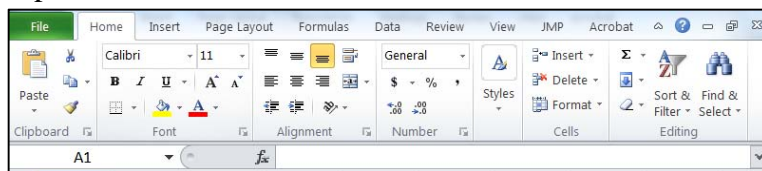


## Using Formulas, Functions, and Data Analysis Tools Excel 2010 Tutorial

Excel file for use with this tutorial	File Location
Tutor1Data.xlsx	<a href="http://faculty.ung.edu/kmelton/Data/Tutor1Data.xlsx">http://faculty.ung.edu/kmelton/Data/Tutor1Data.xlsx</a>

### Introduction:

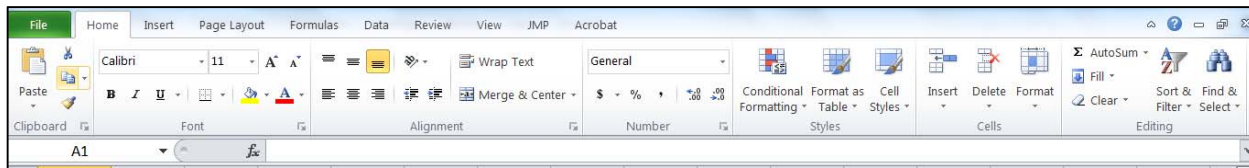
These instructions are designed for Excel 2010. When you open Excel, the “ribbon” at the top of the page will contain several tabs with options below that. The number of tabs and the layout of the items below the tab name will depend on the size of your screen and other options that you have installed on your machine. At a minimum, you should see the File, Home, Insert, Page Layout, Formulas, Data, Review, and View Tab. The next two figures show the contents of the Home tab for a smaller and a larger screen. Note the three sections as you move down the page: Tabs, Options, and the line for input.



← Tabs

← Options available for that Tab

← Formula Bar



We will move between Tabs to find the commands that we need, and we will use the Formula Bar to see what the data, formulas, and functions that are/were entered.

The tutorial will work with student grades to provide a context for the calculations.

### Formulas (and cell references):

Formulas allow us to do math with the values in our spreadsheet. In addition, we can use formulas combined with functions to complete more complicated calculations. The way that we signal Excel that we are entering a formula or a function is to start our entry with an =. Formulas involve addition (+), subtraction (-), multiplication (\*), division (/), and exponentiation (^).

Excel knows the order of precedence for mathematical operations, so you will need to be careful when you enter formulas. For example: If you tell Excel to find  $2 + 3 * 4$  by typing  $=2+3*4$ , Excel will recognize that multiplication is completed prior to addition and will find  $3*4$  first and then add 2 to get 14. If you wanted Excel to do the addition first, you would need to say  $=(2+3)*4$ ; then Excel would return a value of 20. The phrase “Please Excuse My Dear Aunt Sally” will help you remember the order as parentheses first, followed by exponentiation, multiplication, division, addition, and finally subtraction.

Open the Tutor1Data. You should see the grades for JJ Jones in Row 1. During the semester, JJ took two tests, a final, and two quizzes. Final grades have not been posted, so JJ is left to calculate the course average. Since there are five grades, JJ decides to use Excel to add up the grades and divide by five. JJ clicks on cell H2 so that the result will appear in that cell. The “hard” way to get the answer (and the way that is more error prone) would be to enter the following:

$$=(68+96+64+90+84)/5 \text{ and hit enter. Excel would return 80.4.}$$

A better way would be to use cell references instead of typing the actual grades:

$$=(B2+C2+D2+E2+F2)/5 \text{ and hit enter. Excel still returns 80.4.}$$

You may think that typing cell references is more prone to error than typing the grades; but you don't need to type the cell references, instead you can click on the cell when you want to use the value in that cell. So you would type =(, then you would click on cell B2, type +, click on C2, type +, click on D2, type +, click on E2, type +, click on F2, type )/5 and hit enter.

PMT									
= (B2+C2+D2+E2+F2)/5									
A	B	C	D	E	F	G	H	I	J
Student ID	Test 1	Test 2	Final	Quiz 1	Quiz 2				
JJ Jones	68	96	64	90	84		= (B2+C2+D2+E2+F2)/5		

The advantages of using cell references are 1) Excel highlights which cells were selected, 2) if the value in one of the cells changes, the results are updated when the value is revised and 3) you can duplicate the action to other similar cases.

Remembering that the syllabus says that each test counts 25%, the final counts 30% and each quiz counts 10%, JJ needs use a weighted average rather than the simple average done above. This could be done as:

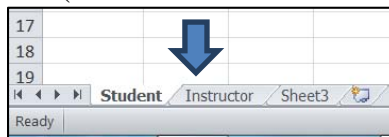
$$=.25*B2+.25*C2+.3*D2+.1*E2+.1*F2 \text{ and hit enter to see 77.6.}$$

(note that you must use the \* to signify multiplication)

Or... the weights could be placed in another row and then, the formula could click on the cells:

PMT									
=B3*B2+C3*C2+D3*D2+E3*E2+F3*F2									
1	A	B	C	D	E	F	G	H	I
2	Student ID	Test 1	Test 2	Final	Quiz 1	Quiz 2			
2	JJ Jones	68	96	64	90	84		80.4	
3	Weights	0.25	0.25	0.3	0.1	0.1		=B3*B2+C3*C2+D3*D2+E3*E2+F3*F2	

Now change the perspective slightly—from a single student's perspective to the instructor's perspective. Go to the bottom of the spreadsheet and click on the word “Instructor” to move to a sheet with grades for each student in the class (note: JJ Jones is Student 13).



The first student's average can be obtained in the same way that JJ's was on the previous sheet:

PMT									
=B1*B3+C1*C3+D1*D3+E1*E3+F1*F3									
1	A	B	C	D	E	F	G	H	I
2	Weights	0.25	0.25	0.3	0.1	0.1			
3	Student ID	1	78	78	76	87	77	=B1*B3+C1*C3+D1*D3+E1*E3+F1*F3	

The same formula could be entered for each student...but this would not take advantage of Excel's capabilities. Basically, we want to do the same thing for each student—multiply the grade by the weight for each assignment and then add these up to obtain the average. Excel allows us to copy and paste formulas rather than entering them over and over. If we copy the formula that is in H3 into H4, we see #VALUE in the cell; and if we click on H3 and then the formula bar, we see the following:

	A	B	C	D	E	F	G	H
1	Weights	0.25	0.25	0.3	0.1	0.1		
2	Student ID	Test 1	Test 2	Final	Quiz 1	Quiz 2		Average
3	1	78	78	76	87	77		78.2
4	2	76	71	75	79	85		#VALUE!

This is not quite what we want. Excel has used the correct columns but has increased the row reference by one for every entry in the formula—it has picked the correct grades, but not the weights. So far the cell references that we have used (e.g., B1, B3) have been what are called “relative references.” To us, the formula that produced 78.2 looked like  $=B1*B3+C1*C3+D1*D3+E1*E3+F1*F3$ —but Excel interpreted this to get the answer that goes in H3 multiply the entry that is six columns to the left of H3 and two rows up from H3 by the entry that is six columns to the left H3 in the same row as H3 and do the same thing with each of the other terms in the equation. When the same logic is used for H4, two rows above H4 is row 2 so Excel tries to use the entries in Row 2 as the weights; and Excel uses the grades for the new student since they are in Row 4. We really want Excel to go to Row 1 to get weights for every student. To force Excel to see a reference as an absolute spot on the page, we need to use an “absolute reference.” The \$ in front of a reference signals Excel that we are using an absolute reference. If we want to force Excel to go to cell B1, we would say \$B\$1. Note: We don't have part of the reference can be absolute and part relative. For example, in this problem we don't need to make the column reference absolute—just the row. So we need B\$1, C\$1, ... when we are talking about the weights.

If we go back and correct our formula for Student 1 as follows:

	A	B	C	D	E	F	G	H
1	Weights	0.25	0.25	0.3	0.1	0.1		
2	Student ID	Test 1	Test 2	Final	Quiz 1	Quiz 2		Average
3	1	78	78	76	87	77		78.2
4	2	76	71	75	79	85		#VALUE!

Copying the formula to row 4 will produce the desired result:

	A	B	C	D	E	F	G	H
1	Weights	0.25	0.25	0.3	0.1	0.1		
2	Student ID	Test 1	Test 2	Final	Quiz 1	Quiz 2		Average
3	1	78	78	76	87	77		78.2
4	2	76	71	75	79	85		77.2

Once we are sure that we have the formula correct, we can copy the formula down the column and have the average for all 50 students in the class. The easiest way to copy the formula is to click on the

first cell with the formula (H3), move our cursor to the lower right corner of the cell until we see a black plus sign, and then hold the mouse button down as we drag down the column.

	A	B	C	D	E	F	G	H
1	Weights	0.25	0.25	0.3	0.1	0.1		
2	Student ID	Test 1	Test 2	Final	Quiz 1	Quiz 2		Average
3	1	78	78	76	87	77		78.2
4	2	76	71	75	79	85		75.65

We see that all of the averages have been calculated, and note that Student 13's average matches what we found when we worked with it on the other sheet.

	A	B	C	D	E	F	G	H
1	Weights	0.25	0.25	0.3	0.1	0.1		
2	Student ID	Test 1	Test 2	Final	Quiz 1	Quiz 2		Average
3	1	78	78	76	87	77		78.2
4	2	76	71	75	79	85		75.65
5	3	92	96	88	87	84		90.5
6	4	88	77	73	81	73		78.55
7	5	72	98	71	82	83		80.3
8	6	61	72	65	77	75		67.95
9	7	78	85	73	86	82		79.45
10	8	87	81	80	92	72		82.4
11	9	63	80	72	86	84		74.35
12	10	71	87	79	75	79		78.6
13	11	79	81	77	87	89		80.7
14	12	64	82	72	91	69		74.1
15	13	68	96	64	90	84		77.6
16	14	67	81	82	86	85		78.7
17	15	67	93	78	79	86		79.9

## Functions:

For those of you with more experience in Excel, you probably recognize that there are easier ways to find the quantities that we just found. Functions are commands that are built into Excel that allow some of the more common tasks to be completed without having to “program” each step. Some of the most common statistical functions help us find statistics such as the mean, median, mode, standard deviation, minimum value, and maximum value. We also used Excel functions to find probabilities and values of a random variable associated with a given probability. Excel functions also allow for logical comparisons (e.g., determining if a value meets some characteristic or not).

Go back to the Student sheet (click on Student at the bottom left of the page). When JJ wanted to find the unweighted average of the grades, we could have used the “sum” and “count” functions in Excel to find the two numbers we needed to calculate the average; or we could have skipped all of the intermediate steps by using the “average” function. The following functions would be used:

=sum(B2:F2) to find the sum of the values in cells B2 through F2

=count(B2:F2) to determine the number of observations

And we could combine the sum and count functions as =sum(B2:F2)/count(B2:F2) to find the average; or bypass all of this to find the average with:

=average(B2:F2) to find the unweighted average of the values in cells B2 through F2

Likewise, we could use the “sumproduct” function to find the weighted average. This multiplies corresponding entries from two vectors and then adds the results. The formula here is:

=sumproduct(B3:F3,B2:F2)


The spreadsheet with formulas and with results are shown below:

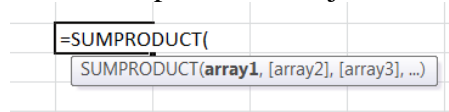
	A	B	C	D	E	F	G	H	I	J	K
1	Student ID	Test 1	Test 2	Final	Quiz 1	Quiz 2					
2	JJ Jones	68	96	64	90	84		=(B2+C2+D2+E2+F2)/5	Unweighted	=SUM(B2:F2)/COUNT(B2:F2)	=AVERAGE(B2:F2)
3	Weights	0.25	0.25	0.3	0.1	0.1		=B3*B2+C3*C2+D3*D2+E3*E2+F3*F2	Weighted	=SUMPRODUCT(B3:F3,B2:F2)	

	A	B	C	D	E	F	G	H	I	J	K
1	Student ID	Test 1	Test 2	Final	Quiz 1	Quiz 2					
2	JJ Jones	68	96	64	90	84		80.4	Unweighted	80.4	80.4
3	Weights	0.25	0.25	0.3	0.1	0.1		77.6	Weighted	77.6	

Moving to the Instructor sheet, we could use the sumproduct function with absolute references for the weights to find all of the students' averages with a single formula copied down the column. Since cell I3 is the active cell, the formula bar below shows the function for this entered for this cell:

	A	B	C	D	E	F	G	H	I	J
1	Weights	0.25	0.25	0.3	0.1	0.1				using sumproduct
2	Student ID	Test 1	Test 2	Final	Quiz 1	Quiz 2		Average	Average	
3	1	78	78	76	87	77		78.2	78.2	
4	2	76	71	75	79	85		75.65	75.65	
5	3	92	96	88	87	84		90.5	90.5	

There are hundreds of functions included in Excel and a few ways to get help. If you click on the cell where you want the result and then click on the  in the formula row, this will bring up a dialogue box that will help identify functions and select the data. Also, if you know the first few letters of the function name, you can click in the destination cell, type = and start typing the name of the function. Excel will provide a list of functions that starts with the letters that you have typed. Once you select a single function, Excel will provide hints just below where you are typing:



Some of the most common Excel 2010 functions we will use are:

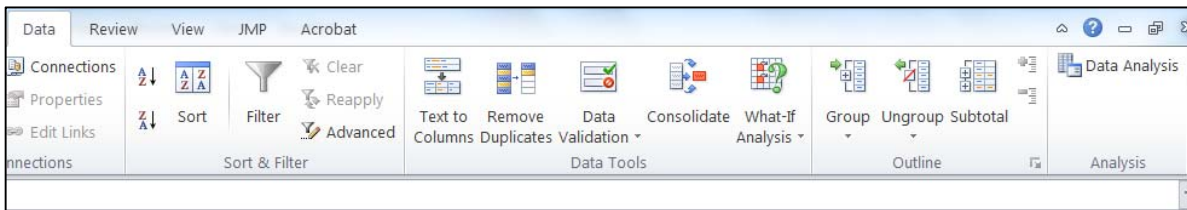
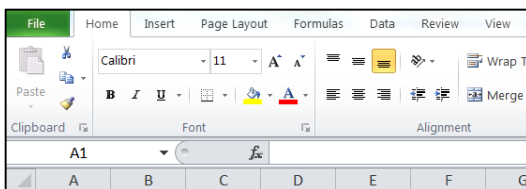
- Summary statistics: =average() =sum() =count() =stdev.s() =min() =max()
- Working with the Standard Normal Distribution (Z): =norm.s.dist() =norm.s.inv()
- Working with other Normal Distributions: =norm.dist() =norm.inv()
- Working with the t Distribution: =t.dist() =t.inv()
- Working with the F Distribution: =f.dist.rt()
- For comparisons: =if()

**Caution:** There are some functions that were new to Excel 2010 (most include the . in the name) that look very much like functions from earlier versions of Excel but some provide VERY different results! The older function names have a yellow triangle next to them in the drop down suggestions provided by Excel 2010. See example below:

Quiz 1	Quiz 2	Quiz 3	Quiz 4	Average	Average
78	76	87	77	78.2	78.2
71	75	79	85	75.65	75.65
96	88	87	84	90.5	90.5
77	73	81	73	78.55	78.55
This function is available for compatibility with Excel 2007 and earlier. Estimates standard deviation based on a sample (ignores logical values and text in the sample)					
85	73	86	82	79.45	79.45

## Data Analysis Tools

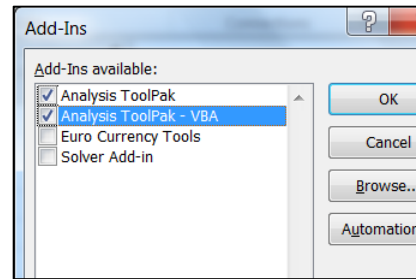
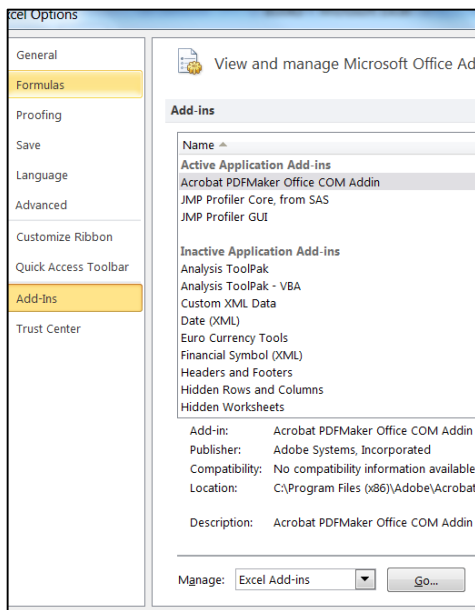
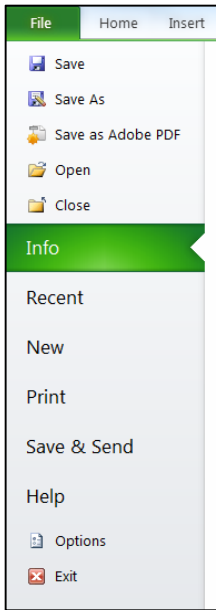
The Data Analysis Tools will be key to the work we do in the regression part of the course and these tools are **not** available in the Mac version of Excel. Even in a Windows environment, they may not appear to be available...but they are there! Depending on how Excel is installed on your machine, you may (or may not) see the Data Analysis tools initially. To see if they are visible on your computer, click on the Data tab at the top of the screen and look to the far right of the ribbon. If you see Data Analysis (as you do in the second figure below), you are ready to go.



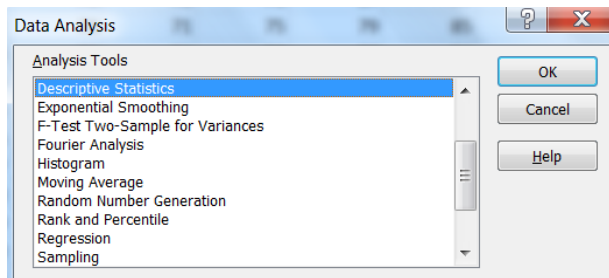
If the Data Analysis tools are not visible, you will need to make them visible. The good news is that you should not have to do this but one time—once you make them visible they should be there when you return to Excel in the future (on the same computer).

Note: There are three screen images on the next page that accompany the first five steps listed here.

- Step 1: Go to the File Tab and select Options near the bottom of the list.
- Step 2: Click on Add-ins in the list on the left.
- Step 3: Look at the bottom of the right frame of the Excel Options dialogue box and make sure that Excel Add-ins is selected in the “Manage:” box, then click Go...
- Step 4: In the next dialogue box, check the boxes next to Analysis ToolPak and Analysis ToolPak – VBA. Leave the other boxes the way you found them.
- Step 5: Click OK
- Step 6: Go back to the Data tab and confirm that the Data Analysis tools are listed on the far right.

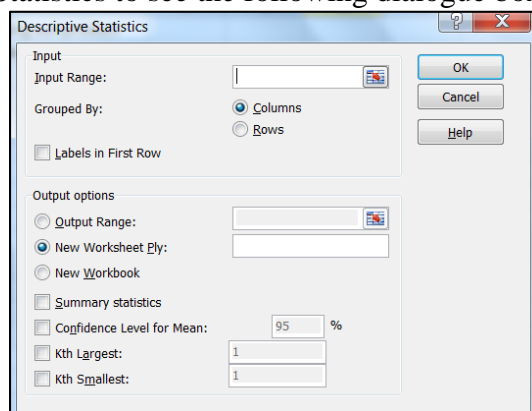


When you click on Data Analysis on the Data tab, you will see a dialogue box with a number of statistical tools. The ones that we will use most are Descriptive Statistics, Histogram, and Regression.



To illustrate one of these tools, let's go back to the grades data. Suppose the instructor wanted to provide some summary statistics to describe how students did on each of the graded assignments during the semester. We could use lots of formulas or lots of functions; but we can get the information much easier by using the Descriptive Statistics option in the Data Analysis tools.

Go to the Instructor sheet (accessed through the lower left of the spreadsheet), and then select Data Analysis and the Descriptive Statistics to see the following dialogue box.

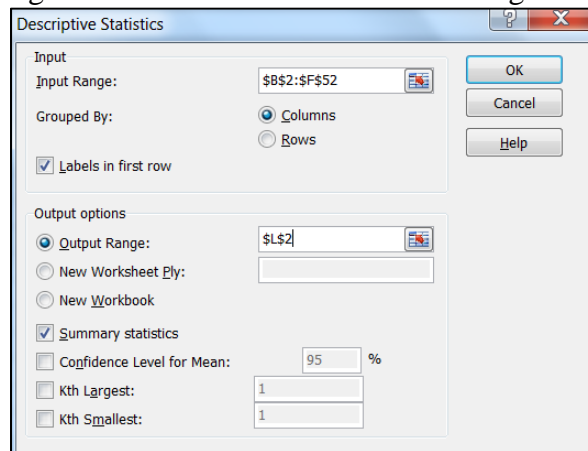


Using this approach we can obtain summary statistics for all five graded assignments at the same time. For input range, we will click in the white box and then drag over cells B2 through F52. This includes the variable names and the grades for all 50 students on each of the assignments.

Since each column includes the data for one graded assignment, the selection of Grouped by Columns is correct. The box next to Labels in the First Row is where you tell Excel that you have the variable names in the first row that you selected—be sure to check it since you did select the variable names. [If you fail to select this box and you have variable names in the first row that you selected, Excel will give an error message. If you selected the data without selecting the variable names (not recommended), do not check the box next to Labels in the First Row unless you want the calculations to be wrong!]

Next, you need to tell Excel where to put the output. I like to put my output on the same sheet as my data. To do this, click the button to the left of “Output Range” and click in the white box to the right of “Output Range.” You designate the cell that will be used as the upper left corner for the output in the white box. You can do this by clicking on the destination cell. For this example, I am using cell L2.

Finally, remember to tell Excel that you want the Summary Statistics by clicking to the left of this option. Your completed dialog box should look like the following. If so, click OK.



	L	M	N	O	P	Q	R	S	T	U
	<i>Test 1</i>		<i>Test 2</i>		<i>Final</i>		<i>Quiz 1</i>		<i>Quiz 2</i>	
Mean	74.38	Mean	84.78	Mean	74.06	Mean	86.02	Mean	80.04	
Standard Error	1.161944	Standard Error	1.149371	Standard Error	0.675767	Standard Error	0.678829	Standard Error	0.787639	
Median	73.5	Median	84.5	Median	74.5	Median	87	Median	80	
Mode	78	Mode	80	Mode	72	Mode	92	Mode	77	
Standard Deviation	8.216186	Standard Deviation	8.127278	Standard Deviation	4.778395	Standard Deviation	4.800043	Standard Deviation	5.56945	
Sample Variance	67.50571	Sample Variance	66.05265	Sample Variance	22.83306	Sample Variance	23.04041	Sample Variance	31.01878	
Kurtosis	-0.41706	Kurtosis	-0.54164	Kurtosis	0.271135	Kurtosis	0.35744	Kurtosis	-0.61308	
Skewness	0.416531	Skewness	-0.02559	Skewness	0.206846	Skewness	-0.7694	Skewness	-0.21564	
Range	33	Range	35	Range	24	Range	21	Range	24	
Minimum	60	Minimum	65	Minimum	64	Minimum	72	Minimum	68	
Maximum	93	Maximum	100	Maximum	88	Maximum	93	Maximum	92	
Sum	3719	Sum	4239	Sum	3703	Sum	4301	Sum	4002	
Count	50	Count	50	Count	50	Count	50	Count	50	



Some observations about this data:

- The highest individual score was scored on Test 2.
- The highest class average was on Quiz 1.
- The lowest individual score was on Test 1.
- The lowest class average was on the Final.
- The biggest difference between the highest and lowest individual scores was on Test 2 and the smallest difference was on Quiz 1.
- The most variability in scores was on Test 1 and the least was on the Final.

Some words about the using this approach:

- Excel assumes that the data are from a sample (rather than an population).
- If there are missing values—cells with no input, these cells will be ignored.
- If there are cells with non-numeric entries—other than the variable names, you will receive an error message.
- If you selected the data without selecting the variable names and you checked the box next to the Labels in First Row, the data in the first row will be ignored and the results will be calculated with the rest of the rows—giving incorrect answers!
- If you go back and change a number in the original data set, these answers will NOT be updated. You must run the analysis again if you want to update the results.
- If there are multiple modes, Excel will not let you know this...and the order that the data are entered can impact which mode is reported!