Conservation Analysis: Current and Future Work

Jon Brown

Contents

- Overview of Work
- Analysis of Current Work
- Future Recommendations
- Review and Conclusions

Overview of Current Work

Overview of Current Work: Foot and Mouth Disease Reduction Effort

Through monitoring efforts by the NPS, in Bryce National Park, it has been determined that 15% of sheep have foot and mouth disease.

Based on this, Yellowstone has implemented a program to reduce the spread of foot and mouth disease in the sheep population.

<u>Stated Goals</u>

- Monitor sheep population for foot and mouth disease.

- Be able to determine at least a 5% difference in the affected population

- Have a confidence interval of 90%

With the goal stated, there are some questions to keep in mind:

Is it accurate?
- Does the answer given accurately reflect the state of the population?

Is it effective?
- Do the results meet the stated goals?

Is it efficient?
- Does the program meet the goals in a cost- and time-effective manner?

Overview of Current Work: Determination of Sample Size

First, the ideal sample size needs to be determined. This can be determined using our goals and a simple online calulator.

The calculator chosen was provided by Optimizely, an provider of web experimentation software.

By using the goals of an initial 15% baseline, a minimum detectable effect of 33% (based on the 5% reduction), and a confidence interval of 90%, the sample size would be 520 sheep.

Baseline Conversion Rate

15 %

Your control group's expected conversion rate. [?]

Minimum Detectable Effect

33 %

The minimum relative change in conversion rate you would like to be able to detect. [?]

Statistical Significance

90%

95% is an accepted standard for statistical significance, although Optimizely allows you to set your own threshold for significance based on your risk tolerance. [?]

Sample Size per Variation

520

The sample size has been determined, and we need to answer some questions of the accuracy of the program.

- How do we know that the sample size is correct?

- Are we sure that we are getting an accurate picture of the population?

- What does the effect of recounting a particular sheep do to the outcome?

Thankfully, we know the answers to those questions.

- Sample size is a well-known statistical measure. This measure is specifically for A/B testing, so we could check the calculator against know formulas.

- Our stated definition of accurate is 90%. So our sample size will give us a fairly accurate measure (though not completely).

- The goal is to determine a reduction in foot and mouth disease. Therefore, as long as our selection is random, recounting a sheep doesn't affect the outcome, since a sheep could contract the disease between observations.

Now that the accuracy of the test is established, we need to ask questions on whether the test meets our needs.

- Is A/B Testing the best test to use?

- How do we know we're measuring the effects of the program, and not some other effect?

- Will we actually see anything from this analysis?

These are important questions, so let's go through each.

- A/B Testing is a good fit for categorical surveys. Since we only have two categories, A/B testing gives a simple, straightforward, and reliable method of testing our observations.

- We have two categories, and an A/B test works well for two categories. While not complete, by structuring the observations and test this way, we can be (90%) certain that the effect is from the program.

- This will be answered in the following slides.

520 sheep is a lot of sheep. Can we really collect 520 samples from sheep within a reasonable time frame?

Yes, we can. From the data in observations.csv, we note that the number of observations range from ~150-500 sheep per week.

From this, we can see that the survey will take from 1-3.5 weeks, depending on the park. Add in 2 weeks for analysis, report creation, review, and submission, and the survey will be completed for all 4 parks in approximately 6 weeks.

Observations of Sheep per Week
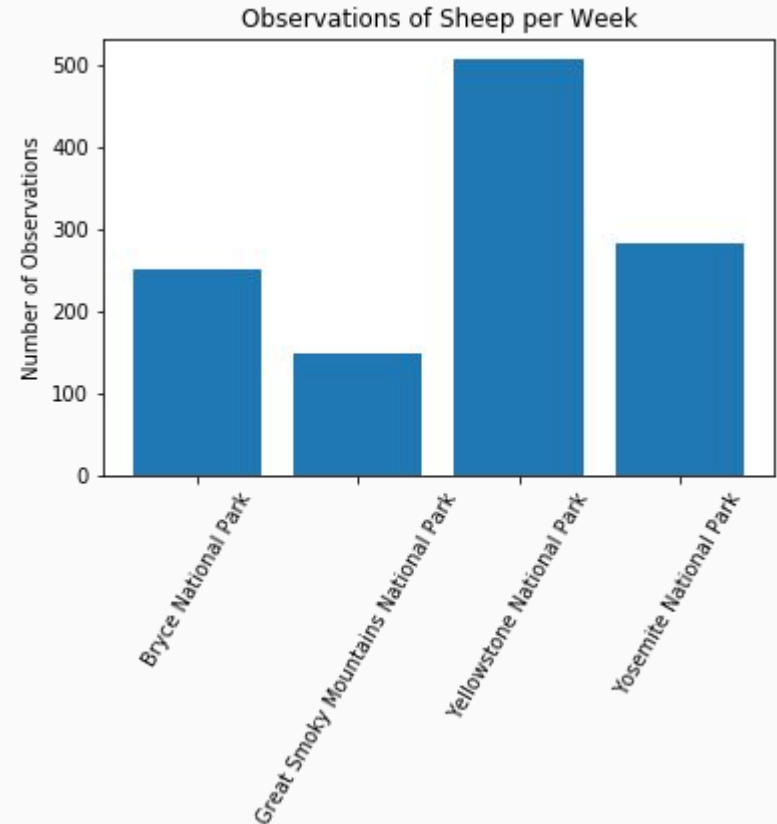
Alright, so we can complete the survey in a reasonable amount of time. Can we perform it within a reasonable amount of cost? Specifically:

- How much do we need to set aside for this?

- Based on the cost, what's the return on our investment from this test?

- Will this help us keep or expand our funding in the future?

These are important questions, so let's go through each.

- Nothing! Since a program is already implemented for monitoring the spread of the disease, it's simply an extra task within the current program.

- The return isn't monetary, however, still valuable. At the end of the survey we will have a verification of whether the reduction effort succeeded or failed.

- It very well could, if we frame the results correctly. This will be covered in the Recommendations section.

Future Recommendations

Future Recommendations: How to Capitalize on Initial Success

Currently, we have a low-cost program that provides valuable information for our parks. Which is fantastic. However, we also want to see how to leverage this success in future programs and budget proposals. Specifically:

- How to determine where the money would have the most impact?
- Is there any particular category or species of animal that is more at risk than others?
- How are we sure that using it on a particular category/species is the correct decision?

All of these are great, important questions. There are several pieces of information we will need. Specifically:

- List of all observed species within the National Park System
- Each species has a listing of it's category, scientific and common name
- A listing of the conservation status for each species

Thankfully, species.csv contains all of this information in a single file.

Future Recommendations: Where to focus the money?

If we're going to ask for money, we need to show a plan to use it. So, where to start?

Obviously, we need to use the money for conservation, but why choose some species over others?

By grouping the data by conservation status, we can see that by focusing on Species of Concern, In Recovery, Threatened and Endangered species, we can provide the maximum impact for every dollar spent.

Conservation Status by Species
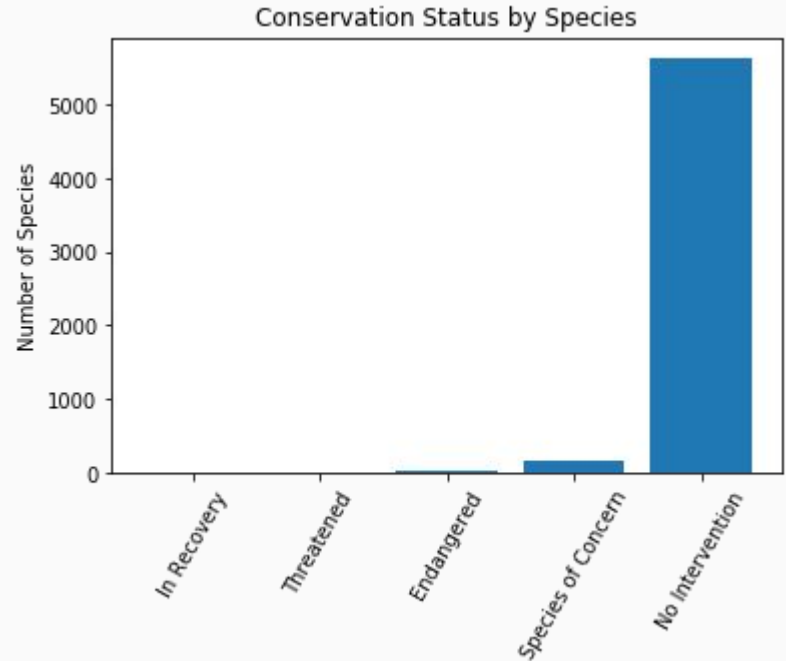
Future Recommendations: Can we be even more precise?

So, saying that we should focus conservation efforts on the species that need conservation is not a groundbreaking idea. Can we determine whether a certain category within the group need more help than the others?

Since our data is divided into categories (mammal, reptile, bird, etc along with proctection status) we can use a Chi Squared test to determine if there is a difference, and if so, between which groups.

From the data, we've created a couple of contingency tables comparing whether Mammals are more likely than Birds or Reptiles to be endangered.

	Protected	Not Protected
Mammals	30	146
Bird	75	413

	Protected	Not Protected
Mammals	30	146
Reptile	5	73

Future Recommendations: Can we be even more precise?

Now that we have the contingency table, we can see whether any difference between the categories is significant or simply chance.

In our first comparison between Mammals and Birds, we can see the test gives us a P-Value of 0.68. Using the standard P-value significance of $P \leq 0.05$, we can safely say the difference is probably just chance. Or in other words, we should not prefer Mammals over Birds in our conservation efforts.

Similarly, with Mammals and Reptiles, we arrive at a P-Value of 0.03. From this, we might want to give preference to Mammals over Reptiles in conservation efforts.

	Protected	Not Protected
Mammals	30	146
Bird	75	413

P-Value: 0.68

	Protected	Not Protected
Mammals	30	146
Reptile	5	73

P-Value: 0.03

Conclusions

Obviously, we could continue the analysis *ad nauseum*. The main issue at this point is knowing what we can take away from this analysis. Specifically:

- Are we able to see the difference we want in the Foot and Mouth Reduction Program?
- Going forward, where should we focus our efforts?

For the Foot and Mouth Reduction Program, by following the current program in place, we can expect to have an analysis of the results completed in ~3 weeks at Yellowstone, with ~6 weeks being the time frame for all 4 parks listed.

For further conservation efforts, it may be that Mammals should be focuses on over other categories, such as Reptiles. However, the analysis is only preliminary, and more work should be done to determine whether Mammals are more at risk to need protection than other categories of animals.

Conclusions: Where Do We Go From Here?

Data analysis is valuable, but only if it leads to definite goal that will provide improvement. So, from our analysis, these are the recommendations we have:

1. Continue and expand the Foot and Mouth Reduction Program, since it will give an easily quantifiable "win" that can be leveraged in future budget proposals and negotiations.

2. For future conservation, further analysis on exactly which groups need to be focused on in conservation efforts to maximize the impact. Analysis should be tailored to desired outcomes of proposed conservation efforts.

Thank You!