

**DATA, DISTRIBUTIONS AND HYPOTHESES:  
EXPLORING DIVERSITY AND DISTURBANCE IN THE TALLGRASS PRAIRIE**

**BIOL 585, Fall 2011**

N. Emery, H. Dalgleish and S. Stevens

**Week 3: Data Analysis and Interpretation**

**LAB OVERVIEW**

Last week you collected vegetation data from two different prairie restoration parcels at Prophetstown State Park that are burned in different seasons: the western parcel receives spring burn treatments while the eastern parcel is burned in the fall. Both parcels are relatively new restorations and received their first burn within the last 12 months. Recall that there are four questions that were provided to you, and one that you generated on your own, that we aim to address with this data set:

- i. How does the relative abundance of different plant functional groups differ between the spring- and fall-burned parcels at Prophetstown?
- ii. How do plant richness and diversity differ between the spring- and fall-burned parcels at Prophetstown?
- iii. How do the relative abundances of warm-season and cool-season grasses at Prophetstown compare to the patterns at Konza reported in Towne & Kemp (2003)?
- iv. How do plant richness and diversity at Prophetstown compare to the patterns at Konza?
- v. Your own question!

**In your writing assignment for Week 1 you generated at least one hypothesis and prediction for each of these questions. The writing assignment for Week 2 asked you to state the comparisons that are necessary to test your predictions.** Today, you will work in pairs to complete some simple graphs and statistical tests that test your hypotheses and predictions.

**Objectives of this lab:**

1. Summarize your data in graphs and figures that reflect the comparisons invoked by your hypotheses and predictions.
2. Conduct simple statistical analyses on population and community vegetation data to test explicit hypotheses and predictions.
3. Interpret the statistical results and infer their biological implications.

**GETTING STARTED: ORGANIZING & PREPARING THE DATA**

Today you will be (1) graphing the Prophetstown data that you collected last week, (2) conducting some simple statistical tests to evaluate a subset of your hypotheses and predictions, (3) interpreting your results and their biological significance. All of your graphs and analyses will be performed in Microsoft Excel. Assemble your answers, figures, and statistical output into a Microsoft Word document and provide an electronic copy of this to your TA at the end of the lab session.

### **Data Prep Step 1: Boot up your computer in Windows**

Hold down the **Option** key, push the **power** button, and *continue to hold the Option key until the icons for the Mac disk and the Windows disk appear*. Double-click the **Windows** icon. (We have to use the Windows OS because the statistical package add-in in Excel is not available for the Mac OS.)

### **Data Prep Step 2: Activate the Analysis ToolPak in Excel**

Open Excel. Under the File menu, select “Options” from the bottom left-hand corner of the window. Highlight “Add-Ins” from the left side of the next window and then click the “Go” button. A small window will open. Check the box next to “Analysis ToolPak” and then click “OK.” You will return to the main spreadsheet window in Excel. Confirm that you have successfully activated the Analysis ToolPak by selecting the “Data” menu. In the toolbar, you should see a group (probably on the right-hand side) called “Analysis” with an icon called “Data Analysis.” If this is visible, you have successfully activated the ToolPak you will need today.

### **Data Prep Step 3: Download the class data sheet from Blackboard**

The class data sheet is located in the Week 3 subfolder of the Prairie Lab Module folder. You may also want to download the Towne & Kemp (2003) paper in the Week 1 folder if you did not bring a printed copy to lab, because you will be making comparisons to some of the data reported in their paper.

### **Data Prep Step 4: Familiarize yourself with the data set**

Take a few moments to look around the spreadsheet to understand how it is organized. Your TA’s have collected data from each team and organized them into this single spreadsheet. Unknown plants have been identified when possible, and the different functional groups have been grouped into adjacent columns and color-coded.

### **Data Prep Step 5: Change the cover class data to midpoint values**

We will follow the same procedure as Towne & Kemp (2003) and convert our cover class data (scored in categories from 1 through 7) into percent cover estimates by replacing each cover class category with its respective midpoint percent cover value. This can be easily done in Excel using the “Find and Replace” option:

- a. First, highlight the data set (only the cells with actual data in them, not the lab section, team, parcel, transect, or plot columns, and not the row headers).
- b. Select the “Home” tab and then the “Find and Select” icon on the right-hand side (this has a pair of binoculars as the icon). Choose “Replace” from the drop-down menu. Select the “Options” button and check the box next to “Match entire cell contents.” Then fill in the “Find what:” and “Replace with” fields according to the table below, clicking the “Replace All” button:

<b>Cover Class Category (enter into "Find what" field)</b>	<b>% Cover range</b>	<b>Midpoint (enter into "Replace with" field)</b>
1	0-1%	0.005
2	1%-5%	0.025
3	5%-25%	0.15

4	25%-50%	0.375
5	50%-75%	0.625
6	75%-95%	0.85
7	95%-100%	0.975

**Data Prep Step 6: Calculate the cumulative cover of each functional group in each plot**

On the far right side of your spreadsheet, there are 5 empty columns, followed by a column with the diversity index for each plot (calculated using Shannon’s Diversity Index,  $H'$ , which is a bit tricky to calculate in Excel so it has been done for you). The first four columns are labeled  $C_4$  grass total,  $C_3$  grass total, Forb total, and Woody total. We want these columns to represent the total cover of each functional group in each plot. You can calculate this total in Excel using a simple formula that automatically calculates the sum of all plant species that fall into each functional category in each row. The steps for calculating the total cover of  $C_4$  grasses in each plot are below (apply the same procedure for the other functional groups):

- a. Select the top empty cell in the “ $C_4$  grass total” column (that is, the one immediately below the header.)
- b. In the formula bar at the top of the spreadsheet (the empty field next to  $f_x$ ), type the following:

=SUM(

- c. Use the cursor to highlight the cells containing the values that you want to add together; in this case, you will select the cells in that row with the cover values for all  $C_4$  grasses (color-coded in light red).
- d. Add a final parenthesis at the end of your formula in the formula bar. Your complete formula should look like this:

=SUM(F3:J3)

- e. Push return on your keyboard. There should now be a value in the cell that is the sum of all the  $C_4$  grasses in the first plot.
- f. Select the cell with the original formula, and that now displays the sum of the  $C_4$  grasses in the first plot. A dark black box will appear around the cell. Carefully select the bottom right-hand corner of this box and drag it down the column to the last row of data. Excel will automatically extend this formula to calculate the total cover of  $C_4$  grasses for each row (that is, each plot).

**Data Prep Step 7: Calculate the plant species richness in each plot**

The next empty column on the right side of your data set is for “Richness.” Remember that species richness is the number of species in a plot and does not take the relative abundance of each species into account. To calculate species richness for each plot, we want to write a formula that simply counts the number of cells that contain a value greater than zero. Excel has another nifty formula that makes this easy:

- a. Select the top empty cell in the “Richness” column (that is, the one immediately below the header.)
- b. In the formula bar at the top of the spreadsheet, type the following:

=COUNTIFS(

- c. Use the cursor to highlight all of the values for the species in that plot (including the zeroes!).
- d. Your formula should now read  

$$=COUNTIFS(F3:AU3$$
- e. Now you need to indicate that within the range you just selected, you only want to count the cells that have non-zero values:  

$$=COUNTIFS(F3:AU3, ">0")$$
- f. Extend this formula to all values in the column using the same procedure described above in Step 5e.

### SIMPLE STATISTICAL ANALYSES OF PROPHETSTOWN VEGETATION DATA

You are now (finally!) ready to do some simple analyses to test the hypotheses and predictions that you generated for the questions listed above.

#### Data Analysis Step 1: Statement of hypotheses and predictions

Work with your partner to identify one hypothesis and one prediction for each question. Of the questions that you generated yourselves, select one and establish the hypothesis and prediction you want to test for it as well. After reaching a consensus, write down the question, one hypothesis, *the corresponding null hypothesis*, and one prediction for all five questions.

#### Data Analysis Step 2: Identify the relevant comparisons

Write down the comparison that you need to make in order to test your hypotheses. *These comparisons should follow directly from your predictions.* Do not worry about the statistical comparison here – just identify the groups that you want to compare. **NOTE: for the purposes of this lab, keep it simple! Limit your comparisons to simple contrasts that can be tested with t-tests or chi-square tests (described below); for example, you might want to compare C<sub>3</sub> grasses in the spring burned parcel to the C<sub>3</sub> grasses in the fall-burned parcel, OR C<sub>3</sub> grasses to C<sub>4</sub> grasses in the same parcel. Check with your TA if you are concerned that your comparisons are too complicated for the goals of this lab.**

**\*\*\*In the interest of time, you will not be conducting statistical analyses for every one of your 5 questions. Instead, pick a minimum of 2 questions that are most interesting to you, and proceed through the following steps for each of them. If you finish early, go ahead and work through the graphs and statistics for additional questions.\*\*\***

#### Data Analysis Step 3: Graph the data

Create a simple graph that performs the comparisons you want to make with your data.

A few tips if you are not familiar with creating figures in Excel:

- a. You may need to calculate averages to make these graphs. This can be done using the following formula:

$$=AVERAGE()$$

Insert the cells that you want to average across inserted in between the parentheses (you can just highlight the relevant cells, as described above). Similarly, you can calculate the standard deviation of a sample using:

$$=STDEV()$$

- and (again) entering the cells containing the relevant data in the parentheses.
- Depending on the calculation that you want to implement, it may help to SORT the data first. You can do this by highlighting *everything in the data set below row 2* (that is, everything except the headers), selecting the “Data” tab from the menu bar, and then choosing the “Sort” icon. Use the drop-down menu to select the best way to sort your data for the calculation or comparison you want to make.
  - Create a graph using the options in the “Charts” group under the “Insert” tab. You have had experience doing this in previous labs, but feel free to try other types of figures if you think they would be more appropriate for your comparisons.

**Data Analysis Step 4: Conduct a simple statistical test**

For the purposes of this lab, you should feel comfortable restricting your statistical analyses to t-tests and chi-squared tests. This will limit the sophistication of your analyses and will require us to ignore some of the structure in our data set (like the fact that we used transects). However, you can still adequately address the questions in the lab with simple tests.

Tips for conducting a t-test in Excel:

t-tests are appropriate for comparing two different sample populations (e.g., C<sub>3</sub> grasses in the spring-burned parcel vs. C<sub>3</sub> grass abundance in the fall-burned parcel). To conduct a t-test in Excel:

- Under the “Data” tab, select the “Data Analysis” icon (this is probably on the far right side of the menu bar)
- Scroll down and select “t-Test: Two-sample Assuming Unequal Variances” and click the “OK” button.
- In the next window, input the range of cells with the first sample population in the field for “Variable 1 Range:” (you can click on the small spreadsheet icon next to this field to go to the data set and highlight the samples). Repeat this procedure to enter the second sample population into the field for “Variable 2 Range:”. Make sure the alpha value is set to 0.05 and click “OK”. A statistical table with the result of your test will be displayed.

Tips for conducting a chi-square test in Excel:

Chi-square tests allow us to infer whether a sample population conforms to a specified or “expected” distribution. For example, in this lab you may choose to use a chi-square test to compare the average abundances of C<sub>3</sub> and C<sub>4</sub> grasses at Prophetstown to the average abundances of the C<sub>3</sub> and C<sub>4</sub> grasses at Konza Prairie (as reported in Towne & Kemp 2003). In this case, you could define the Konza Prairie data set as your “expected” abundances and use a chi-square test to test if the Prophetstown abundances differ from that expectation. To conduct a chi-square test in Excel:

- Create a 2 x 2 table for the comparisons you are interested in testing. For example, your table would have the following structure to test the prediction that the relative abundance of C<sub>3</sub> and C<sub>4</sub> grasses at Prophetstown differs from the average abundance of the C<sub>3</sub> and C<sub>4</sub> grasses at Konza Prairie:

	C3 grasses	C4 grasses
Konza		
Prophetstown		

Each cell in the table would have the average abundance of each grass type at each site.

- b. Select an empty cell in your spreadsheet and type the following formula:  
=CHISQ.TEST(  
Then highlight the “actual range”, or data that our class collected, from your table (here, this would be the two cells in the row for Prophetstown), followed by a comma, and then highlight the “expected range” (here, the two cells in the row for Konza). Then add the final parenthesis.
- c. Push return. The cell will now display a p value that reflects the probability that the Prophetstown data comes from the same distribution as the Konza data.

### **Data Analysis Step 5: Interpret your results**

What do these statistical results mean? Translate these results into biological terms, and relate them to the graphs you developed. Write 1-2 sentences summarizing the statistical output and the biological implications of these results. Then, state if these results support or refute your hypothesis. If you reject your hypothesis, state an alternative hypothesis that is consistent with your results.

### **REFERENCES CITED**

<p><b>Towne, G.E. and K.E. Kemp. 2003. Vegetation dynamics from annually burning tallgrass prairie in different seasons. <i>Journal of Range Management</i> 56:185-192.</b></p>
---