Unfreeze Panes.**

  Note – The freeze panes feature is for viewing only. Formatting the printing of a worksheet so that a selection of top rows and left hand columns appears on every page is done using additional features in the Sheet Options group of the Page Layout tab. More on this later (see section 2.7).

- **Add, Remove, and Renaming Worksheets**

  At the bottom of your excel file are tabs for each worksheet in the file. The default is 3 worksheets named "Sheet1", "Sheet2", and "Sheet3".

  To add a worksheet: (1) Position your cursor on the tab for that worksheet
  (2) Right click (3) From the drop down menu, choose **Insert**.

  To remove a worksheet: (1) Position your cursor on the tab for that worksheet.
  (2) Right click (3) From the drop-down menu, choose **Delete**.

  To rename a worksheet: (1) Position your cursor on the tab for that worksheet.
  (2) Right click (3) From the drop-down menu, choose **Rename**.

- **Moving/Copying Worksheets**

  As your worksheets accumulated, you might want to rearrange their order. Alternatively, you might want to work with a copy of a worksheet.

  (1) Position your cursor on the tab for the worksheet to be moved or copied

  (2) Right click

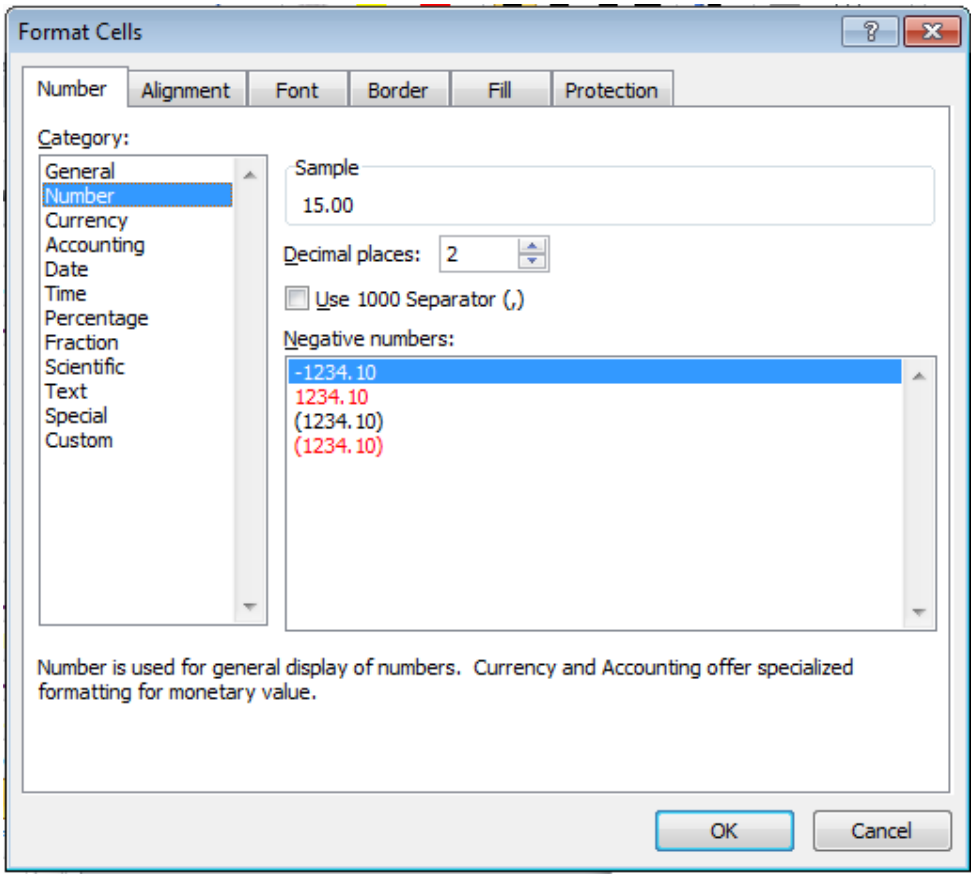
  (3) From the drop down menu, choose **Move or Copy**.

2.3 Formatting Cells

As previously noted, Excel has formatting options for the display of spreadsheet information so that it is readable to us! (eg - dates, times, percentages, or dollars!) To format cells, columns of cells, or multiple columns of cells:

(1) Highlight the cells

(2) Click the **Format** button of **Cells** group of the **Home** tab. Click **Format Cells** from the drop down menu. This will open the dialog box below. It has several tabs. You will be positioned in the **Number** tab.



(3) Choose the tab and category that you want to format. Then select from the drop down menus that are provided.

(4) The other tabs (**Alignment**, **Font**, **Border**, **Fill**, and **Protection**) can be accessed to change the font of text entries, to align entries on the right, left, or center of cells, etc. Try it!

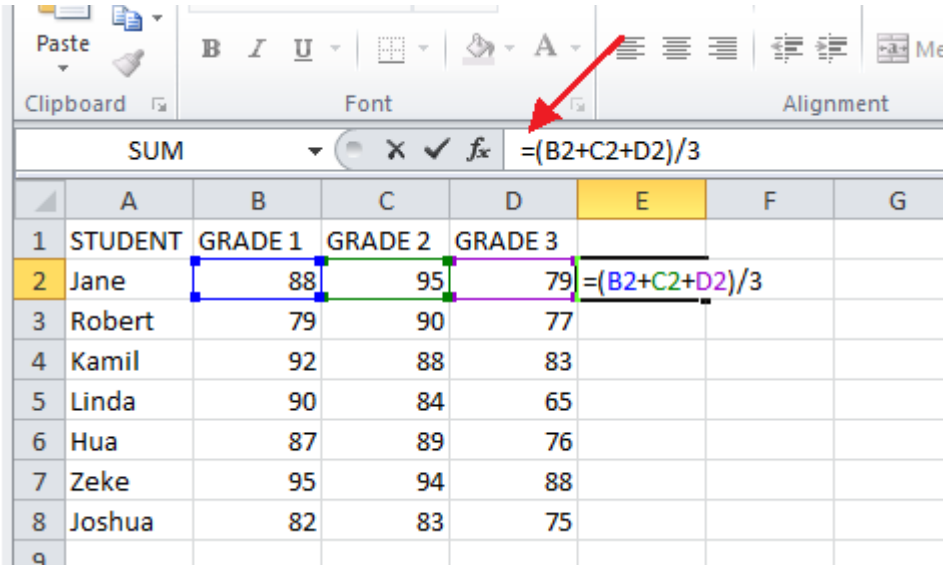
2.4 Formulas and Functions

The creation of new variables (or fields) that are the result of calculations is easy.

Step 1: Highlight the first cell where the result is to be stored

Step 2: Position your cursor in the f_x dialog box that is located just above the column labels "A", "B", "C"

Step 3: Always begin your entry with an equal sign "="



	A	B	C	D	E	F	G
1	STUDENT	GRADE 1	GRADE 2	GRADE 3			
2	Jane	88	95	79	=(B2+C2+D2)/3		
3	Robert	79	90	77			
4	Kamil	92	88	83			
5	Linda	90	84	65			
6	Hua	87	89	76			
7	Zeke	95	94	88			
8	Joshua	82	83	75			

Example We wish to create a new variable that is the average of the values in columns B, C, and D. And we want the result to be stored in column E.

(1) cell E2 is selected (note the bold border of cell E2)
(2) now position your cursor in the dialog box to the right of f_x
(3) ***=(B2+C2+D2)/3*** is entered into this dialog box.
(4) ***Note – An alternative approach that is less error prone is to instead enter =sum(B2:D2)/3. In doing this you could also use highlight and drag over the cells B2, C2, and D2.***

Step 4: Replicate your calculation for every other row.

At this point, you have done the calculation of the new variable for just one record, in this case the record for Jane in row 2. Now you want to repeat this calculation for the other rows in your spreadsheet (Robert, Kamil. etc)

(1): Highlight the cell that has the first result; this is cell E2 in this example.

(2): From the **Clipboard** group of the **Home** tab, choose **Copy** (or CTRL-C).

(3): Highlight all the destination cells; these will be cells E3, E4, and so on down to the last row in your data set.

(4): From **Clipboard** group of the **Home** tab, choose **Paste** (or CTRL-V).

Alternatively

(1): Highlight the cell that has the first result; this is cell E2 in this example.

(2): Click the bottom right corner of this cell.

(3): Now drag down through E3, E4, etc to the last row in your data set

MS Excel has a Selection of "Built in" Functions

These save typing by hand. In our example above, instead of typing the formula "=(B2+C2+D2)/3" in cell E2, we could have accomplished the same operation by typing "=AVERAGE(B2:D2)". The following is a ***partial listing*** of the available functions. A ***complete listing*** of functions can be found in the **Function Library** group of the **Formulas** tab.

Function	Example	Description
SUM	=SUM(A1:100)	finds the sum of cells A1 through A100
AVERAGE	=AVERAGE(B1:B10)	finds the average of 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 cell D10
TODAY	=TODAY()	returns the current date (leave the parentheses empty)

Note – If you want to use the value of a certain cell as a fixed parameter in your function, add a dollar sign before both the column character and the row number, eg A1.

2.5 Sorting

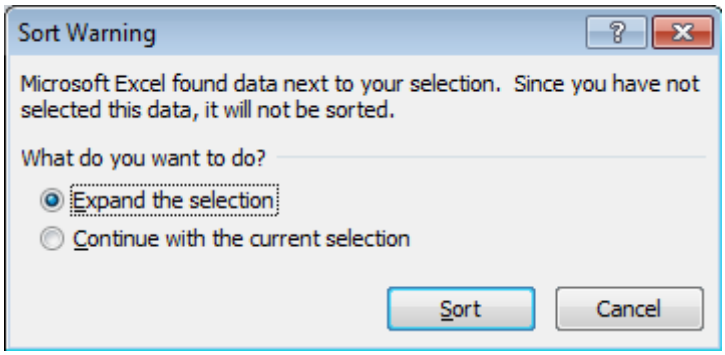
Excel lets you sort the data in your spreadsheet by the entry in one column while retaining the integrity of the entire profile for each record. Different from the older versions, Excel 2010 has divided the sorting command to three parts: basic sorts, custom sorts, and filtering.

Basic Sorts - To execute a basic descending or ascending sort based on one column.

(1) Highlight the column that the sort is based on.

(2) Choose **Sort Ascending** (A-Z) or **Sort Descending** (Z-A) in the menu of the **Sort & Filter** button in **Editing** group of the **Home** tab.

(3) The following dialog box will show up. Choose **Expand the selection**, and then click **Sort**.



Custom Sorts - To sort on the basis of more than one column.

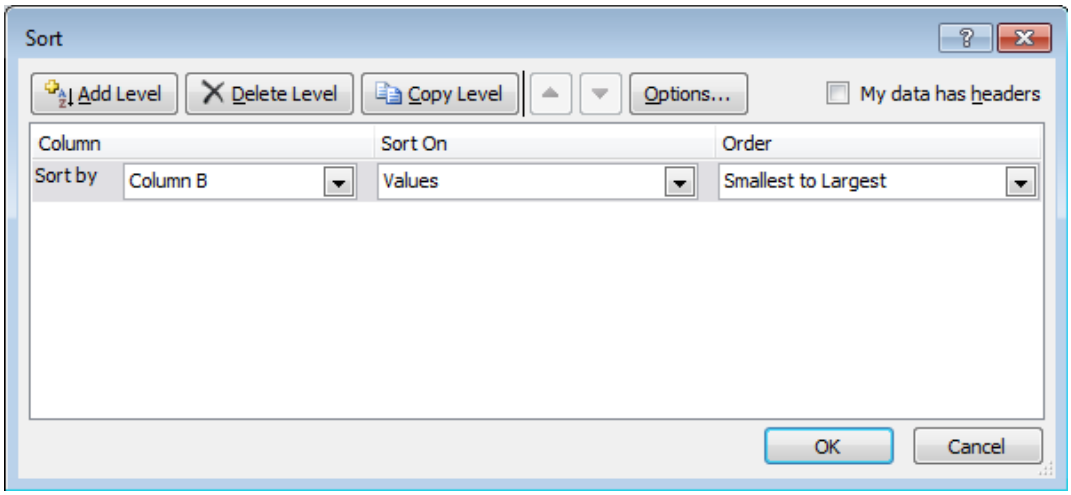
(1) Before sorting, highlight **ALL** of the cells that are to be sorted; **this is usually the entire worksheet.**

(2) Choose **Custom Sort** in the menu of the **Sort & Filter** button in **Editing** group of the **Home** tab.

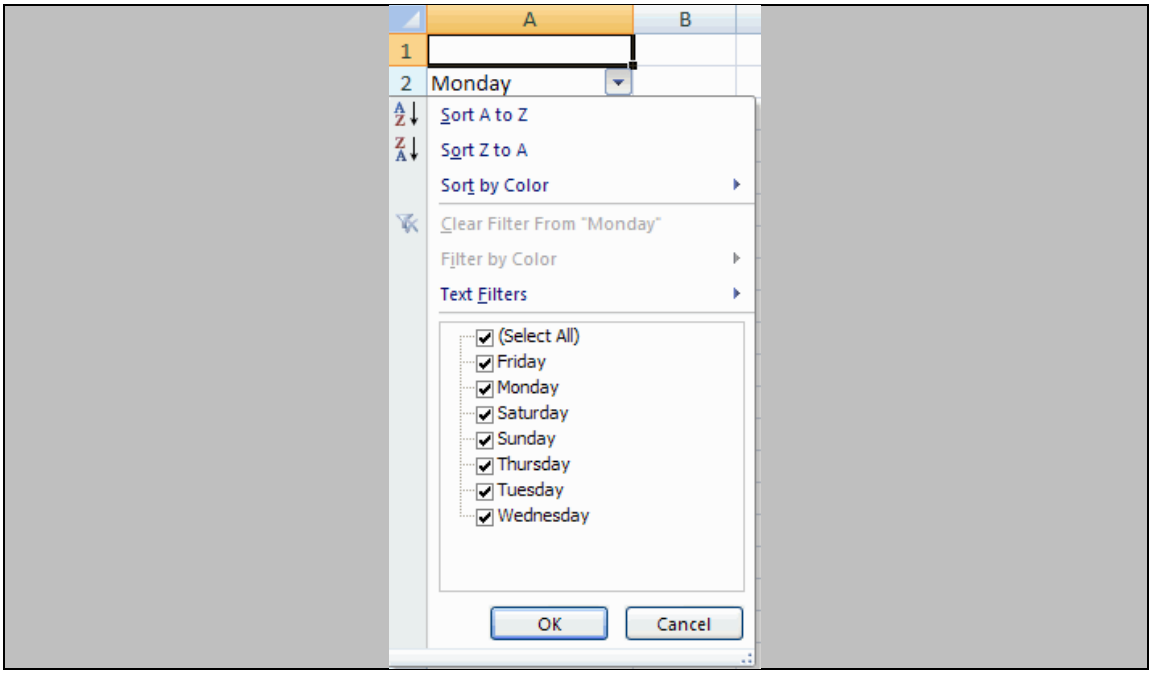
(3) Choose the column, the basis, and the order you want to sort by first.

(4) Click the **Add Level** button and choose the three items you want to sort by second.

(5) Click **OK** when you have added all the columns in the order you want to sort by.


Sort
Add Level
Delete Level
Copy Level
Options...
My data has headers
Column
Sort On
Order
Sort by
Column B
Values
Smallest to Largest
OK
Cancel

Filtering - To display only data that meets certain criteria.
(1) Click the column or columns that contain the data you wish to filter.
(2) Choose **Filter** in the menu of the **Sort & Filter** button in **Editing** group of the **Home** tab.
(3) Click **Text Filter** in the menu of the **Arrow** at the bottom of the first cell.
(4) Click the **Words** you wish to Filter



(5) To clear the filter, choose **Clear** in the menu of the **Sort & Filter** button in **Editing** group of the **Home** tab.

2.6 Auto-filling and Fill Series

Auto-Filling

Excel has an auto-filling feature that lets you replicate a given entry into multiple cells in a column. This can be very handy.

Example Suppose you would like to replicate the A2 cell entry of "2009" into cells A3 through A150.

Step 1: Enter 2009 in cell A2.

Step 2: Position your cursor at the lower right corner of this cell. A small black square should appear.

Step 3: Click and drag down A3, A4 and so on to A150.

Step 4: When you release the mouse, notice that all highlighted cells now contain 2009, and a small **Auto-Fill Options** button appears. This brings us to Fill Series.

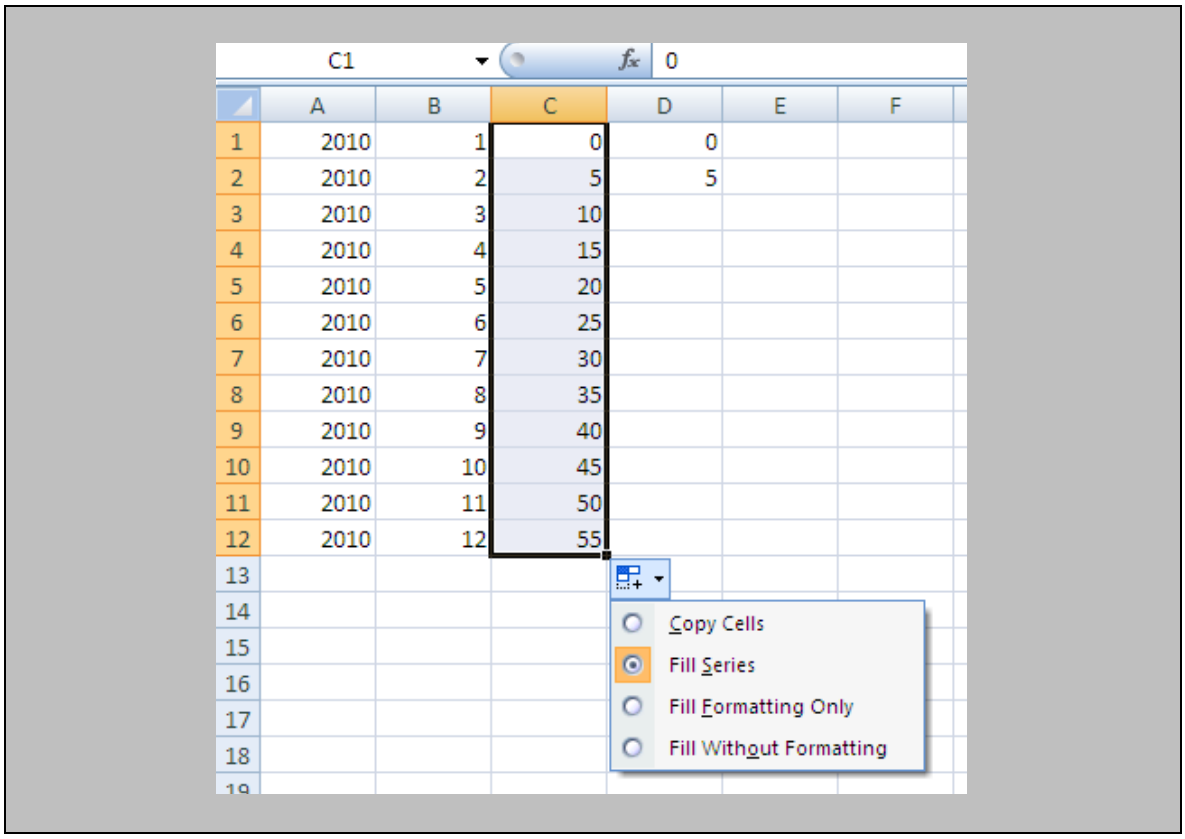
Fill Series

If you click on this button, you will see a short menu including **Copy Cells, Fill Series**, and other options.

Choosing **Copy Cells** will result in all cells being filled with the year 2009, as you entered in the first cell.

Excel can also save you time if you need to enter a regular series of numbers, days of the week, etc. Choosing the **Fill Series** option in the example above will result in the series 2009, 2010, 2011, etc., adding 1 to each successive cell.

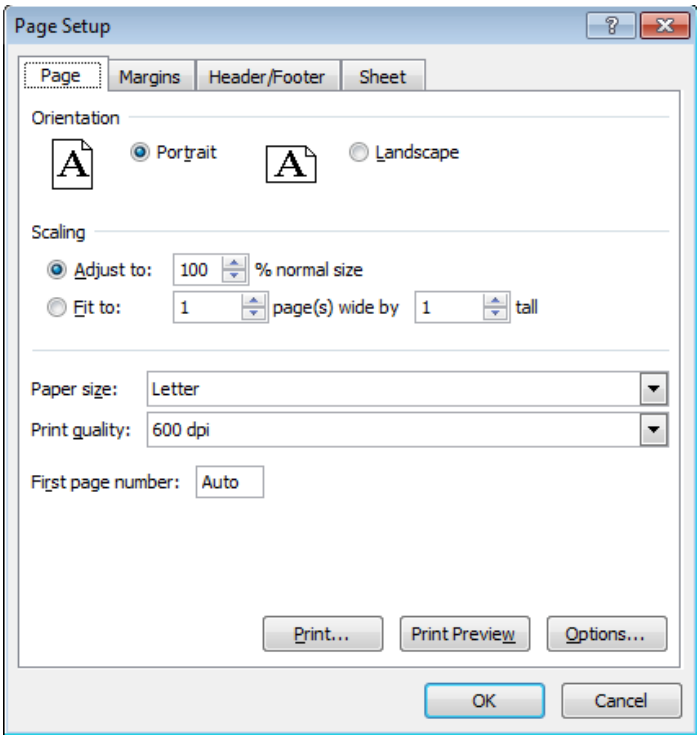
Example – Suppose you want to add 5 to each successive cell down a column, starting with 0. To do this, enter 0 in the top row, 5 in the second row. Then highlight the two cells, click on the lower right corner of the section, as above, and drag down the column. Excel will fill the series, adding 5 to each successive cell. In this example, what happens if you choose the option **Copy Cells**?



A	B	C	D
2010	1	0	0
2010	2	5	5
2010	3	10	
2010	4	15	
2010	5	20	
2010	6	25	
2010	7	30	
2010	8	35	
2010	9	40	
2010	10	45	
2010	11	50	
2010	12	55	

2.7 Page Setup and Printing

Before you do any printing, specify your page layout. Click the **Dialog Box** arrow in the **Page Setup** group of the **Page Layout** tab. Four tabs with a variety of menus will appear:



Page Tab: Choose page orientation (**Portrait** or **Landscape**) If you do not have too many columns, choose the option **Fit to 1 page wide** so that all of your variables appear on one page.

Header/Footer Tab: Use this tab to specify custom headers and footers. A good practice is to use headers and footers to document your name, date, file name, analysis code program names, etc.

Sheet Tab: **Tip!!** Use this tab to choose rows to be repeated at the top of each printed page and columns to appear at the left of each printed page. The **Sheet** tab also allows you to choose whether or not to show **Gridlines** in your printed table.

The **Print Preview** button will let you see just what your worksheet will look like when printed. After pressing **Print Preview**, you will see further options including **Page Break Preview;** this allows you to click and drag on dotted lines that define the borders of the print area.

When you are satisfied with the appearance of your worksheet, select **Print** in the drop down menu of the **Microsoft Office Button**.

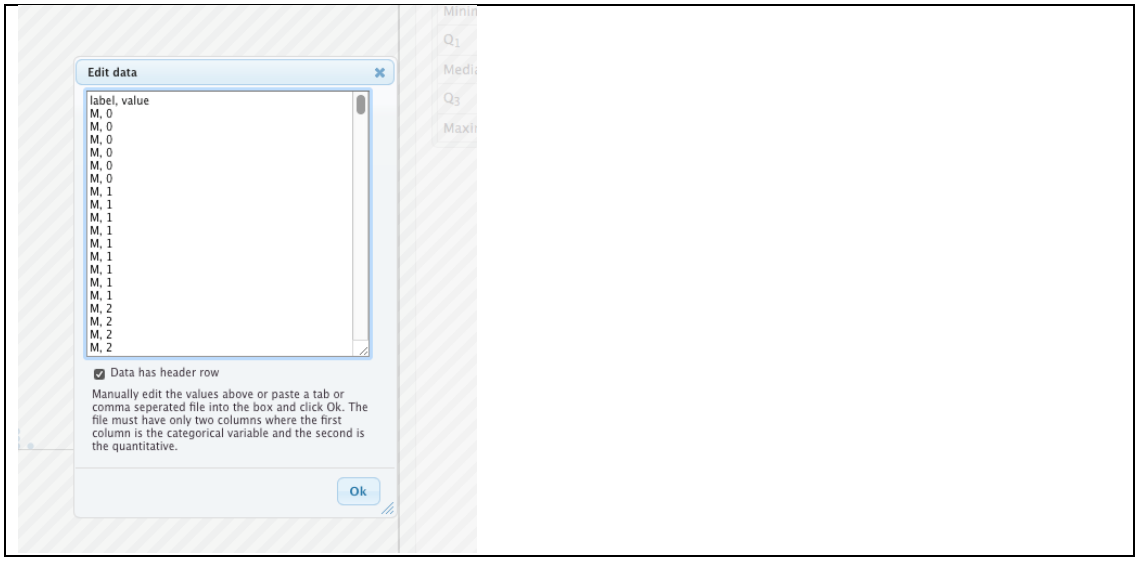
2.8 Handy for BIOSTATS 540 – How to Concatenate

Introduction – Why do I want to Concatenate?

In BIOSTATS 540, a number of online statistical software applications are introduced. These are terrific in that you can enter your data directly! The problem is that, sometimes, the required format of data entry is awkward. Two such online statistical software applications are:

1. Shodor Interactivate Box Plot
   (http://www.shodor.org/interactivate/activities/BoxPlot/)

2. StatKey
   http://www.lock5stat.com/StatKey/



What a nuisance! **In both instances, when the data consists of two variables, one qualitative and one quantitative, each row must contain the two values and they must be separated by a comma.**

Solution – Enter your data into Excel for pasting into Shodor or StatKey or whatever else

Example –
Suppose we want to enter the following data into our online statistical software application

Male, 44
Male, 16
Female, 37

Step 1
Launch Excel. Into Column A of your worksheet put your values of your first variable. For example, this might look like:
COLUMN A
Male
Male
Female

Step 2
Into Column B of your worksheet put your values of your second variable. For example, this might look like:
COLUMN B
44
16
37

Step 3
Now position your cursor in COLUMN C, row 1. Into the formula box, type the following, taking care not to forget the equal sign:

= concatenate(A1,",",B1)

You should now see in Column C, row 1:
COLUMN C
Male, 44

Step 4
Using copy>paste, repeat for the remaining rows of data.

All set!
Paste column C where you need it (StatKey or Shodor, etc)

3. Data Set Creation Basics

Example ("ICU Example")–

Recall from BIOSTATS 540, the study of 25 consecutive patients entering the general medical/surgical intensive care unit at a large urban hospital. For each patient, the following data were collected.

Variable	Description	Code
ID	Confidential Patient Identifier	
AGE	Age (years)	numeric
TYPE_ADM	Type of Admission	1 = emergency 0 = elective
ICU_TYPE	ICU Type	1 = medical 2=surgical 3=cardiac 4=other
SBP	Systolic Blood Pressure (mm Hg)	numeric
ICU_LOS	Number of days in ICU	integer
VIT_STAT	Vital Status at Discharge	1=dead 0=alive

We'll use this example of 25 observations to illustrate the steps and recommendations for data set creation using MS Excel. The data are on the next page (p. 25).

Example ("ICU Example") – Data.

ID	Age	Type_Adm	ICU_Type	SBP	ICU_LOS	Vit_Stat
1	15	1	1	100	4	0
2	31	1	2	120	1	0
3	75	0	1	140	13	1
4	52	0	1	110	1	0
5	84	0	4	80	6	0
6	19	1	1	130	2	0
7	79	0	1	90	7	0
8	74	1	4	60	1	1
9	78	0	1	90	28	0
10	76	1	1	130	7	0
11	29	1	2	90	13	0
12	39	0	2	130	1	0
13	53	1	3	250	11	0
14	76	1	3	80	3	1
15	56	1	3	105	5	1
16	85	1	1	145	4	0
17	65	1	1	70	10	0
18	53	0	2	130	2	0
19	75	0	3	80	34	1
20	77	0	1	130	20	0
21	52	0	2	210	3	0
22	19	0	1	80	1	1
23	34	0	3	90	3	0
24	56	0	1	185	3	1
25	71	0	2	140	1	1

3.1 Design Your Database First

Before entering data into an excel spreadsheet, or any database application, the file's structure must be defined first. Exactly how this is done varies by software type, but the following components are available in good database software:

- **Name of field (note – this is also your variable name)** -- a single word name used as a shorthand reference for a field

  Keep it short (8 characters or under is recommended, though not required)

  Avoid special characters such as '#', '-', '*', '.', and ','. While some software will allow these special characters in a name, others will not, creating problems when you transfer data between formats. Avoid spaces in a name for the same reason. Use an underbar (_) in place of a space or alternating between upper and lower case.

- **Label for field** -- (optional) a longer description of data stored in the field.

- **Type of field** -- there are 2 basic types of fields that dictate the manner in which data is stored: character and numeric.

  Numeric -- containing only numbers

  Text or **Character** -- allowing letters, numbers, other keyboard characters

  Other field formats are often available, too.

  Logical -- Yes/No or True/False

  Date -- containing dates in a specified format
  (in some programs dates are stored as character data, in others, numeric)

  Some programs have other special field types – currency, percent, phone numbers, SSN, ...

  Note - In some software these are considered formats rather than field types.

- **Format for field** -- specifies the number of digits or spaces available for entering and displaying data, or other specialized formats.

  Numeric formats specify the number of digits before and after the decimal place

  Character formats typically define the number of spaces or columns needed

  Date and Date/Time formats specify the order (month/day vs day/month) and presentation of data, e.g., 07JUN2001 vs 06/07/2001

Data type

It is necessary to define the data type for each field. In Excel this is **formatting cells** (see p 11)

- Numeric and character data are stored quite differently, and you should be clear ahead of time, as to data the type required.

- Numbers can be stored in character fields. **Don't do this!!** It can cause great confusion later in the data management process if you think you have numeric data and attempt computations, when the field was defined as character.

- It is not always obvious when data should be numeric, and when character. For example, while a phone number or social security number could be entered as numeric data, you will never want to compute with these numbers -- they serve as ID or identifier types of variables. By entering these as character data, you can include hyphens (e.g., 545-1000 or 999-99-9999), and when printed they will appear in a familiar format.

- There are also occasions when numbers are clearly codes and can be entered as character data since you will never compute with these numbers (such as 1=White, 2=Black, 3=Asian, 4=other). However in some situations it may be advantageous to enter these as numeric data. Some statistical applications (e.g., Minitab) will not allow character variables in analyses -- even if the variable is used solely to define groups. If you know this is true of the software you will be using for analysis – plan accordingly. If you have defined a variable as character and need numeric data or vice versa, it is always possible to convert the data, or create a new variable in the required format from the values of the current one, but planning ahead saves work.

- **Pay attention to dates** – some applications store dates as character data, and others as numeric. This affects how information is transferred between programs, and you will often need to do some special programming to handle dates. This is particularly important if you will be using dates to compute durations (e.g., length of stay in hospital, time between patient interviews, …)

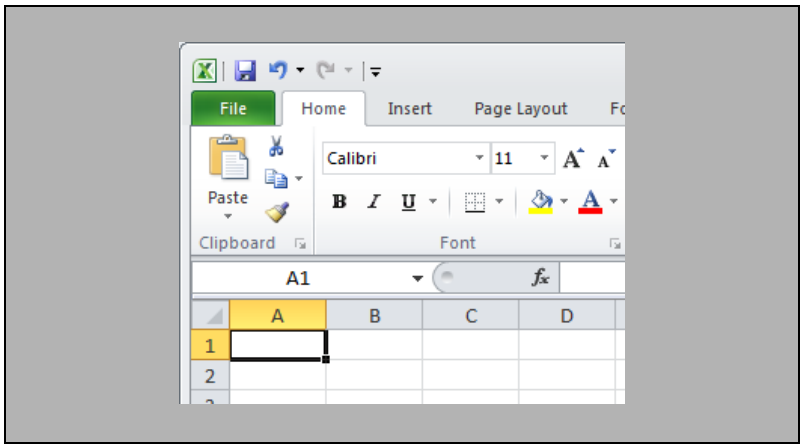
3.2 Data Entry

The steps are best explained in an example.

ICU Example -

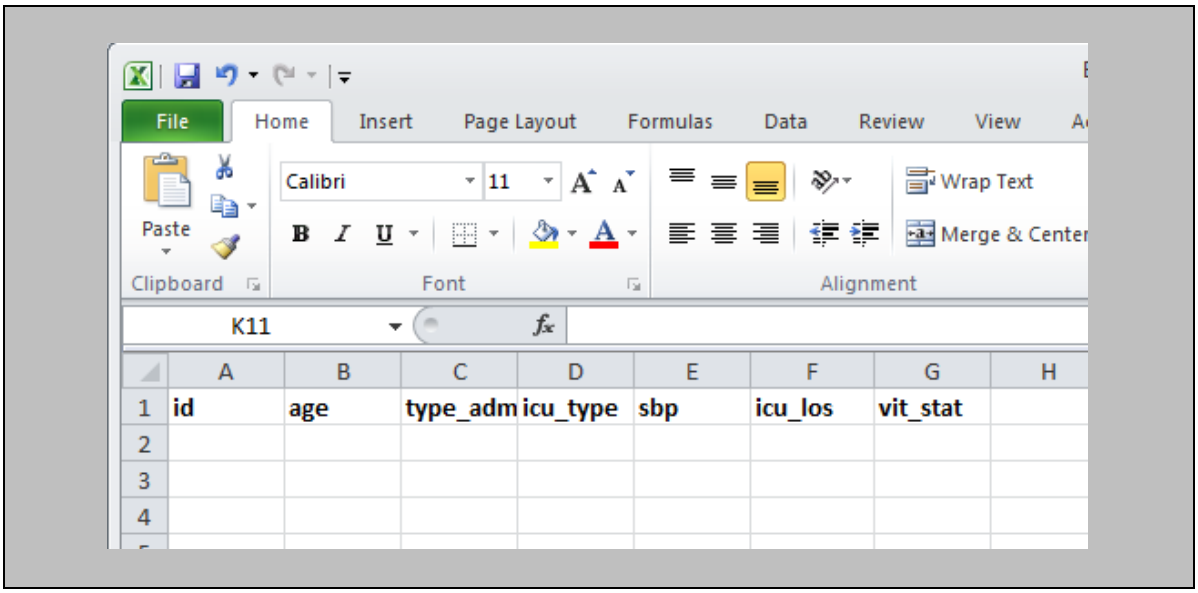
Step 1: Launch MS Excel

You should see an empty spreadsheet and the cell A1 with a bold border.
Cell A1 is the active cell; your cursor (you can't see it actually) is positioned here.



Step 2: Enter the column headings. Again, these are your fields and correspond to your variable names.

Proceeding horizontally across the first row, type the variable names in cells A1, B1, …, G1. Use the right arrow key after each entry so that your cursor moves right along the horizontal. You should now have the following.



Design ..... Data Collection ..... Data Management ..... Data Summarization ..... Statistical Analysis ..... Reporting

Step 3: Enter your data, column by column

To do this, begin by highlighting cell A2. Type a "1" in this cell (this is value of ID for the first record) and then press the down arrow key. Having pressed the down arrow key, you can now use the enter key after each data entry. When you are done, you should now have the following; **note – Only a partial picture is shown here**.

	A	B	C	D	E	F	G	H
1	id	age	type_adm	icu_type	sbp	icu_los	vit_stat	agedays
2	1	15	1	1	100	4	0	5478.75
3	2	31	1	2	120	1	0	11322.75
4	3	75	0	1	140	13	1	27393.75
5	4	52	0	1	110	1	0	18993
6	5	84	0	4	80	6	0	30681
7	6	19	1	1	130	2	0	6939.75
8	7	79	0	1	90	7	0	28854.75
9	8	74	1	4	60	1	1	27028.5
10	9	78	0	1	90	28	0	28489.5
11	10	76	1	1	130	7	0	27759
12	11	29	1	2	90	13	0	10592.25
13	12	39	0	2	130	1	0	14244.75
14	13	53	1	3	250	11	0	19358.25
15	14	76	1	3	80	3	1	27759
16	15	56	1	3	105	5	1	20454
17	16	85	1	1	145	4	0	31046.25

3.3 Formatting Fields, Field Names and Format Type

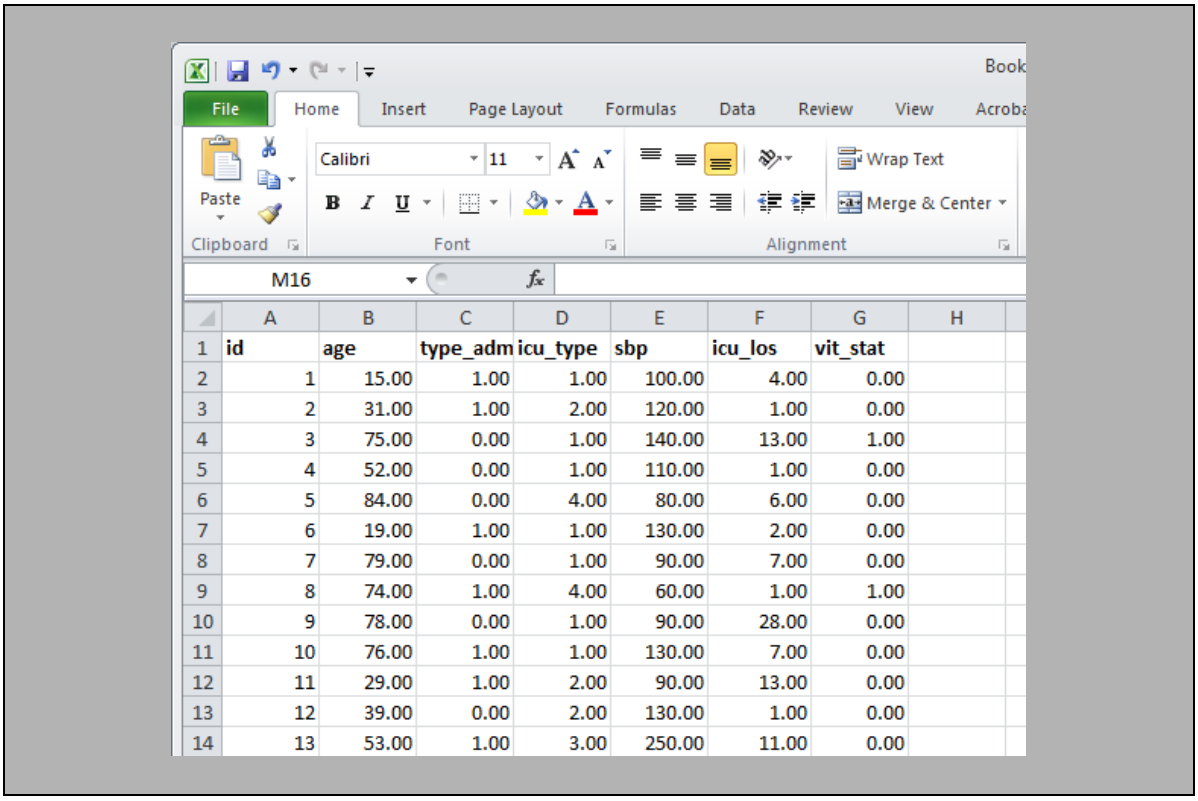
Step 4: Assign format types using instructions on page 13
From the **Cells** group of the **Home** tab, click on **FORMAT**. From the drop down menu, click **Format Cells**.

Example ("ICU Example")–

The following are reasonable choices. **Tip!** Note that, except for the variable ID, I chose to format each variable as numeric. This makes programming convenient, as it spares having to remember special conventions in working with character fields.

Column	Variable	Format cells category:	Notes
A	**ID**	General	
B	**AGE**	Number	At right in the decimal places box, choose "2"
C	**TYPE_ADM**	Number	
D	**ICU_TYPE**	Number	
E	**SBP**	Number	
F	**ICU_LOS**	Number	
G	**VIT_STAT**	Number	

You should now have the following; **note – A partial picture is shown here**.



	A	B	C	D	E	F	G	H
1	id	age	type_adm	icu_type	sbp	icu_los	vit_stat	
2	1	15.00	1.00	1.00	100.00	4.00	0.00	
3	2	31.00	1.00	2.00	120.00	1.00	0.00	
4	3	75.00	0.00	1.00	140.00	13.00	1.00	
5	4	52.00	0.00	1.00	110.00	1.00	0.00	
6	5	84.00	0.00	4.00	80.00	6.00	0.00	
7	6	19.00	1.00	1.00	130.00	2.00	0.00	
8	7	79.00	0.00	1.00	90.00	7.00	0.00	
9	8	74.00	1.00	4.00	60.00	1.00	1.00	
10	9	78.00	0.00	1.00	90.00	28.00	0.00	
11	10	76.00	1.00	1.00	130.00	7.00	0.00	
12	11	29.00	1.00	2.00	90.00	13.00	0.00	
13	12	39.00	0.00	2.00	130.00	1.00	0.00	
14	13	53.00	1.00	3.00	250.00	11.00	0.00	

3.4. Creating New Variables Using Formulae and Functions

Suggestion: As a general rule, it is recommended that you do *not* do variable creations in Excel. The advice is to reserve Excel file work for solely the raw data and that all new variable creation occurs in the software you are using for data analysis (e.g. R or Stata)

See again section 2.4, beginning on page 14.

Example ("ICU Example")–
Here we will illustrate the creation of a new variable. Suppose we want to create a new variable called AGEDAYS with the following definition:

AGEDAYS = AGE * 365.25

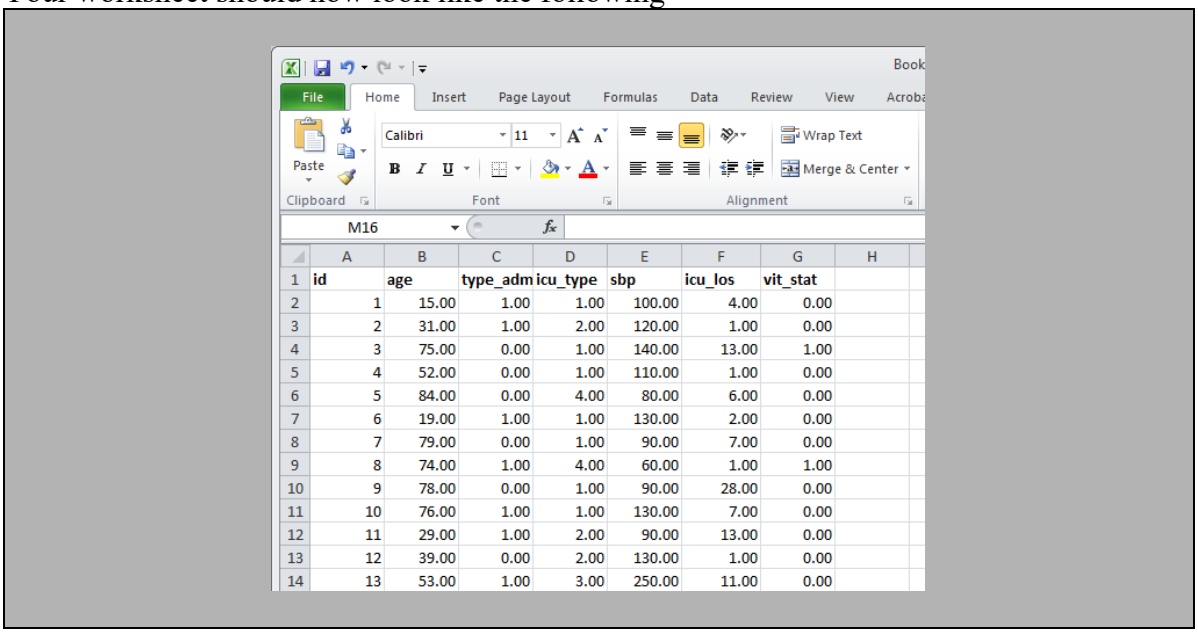
Step 5: Create AGEDAYS in Column H.
(1) In cell H1, enter the variable name **agedaysa**

(2) In cell H2, enter the calculation of agedays for the first record by typing
= B2*365.25 Press enter. You should see the result **5478.75** in cell H2

(3) Highlight cell H2. Copy the cell using the **Copy** button in the **Home** tab or CTRL-C. The border of cell H2 should now be dashed and vibrating!!

(4) Highlight cells H3 through H26. Choose **Paste** in the **Home** button tab or CTRL-V.

Your worksheet should now look like the following



	A	B	C	D	E	F	G	H
1	id	age	type_adm	icu_type	sbp	icu_los	vit_stat	
2	1	15.00	1.00	1.00	100.00	4.00	0.00	
3	2	31.00	1.00	2.00	120.00	1.00	0.00	
4	3	75.00	0.00	1.00	140.00	13.00	1.00	
5	4	52.00	0.00	1.00	110.00	1.00	0.00	
6	5	84.00	0.00	4.00	80.00	6.00	0.00	
7	6	19.00	1.00	1.00	130.00	2.00	0.00	
8	7	79.00	0.00	1.00	90.00	7.00	0.00	
9	8	74.00	1.00	4.00	60.00	1.00	1.00	
10	9	78.00	0.00	1.00	90.00	28.00	0.00	
11	10	76.00	1.00	1.00	130.00	7.00	0.00	
12	11	29.00	1.00	2.00	90.00	13.00	0.00	
13	12	39.00	0.00	2.00	130.00	1.00	0.00	
14	13	53.00	1.00	3.00	250.00	11.00	0.00	

3.5. Documentation with a Data Dictionary/Coding Manual

Include in your Excel file a worksheet that is a coding manual for the data. Document in the coding manual are variable names, labels, type, value labels and a notes column.

Tip! Be sure to include missing value codes.

Example –

I19 f_x

	A	B	C	D	E
1	**Coding Manual ICU Study (n=25)**				
2	version 9-20-2010				
3					
4	Variable	Label	Type	Coding	Remarks
5	**id**	Patient ID	character	1, 2, etc	
6	**age**	Age at admission, years	numeric	.=missing	
7	**type_adm**	admission type	numeric	1=emergency 0=elective .=missing	
8	**icu_type**	ICU type	numeric	1=medical 2=surgical 3=cardiac 4=other .=missing	
9	**sbp**	systolic blood pressure, mm Hg	numeric	.=missing	
10	**icu_los**	Length of Stay ICU, days	numeric	.=missing	
11	**vit_stat**	Status at Discharge	numeric	1=dead 0=alive .=missing	
12	**agedays**	New variable for laughs	numeric	age*365.25	
13					

Tip!! How to get the carriage returns *within* a cell -

Notice that the entry in cell D7 has carriage returns. This was done as follows.

(1) Position cursor in cell D7

(2) After typing 1=emergency, do NOT press the enter key. Instead, press **ALT-ENTER**

3.6. Saving and Exiting

Before exiting, let's give names to the worksheets and reorder them.

Recall how to name a worksheet
(1) Position cursor on the tab
(2) Right click
(3) Select **Rename**

Recall how to rearrange the worksheets
(1) Position cursor on the worksheet you want to move
(2) Right click
(3) Select **Move/Copy**

Example –