

## Confidence Intervals in Excel



This handout uses the following data sets:

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

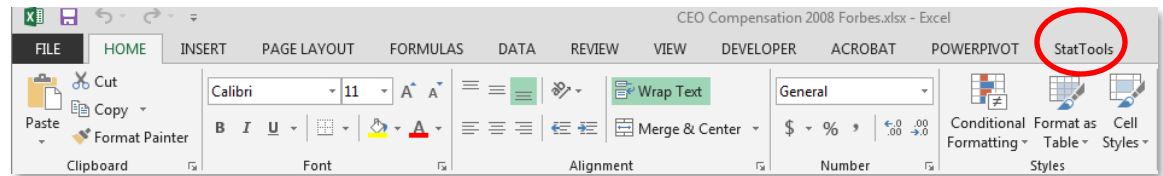
For help with Excel, go to: <http://office.microsoft.com/en-us/excel-help>

In this handout:

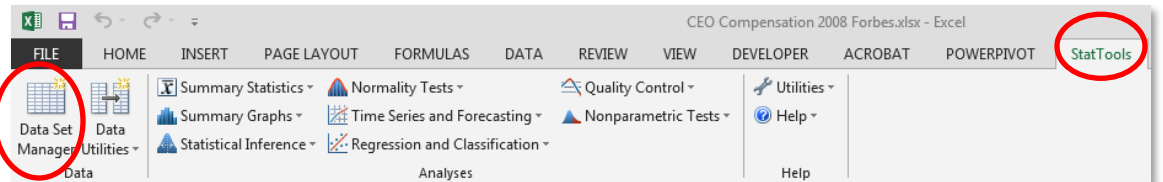
	Topic	Data	Page
0	<a href="#">Getting started with StatTools</a>		2
1	Confidence interval for population mean		3
1.a.	<a href="#">Constructing confidence interval</a>	[1.]	3
1.b.	<a href="#">Selecting appropriate sample size</a>	[1.]	4
2.	Confidence interval for the difference in population means		5
2.a.	<a href="#">Constructing confidence interval</a>	[1.]	5
2.b.	<a href="#">Constructing confidence interval: matched samples</a>	[2.]	8

## 0. Getting Started With StatTools.

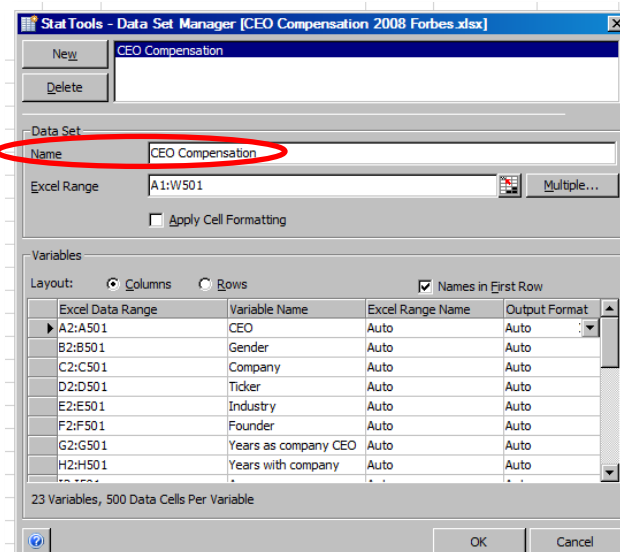
- Make sure StatTools is running.



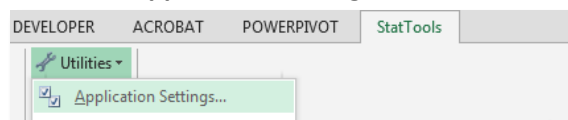
- Highlight all the data (Columns A through W).
- StatTools → Data Set Manager



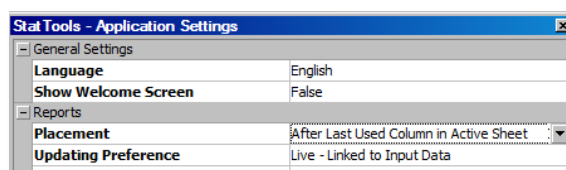
- Click Yes. Change data set's **Name** to "CEO Compensation." Make sure **Excel Range** covers columns A through W (by default, the right range should be already selected). Click OK.



- 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.
- Placement of reports:
  - StatTools → Utilities → Application Settings.



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

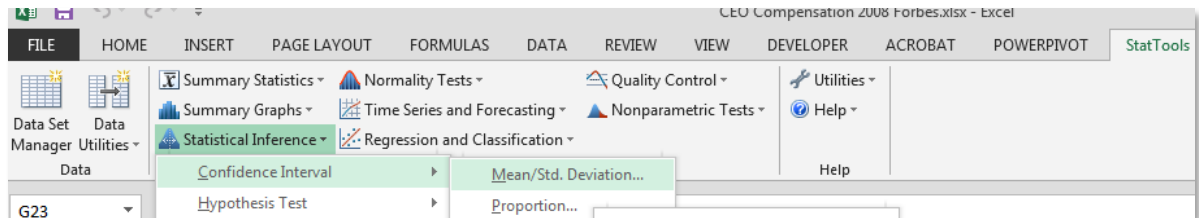


# 1. CONFIDENCE INTERVAL FOR POPULATION MEAN ( $\mu$ )

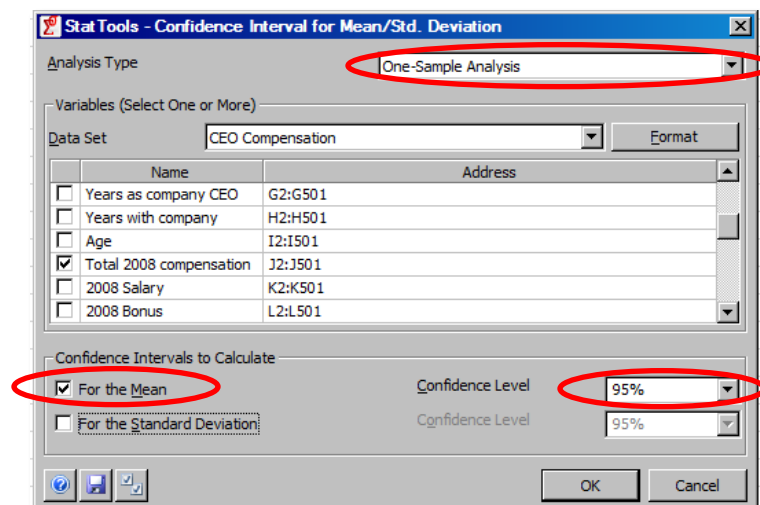
## a) Constructing confidence interval.

Suppose we want to construct a 95% confidence interval for the *average* Total Compensation of *all* CEOs in the US.

- StatTools → Statistical Inference → Confidence Interval → 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 left, remove the check mark from “For the Standard Deviation.”
- On the bottom right, make sure **Confidence Level** is set to 95% (or the level desired). Click OK.



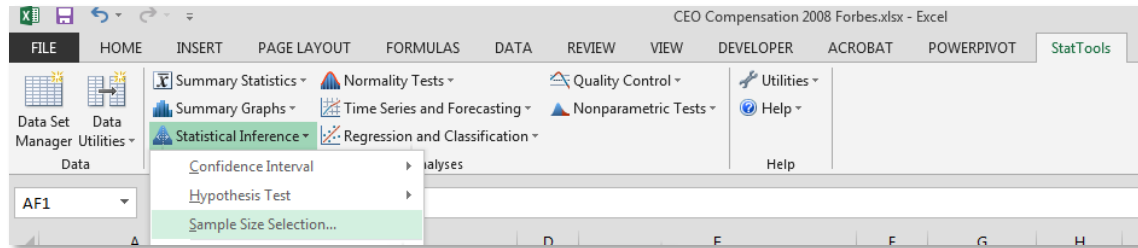
- The following output will appear. I have highlighted the confidence interval in orange.

	Total 2008 compensation
Conf. Intervals (One-Sample)	CEO Compensation
Sample Size	499
Sample Mean	11.43
Sample Std Dev	29.88
Confidence Level (Mean)	95.0%
Degrees of Freedom	498
Lower Limit	8.80
Upper Limit	14.06

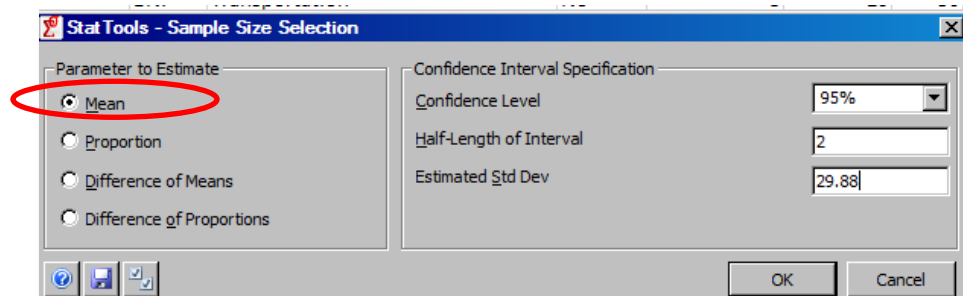
## b) Selecting appropriate sample size.

Suppose we need to find what the appropriate sample size ( $n$ ) should be used in order for the margin of error to be no greater than 2.

- StatTools → Statistical Inference → Sample Size Selection:



- Make sure that **Parameter to Estimate** is set to “Mean.”
- Make sure **Confidence Level** is set to 95% (or another desired level).
- Half-Length of Interval** is the same as margin of error. Set it to 2 for our example (or whatever margin of error you need to achieve).
- Estimated Std Dev** is the sample standard deviation; in our example it is 29.88.



- Click OK.
- The following output will appear. The appropriate sample size is highlighted in orange.

<i>Sample Size for Mean</i>	
Confidence Level	95.00%
Half-length of Interval	2
Std Dev (estimate)	29.88
Sample Size	858

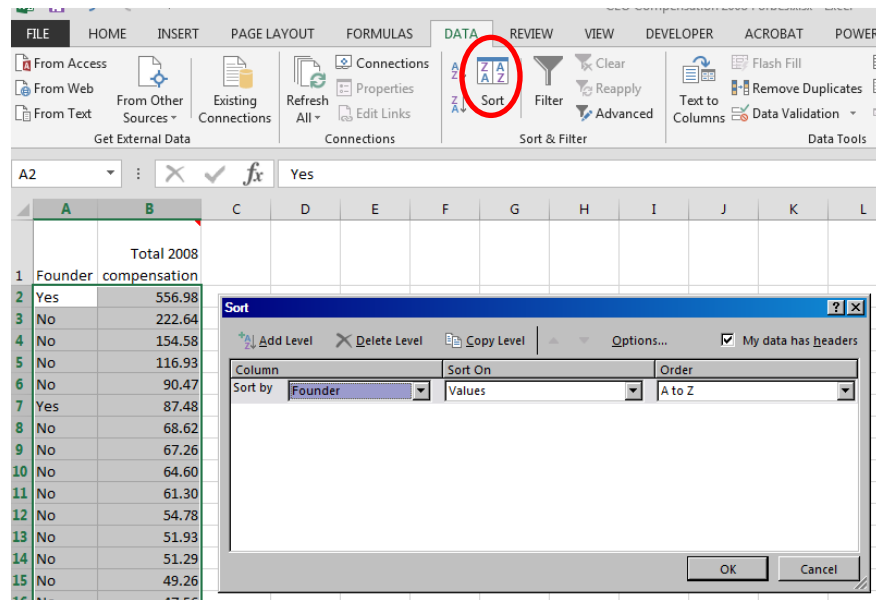
- The answer to the problem formulated on top of this page is that the sample size should be 858 or above.

## 2. CONFIDENCE INTERVAL FOR THE DIFFERENCE IN POPULATION MEANS

### a) Constructing confidence interval.

Suppose we want to construct a 95% confidence interval for the *difference in the average* Total Compensation between *all* CEOs in the US that are founders and that are non-founders.

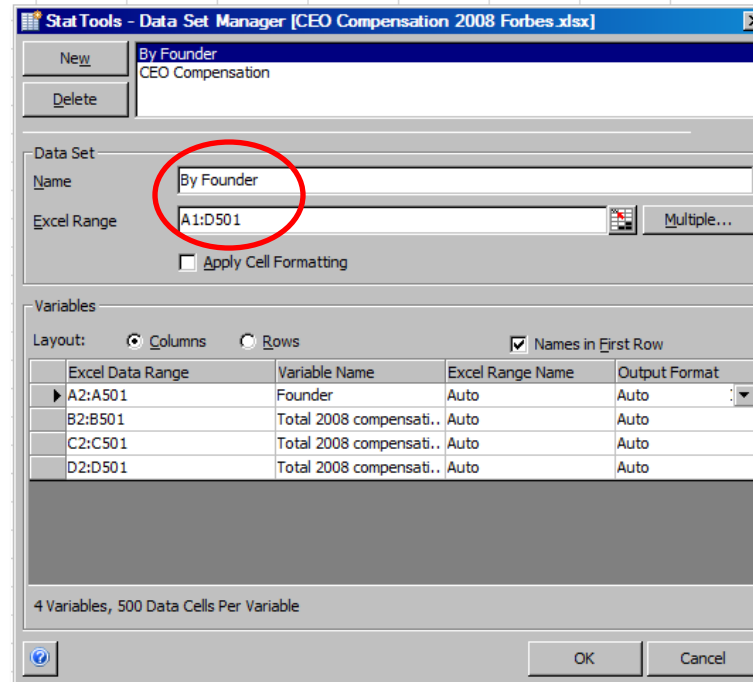
- Take two columns, Column F (*Founder: Yes, No*) and Column J (*Total 2008 Compensation*), copy and paste them in a new spreadsheet. You can highlight Column F, press **CTRL**, highlight Column J; then copy and paste.
- We now need to split the data by *Founder* (Yes, No): *Total 2008 Compensation for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



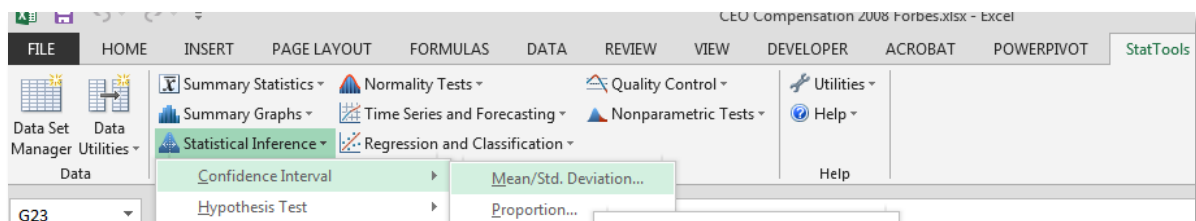
- Then, create two new additional columns: *Total 2008 Compensation-Non-founder* and *Total 2008 Compensation-Founder*. In each column, copy and paste from the original column the compensation values corresponding to Founders (Yes) and Non-Founders (No), respectively. The two new columns will be of different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

	A	B	C	D
		Total 2008 compensation	Total 2008 compensation - Non-founder	Total 2008 compensation - Founder
1	Founder			
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	

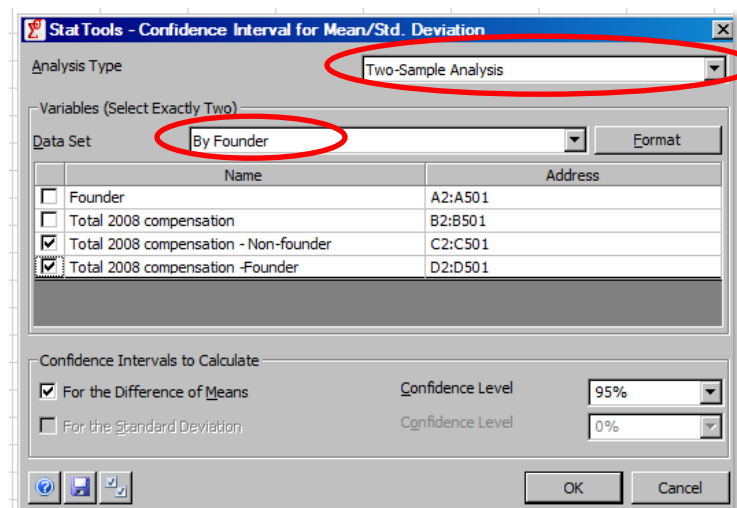
- Highlight the four columns (A through D).
- **StatTools → Data Set Manager →** rename the new dataset, call it “By Founder”, then click OK:



- We can now proceed to constructing the confidence interval for the difference in average compensations between *all* CEOs that are founders and that are non-founders.
- **StatTools → Statistical Inference → Confidence Interval → Mean:**

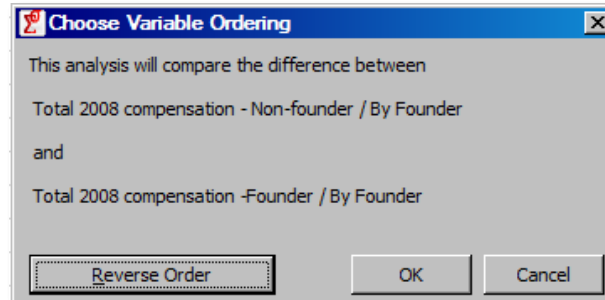


- Make sure that **Analysis Type** is set to “Two-Sample Analysis.” Make sure the **Data Set** is “By Founder.” Put check marks next to “Total 2008 Compensation – Non-founder” and “Total 2008 Compensation – Founder.” On the bottom right, set the confidence level to 95% (or another desired level).



- Click OK.

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



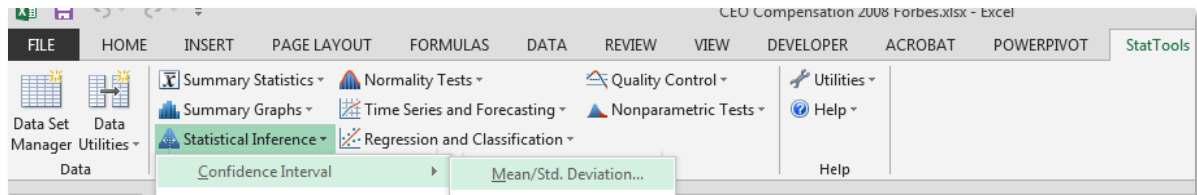
- The following output will appear. I have highlighted in yellow the portion that we need to use (we will always assume Unequal Variances in this course).

	Total 2008 compensation - Non-founder	Total 2008 compensation -Founder
<b>Sample Summaries</b>	<b>By Founder</b>	<b>By Founder</b>
Sample Size	482	17
Sample Mean	10.18	46.95
Sample Std Dev	16.92	133.44
	Equal	Unequal
<b>Conf. Intervals (Difference of Means)</b>	<b>Variances</b>	<b>Variances</b>
Confidence Level	95.0%	95.0%
Sample Mean Difference	-36.77	-36.77
Standard Error of Difference	7.195420011	32.37276946
Degrees of Freedom	497	16
Lower Limit	-50.90943728	-105.3994513
Upper Limit	-22.6350543	31.85495974
<b>Equality of Variances Test</b>		
Ratio of Sample Variances	0.0161	
p-Value	< 0.0001	

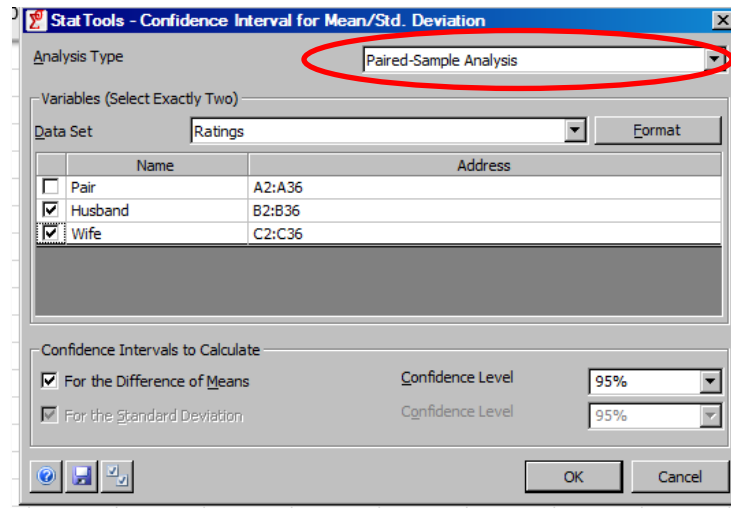
## b) Constructing confidence interval: matched samples.

Using the dataset **Sales Presentation Ratings.xlsx**, suppose we are interested in determining whether there is a significance *difference in the average* responses between husbands and their wives. We can answer such a question using confidence interval for the difference in mean responses using, say, 95% confidence level.

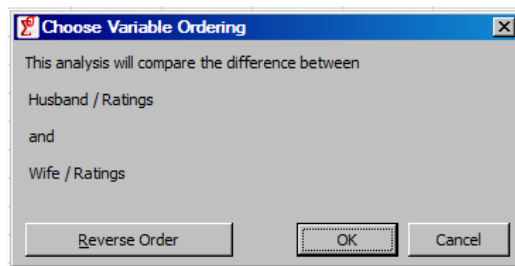
- **StatTools** → **Data Set Manager** → Rename the data to “Ratings.”
- **StatTools** → **Statistical Inference** → **Confidence Interval** → **Mean**:



- In the **Analysis Type**, select “Paired-Sample Analysis.”
- Put a check mark next to *Husband* and *Wife*. Make sure the confidence level is set to 95% (or another desired level). Click OK.



- The following screen will appear. You don't need to reverse the order. The resulting confidence interval that will be produced will be for  $\mu(\text{Husband}) - \mu(\text{Wife})$ .



- The following output will be produced. I have highlighted the confidence interval in orange.

Conf. Intervals (Paired-Sample)	Husband - Wife
Sample Size	35
Sample Mean	1.628571429
Sample Std Dev	1.66425596
Confidence Level	95.0%
Degrees of Freedom	34
Lower Limit	1.056879504
Upper Limit	2.200263353