Statistics Project Technology Guide

Discovery Projects in Introductory Statistics

Brad S. Bailey, Ph.D.
Dianna J. Spence, Ph.D.

Development of this resource supported by NSF grant award DUE-1021584

Table of Contents

Introduction

This is a technology guide for performing t-tests and linear regression analysis in Excel 2007 and the TI-83 and TI-84 family of calculators. There are several ways to use these technologies to perform these types of analysis; in what follows we will demonstrate one or two of methods that we consider to be the most common ways of performing these analyses. This is NOT a statistics textbook; indeed, very little of what follows will pertain to statistical theory or interpretation of statistics. This is strictly a manual for using Excel 2007 and the TI calculators to find these statistics.

T-tests

In this chapter we will focus on running different types of t-tests in Microsoft Excel 2007 and in the TI-83, -84 family of calculators. We will start with Excel.

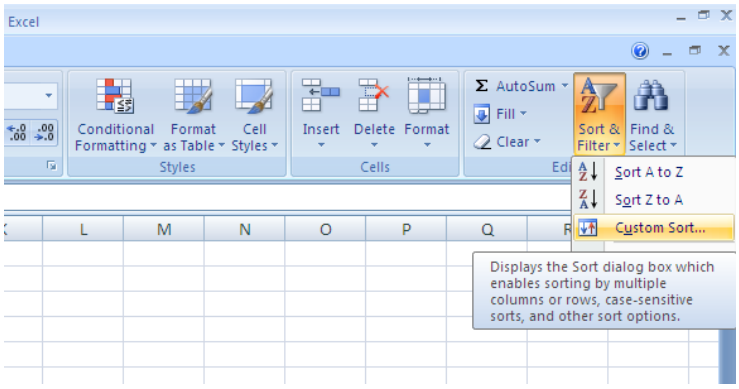
T-tests with Excel

Before we can run a t-test in Excel, it is often necessary to first sort or arrange the data. So we dedicate the first section to that topic.

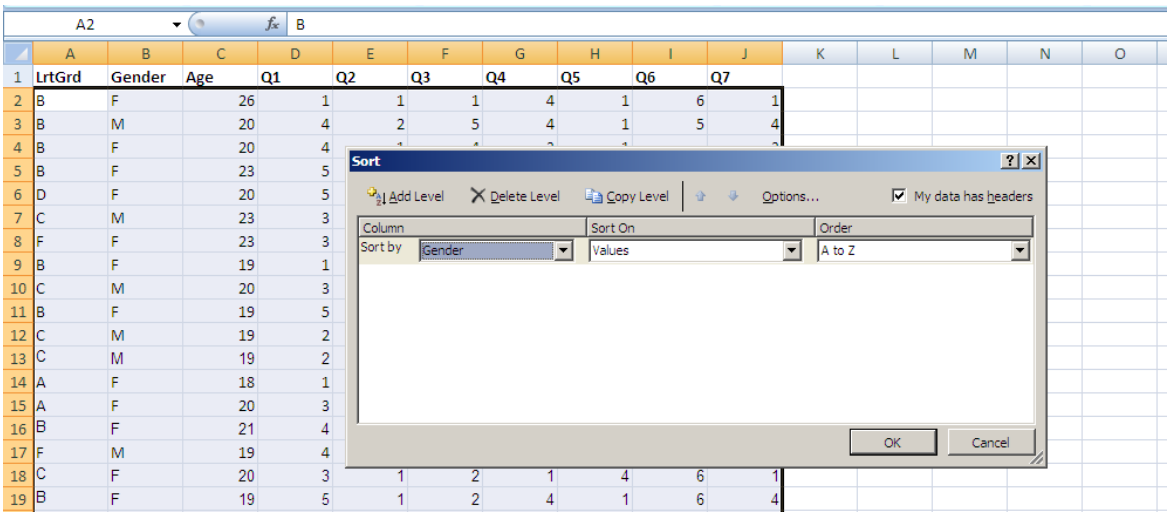
Sorting & Arranging Data in Excel

Often when working with real data in Excel, the rows of a worksheet represent the data for an individual, such as all the responses from a particular person, and the columns represent all the responses to a particular question. As such, it is often beneficial to sort the data according to a grouping variable, such as Gender or Major.

To do this, highlight the entire area containing your data, then select Sort & Filter followed by Custom Sort.



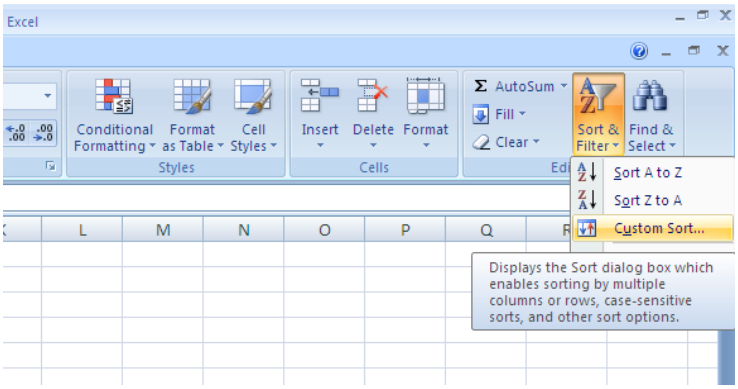
Pick the column that contains the grouping variable. When you hit OK, Excel will sort the data according the grouping variable, keeping all the responses of an individual in a row together.



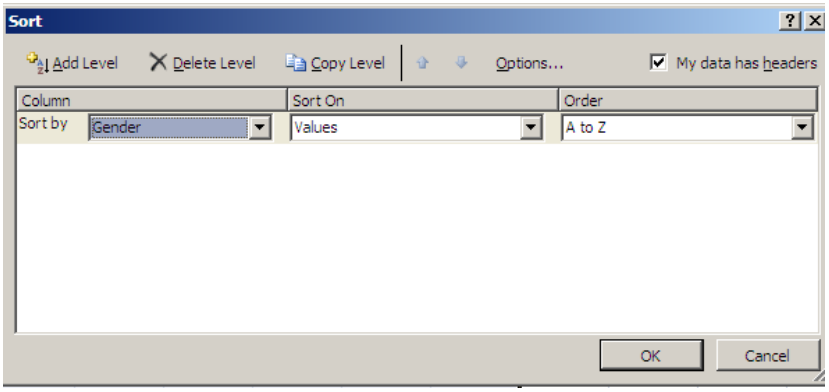
In the example demonstrated below, the grouping variable we picked was Gender. Because we had highlighted the entire data field, the rows still represent the responses of a particular individual. If we had left any columns un-highlighted those columns would not have been included in the sort and those responses would no longer be paired with the subject that provided them.

STEPS FOR SORTING DATA IN EXCEL

1. **Highlight ALL the cells containing data.**
2. **Under Home click on Sort & Filter and select Custom Sort.**



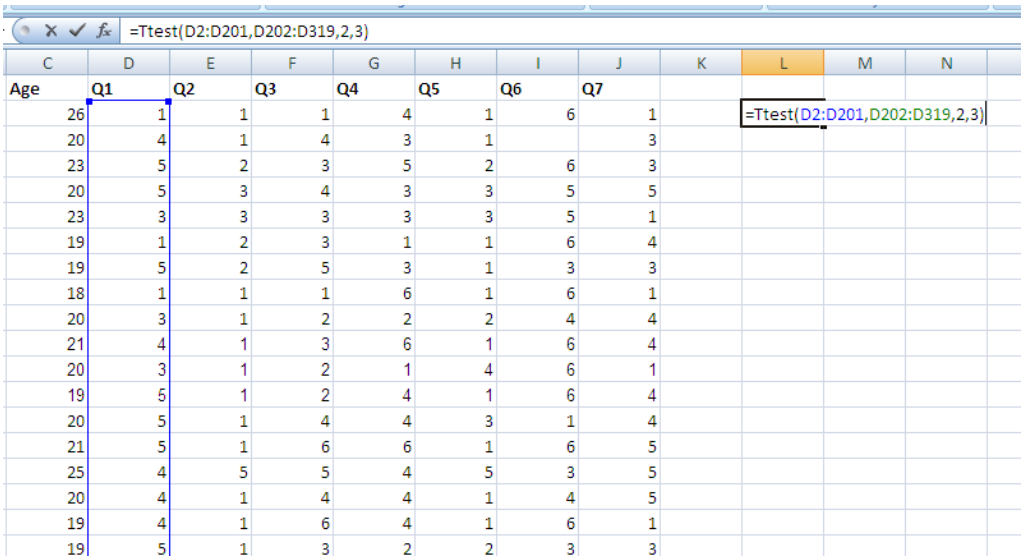
3. **Specify the variable you want to sort by, hit OK.**



In subsequent sections of this manual we will assume that students are aware of how to sort data in Excel.

Two-sample t-tests (independent samples)

Suppose in the following data set we're wondering if the genders have a different mean response to Q1. This is a setting for 2-sample t-test. The first way we're going to demonstrate how to do this is with the TTEST command, which is typed directly into a cell on the Excel worksheet.



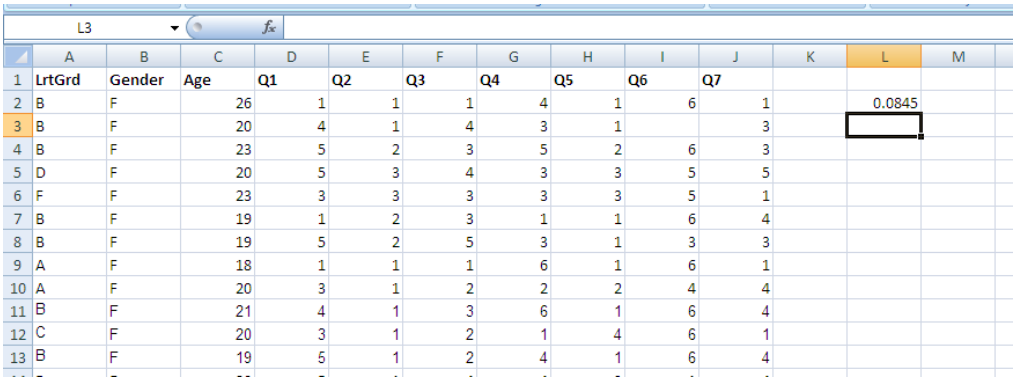
Age	Q1	Q2	Q3	Q4	Q5	Q6	Q7					
26	1	1	1	4	1	6	1	=Ttest(D2:D201,D202:D319,2,3)				
20	4	1	4	3	1		3					
23	5	2	3	5	2	6	3					
20	5	3	4	3	3	5	5					
23	3	3	3	3	3	5	1					
19	1	2	3	1	1	6	4					
19	5	2	5	3	1	3	3					
18	1	1	1	6	1	6	1					
20	3	1	2	2	2	4	4					
21	4	1	3	6	1	6	4					
20	3	1	2	1	4	6	1					
19	5	1	2	4	1	6	4					
20	5	1	4	4	3	1	4					
21	5	1	6	6	1	6	5					
25	4	5	5	4	5	3	5					
20	4	1	4	4	1	4	5					
19	4	1	6	4	1	6	1					
19	5	1	3	2	2	3	3					

The command to have Excel compute a p-value for a two sample t-test is as follows:

=TTEST(DATA1, DATA2, TAILS, 3)

See the portion of the Excel worksheet above.

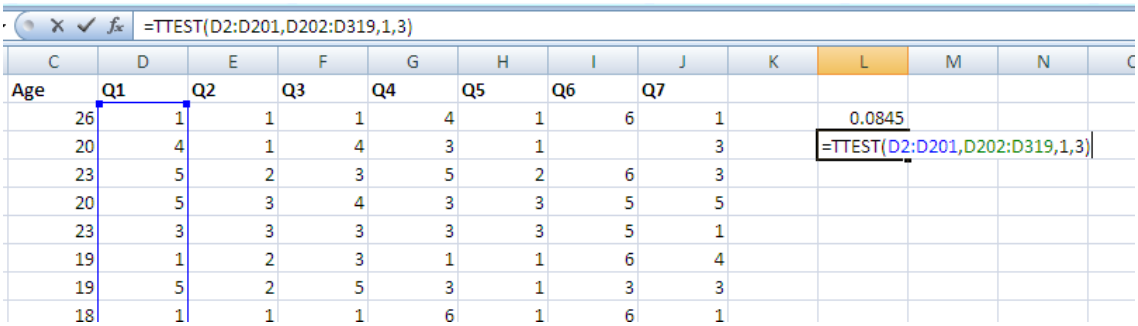
The output is a p-value for the t-test. If you put a 2 for the number of TAILS, the p-value will be a two-sided p-value that corresponds to the alternative hypothesis that $\mu_1 \neq \mu_2$.



LrtGrd	Gender	Age	Q1	Q2	Q3	Q4	Q5	Q6	Q7		
B	F	26	1	1	1	4	1	6	1	0.0845	
B	F	20	4	1	4	3	1		3		
B	F	23	5	2	3	5	2	6	3		
D	F	20	5	3	4	3	3	5	5		
F	F	23	3	3	3	3	3	5	1		
B	F	19	1	2	3	1	1	6	4		
B	F	19	5	2	5	3	1	3	3		
A	F	18	1	1	1	6	1	6	1		
A	F	20	3	1	2	2	2	4	4		
B	F	21	4	1	3	6	1	6	4		
C	F	20	3	1	2	1	4	6	1		
B	F	19	5	1	2	4	1	6	4		

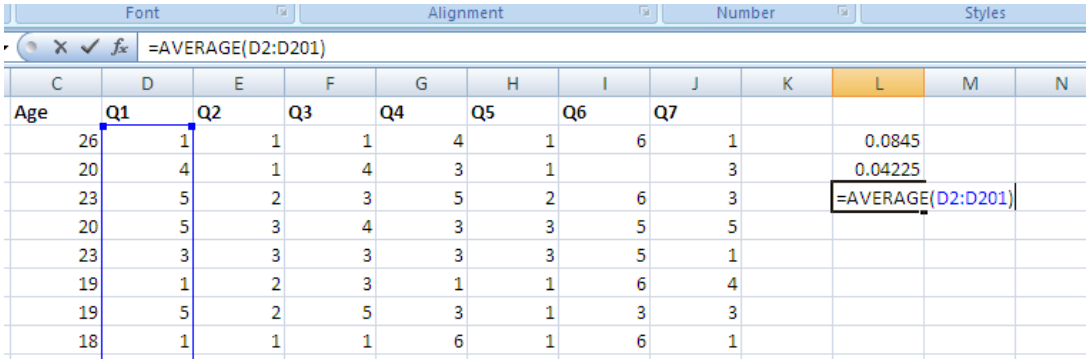
If you enter 1 for the number of tails the output will be the one-sided p-value that corresponds to one of $\mu_1 > \mu_2$ or $\mu_1 < \mu_2$.

Excel does not ask you which of these alternative hypotheses you're testing against; instead it always returns the smaller of the two possible p-values. Looking at our sample data, if the original question had been "Do females have a higher mean response to Q1 than males?", then we would have chosen a 1-tailed t-test.



Suppose you find a low p-value for a one-sided two-sample t-test, how do we decide which of the above alternative hypotheses this hypothesis test supports?

Check to see which population has the higher sample mean, using the AVERAGE function of Excel.

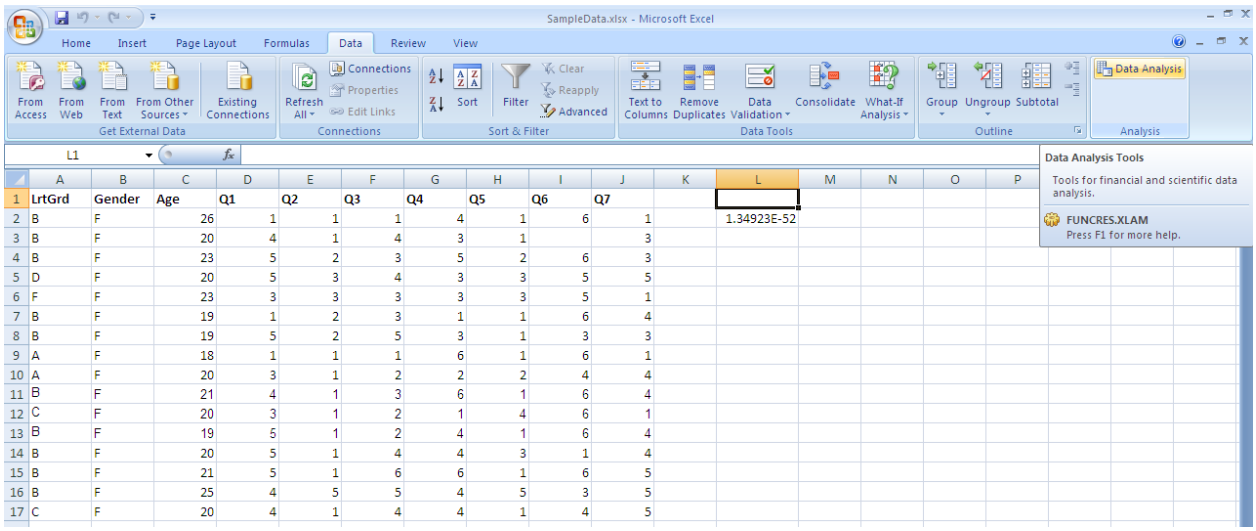


STEPS FOR RUNNING A TWO SAMPLE T-TEST IN EXCEL USING TTEST COMMAND

1. **Sort or group your data.**
2. **Use the TTEST command to calculate the appropriate p-value; for the last entry enter a 3.**
3. **If you find a low p-value use AVERAGE to see which sample has the high/low mean.**

Loading the Analysis Toolpak

The next method we will demonstrate involves using the Analysis ToolPak, which is not part of the standard loading of Excel 2007. If it is loaded in your version of Excel it should appear as part of the Data tab.

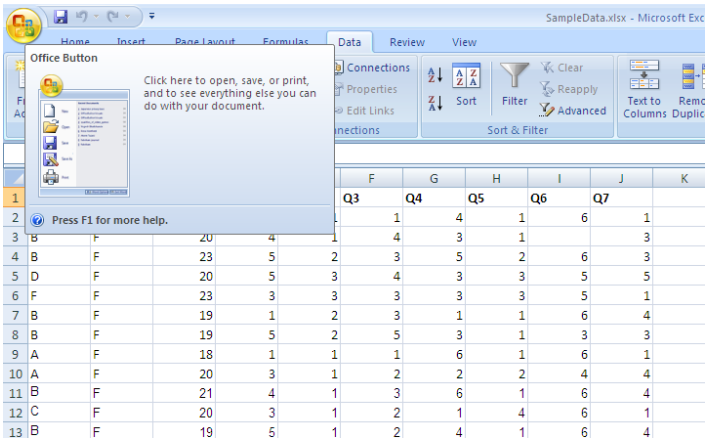


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	LrtGrd	Gender	Age	Q1	Q2	Q3	Q4	Q5	Q6	Q7						
2	B	F	26	1	1	1	4	1	6	1		1.34923E-52				
3	B	F	20	4	1	4	3	1		3						
4	B	F	23	5	2	3	5	2	6	3						
5	D	F	20	5	3	4	3	3	5	5						
6	F	F	23	3	3	3	3	3	5	1						
7	B	F	19	1	2	3	1	1	6	4						
8	B	F	19	5	2	5	3	1	3	3						
9	A	F	18	1	1	1	6	1	6	1						
10	A	F	20	3	1	2	2	2	4	4						
11	B	F	21	4	1	3	6	1	6	4						
12	C	F	20	3	1	2	1	4	6	1						
13	B	F	19	5	1	2	4	1	6	4						
14	B	F	20	5	1	4	4	3	1	4						
15	B	F	21	5	1	6	6	1	6	5						
16	B	F	25	4	5	5	4	5	3	5						
17	C	F	20	4	1	4	4	1	4	5						

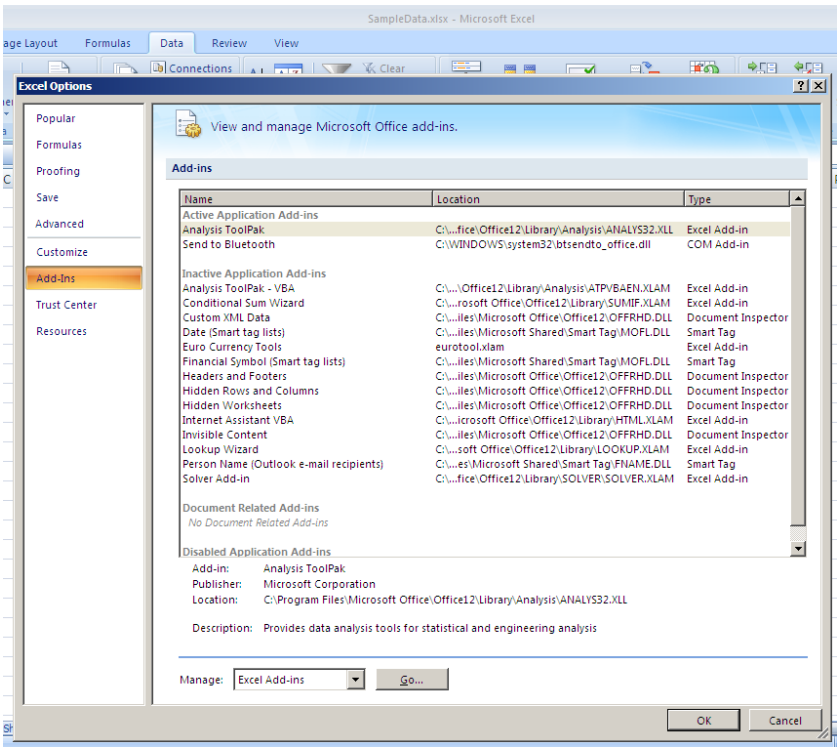
If it's not loaded but you want to use it, follow these steps.

STEPS FOR LOADING ANALYSIS TOOLPAK.

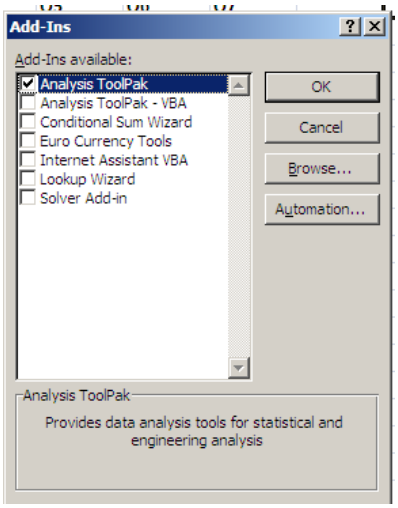
1. **Click on the Office Button.**



2. **Select Excel Options, then select Add-ins.**



3. **Click on the button labeled Go… (if you select OK, the window will close and the ToolPak won't have loaded).**
4. **Hit the checkbox of Analysis ToolPak, then hit OK. The Data Analysis option should now appear on your Data tab.**

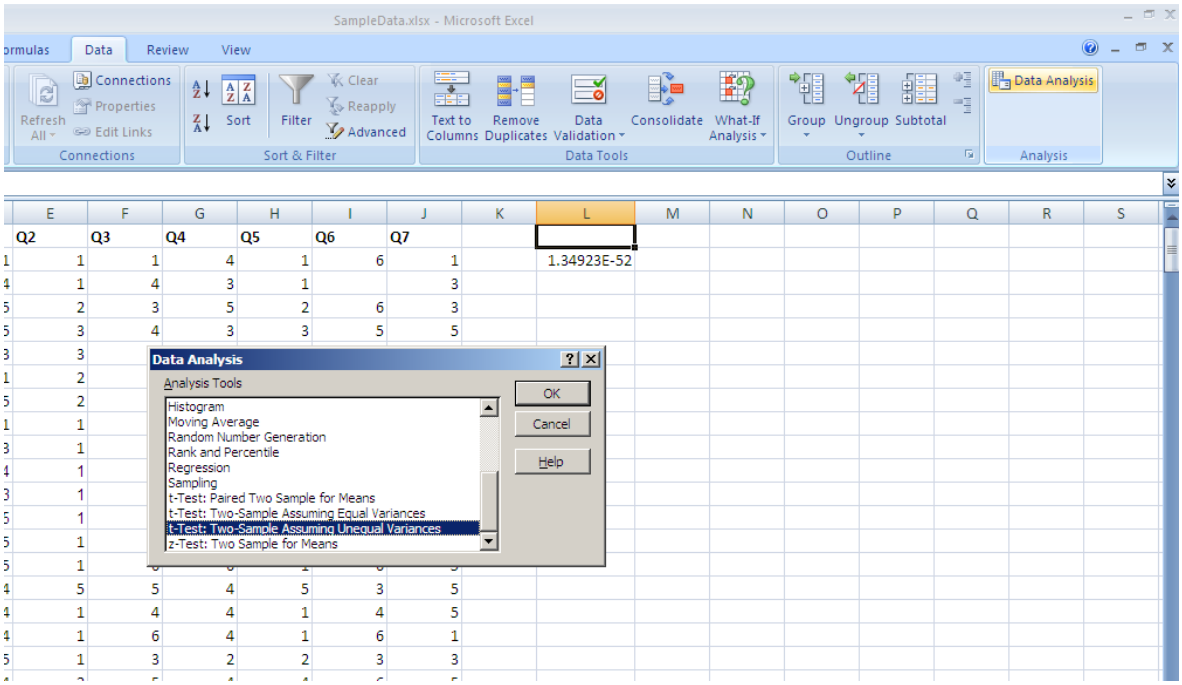


Two –sample T-test using Data Analysis Menu

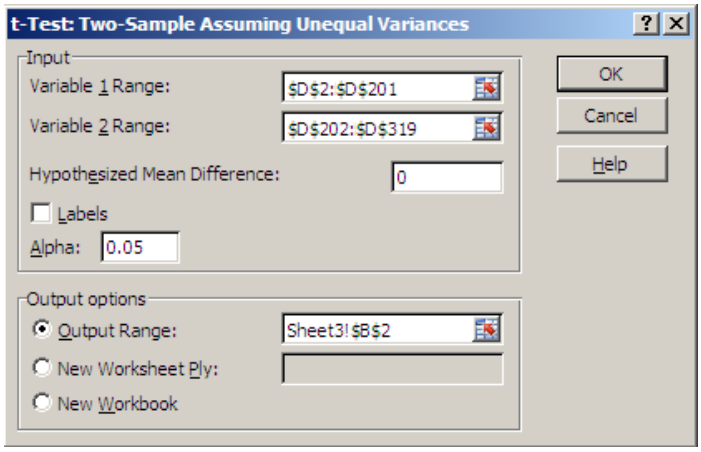
To use this method, the Analysis Toolpak must be loaded, which is explained in the section immediately prior to this one.

STEPS FOR RUNNING TWO-SAMPLE T-TEST USING DATA ANALYSIS MENU

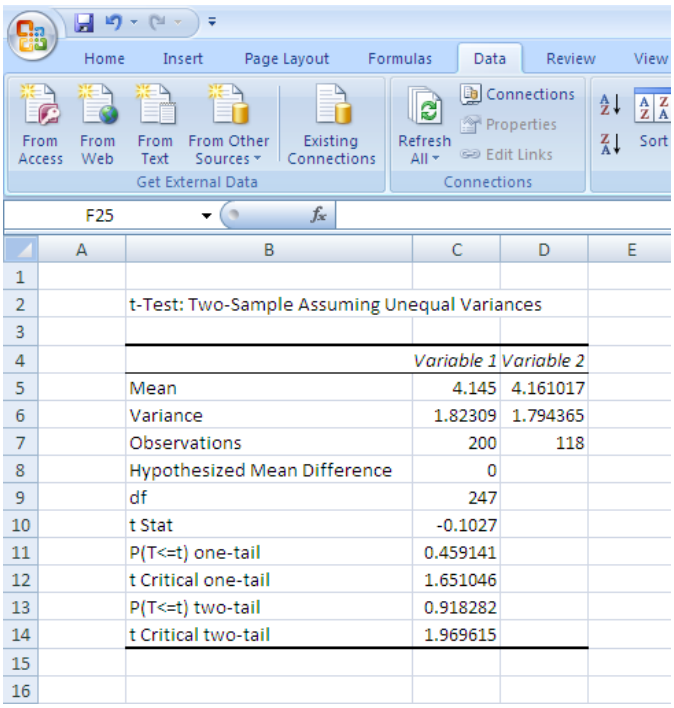
1. **Select Data Analysis under the Data tab. Then find t-test: Two-Sample Assume Unequal Variance.**



2. **Specify where the data is for the two different populations, and give Excel a location where you'd like the output to appear.**



3. **The output includes the t-statistic and the one-tailed and two-tailed p-values. These are in the cells to the left of the cells thusly labeled. (Excel uses a capital P to denote the p-value).**



	A	B	C	D	E
1					
2		t-Test: Two-Sample Assuming Unequal Variances			
3					
4			*Variable 1*	*Variable 2*	
5		Mean	4.145	4.161017	
6		Variance	1.82309	1.794365	
7		Observations	200	118	
8		Hypothesized Mean Difference	0		
9		df	247		
10		t Stat	-0.1027		
11		P(T<=t) one-tail	0.459141		
12		t Critical one-tail	1.651046		
13		P(T<=t) two-tail	0.918282		
14		t Critical two-tail	1.969615		
15					
16					

Matched Pairs t-test (dependent samples)

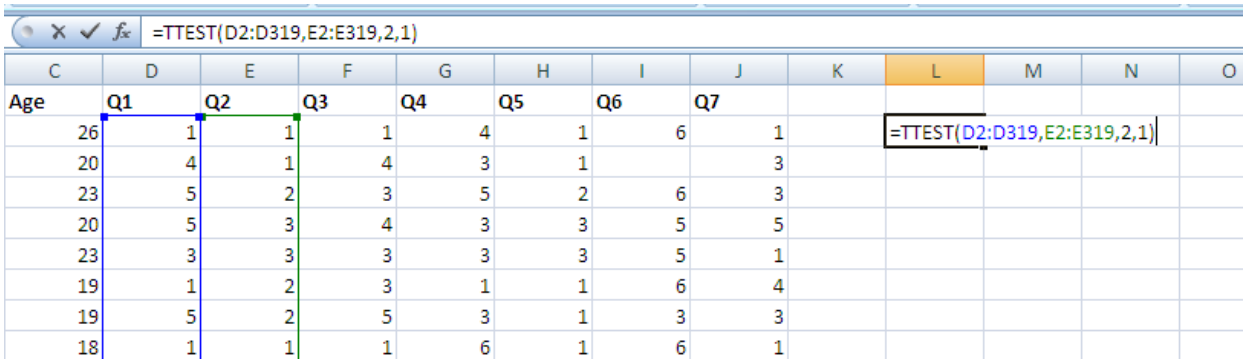
If the research question is whether there's a difference in measurements for the same individuals taken at different times or two different questions, this is a setting for the Matched Pairs t-test.

Suppose in our sample data, the question is "Do students tend to give higher responses to Q1 than to Q2?" This is a setting for a matched pairs t-test.

The command to have Excel compute a p-value for a matched pair t-test is as follows:

=TTEST(DATA, DATA, TAILS, 1)

See the Excel Work sheet below.



C	D	E	F	G	H	I	J	K	L	M	N	O
Age	Q1	Q2	Q3	Q4	Q5	Q6	Q7					
26	1	1	1	4	1	6	1		=TTEST(D2:D319,E2:E319,2,1)			
20	4	1	4	3	1		3					
23	5	2	3	5	2	6	3					
20	5	3	4	3	3	5	5					
23	3	3	3	3	3	5	1					
19	1	2	3	1	1	6	4					
19	5	2	5	3	1	3	3					
18	1	1	1	6	1	6	1					

STEPS FOR RUNNING A MATCHED PAIRS T-TEST USING T-TEST COMMAND.

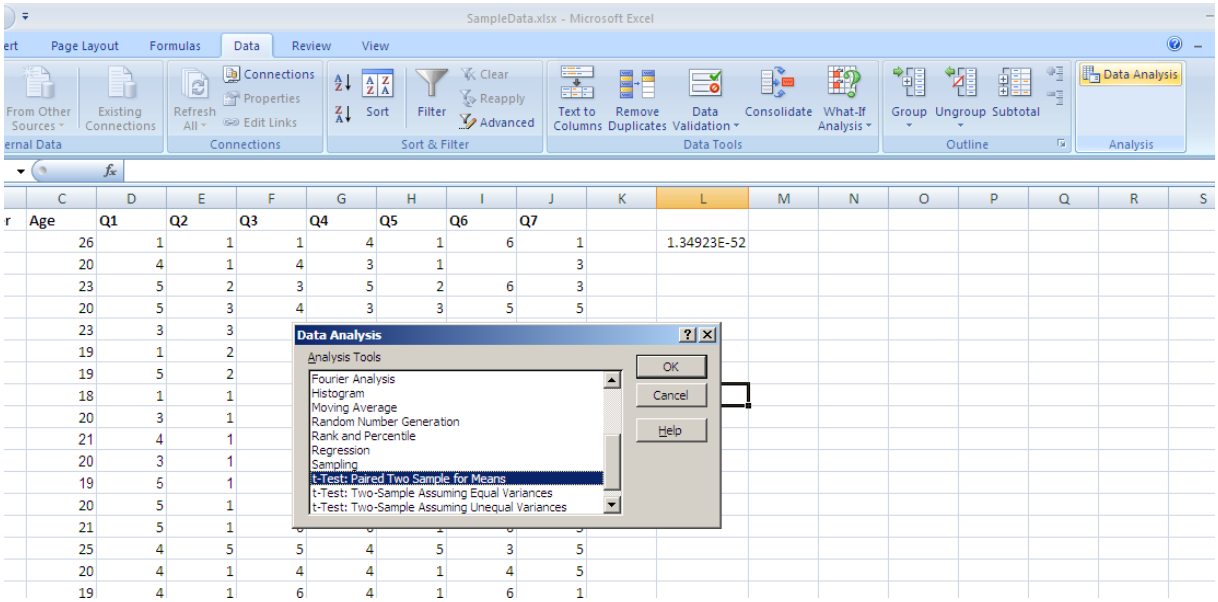
1. **Sort or group your data.**
2. **Use the TTEST command to calculate the appropriate p-value; for the last option, enter a 1.**

3. **If you find a low p-value in an empty column calculate the difference for each subject, then use AVERAGE to find the mean of the differences. Looking at the sign of this and considering the order of subtraction will tell you which response tends to be higher/lower.**

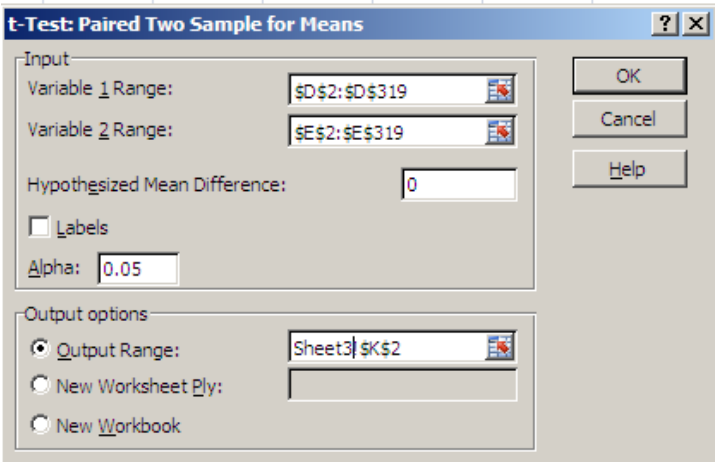
It is also possible to run a matched pairs t-test using the Data Analysis menu.

STEPS FOR RUNNING A MATCHED PAIRS T-TEST USING THE DATA ANALYSIS MENU.

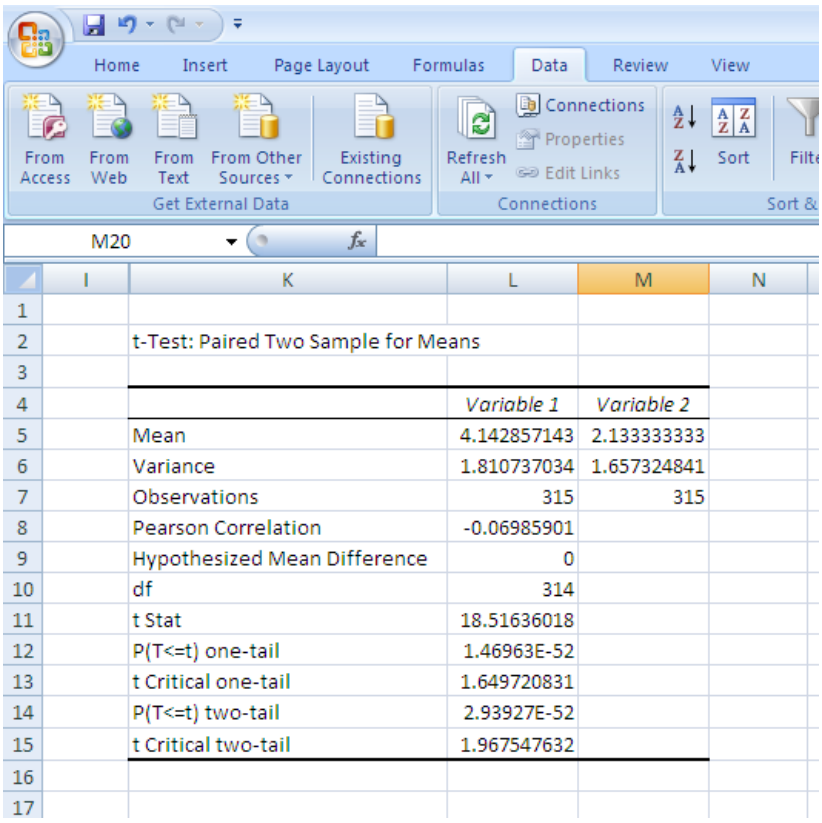
1. **Select Data Analysis on the Data tab, then find t-test: Paired Two Sample for Means.**



2. **Specify the cells containing the data and where you would like Excel to place the output. Hit OK.**



3. **The output includes the t-statistic and the p-values for both the one-tailed and two-tailed hypothesis tests.**



	I	K	L	M	N
1					
2		t-Test: Paired Two Sample for Means			
3					
4			*Variable 1*	*Variable 2*	
5		Mean	4.142857143	2.133333333	
6		Variance	1.810737034	1.657324841	
7		Observations	315	315	
8		Pearson Correlation	-0.06985901		
9		Hypothesized Mean Difference	0		
10		df	314		
11		t Stat	18.51636018		
12		P(T<=t) one-tail	1.46963E-52		
13		t Critical one-tail	1.649720831		
14		P(T<=t) two-tail	2.93927E-52		
15		t Critical two-tail	1.967547632		
16					
17					

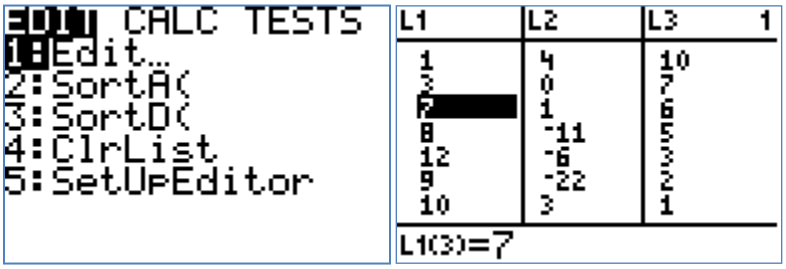
T-tests with TI calculator

Using the TI calculators, we often must first enter our data into a list. Hit the STAT button. The first option under the STAT menu is the EDIT. If you hit ENTER, you'll go into the list editor, where you can enter data into one of six lists (labeled L1 through L6).

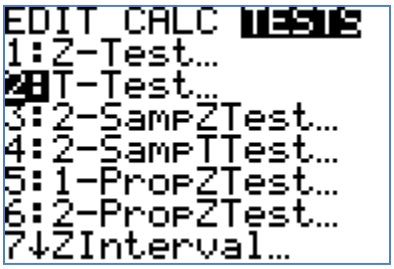
One-sample t-test

STEPS FOR RUNNING A T-TEST IN TI-83, -84 BY ENTERING DATA.

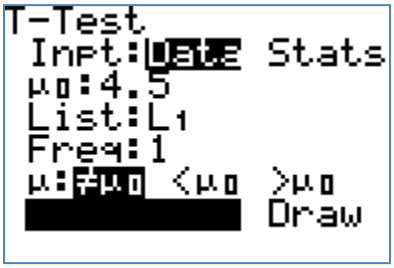
1. **Enter the data into a list. Hit the STAT button, then select EDIT, create your list.**



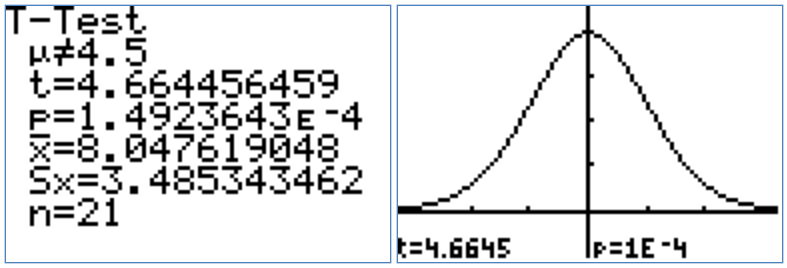
2. **Under STAT, select TESTS, then move down to T-Test.**



3. **For Inpt, select Data.**



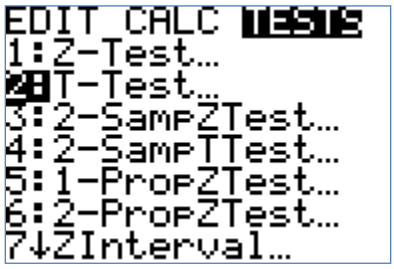
4. **Set the mean from the null hypothesis.**
5. **Select the list where you stored the data (Hint: for List 1, hit 2nd then 1).**
6. **Leave Freq set at 1.**
7. **Pick the appropriate alternative hypothesis.**
8. **Choose either Calculate or Draw, the hit enter. The output will look like one of the two screens below.**



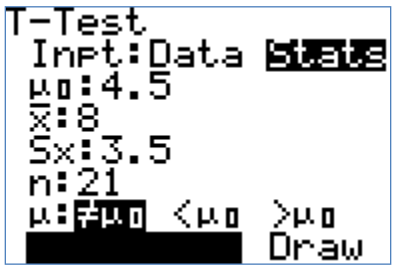
Because you are able to specify what form the alternative hypothesis takes, the p-value you find using the calculator is ready to be interpreted.

STEPS FOR RUNNING A T-TEST IN TI-83, -84 USING SUMMARY STATISTICS.

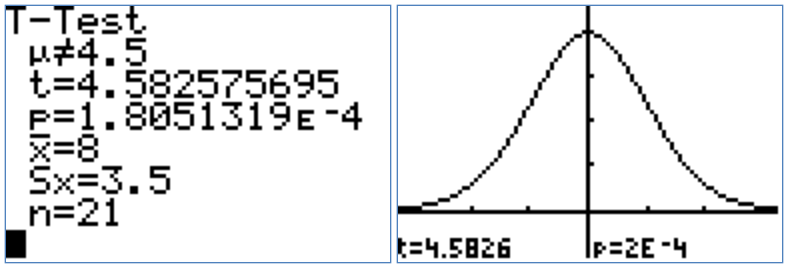
1. **Under STAT, select TESTS, then move down to T-Test.**



2. **For Inpt, select Stats**



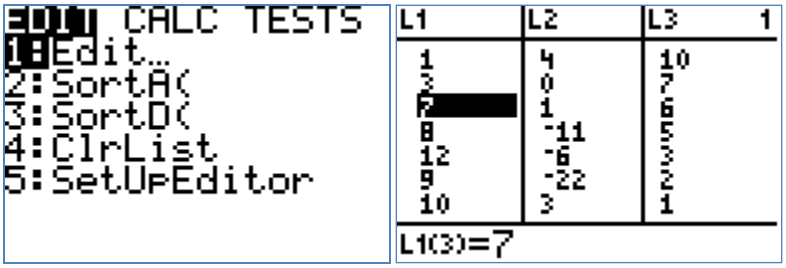
3. **Set the mean from the null hypothesis.**
4. **Enter the summary statistics (sample mean, standard deviation and size).**
5. **Pick the appropriate alternative hypothesis.**
6. **Choose either Calculate or Draw, the hit enter. The output will look like one of the two screens below.**



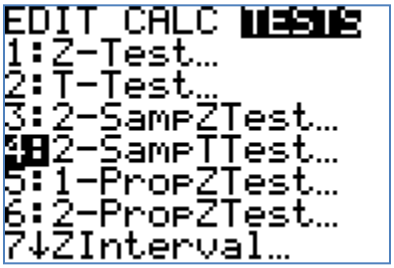
Two-sample t-test (independent samples)

STEPS FOR RUNNING A 2-SAMPLE T-TEST IN TI-83, -84 BY ENTERING DATA.

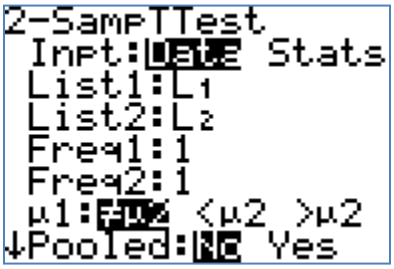
1. **Enter the data into a list. Hit the STAT button, then select EDIT, create your lists. The two populations must be in separate lists.**



2. **Under STAT, select TESTS, then move down to 2-SampTTest.**

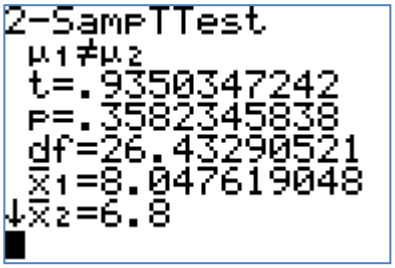


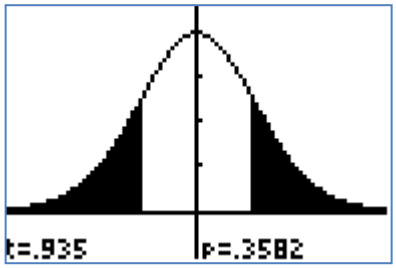
3. **For Inpt, select Data.**



4. **Select the lists where you stored the data (Hint: for List 1, hit 2nd then 1).**
5. **Skip over the Freq1 and Freq2.**
6. **Pick the appropriate alternative hypothesis.**

7. **Leave Pooled set at No.**
8. **Choose either Calculate or Draw, the hit enter. The output will look like one of the two screens below.**

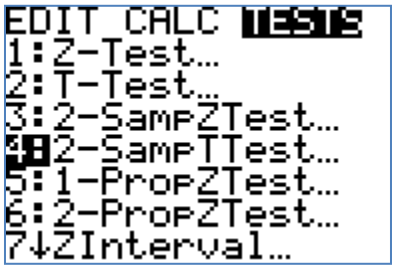




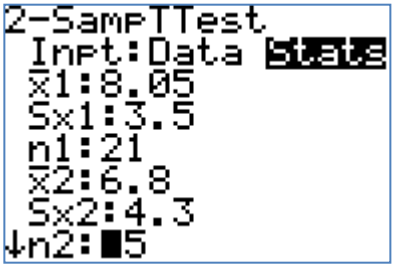
Because you are able to specify what form the alternative hypothesis takes, the p-value you find using the calculator is ready to be interpreted.

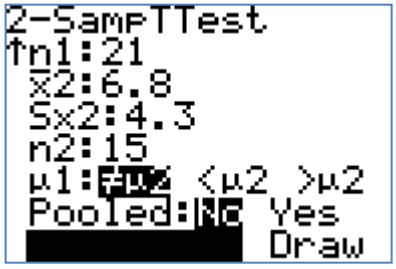
STEPS FOR RUNNING A T-TEST IN TI-83, -84 USING SUMMARY STATISTICS.

1. **Under STAT, select TESTS, then move down to 2-SampTTest.**

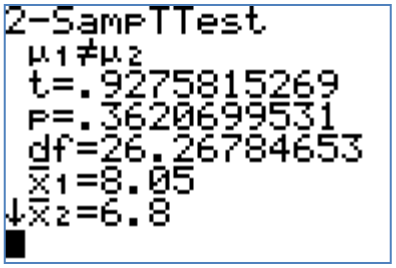


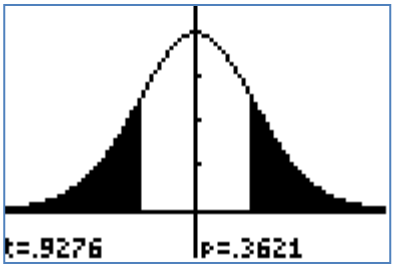
2. **For Inpt, select Stats**





3. **Set the mean from the null hypothesis.**
4. **Enter the summary statistics (sample mean, standard deviation and size) for both populations.**
5. **Pick the appropriate alternative hypothesis.**
6. **Choose either Calculate or Draw, the hit enter. The output will look like one of the two screens below.**

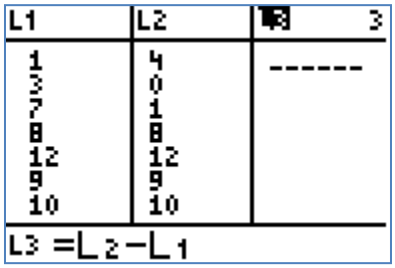




Matched Pairs t-test (dependent samples)

Recall that a Matched Pairs t-test is a 1-sample t-test comparing the mean difference between the scores with 0. That's how the TI calculators are used to run a matched pairs t-test. You can calculate the differences and record them in a list but if you have both values already in a list in the calculator, the calculator can find the differences. Here's how:

Use the arrows to move to the label for a list *other* than the two with your data stored. From here you can type in a formula involving the other lists. In this instance, you just need to subtract one like from another.



***WARNING*: Unlike Excel, the calculator will not automatically update data that is created by applying a formula to other data when the original data is modified. If you change an entry among the other data, you must RE-CALCULATE the third list.**

STEPS FOR RUNNING MATCHED PAIRS T-TEST IN TI-83

1. **Run a one sample t-test on the difference between the values for each subject. The null hypothesis should be 0.**

Linear Regression

Linear Regression using Excel

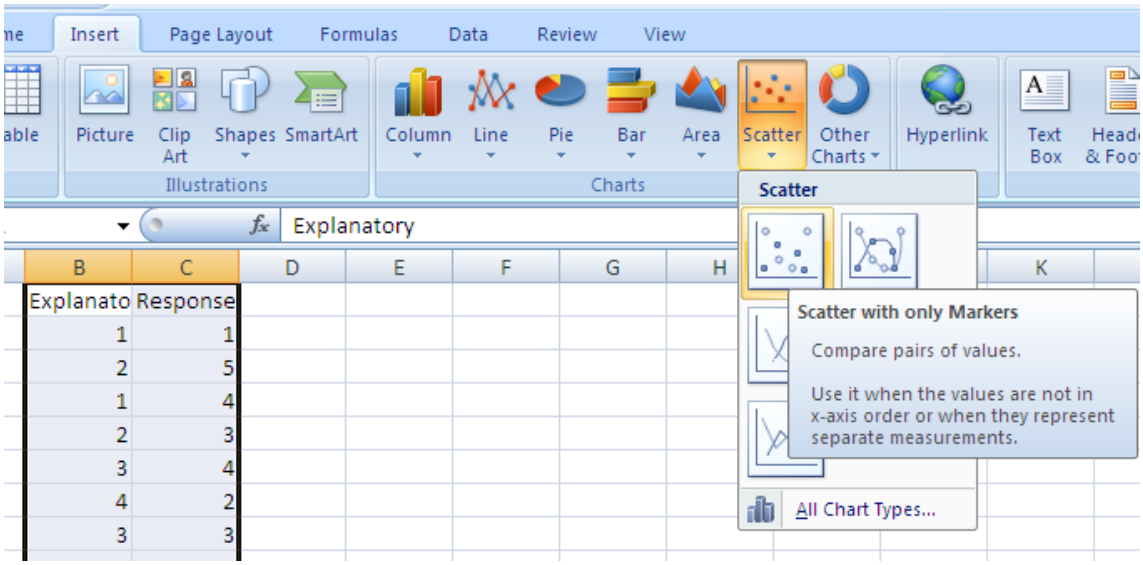
There are several ways of find the equation for a least squares regression line and the coefficients of correlation and determination using Excel. We will detail two methods.

Graphing Method

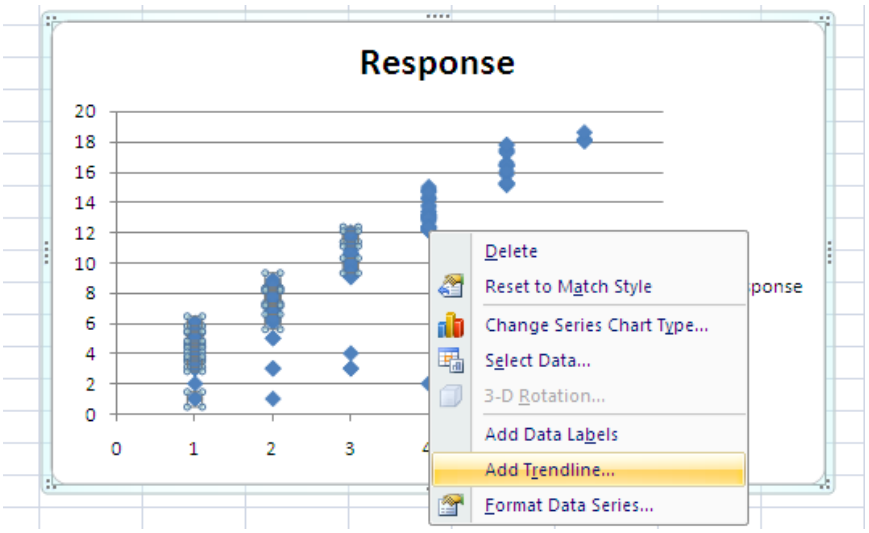
First create the scatterplot. The easiest way to do this is to arrange the data in two adjacent columns with the explanatory variable in the left column. This is not necessary, but it does simplify the process greatly. Select both columns of data and then on the Insert tab select Scatter. Once the scatterplot appears on the screen, left-click on one point, then right-click. Select Add trendline; in the next window, specify a linear equation, then hit the check boxes for Display Equation and Display R-square value. When you close this dialogue box, the regression equation and coefficient of determination, R^2, will appear on the graph next to the trend line. They are often difficult to read because they are on top of the data points, this can be fixed by clicking on the text box containing the equation and moving them to the side. Once you find R^2, compute the correlation coefficient by taking the square root of R^2 and matching the sign to that of the slope of the regression equation.

STEPS TO CREATING SCATTERPLOT AND FINDING REGRESSION EQUATION, COEFFICIENT OF DETERMINATION

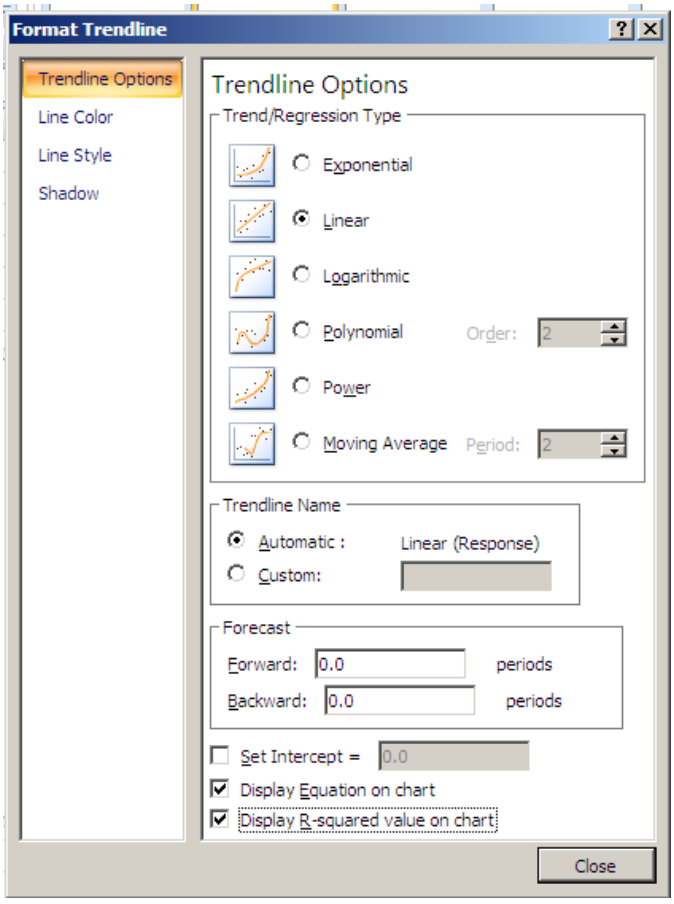
1. **Create the scatterplot.**



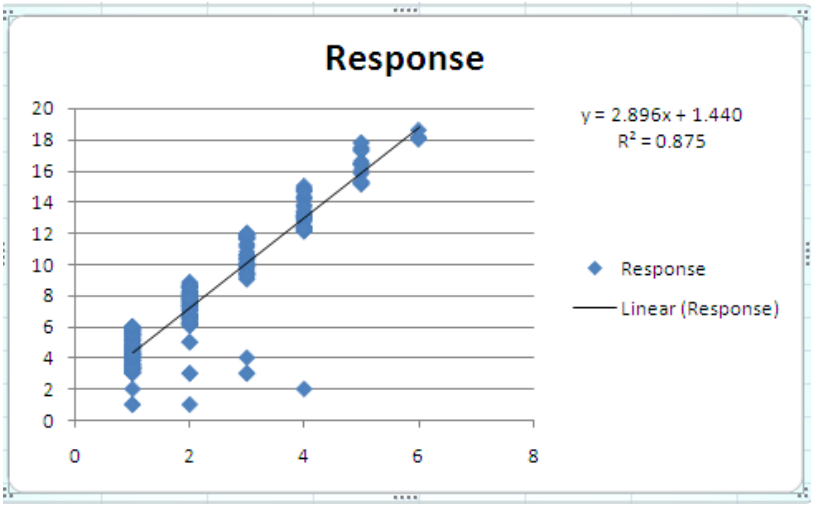
2. **Right-click on a data point and select Add Trendline.**



3. **Select Linear (this is the default), check Display Equation and Display R-squared. Click Close.**



4. **The regression equation and coefficient of determination now appear on your scatterplot.**
   a. **Find correlation coefficient. In the example below, the correlation coefficient is about 0.935.**

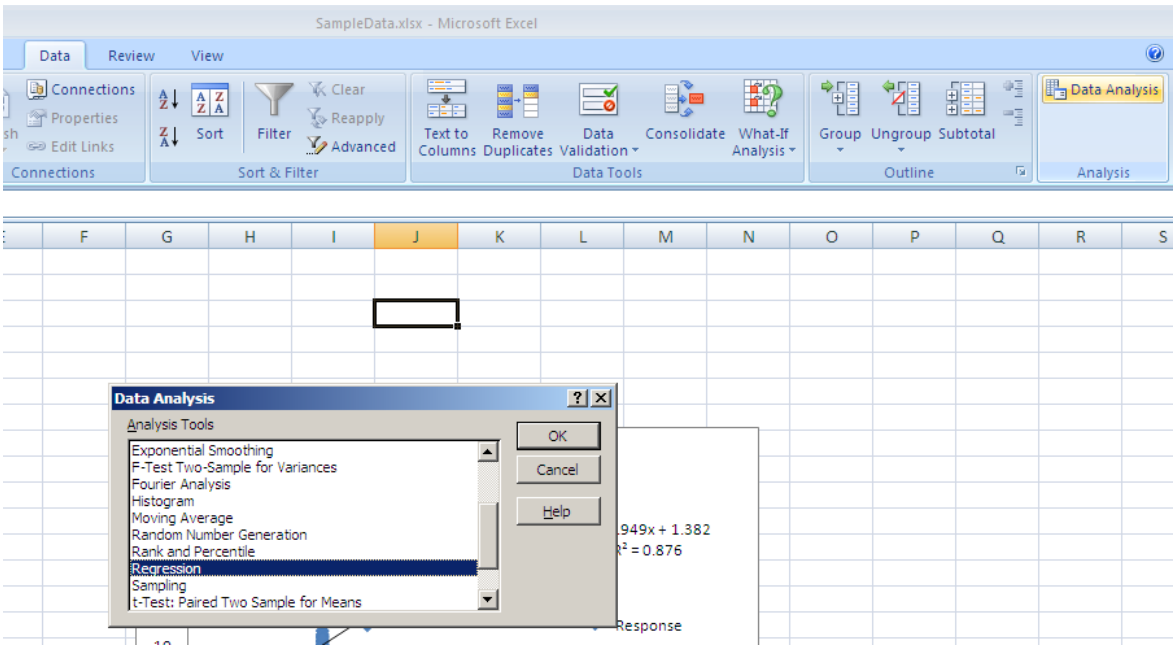


Data Analysis Method

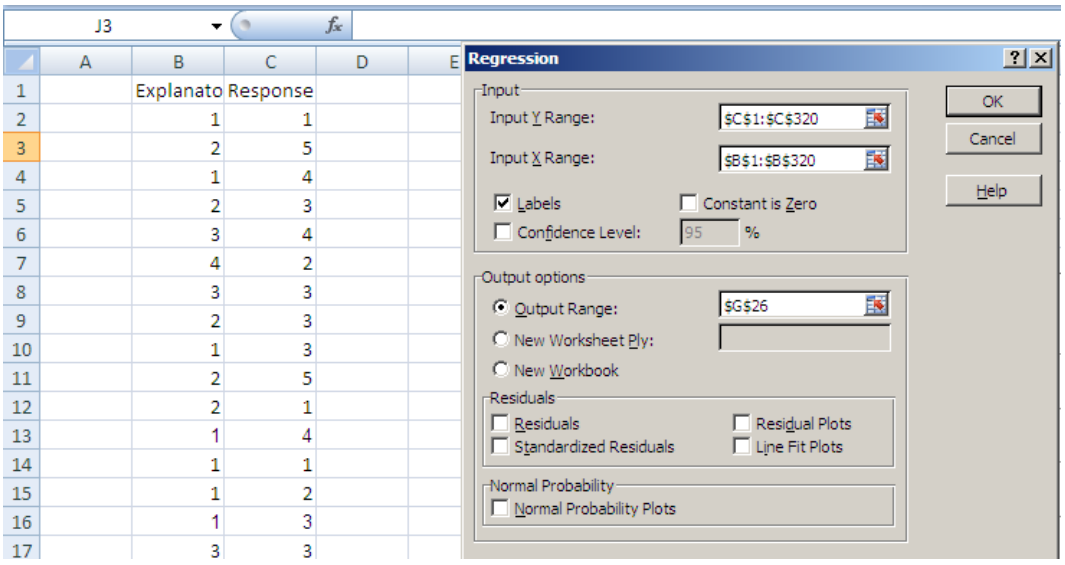
To use this method, the Analysis Toolpak must be loaded. If it is not loaded in the version of Excel you're using, see the section in the previous chapter explaining how to load the Analysis Toolpak.

STEPS FOR FINDING REGRESSION EQUATION AND CORRELATION COEFFICIENT USING DATA ANALYSIS

1. **On the Data tab, select Analysis and find Regression.**



2. **Specify the cells containing the data (if you include the column heading, be sure to select Labels too). Specify a place for the output to appear (the default is on another sheet, which is also fine).**



3. **When you hit OK, the output will appear.**

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.934638561
R Square	0.873549239
Adjusted R Square	0.873080903
Standard Error	1.447032524
Observations	272

ANOVA

	df	*SS*	*MS*	*F*	*gnificance F*
Regression	1	3905.587	3905.587	1865.218	3E-123
Residual	270	565.3538	2.093903		
Total	271	4470.941			

	Coefficients	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	1.349966279	0.172717	7.816063	1.22E-13	1.009923	1.690009	1.009923	1.690009
Explanatory	2.922485213	0.067669	43.18818	3E-123	2.78926	3.05571	2.78926	3.05571

4. **The regression equation is found by looking at the numbers in the column labeled *Coefficients*.**
   a. **Here the equation is roughly $\hat{y} = 1.35 + 2.92x$.**

5. **Find the magnitude of the correlation coefficient is next to the cell Multiple R. The sign is the same as the coefficient of the explanatory variable.**
   a. **Here the correlation coefficient is +0.935.**

Linear Regression using TI calculator

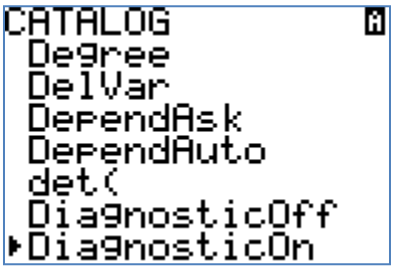
In this section we will explain how to run a linear regression analysis in the TI-83, -84 family of calculators.

Turning Diagnostics On.

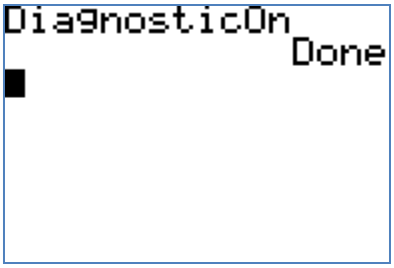
With the factory default settings in place, the calculator does not produce the correlation coefficient (r) or the coefficient of determination (r^2) when running a linear regression analysis. We want to get these values when we perform this analysis. To have the calculator produce r and r^2, turn on the Diagnostics. Here are the steps to do this.

STEPS TO TURNING ON DIAGNOSTICS

1. **Hit 2nd then 0 (this opens the CATALOG).**
2. **Scroll down to the command DiagnosticOn.**



3. **Hit ENTER to bring the command to the main screen, then hit ENTER again.**

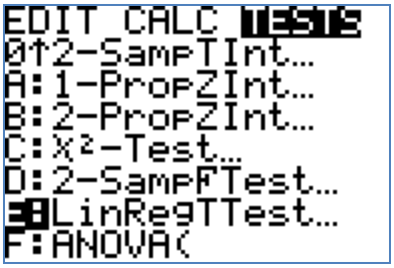


You won't typically have to do this ever again unless for some reason the factory default settings are reset.

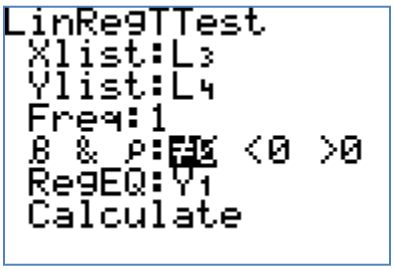
LinRegTTest

STEPS FOR RUNNING LINEAR REGRESSION ANALYSIS IN TI-83 [LinRegTTest]

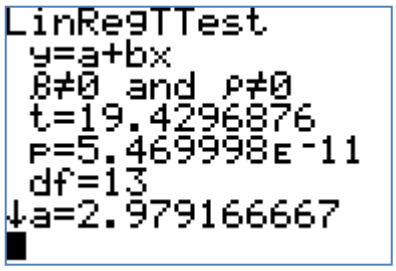
1. **Enter your data into two lists; one for the explanatory variable and another for the response.**
2. **From within the STAT menu, select TESTS, then find LinRegTTest.**

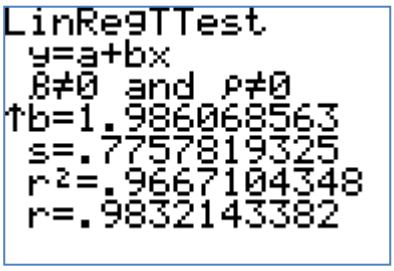


3. **Specify the two lists.**



4. **The next two lines are optional; you may safely skip to Step 5.**
   a. **The first optional line performs the hypothesis test against the null hypothesis that the slope and intercept are 0. If you want to use this you must choose the form of the alternate hypothesis.**
   b. **The second asks where you'd like the linear regression equation stored. To pick a location to store the equation, hit VARS, select Y-VARS then FUNCTION and pick a Y to save the regression equation as.**
5. **Last, ENTER on Calculate. The output looks like this.**



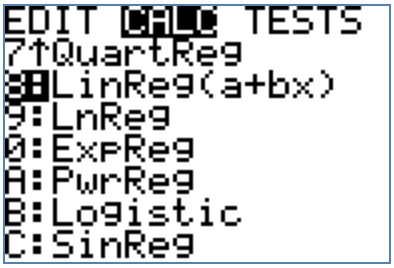


   a. **The equation of the regression line is $\hat{y} = 2.97 + 1.99x$.**
   b. **The correlation coefficient is about 0.983 and the coefficient of determination is about 0.967.**

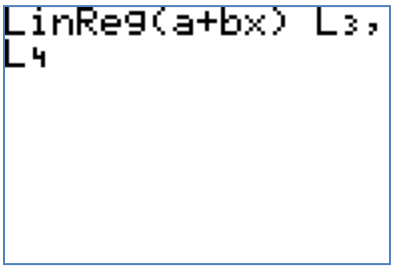
LinReg(ax+b) & LinReg(a+bx)

STEPS FOR RUNNING LINEAR REGRESSION ANALYSIS IN TI-83 USING LinReg(ax+b) OR LinReg(a+bx)

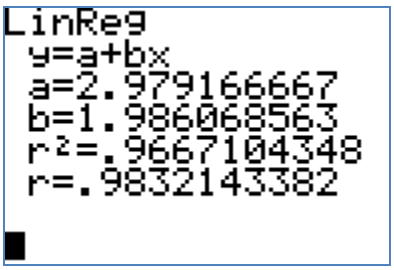
1. **Enter your data into two lists.**
2. **From within the STAT menu, select CALC, then find LinReg(a+bx) [or LinReg(ax +b) if you prefer].**



3. **The command LinReg(a+bx) now appears on the home screen.**
4. **Specify the lists containing the data. Enter the explanatory variable first; separate the lists by a comma.**



5. **Hit ENTER. The output looks like:**



   a. **The regression equation is $\hat{y} = 2.97 + 1.99x$.**
   b. **The correlation coefficient is about 0.983 and the coefficient of determination is about 0.967.**