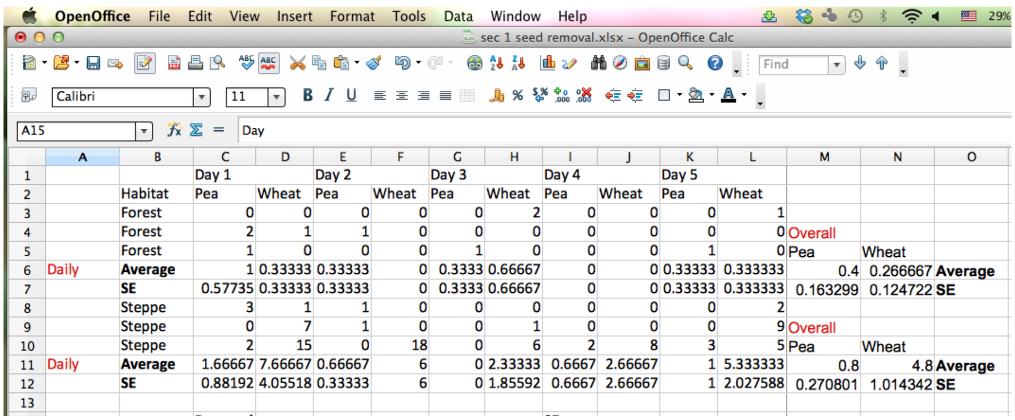
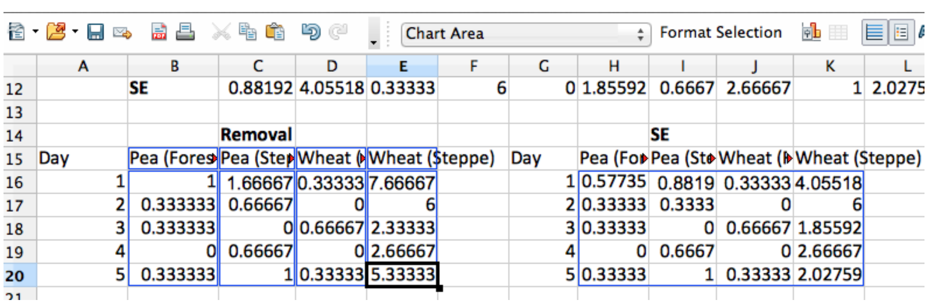
<u>Graphing Seed Removal Data in Open Office</u>

1. First, we need to think about the types of comparisons we will be making so we know what averages and standard errors (SE) to calculate. We were interested in differences in removal rates between the two seed types, so we want to calculate the average removal each day for each seed type. Additionally, we were interested in potential differences in removal among the different habitats (i.e. forest vs. steppe), so we probably want to separate the seeds removed each day by habitat as well. Below, you will see I calculated the average removal and SE for each seed type in each habitat for each day. At the end, I calculated the average removal and SE for each seed type and each habitat ACROSS the 5-day span.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1			Day 1		Day 2		Day 3		Day 4		Day 5				
2		Habitat	Pea	Wheat	Pea	Wheat	Pea	Wheat	Pea	Wheat	Pea	Wheat			
3		Forest	0	0	0	0	0	2	0	0	0	1			
4		Forest	2	1	1	0	0	0	0	0	0	0	Overall		
5		Forest	1	0	0	0	1	0	0	0	1	0	Pea	Wheat	
6	Daily	**Average**	1	0.33333	0.33333	0	0.3333	0.66667	0	0	0.33333	0.333333	0.4	0.266667	**Average**
7		**SE**	0.57735	0.33333	0.33333	0	0.3333	0.66667	0	0	0.33333	0.333333	0.163299	0.124722	**SE**
8		Steppe	3	1	1	0	0	0	0	0	0	2			
9		Steppe	0	7	1	0	0	1	0	0	0	9	Overall		
10		Steppe	2	15	0	18	0	6	2	8	3	5	Pea	Wheat	
11	Daily	**Average**	1.66667	7.66667	0.66667	6	0	2.33333	0.6667	2.66667	1	5.333333	0.8	4.8	**Average**
12		**SE**	0.88192	4.05518	0.33333	6	0	1.85592	0.6667	2.66667	1	2.027588	0.270801	1.014342	**SE**
13															

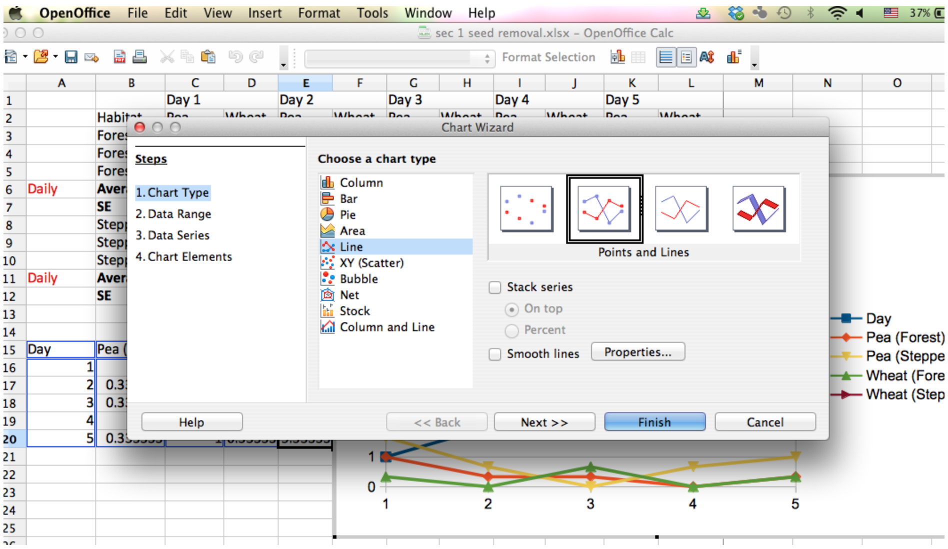
2. When it comes to graphing, having your data organized in a condensed, organized manner will make things easier in the long run. It is worth taking time to organize your data after you calculate averages and SEs. It is important to first decide what you want to show with your graph, and this will determine how you organize your data. I am interested in showing the average removal on each of the 5 days for each combination of seed type. This will allow me to see if there is more removal on a certain day and also to see if a certain type of seed was removed more frequently. Below, I made a heading for each combination of seed type and habitat and copied the average removal. I repeated the process for the SEs. These are the same numbers, just in a condensed format.



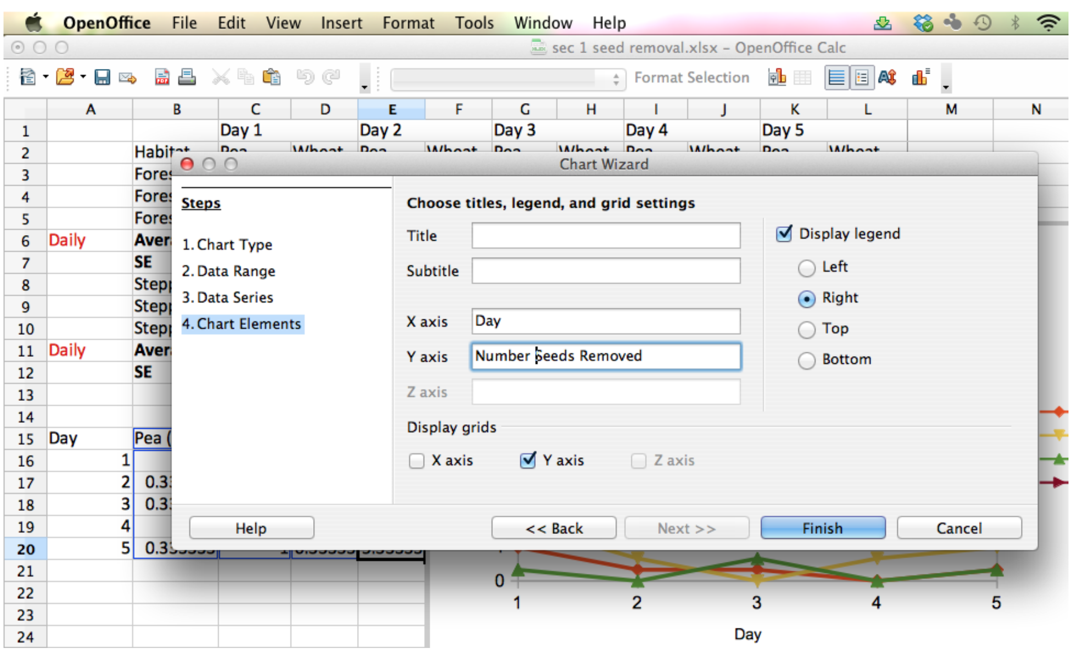
	A	B	C	D	E	F	G	H	I	J	K	L
12		**SE**	0.88192	4.05518	0.33333	6	0	1.85592	0.6667	2.66667	1	2.0275
13												
14			**Removal**					**SE**				
15	Day	Pea (Fores	Pea (Step	Wheat (	Wheat (Steppe)		Day	Pea (Fo	Pea (Ste	Wheat (	Wheat (Steppe)	
16	1	1	1.66667	0.33333	7.66667		1	0.57735	0.8819	0.33333	4.05518	
17	2	0.333333	0.66667	0	6		2	0.33333	0.3333	0	6	
18	3	0.333333	0	0.66667	2.33333		3	0.33333	0	0.66667	1.85592	
19	4	0	0.66667	0	2.66667		4	0	0.6667	0	2.66667	
20	5	0.333333	1	0.33333	5.33333		5	0.33333	1	0.33333	2.02759	

3. Now that the data are arranged, we can begin making the graph. We should highlight the data we want to display (the removal columns) and pick the type of graph we want. A line graph is probably most appropriate for this example. It is also worth stopping at 3. After selecting Line Graph, click next until you get to 3. Data Series. Here, you can make sure your graph displays the appropriate groups.

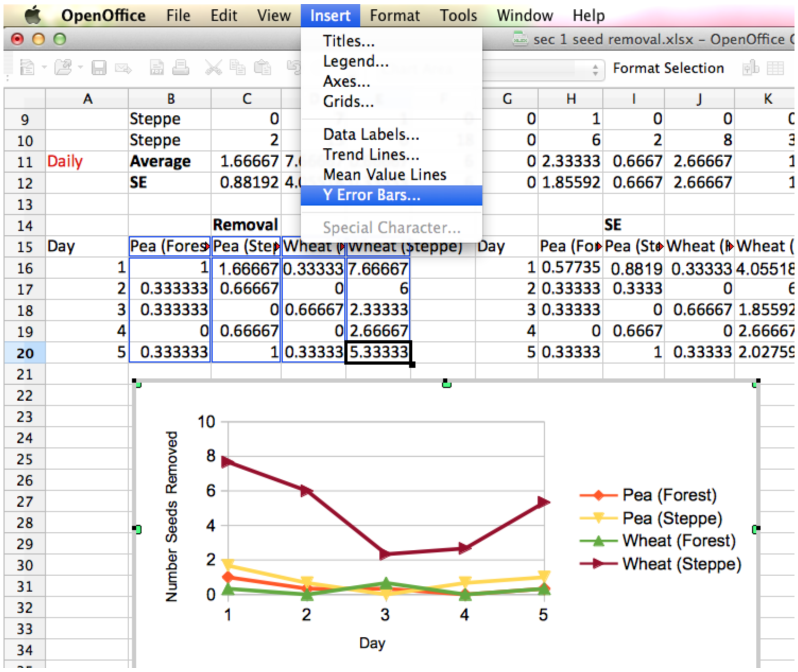
You can see below that I accidentally highlighted the Day column, so the program added in Day as one of my dependent variables. I deleted this series before continuing.



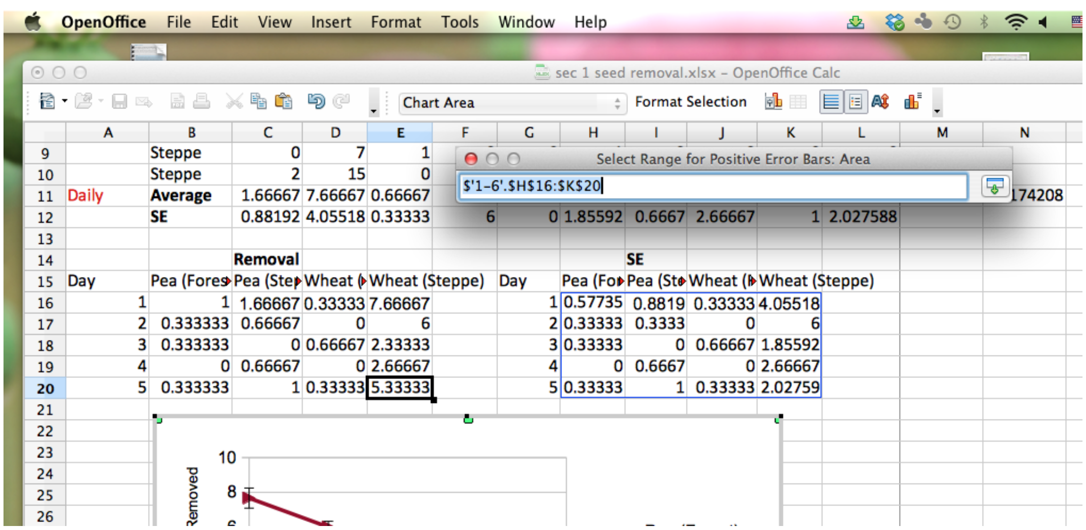
4. You can keep clicking ‘Next’ until you get to 4. Chart Elements. This is where you can add X and Y axis Labels. Once you name your axes, click Finish.



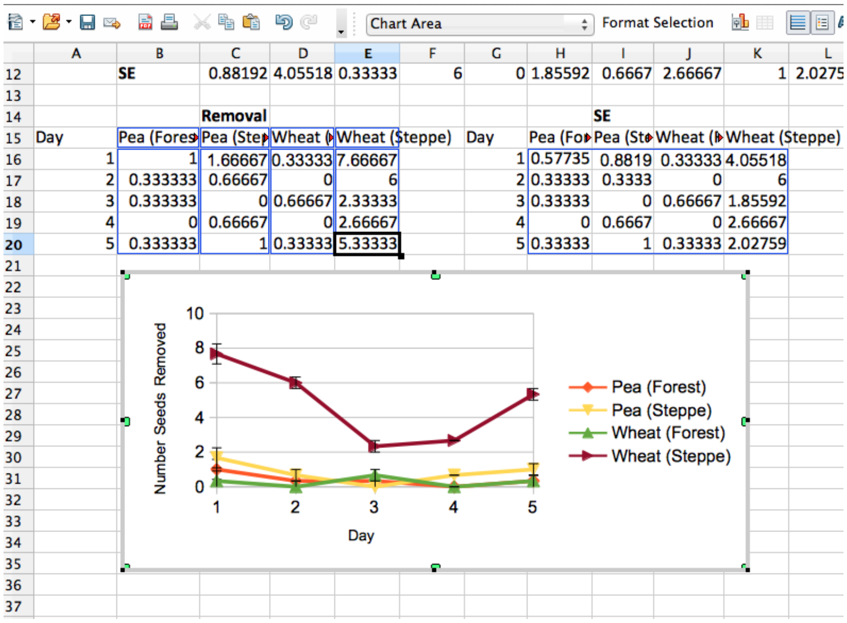
5. You now have a graph, but no error bars. Making sure your graph is selected, go to Insert on the top menu and select Y Error Bars to add your error bars.



6. Since we want our error bars to show the SE, we need to click Cell Range so we can choose the specific SEs we calculated. We do this by highlighting the SEs we calculated for each seed/habitat combination. Make sure they are arranged in the same order as the removal data so you get the correct removal and error matchup.

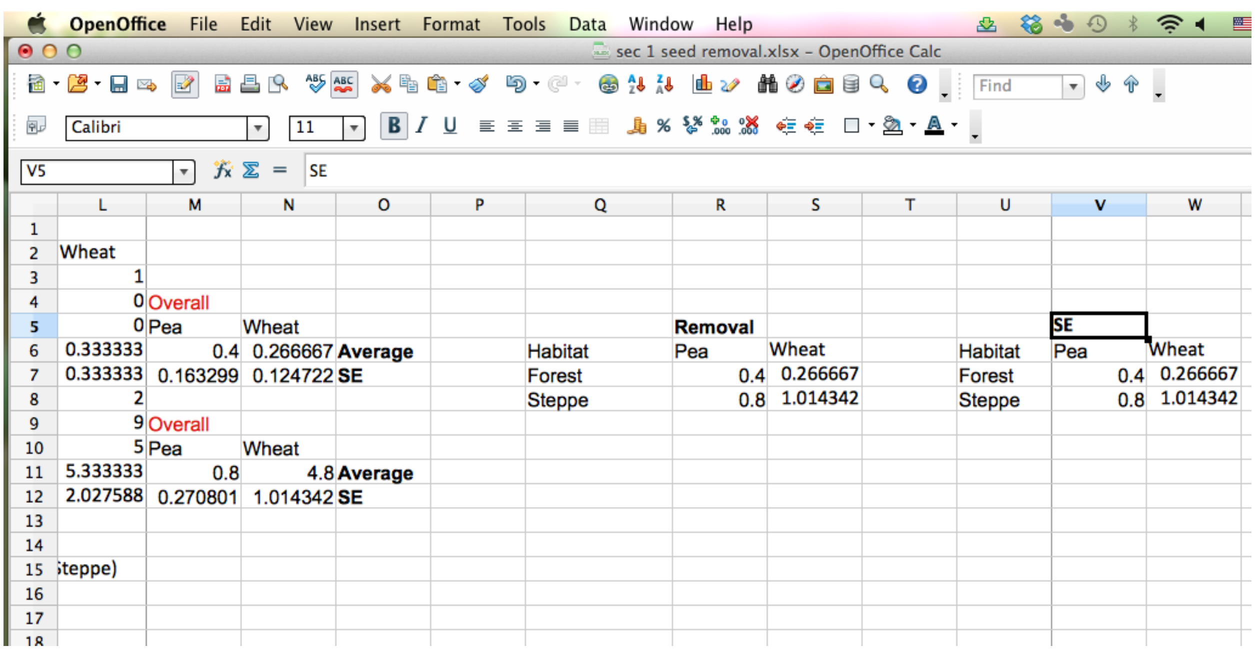


7. Click OK and voila! You have your graph with error bars.



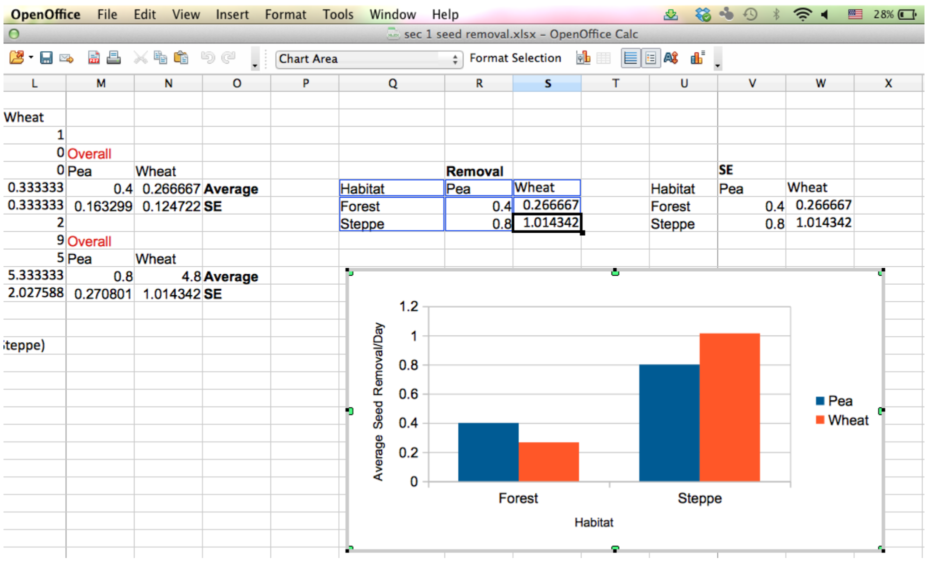
Below, I'll show how to make a bar graph that compares removal rates ACROSS all 5 days for each seed/habitat combination.

1. Again, organize your data.

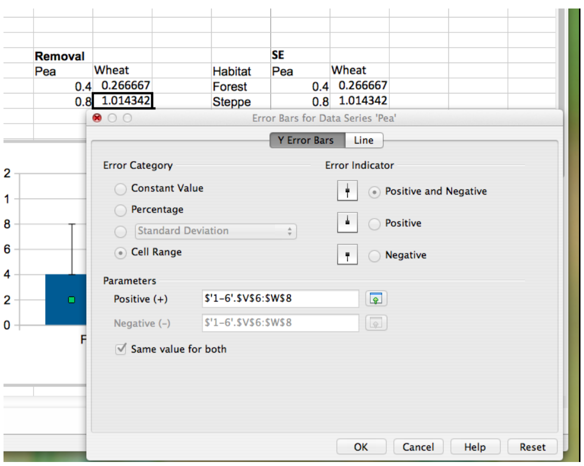


2

2. Highlight the appropriate data (and labels) and insert a new chart. A bar graph would best display these data. Follow the rest of the steps you followed for the line graph (e.g. make sure you have the correct data represented by your series and name your axes).



3. Repeat the same steps as above to insert error bars



4. Now you have another pretty graph!

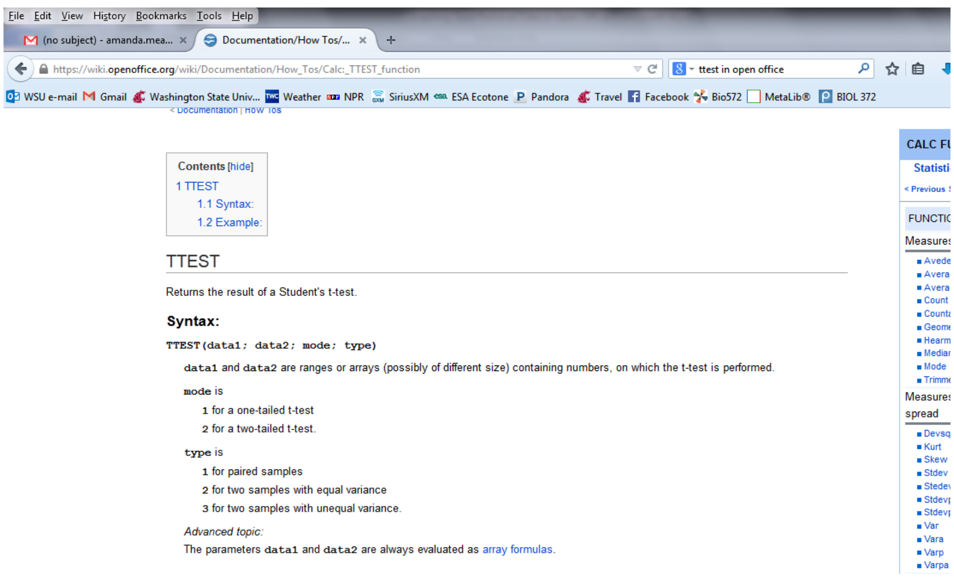
***** Please note that these are just a few examples of acceptable graphs for the paper. Your graphs may vary depending on what your specific hypotheses address. This is just meant to be a tutorial that introduces you to the tools you need to make graphs.** ********

Analyzing Seed Removal Data in Open Office

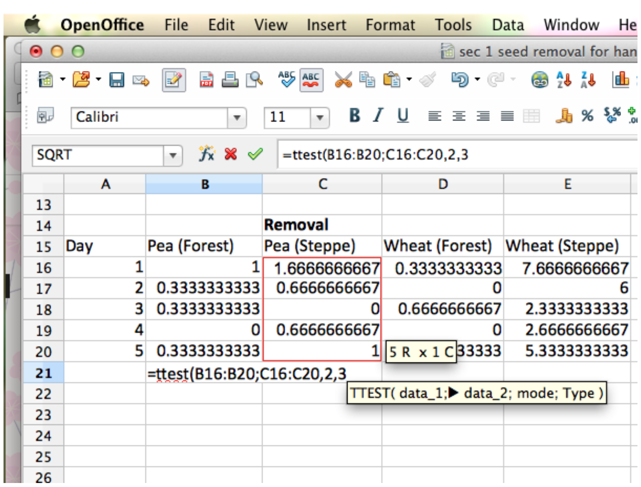
Graphing your results is a good place to start when it comes to making conclusions from the data you collected from the seed removal experiment, but we need to actually run statistics to see if differences among our treatments are statistically significant. When comparing two groups, a t-test is an easy test to use.

Say we want to see if there is a significant difference in removal of pea seeds in the forest vs. the steppe habitat:

1. The function in Open Office for a t-test is simply TTEST(data set 1, data set 2, mode, Type) . Below is a screen shot explaining the syntax.

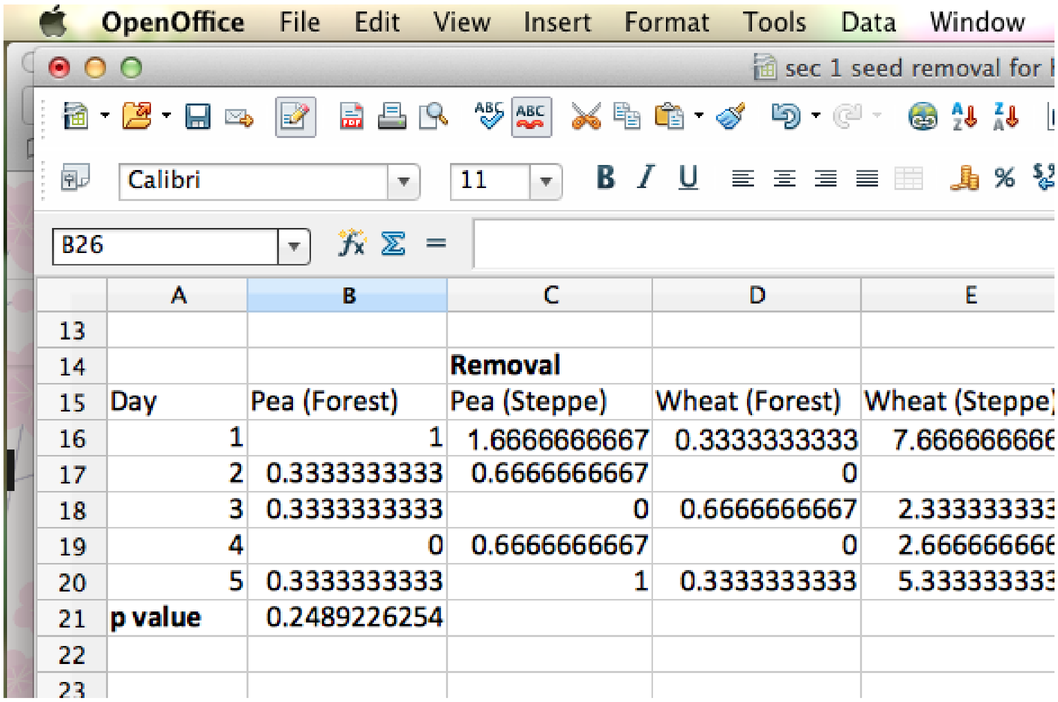


Here is what it would look like in Open Office:



Using mode=2 means we are conducting a two-tailed test, so we are testing if removal from the forest is higher OR lower than in the steppe (a one-tailed test would only test for one direction). Using type=3 means we are allowing our two samples (forest vs. steppe) to have unequal variances.

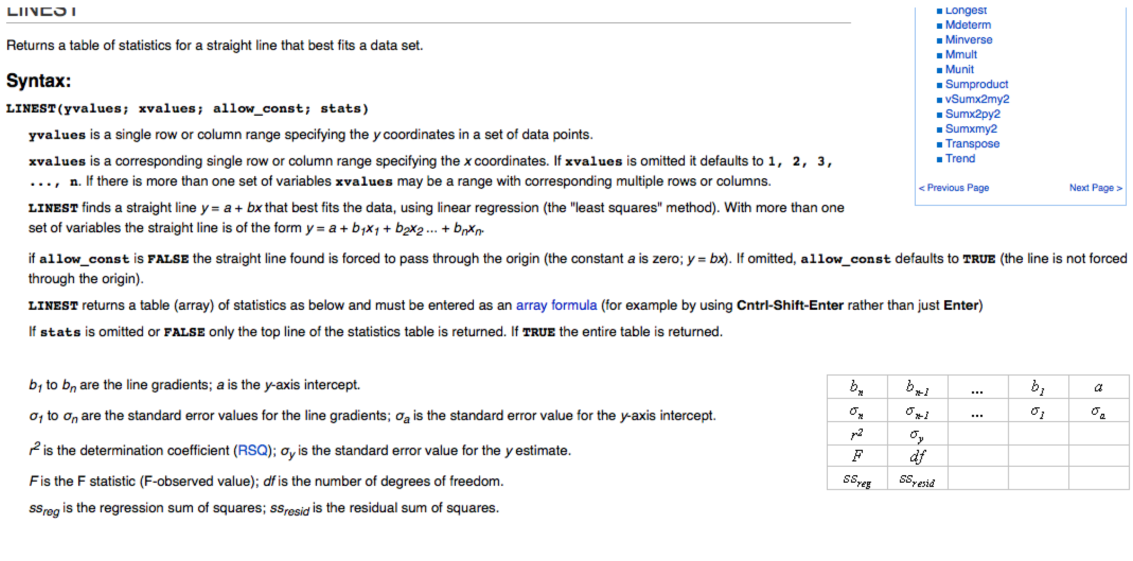
2. Hit enter and Open Office will return a P-value for the t-test. Remember when $p<.05$, you have a significant difference among your treatments. In this example, there is no significant difference.



	A	B	C	D	E
13					
14			**Removal**		
15	Day	Pea (Forest)	Pea (Steppe)	Wheat (Forest)	Wheat (Steppe
16	1	1	1.6666666667	0.3333333333	7.66666666
17	2	0.3333333333	0.6666666667	0	
18	3	0.3333333333	0	0.6666666667	2.33333333
19	4	0	0.6666666667	0	2.66666666
20	5	0.3333333333	1	0.3333333333	5.33333333
21	**p value**	0.2489226254			
22					
23					

What if you want to see if there is a relationship between distance from the center of the forest (i.e. the beginning of our transect) and seed removal?

1. A linear regression would be an appropriate test for this question. You would use the function LINEST for this test. The syntax can be found in the screenshot below. LINEST is an array function, meaning it gives you a table of output, not just a p-value (for example). The table is also in the syntax below. Note that you do not get a p-value in the output, but you do get enough information to calculate one on your own.



LINEST

Returns a table of statistics for a straight line that best fits a data set.

Syntax:

LINEST(yvalues; xvalues; allow_const; stats)

yvalues is a single row or column range specifying the *y* coordinates in a set of data points.

xvalues is a corresponding single row or column range specifying the *x* coordinates. If **xvalues** is omitted it defaults to **1, 2, 3, ..., n**. If there is more than one set of variables **xvalues** may be a range with corresponding multiple rows or columns.

LINEST finds a straight line $y = a + bx$ that best fits the data, using linear regression (the "least squares" method). With more than one set of variables the straight line is of the form $y = a + b_1x_1 + b_2x_2 ... + b_nx_n$.

if **allow_const** is **FALSE** the straight line found is forced to pass through the origin (the constant *a* is zero; $y = bx$). If omitted, **allow_const** defaults to **TRUE** (the line is not forced through the origin).

LINEST returns a table (array) of statistics as below and must be entered as an array formula (for example by using **Cntrl-Shift-Enter** rather than just **Enter**)

If **stats** is omitted or **FALSE** only the top line of the statistics table is returned. If **TRUE** the entire table is returned.

b_1 to b_n are the line gradients; *a* is the *y*-axis intercept.

σ_1 to σ_n are the standard error values for the line gradients; σ_a is the standard error value for the *y*-axis intercept.

r^2 is the determination coefficient (RSQ); σ_y is the standard error value for the *y* estimate.

F is the F statistic (F-observed value); *df* is the number of degrees of freedom.

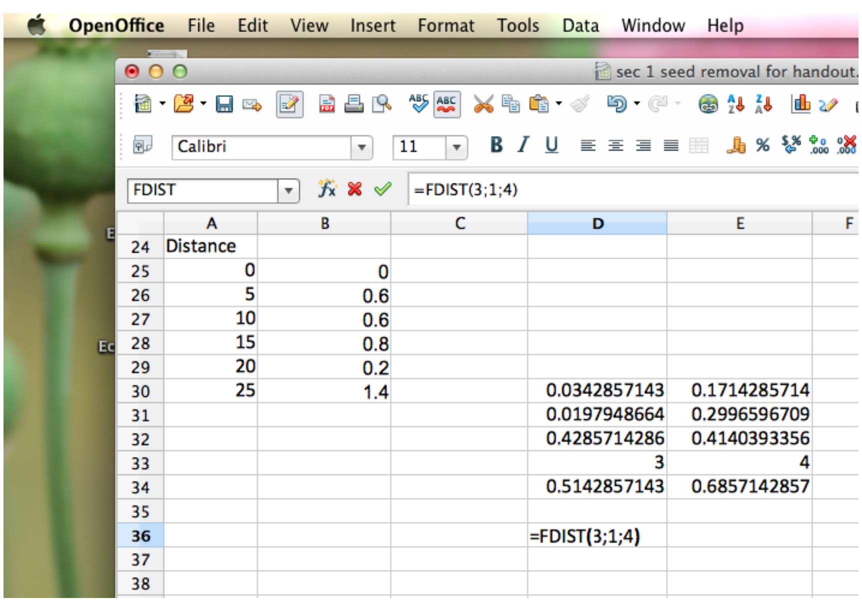
ss_{reg} is the regression sum of squares; ss_{resid} is the residual sum of squares.

b_n	b_{n-1}	...	b_1	a
σ_n	σ_{n-1}	...	σ_1	σ_a
r^2	σ_y			
F	df			
ss_{reg}	ss_{resid}			

- Longest
- Mdeterm
- Minverse
- Mmult
- Munit
- Sumproduct
- vSumx2my2
- Sumx2py2
- Sumxmy2
- Transpose
- Trend

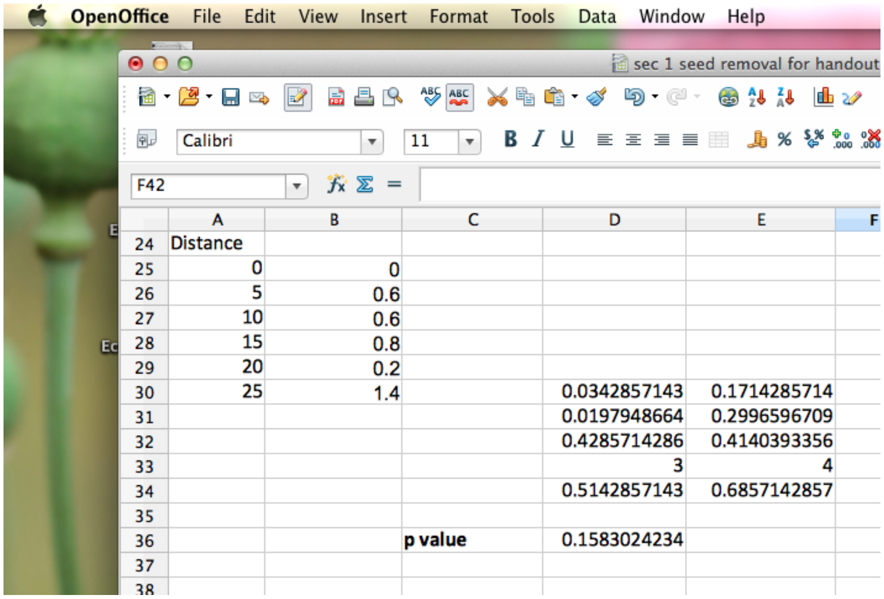
< Previous Page Next Page >

2. To calculate a p-value, use the FDIST function. The syntax is FDIST(F value, df1, df2). Below is a screenshot of the output from the LINEST function and the FDIST syntax. The table gives the F statistic (3) and the denominator degrees of freedom (4). In linear regression the numerator degrees of freedom is the (number of terms in the model)-1, in our case that is 2-1. The two terms are the intercept and slope. So, FDIST(3;1;4) is the appropriate syntax here.



	A	B	C	D	E	F
24	Distance					
25	0	0				
26	5	0.6				
27	10	0.6				
28	15	0.8				
29	20	0.2				
30	25	1.4		0.0342857143	0.1714285714	
31				0.0197948664	0.2996596709	
32				0.4285714286	0.4140393356	
33				3	4	
34				0.5142857143	0.6857142857	
35						
36				=FDIST(3;1;4)		
37						
38						

3. Hit enter and you will have your p-value



	A	B	C	D	E	F
24	Distance					
25	0	0				
26	5	0.6				
27	10	0.6				
28	15	0.8				
29	20	0.2				
30	25	1.4		0.0342857143	0.1714285714	
31				0.0197948664	0.2996596709	
32				0.4285714286	0.4140393356	
33				3	4	
34				0.5142857143	0.6857142857	
35						
36			**p value**	0.1583024234		
37						
38						