Statistical Analysis Guidelines Using Excel

Basic Analysis of Data Using Excel For Phase II-A Projects

Who Should Use This Guideline?

This guideline should be used by projects that are in Phase II-A. These are projects that meet the following criteria:

- In the growth phase of development
- Have basic quantitative measures of knowledge, attitudes or behavior
- Have measured participants pre and post program
- Have not matched pre-post scores
- Have at least 20 units of observation

It is assumed that the user has a basic familiarity with Microsoft Excel, including how to enter data into columns, how to format a worksheet, and how to do simple formulas.

Introduction

These guidelines describe how Microsoft Excel can be used to accomplish a basic analysis of the pre-post quantitative data for a project where it is assumed the pretest and posttest scores are not matched for participants.

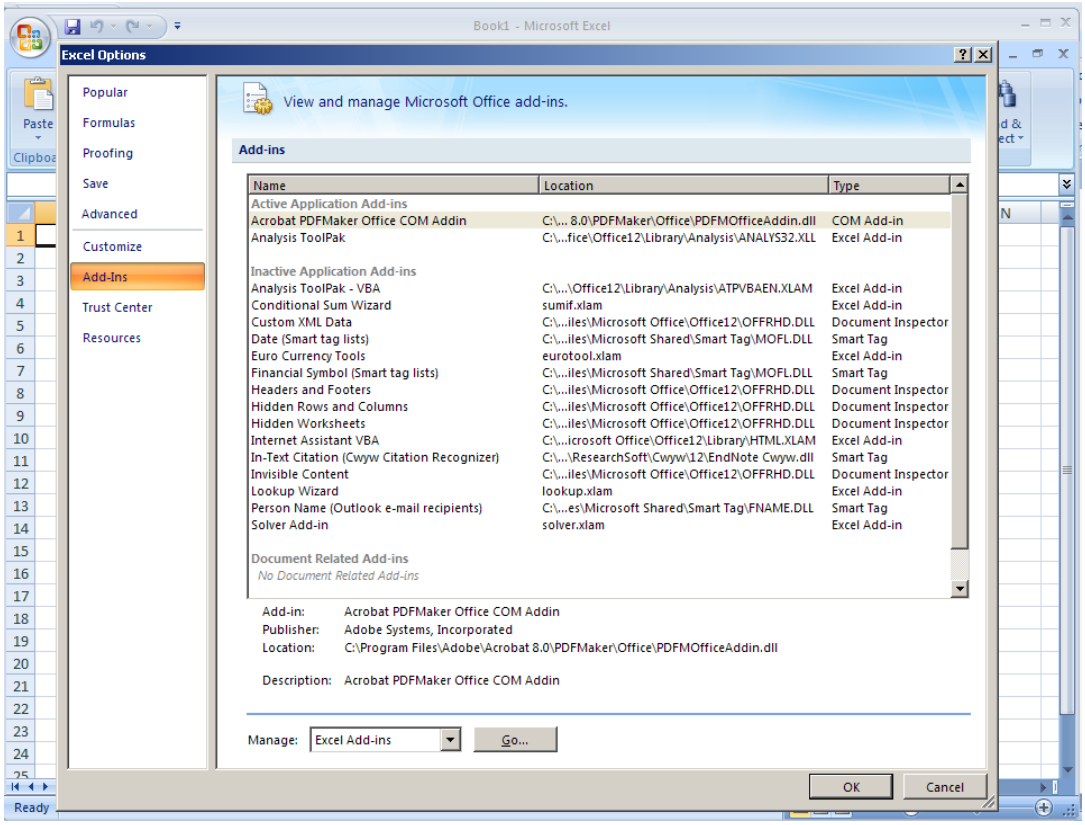
The guidelines describe how to:

- set up Excel for basic statistical analysis
- enter basic data into an Excel spreadsheet
- obtain descriptive statistics
- draw a histogram for each variable
- conduct a statistical test to assess change

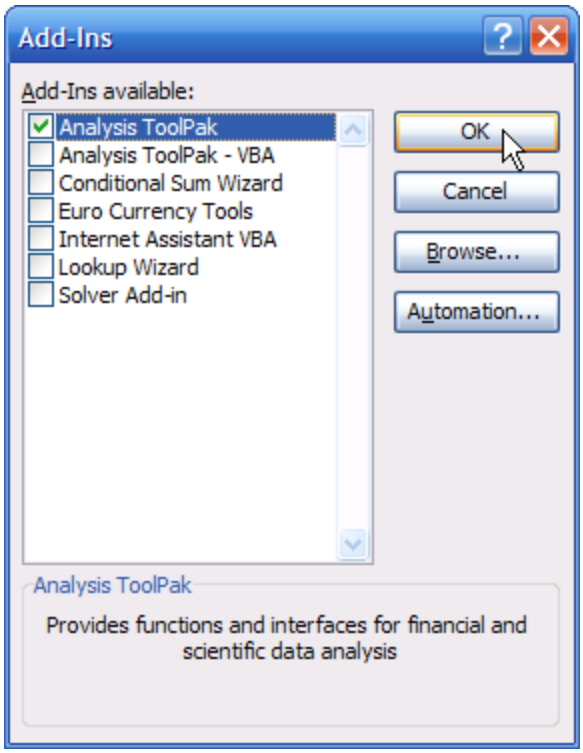
Installing the Analysis ToolPak

Before you can do the analyses described here, you have to first install the Analysis ToolPak that comes with Excel. How do you know if you already have it installed? Open Excel, select the Office button in the top-left corner, select Excel Options, and click Add Ins. If you do not have Analysis ToolPak under Active Application Add Ins then you must install it. If you do, then the Analysis ToolPak has already been added to your version of Excel. To install Analysis ToolPak, follow these steps:

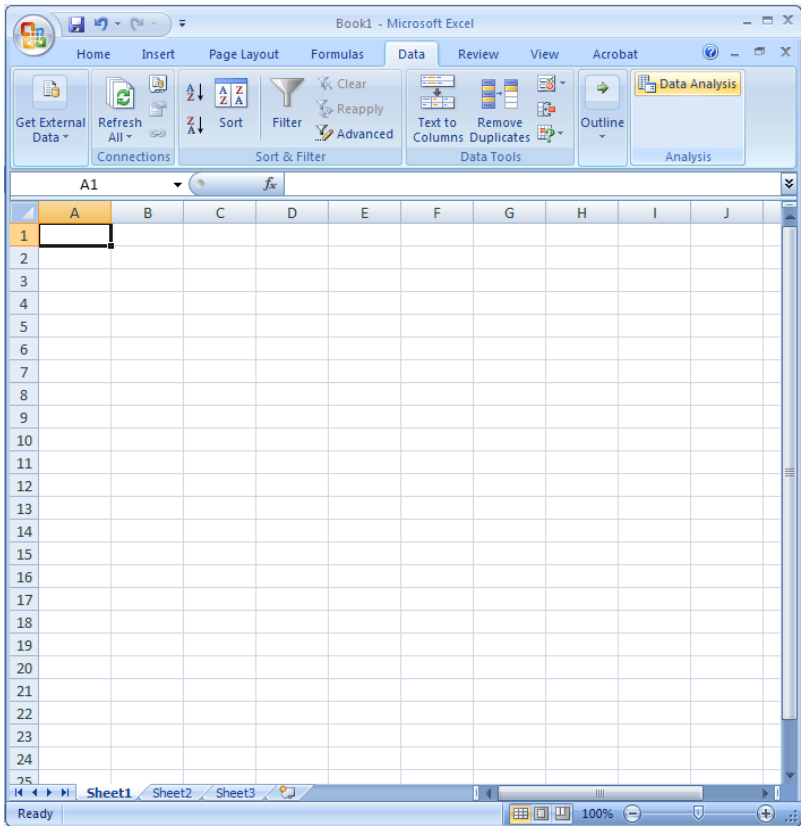
1. Select the Analysis Toolpack option under Inactive Application Add Ins from the Add Ins tab of Excel Options.
2. Make sure that the Excel Add Ins option is selected under the Manage option at the bottom of the box. Click go.



2. Check the Analysis ToolPak item and click OK



To verify that you have added the Analysis ToolPak correctly, select the Data tab. You should now see a Data Analysis… menu item.



Enter Data Into Spreadsheet

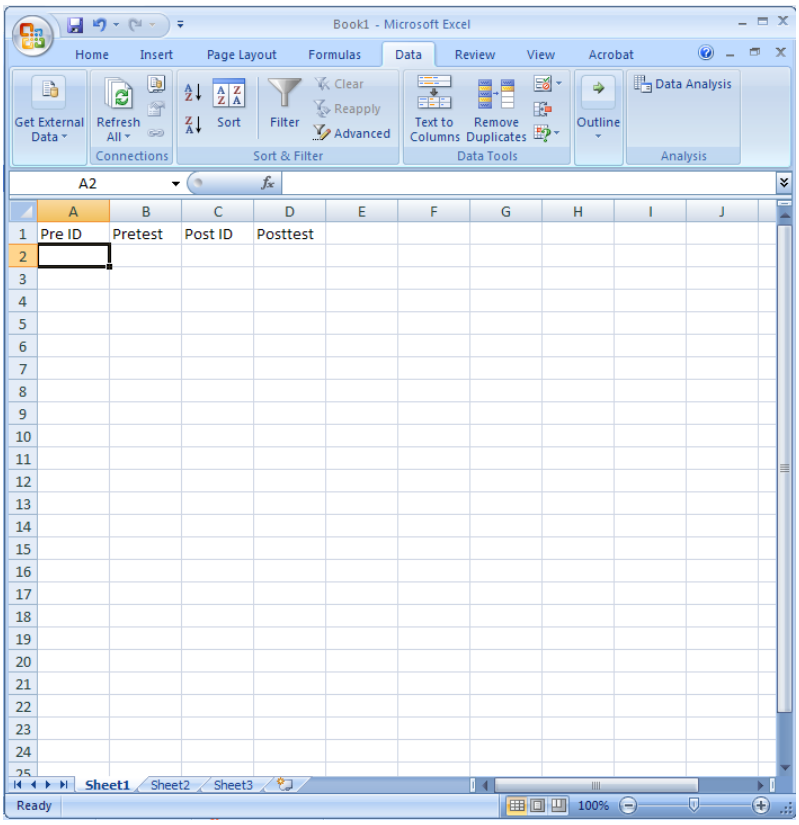
These guidelines only show how to analyze the pretest and posttest data for your project. You may have other data that you would also like to analyze. For instance, you may want to look at descriptive statistics for each response on a measure, explore demographics or your respondents, or do a qualitative analysis of open-ended responses. These tasks are not addressed here. These guidelines simply show you how to analyze your pretest and posttest total scores for your project.

Before entering the data into the spreadsheet we assume that you have already cleaned the data (identified and corrected any inaccurate or missing data- see *Tips for data collection and analysis*), assigned each respondent an ID number (and written it on their instrument), and computed the total pretest score and posttest score for each respondent. For instance, let's say you give a 10-item test of knowledge relevant to your program. We assume that you have scored the answers and have obtained a total score for each person (e.g., a score between 0 and 10).

You will only enter data for respondents from whom you have completed data. If a respondent did not complete the test, you will not enter their data into the spreadsheet.

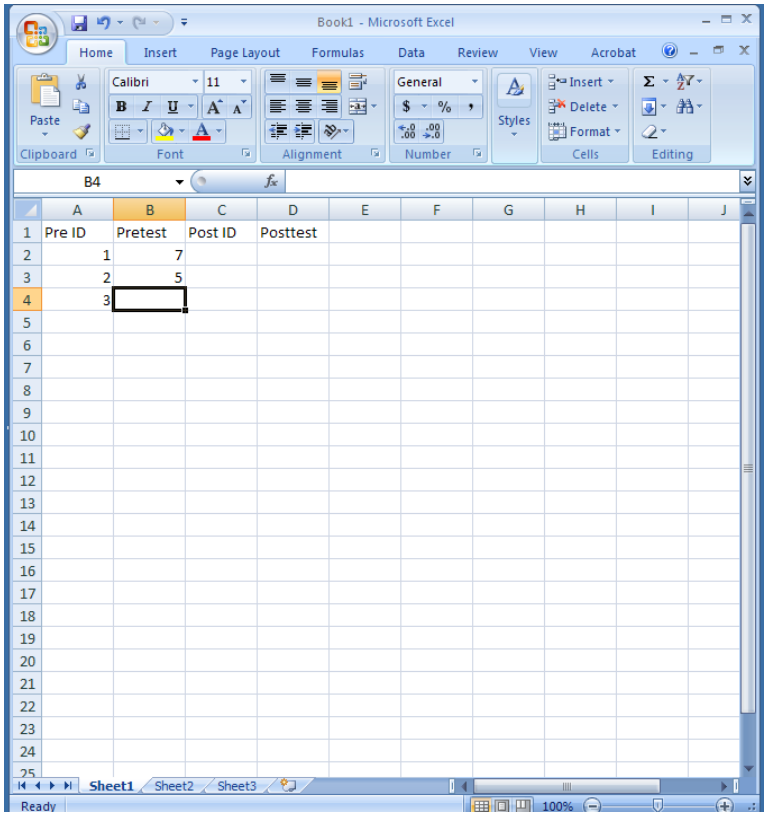
To enter your data into the spreadsheet, do the following:

1. Enter in a label for Pre ID, Pretest, Post ID and Posttest in the first row of the first four columns.

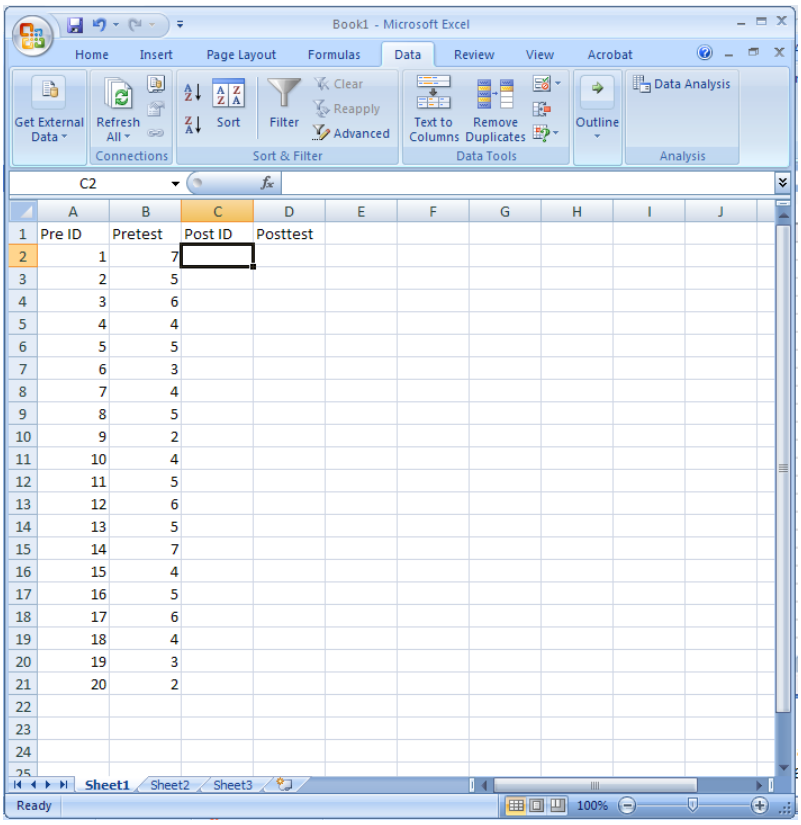


The ID column will be used to enter in the ID number for each respondent. You should make sure this ID number corresponds with the number on their instrument. By doing this, you will assure that you will always be able to track a number in the Excel spreadsheet directly back to the instrument itself.

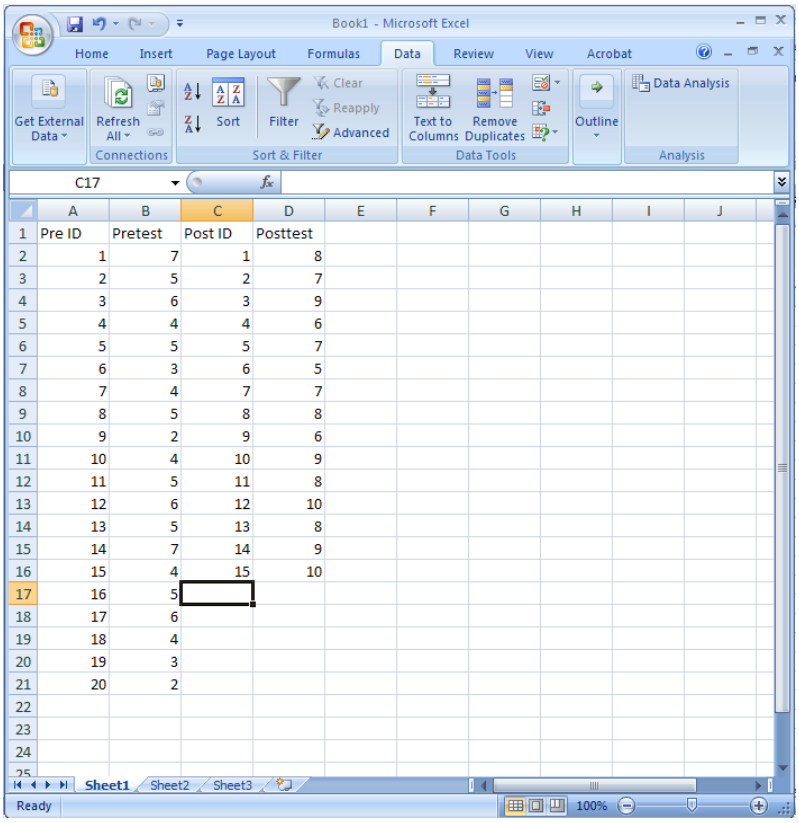
2. For each respondent, enter in their data. For example, let's say the first respondent (ID=1) got a score of 7 out of 10 on the pretest. Put their ID number in the Pre ID and their score in the Pretest columns as shown below. In the example, the second respondent got a score of 5 on the pretest. Continue entering all of the pretest scores.



Imagine that you had pretest scores for 20 participants. Your spreadsheet might look like the following:



Now you can enter in the posttest scores in the next two columns. For instance, you might only have 15 posttest respondents (you may have lost a few people over the course of the program). The complete data might look like the following:

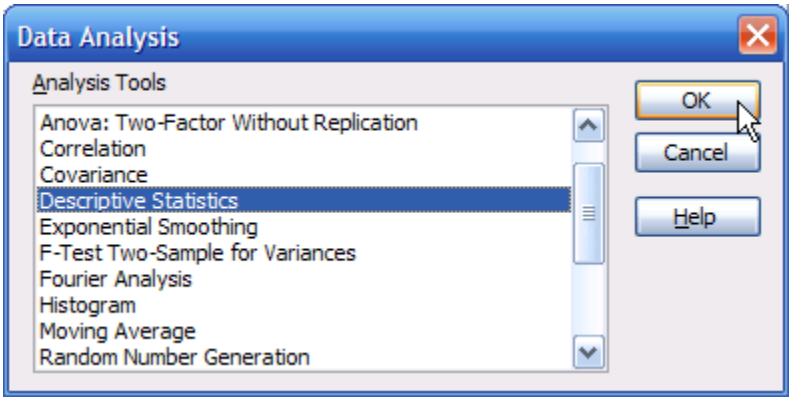


That's all you need to do to enter the data. If you haven't already, you should save the data. Go to File → Save and give the Excel spreadsheet a filename.

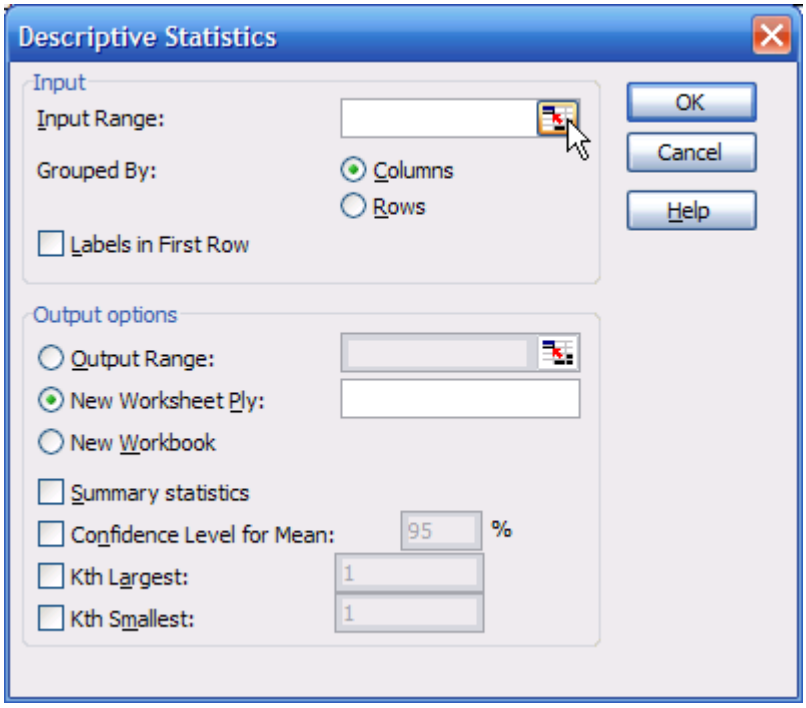
Compute Descriptive Statistics

The first step in analyzing the data is to obtain basic descriptive statistics on both the pretest and posttest. Let's begin with the Pretest variable first. To get descriptive statistics for the pretest, do the following:

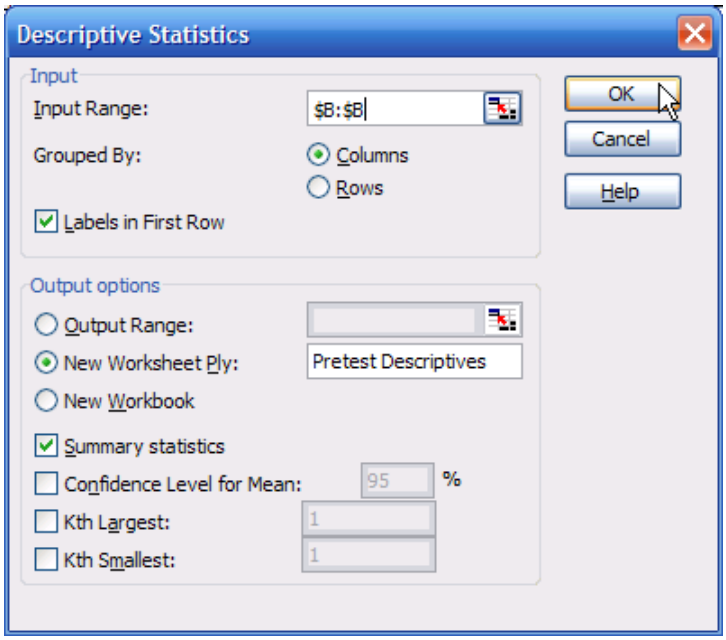
1. Select the Data menu and the Data Analysis… item. A Data Analysis selection box will pop up. Click on Descriptive Statistics and then click OK.



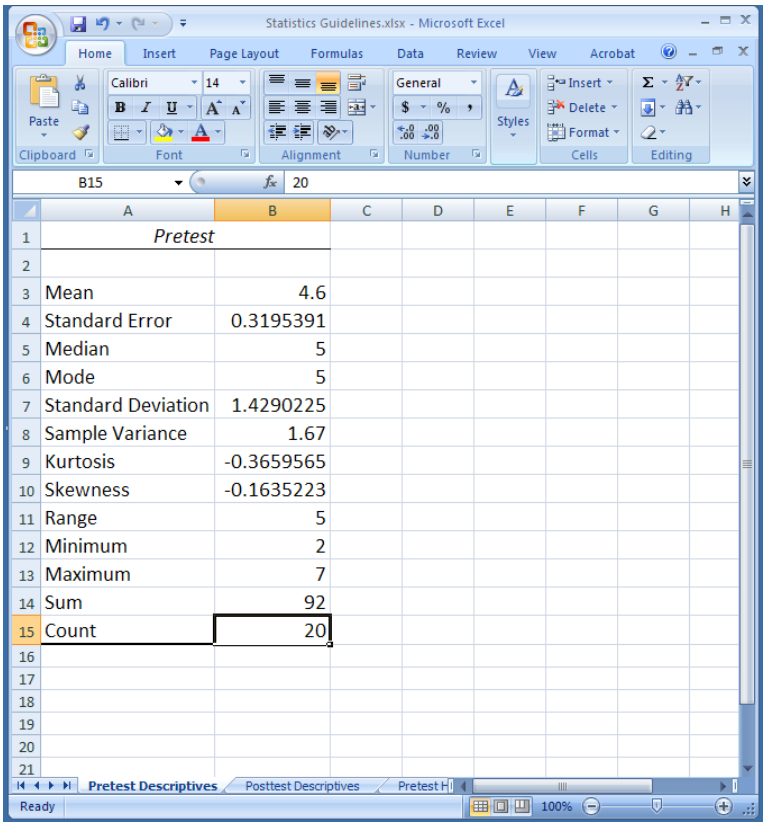
2. On the Descriptive Statistics pop-up window, we have to first specify the input we want analyzed – the pretest column. To do this click on the little spreadsheet icon just to the right of the Input Range box as shown below.



The Descriptive Statistics window will reduce. You can now click on the top of the first column (on the letter B) and then hit the enter key. You should once again see the Descriptive Statistics window, but now it will have the Input Range filled in.



Now, fill in the rest of the box as shown above. Check the 'Labels in First Row' box, select the 'New Worksheet Ply' button and type in the name 'Pretest Descriptives' for the worksheet, and click on the Summary Statistics box. When you have done all this, click OK. You will see a new worksheet pop up with your descriptive statistics on it. It will look something like the following:



	A	B
1	*Pretest*	
2		
3	Mean	4.6
4	Standard Error	0.3195391
5	Median	5
6	Mode	5
7	Standard Deviation	1.4290225
8	Sample Variance	1.67
9	Kurtosis	-0.3659565
10	Skewness	-0.1635223
11	Range	5
12	Minimum	2
13	Maximum	7
14	Sum	92
15	Count	20

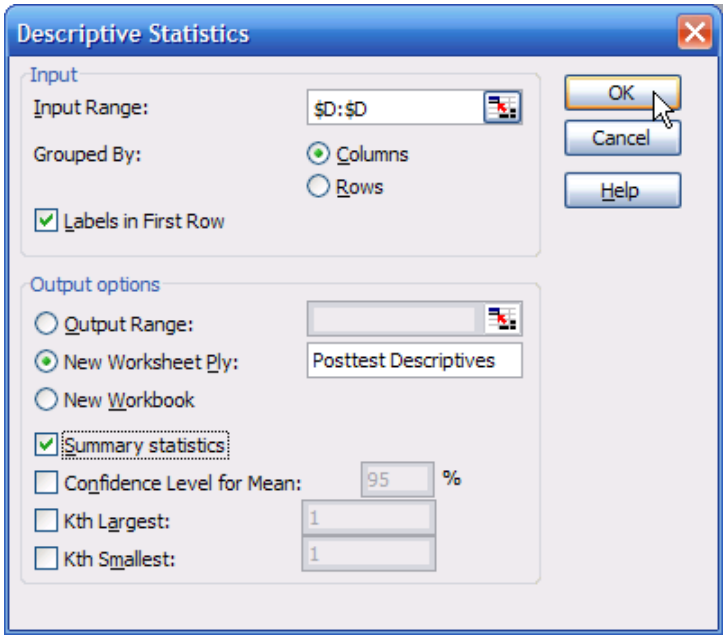
I've made the first column (column A) wider by dragging the vertical divider between column A and B to the right a bit. That way you can see the complete label for each statistic. In this example, we see that for the pretest the Mean (or Average) was 4.6, there were 20 participants (Count = 20) and the data ranged from a Minimum of 2 to a Maximum of 7.

Notice the tabs at the bottom of the screen. The currently selected tab is the one labeled Pretest Descriptives. Where did your data go? They are on the tab labeled Sheet 1. To get back to your data, click on the Sheet 1 tab.

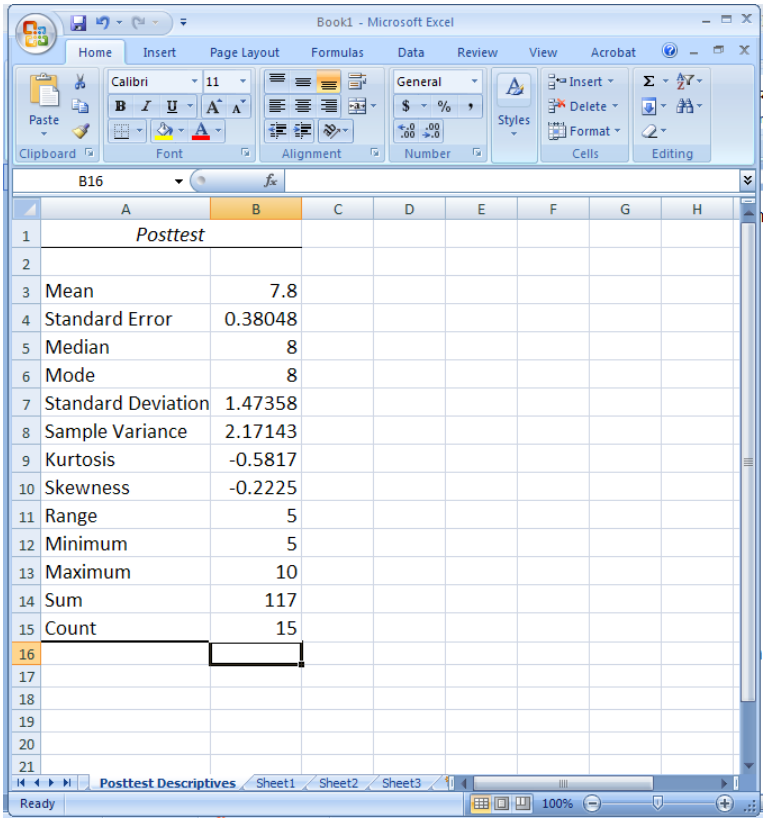
3. Now, let's do the descriptive analysis for the Posttest scores. Be sure you are on the Sheet 1 tab and can see your data. Also, you may want to save your Excel file again (go to File → Save).

The steps for the posttest are identical to the pretest steps, except that you will specify the posttest column for the input range, and name the worksheet ply 'Posttest Descriptives.' Select Tools → Data Analysis… and click on the Descriptives Statistics box and click OK.

After you select the posttest column for the Input Range, and rename the worksheet ply to say 'Posttest Descriptives' your screen should look like the following.



Once you have confirmed that the specifications are correct, hit OK. Your results should look something like the following:



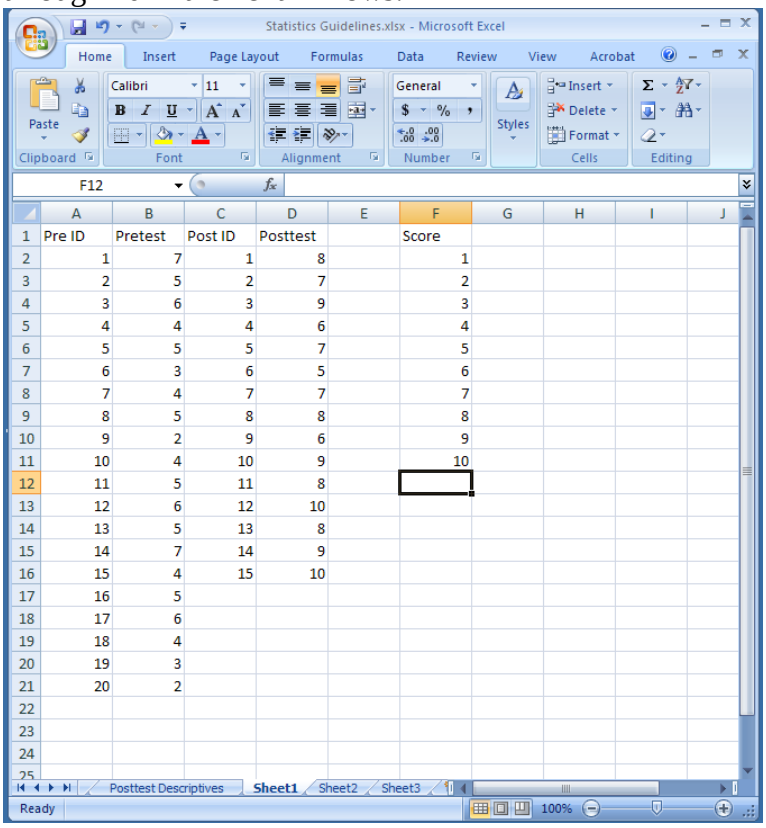
For the posttest, we see that the Mean = 7.8, the range is from a Minimum of 5 to a Maximum of 10 and there were 15 respondents (Count = 15). Again, save your workbook (File → Save). To get back to your data, click on the Sheet 1 tab at the bottom of the workbook.

Compute Histograms

The next step in the analysis is to compute a histogram or bar chart for the pretest and posttest variables. The histogram shows the distribution of each variable – the number of respondents who scored within each range of possible scores. This distribution shows whether the data is normally distributed or not.

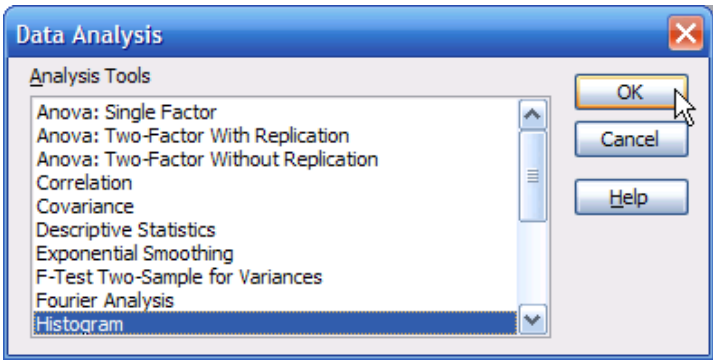
Before we can compute the histogram, we have to set up a table that will tell the program how we would like the bars displayed. If you haven't already done so, click on the Sheet 1 tab to get back to the sheet that has the raw data on it. We know that both the pretest and posttest in our example could have scores that range from 0 (none correct) to 10 (all ten questions correct). Let's construct a histogram that has one bar for each possible score. Now, do the following to get a histogram:

1. Create the information for the bars of the histogram. As in the figure below, in column F type the word ‘Score’ into the first row, and then type the numbers 0 through 10 in the next 11 rows.



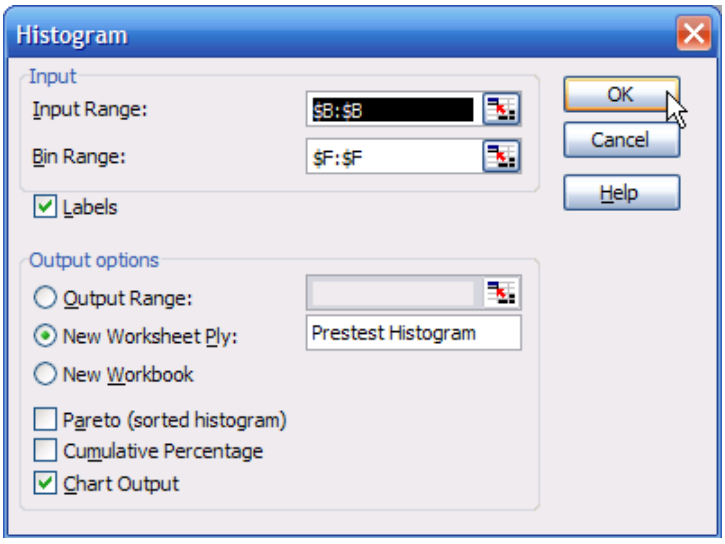
	A	B	C	D	E	F
1	Pre ID	Pretest	Post ID	Posttest		Score
2	1	7	1	8		1
3	2	5	2	7		2
4	3	6	3	9		3
5	4	4	4	6		4
6	5	5	5	7		5
7	6	3	6	5		6
8	7	4	7	7		7
9	8	5	8	8		8
10	9	2	9	6		9
11	10	4	10	9		10
12	11	5	11	8		
13	12	6	12	10		
14	13	5	13	8		
15	14	7	14	9		
16	15	4	15	10		
17	16	5				
18	17	6				
19	18	4				
20	19	3				
21	20	2				
22						
23						
24						

2. Now, select the Data menu, Data Analysis… item and select histogram from the pop-up menu and click OK as in the figure below.

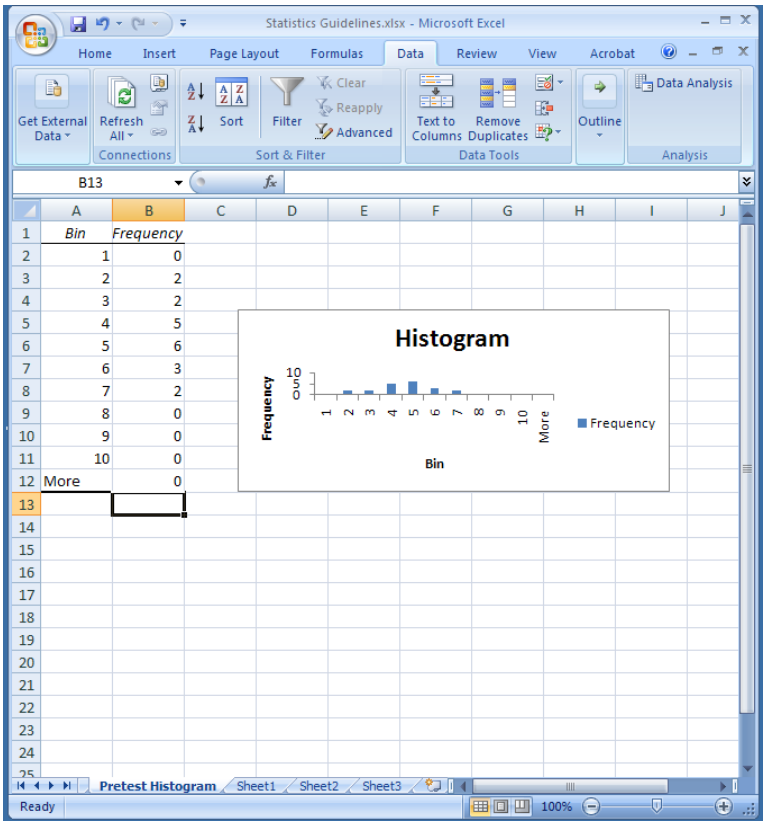


3. Now select the input range for the Pretest variable like you did in the descriptive statistics but do not include the title cell. Next, select the Bin range – that’s the new column you set up in column F with the possible scores shown. Be sure to check the ‘Labels’ box, select ‘New Worksheet Ply’ and enter in the title ‘Pretest Histogram’, and check the box for ‘Chart Output.

Your screen should look like the one below. When you've confirmed that it does, click OK.



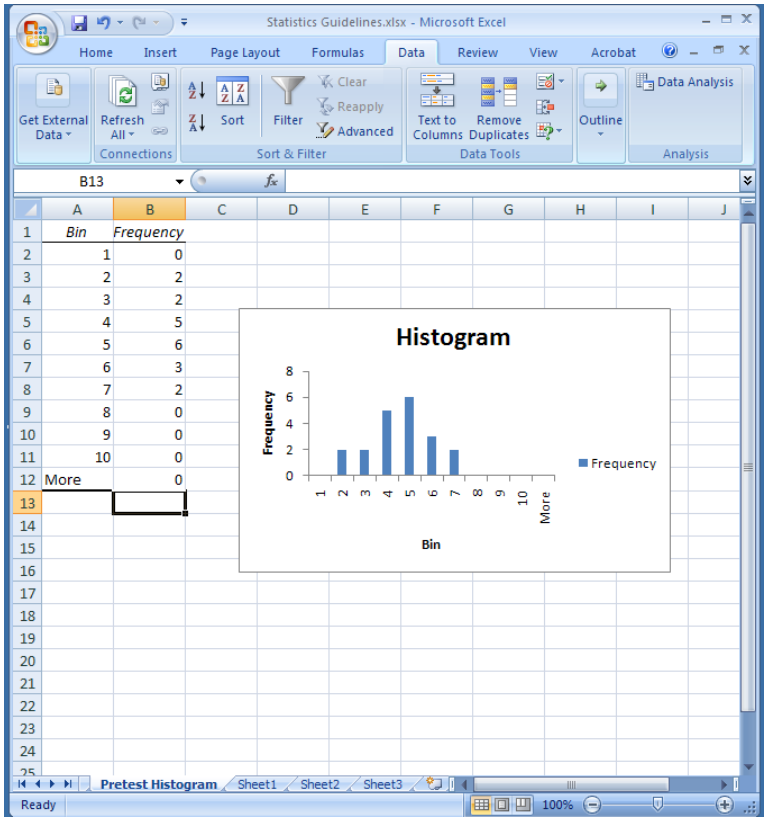
The histogram will show up in a new worksheet and look something like this:



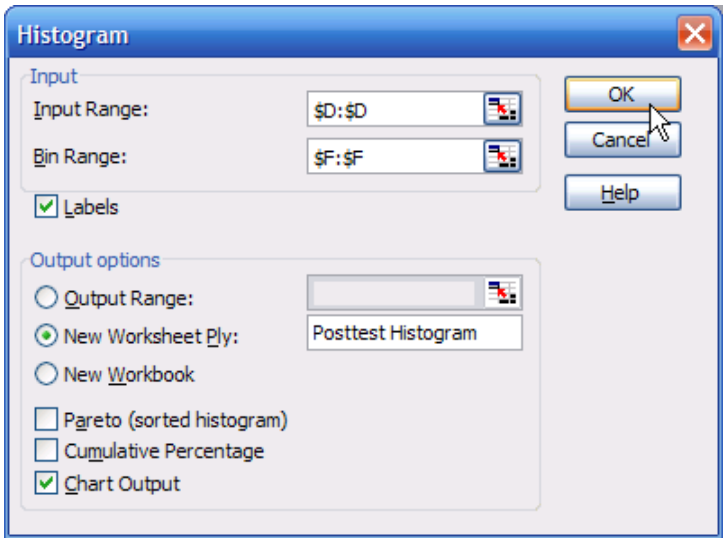
Bin	Frequency
1	0
2	2
3	2
4	5
5	6
6	3
7	2
8	0
9	0
10	0
More	0

Notice that your score column is shown on the left side of the sheet and that the number of participants who received each score is shown in the second column. The histogram or bar chart is shown next to the table. You can alter

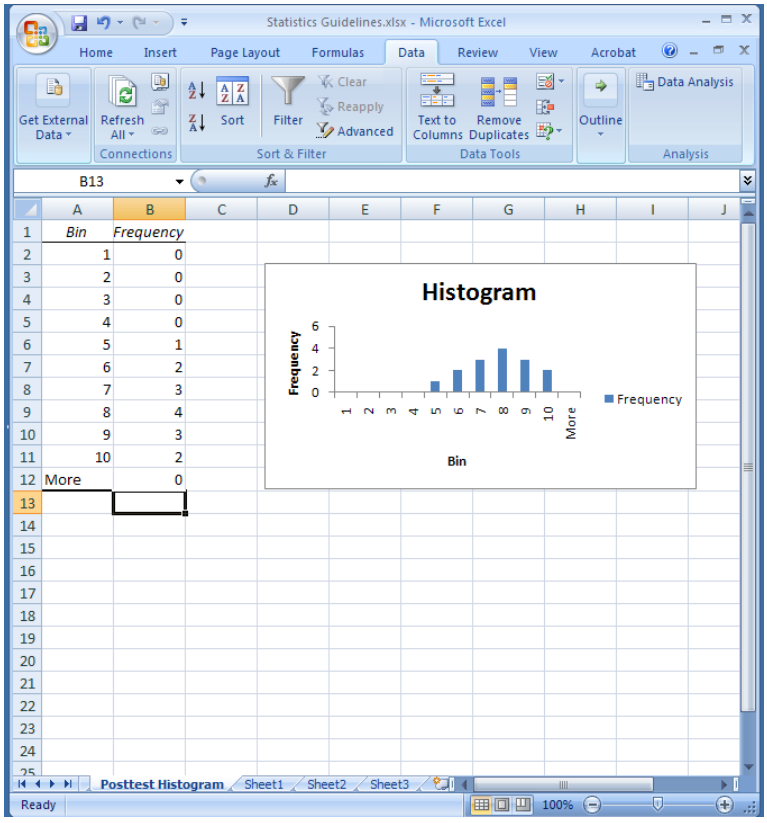
the look of the chart to make it easier to read. Just click on the chart and resize it as you would any graphics object. Here, I've stretched it vertically so we can see the bars more clearly.



4. Repeat steps 2 and 3 above to get the histogram for the posttest. To get back to the data, remember to click on the Sheet 1 tab at the bottom of your screen. We'll use the exact same Bin Range in this histogram so our two graphs will be comparable. Here's what my histogram specification looks like when I applied it to the posttest variable and named the new worksheet appropriately:



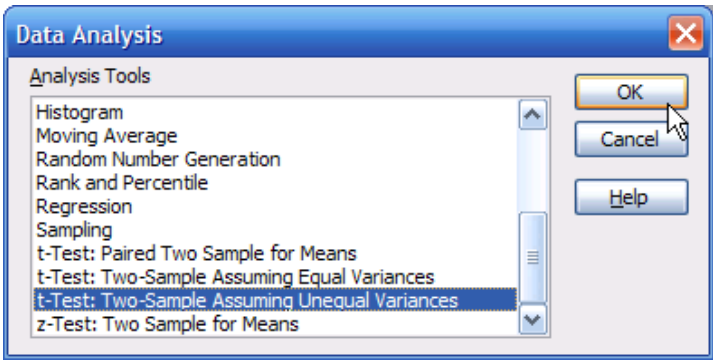
And here's what the histogram looks like, after I adjusted it to make it a little taller:



Test For Change

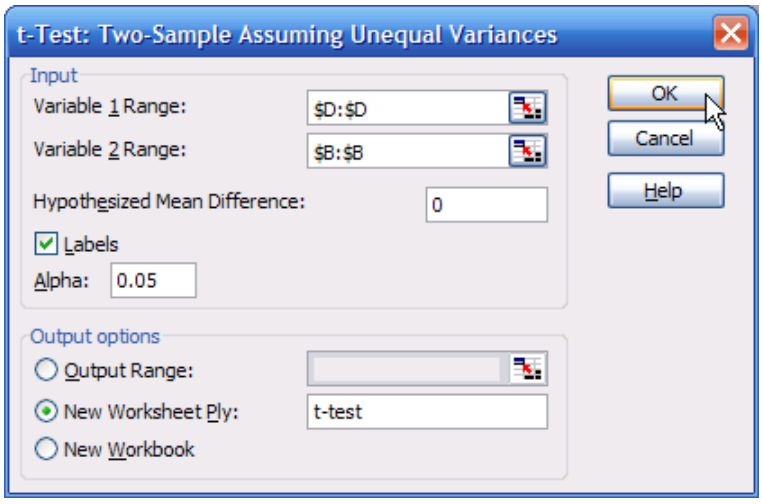
The last step in our analysis is to test whether there is a significant change in scores from pretest to posttest. We will use the t-test to accomplish this analysis. Before conducting the t-test, use the histogram to check whether the data is normally distributed. If the data is not normally distributed, see *Tips for data collection and analysis, On transformations.* The t-test compares the averages of the pretest and posttest to see if they are statistically different from each other. Do the following steps:

1. Select the Tools menu, Data Analysis… item, select the item 't-Test Two-Sample Assuming Unequal Variances' and click OK, as shown in the following figure:



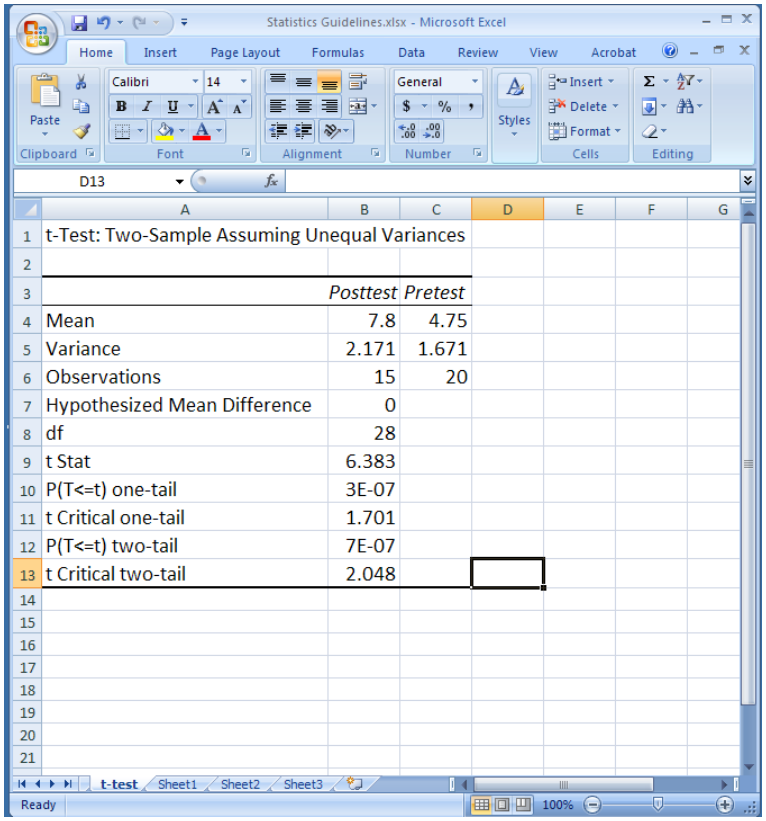
   We are using the Unequal Variance version of the t-test because we saw in the descriptive statistics that the variance of the pretest was 1.67 and the variance of the posttest was 2.17. If these two were closer, we could use the Equal Variance version of the t-test.

2. Specify the t-test input. As in the earlier analyses, you have to specify the input range. But this time, we are comparing the two variables, the pretest and posttest, so we have to specify both. For Variable 1 Range, specify the Posttest column. Then for Variable 2 Range, specify the pretest.



For the Hypothesized Mean Difference, enter 0 (this is called the 'null' hypothesis – we predict that if the program does not work there will be no difference between the pretest and posttest). Click on the Labels box, select New Worksheet Ply and enter the name 't-test'.

When you click OK, you will get the following results:



I made column A wider so that we can read the labels for each result. The results show that the pretest average was 4.75 while the posttest average was

7.8. The t-value is 6.38 with a df=28. This t-value is much higher than the critical value of 2.04 for a .05 level of significance (what we specified for Alpha in the t-test screen above). We would interpret this to indicate that there was a statistically significant increase in the means from pretest to posttest in our project. If the t-value is below the critical value, that means that the change in mean from pretest to posttest is not statistically significant and you cannot reject the null hypothesis.

Using the Results

The results that you have generated can easily be copied and pasted into a word processing program, can be edited as needed, and can be incorporated into reports.