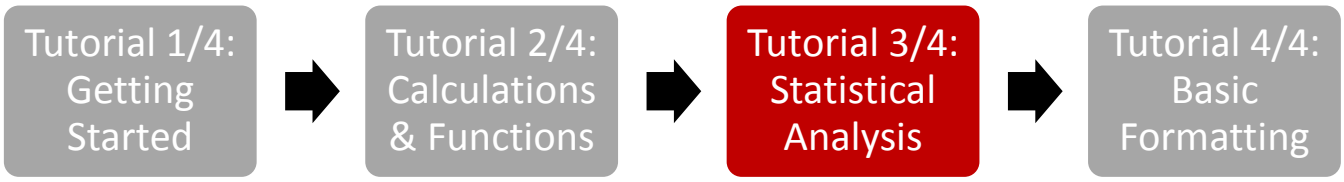
Basic Microsoft Excel skills will be required for some of your Glenn College courses and may prove very helpful for others. This four-part tutorial addresses getting started with Excel, performing calculations and functions, conducting data analysis, and basic formatting.

If you are interested in watching a 10-minute video overview of basic Excel data creation and formatting, here is a YouTube video link: https://www.youtube.com/watch?v=8L1OVkw2ZQ8.



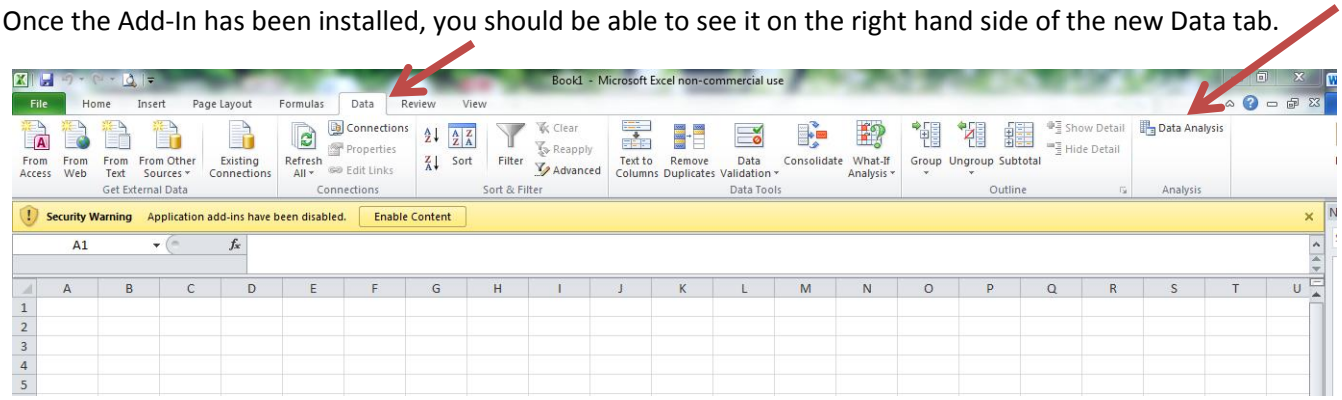
Tutorial 3/4: Statistical Analysis in Excel

This document covers (click each link to skip to that section)

Installing Excel Add-Ins

To create tables and conduct hypothesis testing, you will need the Data Analysis Toolpak Add-In. Here are Microsoft's detailed instructions for installing this Add-In: https://support.office.com/en-us/article/Load-the-Analysis-ToolPak-6a63e598-cd6d-42e3-9317-6b40ba1a66b4

Once the Add-In has been installed, you should be able to see it on the right hand side of the new Data tab.



In Excel, open the file with your data.

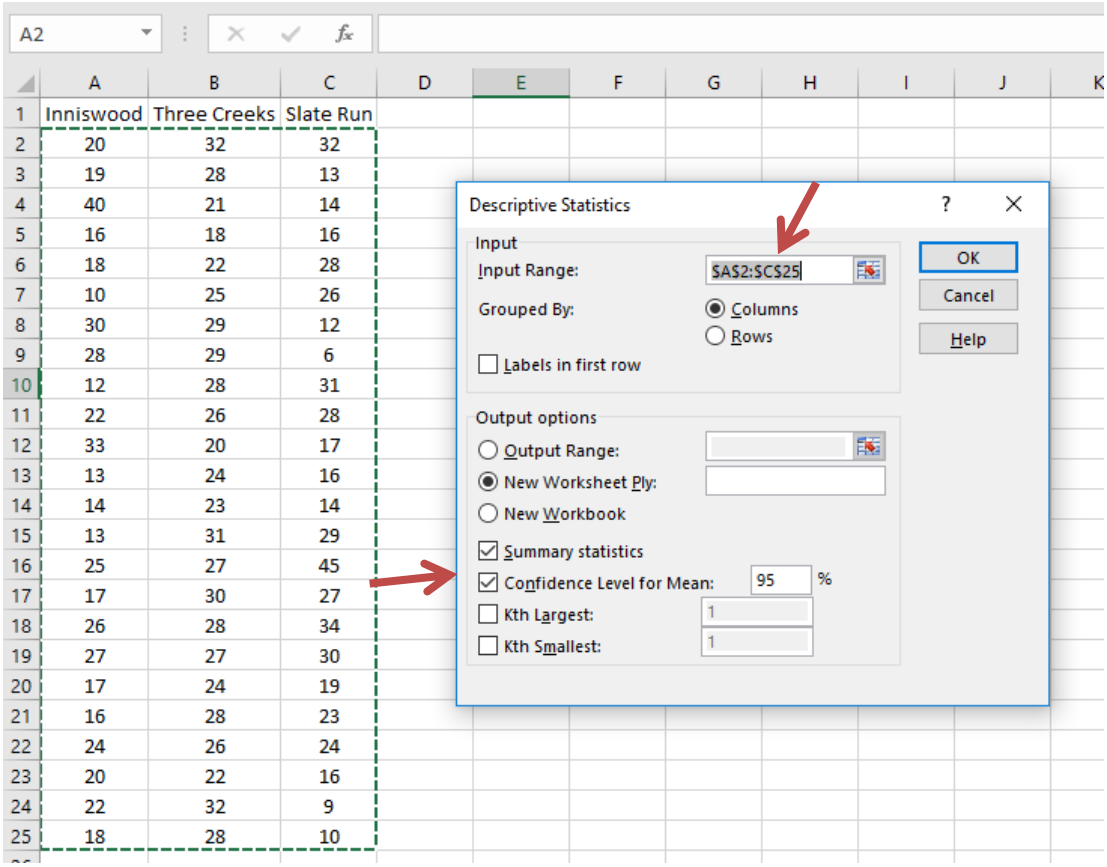
For this tutorial use Tutorial_Blackbirds.xlsx. This is synthetic data about the number of Red Winged Blackbirds spotted in different Metro Parks over 24 randomly selected days.

Creating a Summary Statistics Table

Click the **Data tab** on the "ribbon" of options. Click **Data Analysis** in the Analysis group.

Choose **Descriptive Statistics**. Click **OK**.

For the Input Range, use the mouse to **select all of the data but not the headers** … A2:C25.



	A	B	C
1	Inniswood	Three Creeks	Slate Run
2	20	32	32
3	19	28	13
4	40	21	14
5	16	18	16
6	18	22	28
7	10	25	26
8	30	29	12
9	28	29	6
10	12	28	31
11	22	26	28
12	33	20	17
13	13	24	16
14	14	23	14
15	13	31	29
16	25	27	45
17	17	30	27
18	26	28	34
19	27	27	30
20	17	24	19
21	16	28	23
22	24	26	24
23	20	22	16
24	22	32	9
25	18	28	10

On the same Descriptive Statistics window, under Output options, click **Summary statistics** and **Confidence Level for Mean (95%)**. Click **OK**.

A new worksheet, Sheet1, will appear with three tables of raw summary statistics. It will look like this:

	A	B	C	D	E	F
1	*Column1*		*Column2*		*Column3*	
2						
3	Mean	20.83333	Mean	26.16667	Mean	21.625
4	Standard E	1.478918	Standard E	0.770062	Standard E	1.951228
5	Median	19.5	Median	27	Median	21
6	Mode	20	Mode	28	Mode	16
7	Standard D	7.245188	Standard D	3.772517	Standard D	9.559027
8	Sample Va	52.49275	Sample Va	14.23188	Sample Va	91.375
9	Kurtosis	0.670832	Kurtosis	-0.47447	Kurtosis	-0.17957
10	Skewness	0.851513	Skewness	-0.41678	Skewness	0.429022
11	Range	30	Range	14	Range	39
12	Minimum	10	Minimum	18	Minimum	6
13	Maximum	40	Maximum	32	Maximum	45
14	Sum	500	Sum	628	Sum	519
15	Count	24	Count	24	Count	24
16	Confidenc	3.059375	Confidenc	1.592994	Confidenc	4.036423

Excel does not include the column headings from the data set. It is a good practice to copy/paste these in so that if time passes between creating the table and coming back to it, you will be able to connect the summary statistics with the correct variable.

	A	B	C	D	E	F
1	*Column1*	Inniswood	*Column2*	Three Creeks	*Column3*	Slate Run
2						
3	Mean	20.83333333	Mean	26.16666667	Mean	21.625
4	Standard E	1.47891787	Standard E	0.770061796	Standard E	1.951228246
5	Median	19.5	Median	27	Median	21
6	Mode	20	Mode	28	Mode	16
7	Standard I	7.245188308	Standard I	3.772516939	Standard I	9.559027147
8	Sample Va	52.49275362	Sample Va	14.23188406	Sample Va	91.375
9	Kurtosis	0.670832355	Kurtosis	-0.474470697	Kurtosis	-0.179572735
10	Skewness	0.851512583	Skewness	-0.416777831	Skewness	0.42902181
11	Range	30	Range	14	Range	39
12	Minimum	10	Minimum	18	Minimum	6
13	Maximum	40	Maximum	32	Maximum	45
14	Sum	500	Sum	628	Sum	519
15	Count	24	Count	24	Count	24
16	Confidenc	3.059374708	Confidenc	1.592994194	Confidenc	4.03642316
17						

This table needs formatting! We should never include a table that looks like this in our final policy brief/memo/report.

In addition to the inconsistent decimals, the row labels in this table are in triplicate. *This table needs major reformatting before it can be included in a policy brief, paper, or memo.*

For information on formatting cells, rows, and columns, refer to **Tutorial 4 of 4: Formatting** included in this set of four tutorials.

Hypothesis Testing

The means for Inniswood, Three Creeks, and Slate Run are 20.83, 26.17, and 21.63 respectively. We would like to know if the difference in these means is statistically significant.

The null hypothesis and alternative hypotheses:

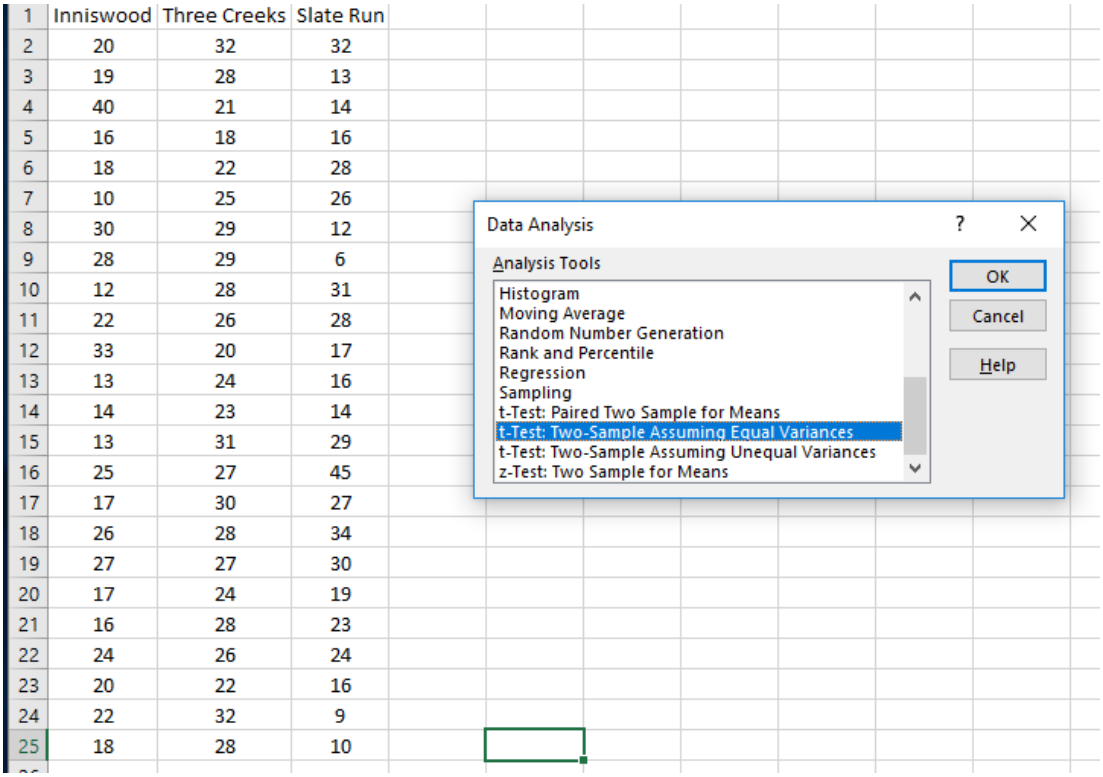
H_0: The mean number of red winged blackbirds at Inniswood = the mean number of red winged blackbirds at Three Creeks

H_a: The mean number of red winged blackbirds at Inniswood and Three Creeks are not equal (they are different).

To test this hypothesis in Excel **return to the raw data worksheet**.

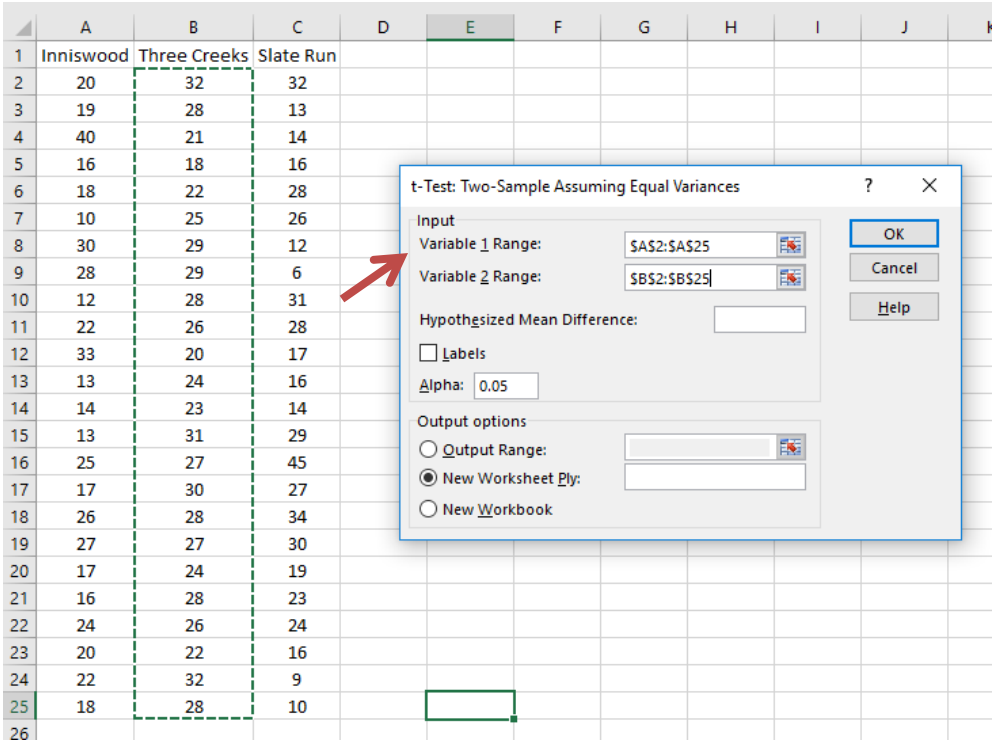
On the **Data tab**, click **Data Analysis** in the Analysis Group

Scroll to the bottom of the list and **select the appropriate test**. Click **OK.**



	Inniswood	Three Creeks	Slate Run
2	20	32	32
3	19	28	13
4	40	21	14
5	16	18	16
6	18	22	28
7	10	25	26
8	30	29	12
9	28	29	6
10	12	28	31
11	22	26	28
12	33	20	17
13	13	24	16
14	14	23	14
15	13	31	29
16	25	27	45
17	17	30	27
18	26	28	34
19	27	27	30
20	17	24	19
21	16	28	23
22	24	26	24
23	20	22	16
24	22	32	9
25	18	28	10

Variable 1 Range should be the data from first column (Inniswood) and Variable 2 Range should be the second column (Three Creeks). Click **OK.**



	A	B	C
1	Inniswood	Three Creeks	Slate Run
2	20	32	32
3	19	28	13
4	40	21	14
5	16	18	16
6	18	22	28
7	10	25	26
8	30	29	12
9	28	29	6
10	12	28	31
11	22	26	28
12	33	20	17
13	13	24	16
14	14	23	14
15	13	31	29
16	25	27	45
17	17	30	27
18	26	28	34
19	27	27	30
20	17	24	19
21	16	28	23
22	24	26	24
23	20	22	16
24	22	32	9
25	18	28	10

Another new worksheet appears with the results of the test. If you were using the Tutorial_Blackbirds.xlsx data, yours should look like this:

	A	B	C	D	E
1	t-Test: Two-Sample Assuming Equal Variances				
2					
3		*Variable 1*	*Variable 2*		
4	Mean	20.83333	26.16667		
5	Variance	52.49275	14.23188		
6	Observati	24	24		
7	Pooled Va	33.36232			
8	Hypothesi	0			
9	df	46			
10	t Stat	-3.19861			
11	P(T<=t) on	0.00125			
12	t Critical o	1.67866			
13	P(T<=t) tw	0.002501			
14	t Critical t	2.012896			

Look at the t Stat… can we reject the null hypothesis, that the mean number of red winged blackbirds at these two parks are the same? In this case, the t-statistic 3.199 > 1.96, so the hypothesis that the two means are equal is rejected.

Repeat this test for Inniswood and Slate Run.

Check that the table in your t-test output worksheet matches the one below:

1	t-Test: Two-Sample Assuming Equal Variances		
2			
3		*Variable 1*	*Variable 2*
4	Mean	20.83333	21.625
5	Variance	52.49275	91.375
6	Observati	24	24
7	Pooled Va	71.93388	
8	Hypothesi	0	
9	df	46	
10	t Stat	-0.32335	
11	P(T<=t) on	0.37395	
12	t Critical o	1.67866	
13	P(T<=t) tw	0.747899	
14	t Critical t	2.012896	

In this case, the t Stat is 0.32 < 1.96. The null hypothesis of equal mean blackbirds between these two parks cannot be rejected.

Since the means between Inniswood and Slate Run are equal and the means between Inniswood and Three Creeks are not, it is not necessary to test the difference in means between Slate Run and Three Creeks… we know the means will be different! [Nevertheless, if you would like the practice, go head and conduct the test. The t Stat should be 2.165.]

For more information about conducting statistical analysis in Microsoft Excel, consult Microsoft's support page, Use the Analysis ToolPak to perform complex data analysis.

https://support.office.com/en-us/article/Use-the-Analysis-ToolPak-to-perform-complex-data-analysis-6c67ccf0-f4a9-487c-8dec-bdb5a2cefab6

Here is a video about calculating descriptive statistics in Excel (including installing the Excel Add-In):
https://www.youtube.com/watch?v=5MFjwM6K5Sg

Note, this video specifies the descriptive statistics appear in the same worksheet as the raw data, rather than in a new, separate worksheet.