Exploration 6.2: Does presence of a dominant species affect native prairie plant diversity in a restored prairie?

Comparing Two Means: Simulation-Based Approach

LEARNING GOALS

- State the null and the alternative hypotheses for an explanatory variable with two categories.
- Implement the 3S strategy to compare two means: Find a **statistic**, **simulate**, and compute the **strength** of evidence against observed study results happening by chance alone.
- Use the **Multiple Means** applet to conduct a simulation of the null hypothesis.
- Find and interpret the standardized statistic and the p-value for a test of two means.
- State a complete conclusion about the alternative hypothesis (and null hypothesis) based on the p-value and/or standardized statistic and the study design, including statistical significance, estimation, generalizability, and causation.

STEP 1: Ask a research question. Dominant C4 grasses like big bluestem are common components of prairie restoration seed mixes. Evidence about the effect of big bluestem on prairie plant diversity is mixed: some studies demonstrate that it effectively excludes invasive competitors and enhances native diversity whereas others find that its competition for light also shades out and depresses diversity of native species. Researchers at Eastern Michigan University investigated whether including big bluestem in prairie restoration seed mixes increased or decreased Simpson's diversity (based on percent cover) of native prairie plant species in a midwestern old field prairie restoration.

STEP 2: Design a study and collect data. The researchers created identical native prairie plant seed mixes with and without big bluestem. They randomly assigned either the big bluestem **present** or the big bluestem **absent** seed mix to each of 36 individual 2x2 meter plots in a former corn and soybean agricultural field that was undergoing prairie restoration. (See Fig. 1.)

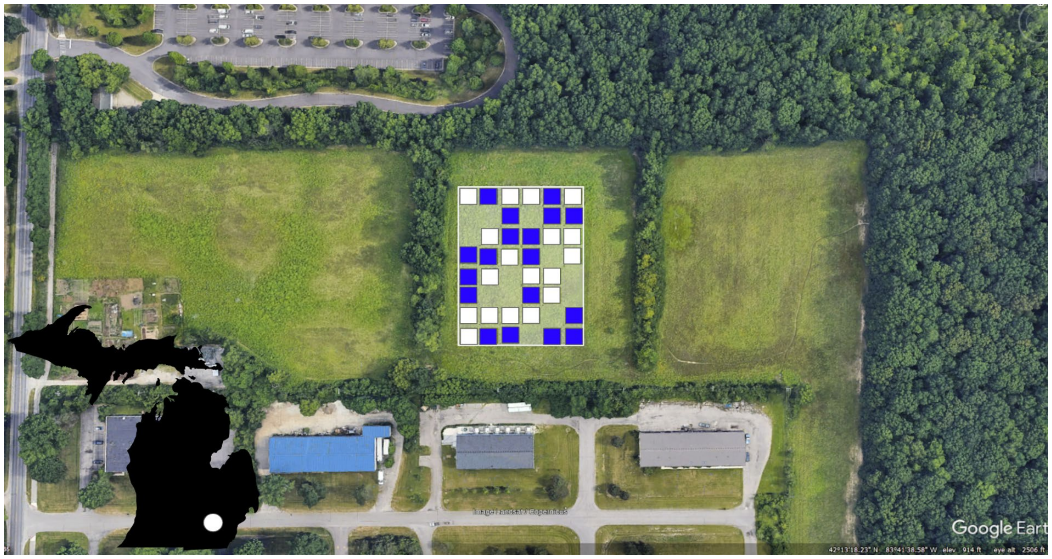

Figure 1. Prairie restoration study design. Blue plots received native prairie plant seed mixes containing big bluestem. White boxes received identical seed mixes without big bluestem.

After four seasons of growth the researchers measured the percent cover for the native species in each plot (See Fig. 2.) and calculated Simpson's diversity index from those measurements using the formula shown below. For each species p_i is equal to the percent cover for that species divided by the sum of species cover.

$$D = 1 - \sum_{i=1}^{N} (p_i)^2$$

While we are not calculating this index for ourselves in this exploration, you can experiment with different numbers of species and percent cover values and see how the index will increase to reflect both richness (how many native species were present) and evenness (how evenly percent cover was distributed among the native species present). Essentially, higher values of Simpson's diversity index (D) reflect higher native prairie plant diversity.

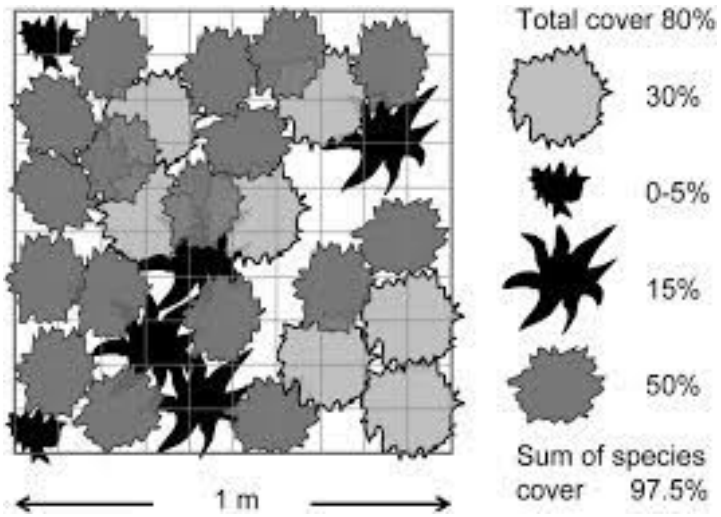

Figure 2. Example of how percent cover was estimated for the native species in this study. Note that cover is estimated for each plant and includes places where plants overlap (so the sum of species cover can exceed total cover).

1. What are the observational (or experimental) units in this study and how many are there?

2. Identify the explanatory variable in this study and classify it as categorical or quantitative.

3. Identify the response variable in this study and classify it as categorical or quantitative.

4. Was this a manipulative experiment or an observational study? Explain how you are deciding.

5. State the null and alternative hypotheses to investigate whether the presence of big bluestem affected native prairie plant diversity in this restoration.

STEP 3: Explore the data.

6. Open the **Multiple Means** applet and click "clear". Enter the Prairie Plant Diversity data in the data field.

a. Notice that the applet creates parallel dotplots, one for each study group. You should also check the **Add boxplots** button to add boxplots to these graphs. Based on these graphs alone, which group (**absent** or **present**) appears to have had the higher mean native species diversity? How are you deciding?

b. Notice also that the applet also computes numerical summaries of the data, such as the mean and standard deviation (SD) for the diversity scores in each group.

i. For the big bluestem **absent** group, record the sample size (n), mean, and SD.

n_{absent} = $\bar{x}_{absent}$ = SD_{absent} =

ii. For the big bluestem **present** group, record the sample size (n), mean, and SD.

$n_{present}$ = $\bar{x}_{present}$ = $SD_{present}$ =

The standard deviation is a measure of variability. Relatively speaking, smaller standard deviation values indicate less variability and a distribution whose data values tend to cluster more closely together, compared to a distribution with a larger standard deviation.

c. Based on the numerical summaries, which group (present or absent) had the higher mean diversity score? Was your observation in (a) correct? Which group had the higher variability in diversity?

STEP 4: Draw inferences beyond the data.

7. What is a possible explanation for why we observed the two groups to have different sample mean diversity indexes?

8. Describe how you might go about deciding whether the observed difference between the two sample means is statistically significant.

The key question is how often random assignment alone would produce a difference in the groups at least as extreme as the difference observed in this study if there really was no genuine effect of big bluestem on plant diversity. We can use *simulation* to investigate how often such an extreme difference would occur by chance (random assignment) alone (if the null hypothesis of no difference/no effect/no association were true). In other words, we will employ the 3S strategy.

1. Statistic

9. A natural statistic for measuring how different the observed group means are from each other is the difference in the mean diversity between the two groups. Report the value of this statistic, as you did above.

2. Simulate*:* You can start by using index cards to perform a tactile simulation of randomly assigning the 36 plots (and their diversity scores) between the two groups, *assuming* that presence of big bluestem has no impact on Simpson's diversity. Because the null hypothesis asserts that plant diversity is not

associated with presence of big bluestem, we will assume that the 36 plots would have had exactly the same diversity as they did, *regardless* of which big bluestem group (present or absent) the plot had been assigned.

10. To conduct this simulation:

    a. How many cards do you need?

    b. What will you write on each card?

    c. To conduct *one repetition* of this simulation, shuffle the stack of 36 cards well and then randomly distribute cards into two stacks of 18: one for the big bluestem present group and one for the big bluestem absent group.

        i. Calculate and report the sample means for each rerandomized group.

        ii. Calculate the difference in group means: Big bluestem present mean minus big bluestem absent mean. Report this value.

    d. Combine this result with your classmates' to create a dotplot that shows the distribution of several possible values of the difference in sample means that could have happened due to pure chance if the big bluestem treatment has no impact on native plant diversity. Sketch the dotplot, being sure to label and scale the horizontal axis.

    e. At about what value is the dotplot centered? Explain why this makes sense. (*Hint*: What are we assuming to be true when we conduct the simulation?)

    f. Where is the observed difference in means from the original study (as reported in #9) on the dotplot? Did this value happen often, somewhat rarely, or very rarely? How are you deciding?

11. You would now like to conduct many, many more repetitions to determine what is typical and what is not typical for the difference in group means, assuming that big bluestem has no impact on plant diversity. We think you would prefer to use a computer applet to do this rather than continue to shuffle cards for a very long time, calculating the difference of group means by hand. Go back to the **Multiple Means** applet, check the **Show Shuffle Options** box, select the **Plot** display, and press **Shuffle Responses**.

    a. Describe what the applet is doing and how this relates to your null hypothesis.

    b. Record the shuffled difference in sample means for the rerandomized groups, as given in the applet output. Is this difference more extreme than the observed difference from the actual study? How are you deciding?

    c. Click on **Shuffle Responses** again and record the simulated difference in sample means for the rerandomized groups. Did it change from #11b?

d. Click on **Re-Randomize** again and record the simulated difference in sample means for the rerandomized groups. Did it change from #11b and #11c?

Now to see many more possible values of the difference in sample means, assuming voice condition has no impact on surprise score, do the following in the **Multiple Means** applet:

- Change **Number of Shuffles** from 1 to 997.
- Press **Shuffle Responses** to produce a total of 1,000 shuffles and rerandomized statistics.

e. Consider the histogram of the 1,000 could-have-been values of difference in sample means, assuming that big bluestem has no effect on plant diversity.

i. What does one dot on the dotplot represent? (*Hint*: Think about what you would have to do to put another dot on the graph.)

ii. Describe the overall shape of the null distribution displayed in this dotplot.

iii. Where does the observed difference in sample means (as reported in #9) fall in this dotplot: near the middle or out in a tail? Are there a lot of dots that are even more extreme than the observed difference, assuming the presence or absence of big bluestem has no impact on plant diversity? How are you deciding?

f. To estimate a p-value, continue with the **Multiple Means** applet. Type in the observed difference in group means (as reported in #9), choose the appropriate alternative hypothesis in the **Count Samples** box, and press **Count**. What is your approximate p-value?

g. Complete the following sentence to provide the interpretation of the p-value.

The p-value of _____ is the probability of observing _____assuming _____.

3. Strength of evidence

12. Based on the p-value, evaluate the strength of evidence provided by the data against the null hypothesis that the big bluestem has no effect on plant diversity: not much evidence, moderate evidence, strong evidence, or very strong evidence?

13. Use the 2SD method to approximate a 95% confidence interval for the difference in long-run mean plant diversity score for plots with and without big bluestem. (*Hints*: Remember the observed value of the difference in group means and obtain an estimate of the SD of the difference in group means from the applet's simulation results. The interval should be *observed difference in means* ± *2SD*.)

STEP 5: Formulate conclusions.

14. ***Significance:*** Summarize your conclusion with regard to strength of evidence in the context of this study.

15. ***Estimation:*** Fill in the following interpretation of what this confidence interval reveals, paying particular attention to whether the interval is entirely positive, entirely negative, or contains zero. (*Hint*: Include the appropriate numbers and then choose the appropriate "direction" (higher or lower) in your interpretation.)

    I'm 95% confident that the long-run mean plant diversity score with the _____ treatment is ____ (higher/lower) to _____(higher/lower) than the long-run mean plant diversity score with the _____ treatment.

16. ***Generalization***
    Were the plots in this study randomly selected from a larger "population" of 2x2 meter plots? Describe the population to which you would feel comfortable generalizing the results of this study.

17. ***Causation***
    Were the plots in the study randomly assigned to a big bluestem treatment? How does this affect the scope of conclusion that you can draw?

STEP 6: Look back and ahead.

18. ***Looking back:*** Did anything about the design and conclusions of this study concern you? In particular, are there things that could have been done to give a better chance finding strong evidence of a true difference between the two groups? Issues you may want to critique include:

    - Any mismatch between the research question and the study design
    - How the experimental units were selected
    - How the treatments were assigned to the experimental units
    - How the measurements were recorded
    - The number of experimental units in the study
    - Whether what we observed is of practical value

19. ***Looking ahead:*** What should the researchers' next steps be to fix the limitations or build on this knowledge?