|  |  |
| --- | --- |
| **Tabletop Lab:** | **Sampling Error and the Scientific Method** |

**Introduction:**

One of the most fundamental processes of science is data collection. After the formulation of a sound hypothesis, care must be taken to obtain information that may be ***qualitative*** (based on characteristics such as color, texture etc.) or ***quantitative*** (based on numerical information such as height, mass, amount etc.). Collecting data can be the most time intensive portion of any experiment and sometimes it can be fascinating and exciting. Consider for instance the famous animal behavior studies of charismatic mega-fauna such as lions and chimps in Africa. Studying such amazing animals is often a big draw for young people to consider a career in science.

More commonly however, and often to the dismay of many budding young scientists, is the discovery that data collection can take hours and hours and not be so glamorous. It takes both ***precision*** (discriminating fine detail) and ***accuracy*** (being as correct and exact as possible) to collect data effectively. Some areas of science are more oriented towards precision, e.g. medicine and pharmacy. In this lab we will look at the nature of a concept that is important in science called **sampling error**.

In the weeks to come we will learn about the diversity of life forms found in the ***biosphere***, i.e. the living portion of the Earth’s surface and in its oceans. We will also learn how scientists estimate the numbers of species for a given taxonomic group, or even for a population of concern. The state insect of Oregon is the Oregon Swallowtail Butterfly and the state mammal is the beaver. How do population biologists estimate, and then assist in the regulation or management if necessary, for a given population? Well to start, they must do some estimation. This may be done via capture/recapture studies. Scientists may distinguish individuals of the population and “track” their movements by capturing and releasing a sample set of animals applying collars, paints on the fur, or tags in the ears. It is not possible to track all individuals for the entire state because this simply would not be practical with the limited personnel and the financial resources of the state. So samples are taken in the form of sightings and/or tagging of individuals for further studies in order to “estimate” or extrapolate the likelihood of the entire population.

The potential for error in making these estimates illustrates the concept of ***sampling error***. Sampling error ***occurs when sampled subsets*** are ***not representative*** of the true population. Another way that you can think of this is that it is an indefinite or distorted view of the actual population. This could lead to serious consequences if the estimates were way off. This could be especially significant when estimating numbers within species that are threatened or endangered. Consider the potential management decisions that would be necessary if a sample of the seriously endangered Siberian tiger or the Black footed ferret were over- or under-estimated.

You will be exploring three different ways to estimate population sizes using small objects such as dried beans, candy, or marbles to represent a population of organisms such as seeds of two separate plant species. These three methods are:

1. Mark – Recapture method
2. Sampling method
3. Transect method

After conducting your experiments and collecting data on your two “species”, you will do a direct count of each population and determine which method was the most accurate in predicting population size.

j0250819

For this lab you will need the following supplies:

* Two sets of dried beans or other small objects of approximately one cup in volume, each set being a different color. Items such as M&Ms and Skittles, two different colors of marbles, or poker chips would be suitable alternatives to using dried beans.
* A measure of string (~4 meters).
* Four popsicle sticks.
* A willing partner
* A bowl or large jar
* A calculator
* A marker
* A bandana (to blindfold the volunteer).

**Important: Do NOT count the beans until instructed to do so.**

1. Describe the QUALITATIVE characteristics of these two “bean” species.
2. If you were to measure the width and length respectively of each bean and then come up with an “average” dimension of each type of seed would you be collecting qualitative or quantitative data?

**Part 1: Mark – Recapture Method**

**I. Introduction:**

The mark – recapture method is well suited to organisms that are mobile and have an extremely large territory to roam in. In this method, individuals are randomly captured, marked in some way, and released back into the habitat. At a later time, a second random capturing event is done and the number of marked individuals that are caught in this second round is recorded and used to extrapolate the ration of marked vs. unmarked individuals that were collected. This technique requires several assumptions:

1. All individuals in the population are just as likely to get captured.
2. There are no changes in population size during the study, i.e. no migrations or reproductive events.
3. The method of capture and markers used in no way affects behavior or likelihood of survival of the individual affected.

The formula used to predict population size is as follows:

Where

* **M** = # of individuals first captured and marked.
* **n** = the total number of individuals caught at the second capturing event (Both marked and unmarked)
* **m** = the number of individuals in at the second capture event that were marked (Marked only)
* **N** = Estimated total population size.

3 out of the 4 variables (M, m, and n) are collected during the mark – recapture study. The 4th variable (N) is then determined by solving the equation using a basic algebraic technique.

Practice performing these calculations before using this method to predict your bean populations.

1. Biologists have begun a study of pronghorn population dynamics. In the intensive first stage, they constructed traps and corrals, and conducted a mark – recapture study. A total of 175 pronghorns were captured and marked. Two weeks later, 165 pronghorns were captured and of that 165, only 130 of them were found to have been marked in the first capture event. Use these numbers to solve the equation for N:

**II. Procedure:**

1. Put each color of bean in separate cups. (Keep each color of bean separate for this method, do not mix them together.) Remove a small handful of beans from one of your cups.
2. Using a marker or small piece of tape, clearly mark all the beans you pulled out. Be sure your mark can be easily seen on all sides of the bean. Record the number of beans you marked (M) in the **Table 1** below.
3. Return your marked beans to your original population and mix thoroughly.
4. Remove another small handful out of your cup. Record the total number of beans grabbed (n) and the number of those beans that are marked (m) in the table below.
5. Using the data collected, calculate the estimated total population size for each bean color using the method presented in the introduction and record it in the table.
6. Repeat steps 1 – 5 on your second color of beans.
7. Determine the predicted overall population size of both bean colors by adding together the two estimated sizes of each color. Record the data below:
   1. Estimated overall population size: \_\_\_\_\_\_\_\_
8. Calculate the percentage of the overall population for each color of bean by dividing the estimated population size (N) by the number you calculated in 6). Record the data in the table.

**Table 1: Mark – Recapture Method**

|  |  |  |
| --- | --- | --- |
|  | **Color 1 Beans** | **Color 2 Beans** |
| **(M) – # of Individuals marked** |  |  |
| **(n) – Total # recaptured** |  |  |
| **(m) – # recaptured that are marked** |  |  |
| **(N) – Estimated population size** |  |  |
| **% of overall population** |  |  |

Use the space to the right to perform your calculations

**Part 2: Sampling Method**

**I. Introduction:**

This type of method is often used for organisms that have large population sizes and would not be harmed by permanently removing some individuals from the habitat. Additionally in order for the highest degree of accuracy, the species in question should be relatively stationary, with little or no migratory habits. If there is a lot of movement in and out of the population, this type of method could be relatively inaccurate, depending on the timing of the sampling event.

**II. Procedure:**

1. Place all of the beans (of both colors) into a bowl or jar and mix thoroughly.
2. Here is where you need a volunteer willing to be blindfolded, while you record data throughout the experiment.
3. The blindfolded volunteer will pull 5 beans at random out of the bowl, while the data collector will record the number of each bean color that was removed. Do not replace the beans as they are removed from the bowl.
4. Record the number of the beans and their respective colors in Columns 2 and 3 in **Table 2** below.
5. Based on this one sample, calculate the estimated percentage of each color of bean by dividing the # of each color by your sample size of 5 total beans. Record this in columns 4 and 5 of the table.

For example, if you pulled 3 white beans and 2 red beans:

Or, 40% of the two bean populations is red.

1. **Repeat** steps 1 – 5 two more times for a total of three trials.
2. **Calculate** the average predicted percentage of each of the bean colors based on the three trials, which have a total sample size of 15. (i.e. add the number of each color from all three trials and divide by the total sample size of 15) Record this data in row 5 of the table. Use the space below to show your calculations.

**T able 2: Sampling Method**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **# of Color 1 Beans** | **# of Color 2 Beans** | **Estimated % of Color 1 Beans** | **Estimated % of Color 2 Beans** |
| **Sample #1** |  |  |  |  |
| **Sample #2** |  |  |  |  |
| **Sample #3** |  |  |  |  |
| **Average Predicted %** |  |  |

**Part 3: Transect Method**

**I. Introduction:**

Transecting is a standard survey technique. A transect is a straight line extending through a habitat. All animals or plants seen within a predetermined distance of the transect line are counted. That distance may be so small that only individuals touching the transect line are counted. Alternatively, the distance could be any individuals within eyeshot while walking the transect line. Factors such as the size of the area, total amount of habitat, visibility, and mobility of the organisms will contribute to deciding your distances used in data collection methods. For our purposes we will only count beans that are touching the transect line.

**II. Procedure**

1. This method is best demonstrated outside in an area that has some natural vegetation cover, but not too high, so don’t choose overly tall grass for instance. A good sampling area would be a local park, but not a manicured lawn or artificially landscaped area. Near a stream or river or other wild area that has public access is ideal.
2. Using string, approximately 4 meters in length, make a square in your chosen area, with one corner approximately in the north, but exact north is not necessary.
3. Use one of the Popsicle sticks to designate the north corner, mark with an N.
4. Do the same with south (S), then either west (W) or east (E).
5. Adjust your string and corners so that you have a regular square of equal dimensions per each side – this may take several adjustments. Exact 90 degree angles are not required, but should be close.
6. Take both colors of beans and scatter them randomly through the sampling square.
7. Tie a string between the two stakes that mark north and south. This now represents your transect line.
8. **Count** how many beans of one color there are directly under the string, any part of a bean under that string counts, it need not be the entire bean.
9. **Record** that number in **Table 3** below.
10. **Repeat** step 8 – 9 for the second color of bean.
11. **Calculate** the percentage of each color that touched your transect line. Your sample size is now the total number of beans counted under the transect line, so much like the calculations perform in part 2, step 5, divide the number of each color by the total number of beans touching the transect line and record the data in the table.
12. **Leave** your square in place to complete Part 4.

**Table 3: Transect Method**

|  |  |  |
| --- | --- | --- |
|  | **Color 1 Beans** | **Color 2 Beans** |
| **# of Individuals touching transect** |  |  |
| **% of sample size** |  |  |

**Part 4: Sampling in the Field – Extension Activity**

**I. Introduction**

Now that we have explored the “theory” of sampling methods, let’s do an estimate that is more similar to what scientists actually carry out.

The square that you assembled in part 3 is what scientists call a ***quadrat***, which has four equal sides. Sounds easy, but you may be surprised how challenging it actually is. A quadrat is a manageable sampling area, often 1 square meter, but it could be one square mile or kilometer. Once the quadrat (four sided polygon) is designated it may be marked with flagging tape, or in this case string to define the boundaries. After the perimeter is designated an official count is made of whatever organisms the researcher(s) are interested in; it could be dandelions in a park, grasshoppers in a field, squirrels in a forest, or wolves in a national park. Because counting and capturing subjects from the animal realm can include complicated permits, we will keep things simple by estimating plants but we will not pluck or otherwise collect them – just observe their numbers, just in case there are any protected members in our sample plot. The ***distribution*** of different species in a given community or area, is affected by many different aspects such as shade, soil type, moisture, etc.

**Species richness** is a concept in biology that addresses the ***number of species*** for a given area or community. This value is simply the number of species (regardless of how many there are) for a given area or community.

**Species abundance** is defined by the ***number of overall members*** in a given population or sample area. An example would be that house sparrows have much higher species abundance than peregrine falcons, the latter of which are rare and require large areas to roam and hunt.

**Species evenness** is a manner of communicating about ***relative abundance*** of individuals for a given species i.e. the number of individuals for a given type in a sample area. For example a prairie has a high species evenness factor because usually for a given area there are only a few or even one species where the predominant species is a grass and has a much higher relative abundance compared to other plants also found on the prairie such as sunflowers.

**Species diversity** considers both richness and evenness. Usually complex mathematical formulas such as the Shannon Index, a logarithmic function, are calculated to compare between different communities. We will not be running such a calculation for this lab, but it is important to know that data related to how many types and how many individuals is used for calculating an index that can be used for comparison reasons. When you hear that the tropical rainforest has high species diversity, this statement can be verified using both the quadrat method and transect samples, i.e. counting along a line, in order to gather data for comparing with other communities. In this way an index is really a basis of comparison; in a way it is like a “control” value for discussions between communities or sample areas, not within a given area.

**II. Procedure**

1. Count the number of **plant species** present in your quadrat**.** Next,systematically count the **number of each species** (not the number of leaves, but number of whole plant specimens), found in your quadrat, starting in the north end and working your way towards the south end of the square scanning back and forth, like a printer does on your home computer. For species such as grass that would be too numerous to count, estimate the percentage of the quadrat that it covers.
2. Even if you don’t know a plant, **describe it** and give it a descriptive name such as [Broad leaf #1], [Grass like plant #1, 2, 3], [Flowering plant with fuzzy leaves] and so forth. Exact names or identification are NOT necessary for this lab, but it may inspire you to learn from guide books in the library if you are interested.
3. As you scan through your quadrat, note any insects or animal marks while you look very carefully through the square for your different types of plants.

**II. Analysis.** Answer the following questions:

1. How many ***different types of plants*** in total did your quadrat contain? This value corresponds to species richness.
2. How many ***total plants*** did your quadrat contain?
3. Which plant had the ***greatest species abundance***?

**Part 5: Direct Counts**

Now it’s time to perform a few final calculations and analyze the results.

1. Count the total number of each color bean you have and record the data below:
   1. Total # of color 1 beans: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   2. Total # of color 2 beans: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   3. Overall # of beans ( a + b): \_\_\_\_\_\_\_\_\_\_\_\_\_
2. For the sampling method in part 2 and the transect method in part 3, we calculated our data in terms of predicted percentages of the overall population, therefore we need to calculate the actual percentage for the sake of accurate comparisons. Record this data below:
   1. % of color 1 beans: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   2. % of color 2 beans: \_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Part 6: Analysis** **Table 4: Combined Data**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Percent of Overall Population** | | | |
| Actual Percentage | Mark – Recapture | Sampling Percentage | Transect Percentage |
| **Color 1 Bean** |  |  |  |  |
| **Color 2 Bean** |  |  |  |  |

For ease of analysis, combine your estimated and actual population sizes, expressed as a percentage of the overall population of beans into **Table 4** and answer the questions below. These questions will help to prepare you for the quiz you will take when logged on to the course website.

**Analysis Questions:**

1. Which method was most accurate at predicting the population size of each bean color?
2. Discuss how the qualitative aspects of your beans affected the outcome of each of your methods. Did shape or size influence the accuracy of your results and what are some ways you could reduce this type of error?
3. Do you think that you would have higher reliability of your predicted percentages with more trials in each of the methods you simulated? Why or why not?
4. Would you have better reliability with a larger sample size per each trial? Explain your answer.
5. Write a hypothesis that would take either question in #5 or #6 into account and run your own experiment. Pick one of the methods and repeat the process, in an attempt to verify or disprove your hypothesis.
6. After running your experiment, write a brief 2-3 sentence analysis/discussion, including your data, and state whether your hypothesis was supported or rejected based on your findings. Be sure to address how increasing the repetitions and/or sample size influenced the accuracy of your estimations.
7. What is the purpose of blindfolding the individual in the sampling method?
8. In the introduction to each method you learned that there are certain assumptions that must be made when performing these types of predictions and calculations. Based on these assumptions and observations you made while carrying out the experiments, discuss some limitations of each method. Include in your discussion examples of a species that would be appropriate and which would not be appropriate to use that method on.
   1. Mark – Recapture Method
   2. Sampling Method
   3. Transect Method
9. What are some obvious limitations of working with transects to estimate species richness and species abundance respectively?
10. How could a researcher improve the confidence level of their sampling method, whether using quadrats or transects to estimate population numbers, in order to address sampling error?
11. List at least three reasons why estimating plant species as “whole plants” i.e. not just by seeds may be difficult. It may be helpful to think of a specific plant species such as Douglas Fir trees or sunflowers in a meadow.