

Hypothesis Testing in Excel



This handout uses three following data sets:

- [1.] [CEO Compensation 2008 Forbes.xlsx](#) for Topics 1, 2a.
- [2.] [Sales Presentation Ratings.xlsx](#) for Topic 2b.

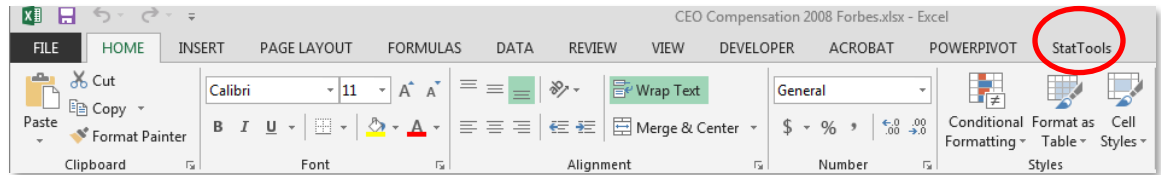
For help with Excel, go to: <http://office.microsoft.com/en-us/excel-help>

In this handout:

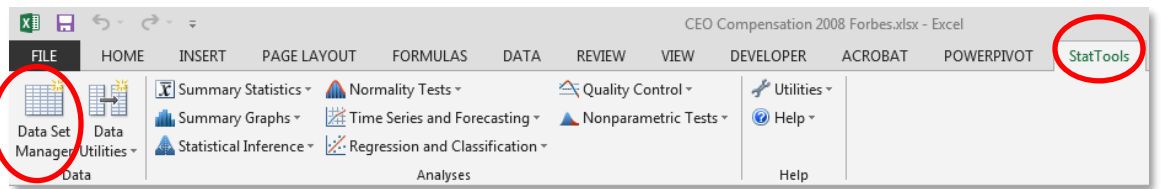
	Topic	Data	Page
0	Getting started with StatTools		2
1	Hypothesis testing for population mean		
	<ul style="list-style-type: none"> Using StatTools 	[1.]	3
2	Hypothesis testing for the difference in population means		
	a. Independent samples	[1.]	4
	<ul style="list-style-type: none"> Using Data Analysis ToolPak 		5
	<ul style="list-style-type: none"> Using StatTools 		6
	b. Matched (paired) samples	[2.]	8
	<ul style="list-style-type: none"> Using Data Analysis ToolPak 		8
	<ul style="list-style-type: none"> Using StatTools 		9

0. Getting Started with StatTools

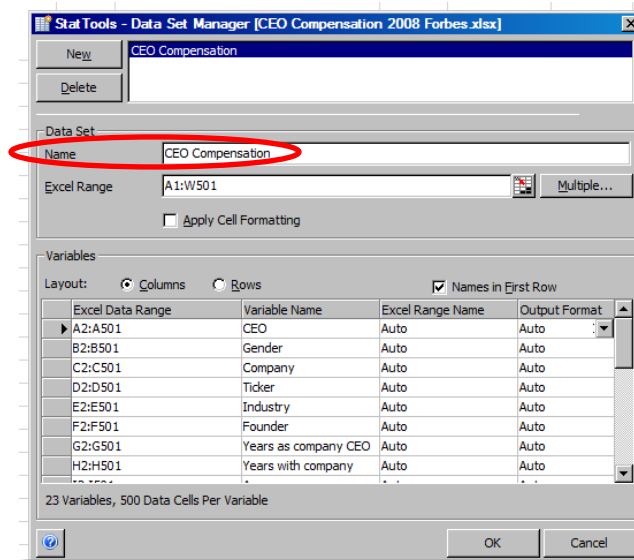
- Make sure StatTools is running. To connect remotely: <https://remoteapp.whitman.syr.edu>



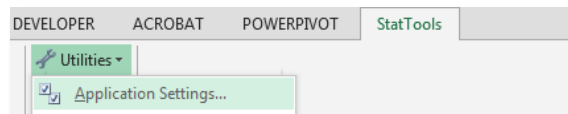
- Highlight all the data (Columns A through W).
- StatTools → Data Set Manager



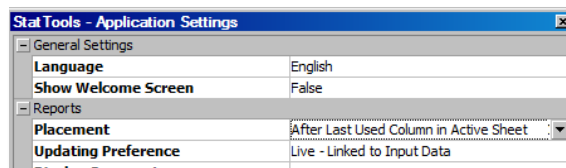
- Click Yes. Change data set's **Name** to "CEO Compensation." Make sure **Excel Range** covers columns A through W (by default, the correct range should be already selected). Click OK.



- You can now use StatTools. You should always perform the above steps whenever you need to do analysis of data using the StatTools add-in.
- Change placement of reports:
 - StatTools → Utilities → Application Settings.



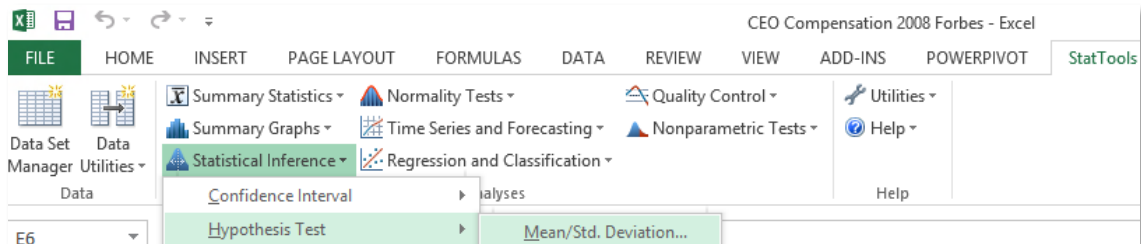
- In **Reports** → **Placement**, select "After Last Used Column in Active Sheet" → OK.



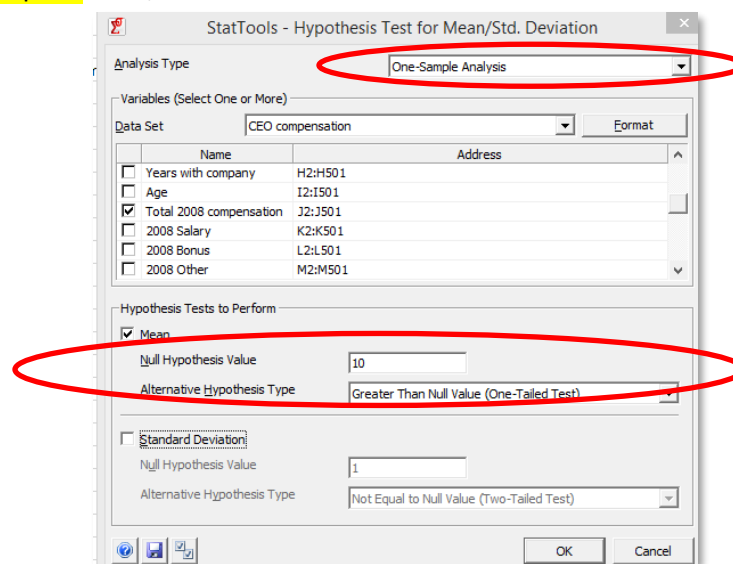
1. HYPOTHESIS TESTING FOR POPULATION MEAN (μ)

Use data set **CEO Compensation 2008 Forbes.xlsx**. Suppose you want to test whether the *average* Total Compensation of *all* CEOs is greater than \$10 million.

- StatTools → Data Set Manager → Rename the data to “CEO Compensation.”
- StatTools → Statistical Inference → Hypothesis Test → Mean:



- Make sure that **Analysis Type** is set to “One-Sample Analysis.” Put a check mark next to the variable *Total 2008 Compensation*. On the bottom, in **Null Hypothesis Value** write **10** and in **Alternative Hypothesis Type** select “Greater Than Null Value (One-Tailed Test),” because our alternative hypothesis is $H_A: \mu > 10$. Also, remove the check mark from **Standard Deviation**. Click OK.



- The following output will appear. The hypothesis is tested using 3 levels of α : 10%, 5%, and 1%.

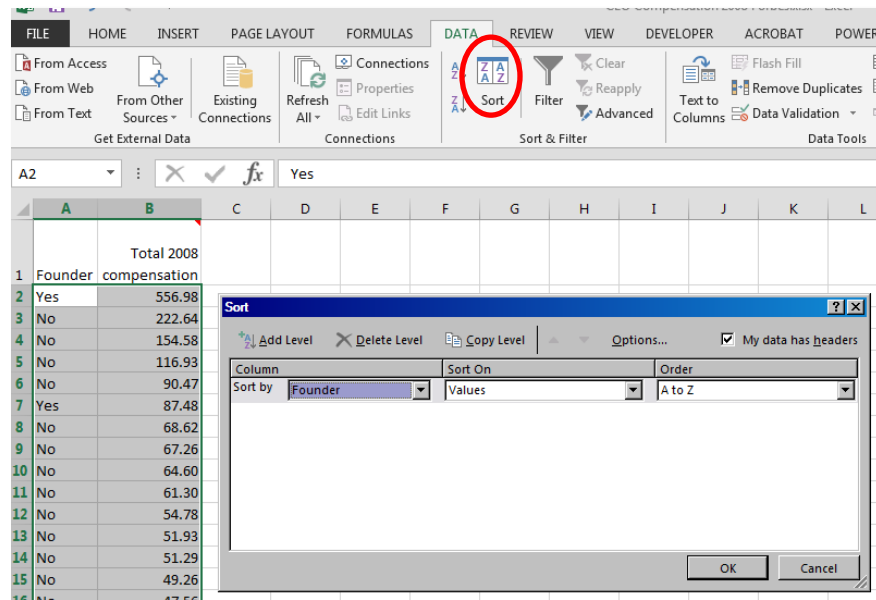
	Total 2008 compensation
Hypothesis Test (One-Sample)	CEO compensation
Sample Size	499
Sample Mean	11.43
Sample Std Dev	29.88
Hypothesized Mean	10
Alternative Hypothesis	> 10
Standard Error of Mean	1.34
Degrees of Freedom	498
t-Test Statistic	1.0689
p-Value	0.1428
Null Hypoth. at 10% Significance	Don't Reject
Null Hypoth. at 5% Significance	Don't Reject
Null Hypoth. at 1% Significance	Don't Reject

2. HYPOTHESIS TESTING FOR THE DIFFERENCE IN POPULATION MEANS ($\mu_1 - \mu_2$)

a) Independent samples

Use data set **CEO Compensation 2008 Forbes.xlsx**. Suppose we want to test whether there is a difference in the average Total Compensations between all CEOs that are founders and that are non-founders.

- Copy and paste Column F (Founder: Yes, No) and Column J (Total 2008 Compensation) into a new spreadsheet. You can highlight Column F, press **CTRL**, highlight Column J; then copy and paste.
- We now need to split the data by Founder (Yes, No): Total 2008 Compensation for Founder=Yes, and Total 2008 Compensation for Founder=No. To do so, sort the two columns by Founder.



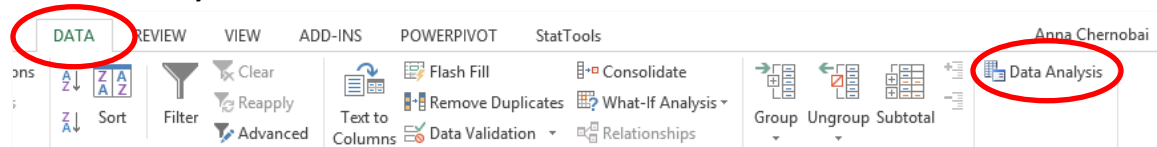
- Then, create two new columns: *Total 2008 Compensation-Non-founder* and *Total 2008 Compensation-Founder*. In each column, copy and paste from Column B the compensation values of Founders (Yes) and Non-Founders (No), respectively. The two new columns will be of different lengths; that's fine.

	A	B	C	D
		Total 2008 compensation	Total 2008 compensation - Non-founder	Total 2008 compensation - Founder
1	Founder	compensation		
2	No	222.64	222.64	556.98
3	No	154.58	154.58	87.48
4	No	116.93	116.93	44.49
5	No	90.47	90.47	36.01
6	No	68.62	68.62	29.75
7	No	67.26	67.26	12.05
8	No	64.60	64.60	8.23
9	No	61.30	61.30	8.17
10	No	54.78	54.78	5.58
11	No	51.93	51.93	1.99
12	No	51.29	51.29	1.95
13	No	49.26	49.26	1.40
14	No	47.56	47.56	1.30
15	No	42.68	42.68	1.28
16	No	42.27	42.27	0.97
17	No	39.26	39.26	0.51
18	No	39.22	39.22	0.00
19	No	38.66	38.66	
20	No	38.12	38.12	

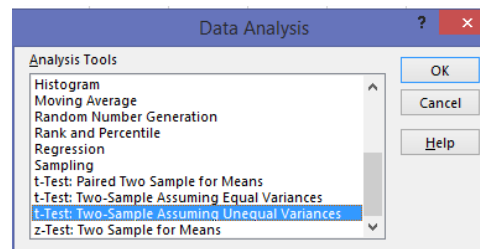
- We can now proceed to hypothesis testing for the difference in population means. There are two ways to do so, as we will see next.

[1.] Using Excel's Data Analysis ToolPak:

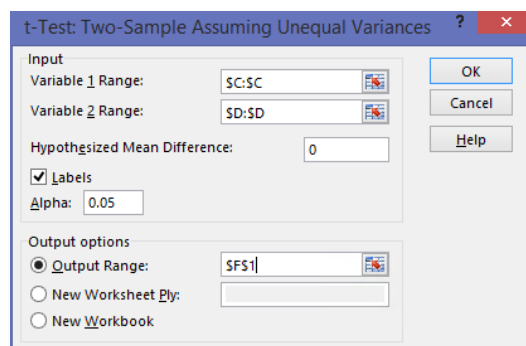
- To activate Data Analysis ToolPak: go to **File** → **Options** → **Add-Ins** → around the top of the list find and click on **Analysis ToolPak**; on the bottom next to **Manage: Excel Add-ins** click on **Go...** and put a check mark next to **Analysis ToolPak** → Click OK.



- Data** → **Data Analysis** → scroll down, select **t-Test: Two-Sample Assuming Unequal Variances**. Click OK.



- Specify the following inputs:
 - In **Variable 1 Range**, click on Column C (*Total 2008 compensation – Non-founder*).
 - In **Variable 2 Range**, click on Column D (*Total 2008 compensation – Founder*).
 - In **Hypothesized Mean Difference** write **0**, because our null and alternative hypotheses are formulated as:
 $H_0: \mu_1 = \mu_2$ (which is the same as: $H_0: \mu_1 - \mu_2 = 0$)
 $H_A: \mu_1 \neq \mu_2$ (which is the same as: $H_A: \mu_1 - \mu_2 \neq 0$)
 - Put a check mark in **Labels**.
 - Specify **Alpha**. Suppose, we need to use $\alpha=5\%$: then write 0.05.
 - In **Output options**:
 - Either select **New Worksheet Ply** if you want your output to be in a separate tab, or
 - Click on **Output Range** and click on cell F1 for the range to place the report so that the upper left corner is in cell F1.
 - Click OK.

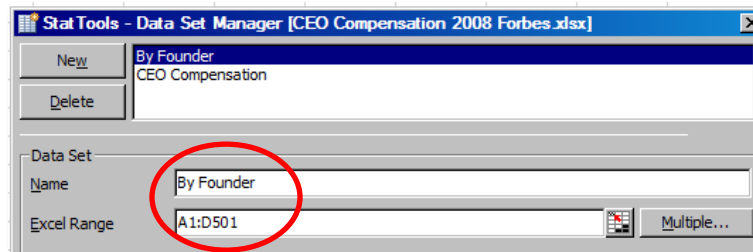


- Expand the columns of the output table to see better all results. The following output will appear:

t-Test: Two-Sample Assuming Unequal Variances		
	Total 2008 compensation - Non-founder	Total 2008 compensation - Founder
Mean	10.17716598	46.94941176
Variance	286.159859	17805.84267
Observations	482	17
Hypothesized Mean Difference	0	
df	16	
t Stat	-1.135900524	
P(T<=t) one-tail	0.136363485	
t Critical one-tail	1.745883676	
P(T<=t) two-tail	0.272726971	
t Critical two-tail	2.119905299	

[2.] Using StatTools:

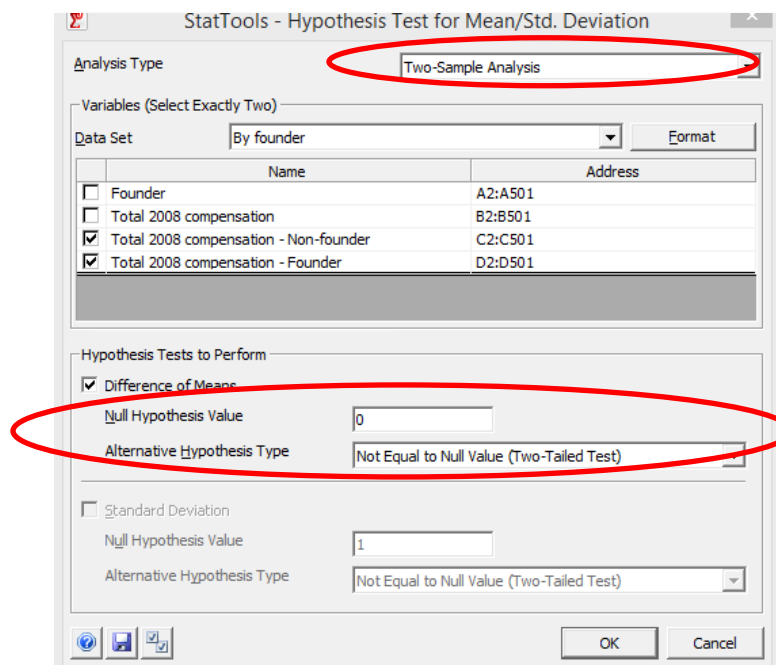
- Highlight the four columns (A through D).
- StatTools** → **Data Set Manager** → Rename the new dataset, call it “By Founder”. Click OK.



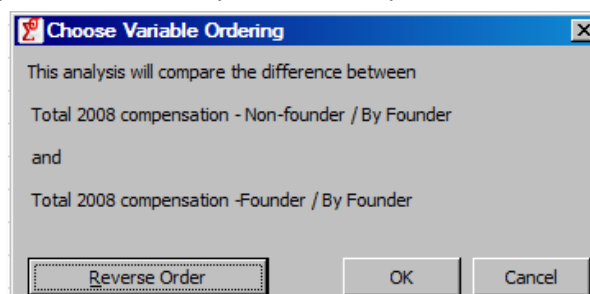
- StatTools** → **Statistical Inference** → **Hypothesis Test** → **Mean**.
- Set **Analysis Type** to “Two-Sample Analysis.” Make sure the **Data Set** is “By Founder.” Put check marks next to *Total 2008 Compensation – Non-founder* and *Total 2008 Compensation – Founder*. In **Null Hypothesis Value** write **0** and set **Alternative Hypothesis Type** to “Not Equal to Null Value (Two-Tailed Test)”, because our null and alternative hypotheses are formulated as:

$$H_0: \mu_1 = \mu_2 \quad (\text{which is the same as: } H_0: \mu_1 - \mu_2 = 0)$$

$$H_A: \mu_1 \neq \mu_2 \quad (\text{which is the same as: } H_A: \mu_1 - \mu_2 \neq 0)$$



- Click OK. The following screen will appear. You don’t need to reverse the order. The resulting hypothesis testing results will be for $\mu(\text{Non-Founder}) - \mu(\text{Founder})$. If you don’t want to reverse the order, click OK.



- The following output will appear. I have highlighted the portion that we need to use (we will always assume Unequal Variances in this course). The hypothesis is tested using 3 levels of α : 10%, 5%, and 1%.

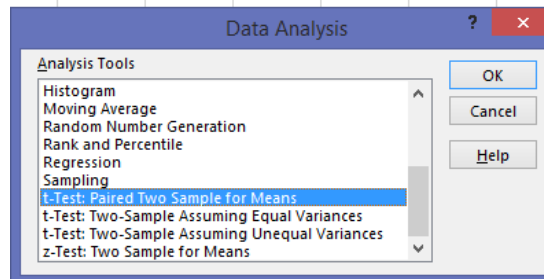
	Total 2008 compensation - Non-founder	Total 2008 compensation - Founder
<i>Sample Summaries</i>	By founder	By founder
Sample Size	482	17
Sample Mean	10.18	46.95
Sample Std Dev	16.92	133.44
	Equal	Unequal
<i>Hypothesis Test (Difference of Means)</i>	Variances	Variances
Hypothesized Mean Difference	0	0
Alternative Hypothesis	$\neq 0$	$\neq 0$
Sample Mean Difference	-36.77	-36.77
Standard Error of Difference	7.195420011	32.37276946
Degrees of Freedom	497	16
t-Test Statistic	-5.1105	-1.1359
p-Value	< 0.0001	0.2727
Null Hypoth. at 10% Significance	Reject	Don't Reject
Null Hypoth. at 5% Significance	Reject	Don't Reject
Null Hypoth. at 1% Significance	Reject	Don't Reject
<i>Equality of Variances Test</i>		
Ratio of Sample Variances	0.0161	
p-Value	< 0.0001	

b) Matched (paired) samples

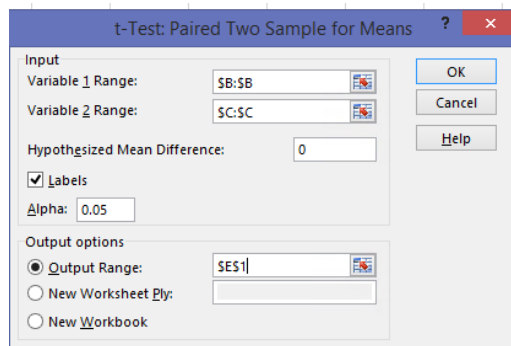
Use the dataset **Sales Presentation Ratings.xlsx**. Suppose we are interested in determining whether there is a significant difference in the average responses between husbands and their wives.

[1.] Using Excel's Data Analysis ToolPak:

- Data → Data Analysis → t-Test: Paired Two Sample for Means → click OK.



- Specify the following inputs:
 - In **Variable 1 Range**, click on Column B (*Husband*).
 - In **Variable 2 Range**, click on Column C (*Wife*).
 - In **Hypothesized Mean Difference** write **0**, because our hypotheses are formulated as:
 - $H_0: \mu_1 = \mu_2$ (which is the same as: $H_0: \mu_1 - \mu_2 = 0$)
 - $H_A: \mu_1 \neq \mu_2$ (which is the same as: $H_A: \mu_1 - \mu_2 \neq 0$)
 - Put a check mark in **Labels**.
 - Specify **Alpha**. Suppose, we need to use $\alpha=5\%$: then write 0.05.
 - In **Output options**:
 - Either select **New Worksheet Ply** if you want your output to be in a separate tab, or
 - Click on **Output Range** and click on cell E1 for the range to place the report so that the upper left corner is in cell E1.
 - Click OK.

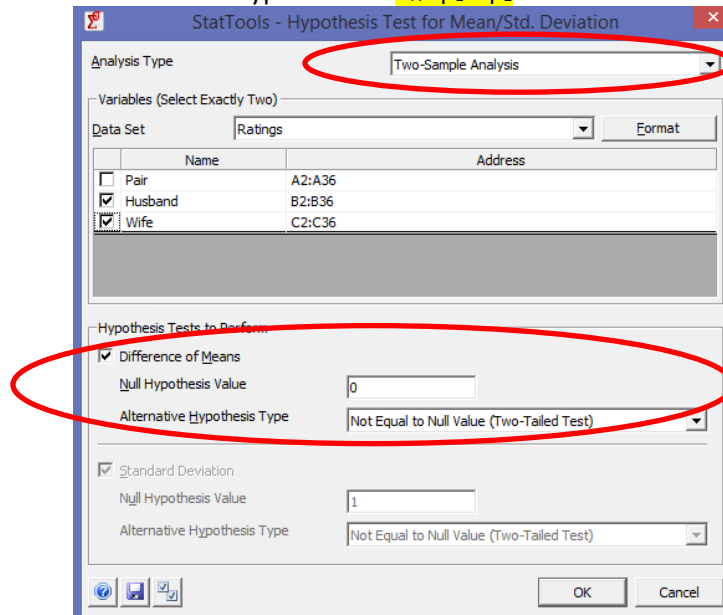


- Expand the columns of the output table to see better the results. The following output will appear:

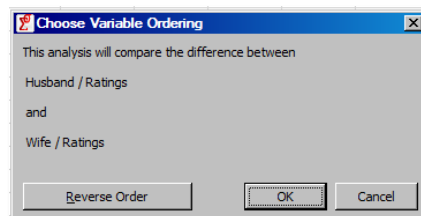
t-Test: Paired Two Sample for Means		
	Husband	Wife
Mean	6.91428571	5.285714
Variance	1.49243697	3.210084
Observations	35	35
Pearson Correlation	0.4415142	
Hypothesized Mean Difference	0	
df	34	
t Stat	5.78922878	
P(T<=t) one-tail	8.0883E-07	
t Critical one-tail	1.69092426	
P(T<=t) two-tail	1.6177E-06	
t Critical two-tail	2.03224451	

[2.] Using StatTools:

- StatTools → Data Set Manager → Rename the data to “Ratings.”
- StatTools → Statistical Inference → Hypothesis Test → Mean.
- In the **Analysis Type**, select “Paired-Sample Analysis.” Put a check mark next to *Husband* and *Wife*. Specify **Null Hypothesis Value** as **0** and **Alternative Hypothesis Type** as “Not Equal to Null Value (Two-Tailed Test),” because our alternative hypothesis is $H_A: \mu_1 \neq \mu_2$ which is the same as $H_A: \mu_1 - \mu_2 \neq 0$. Click OK.



- The following screen will appear. You don't need to reverse the order. The resulting hypothesis testing is for $\mu(\text{Husband}) - \mu(\text{Wife})$.



- The following output will be produced. The hypothesis is tested using 3 levels of α : 10%, 5%, and 1%.

	Husband	Wife
Sample Summaries	Ratings	Ratings
Sample Size	35	35
Sample Mean	6.914	5.286
Sample Std Dev	1.222	1.792
	Equal	Unequal
Hypothesis Test (Difference of Means)	Variances	Variances
Hypothesized Mean Difference	0	0
Alternative Hypothesis	$\neq 0$	$\neq 0$
Sample Mean Difference	1.629	1.629
Standard Error of Difference	0.366548418	0.366548418
Degrees of Freedom	68	59
t-Test Statistic	4.4430	4.4430
p-Value	< 0.0001	< 0.0001
Null Hypoth. at 10% Significance	Reject	Reject
Null Hypoth. at 5% Significance	Reject	Reject
Null Hypoth. at 1% Significance	Reject	Reject
Equality of Variances Test		
Ratio of Sample Variances	0.4649	
p-Value	0.0285	