

Species Analysis

Codecademy

Bryan Wiedeman



SPECIECS_INFO.CSV

2

- The species_info.csv contained a data set of information about 5823 different animals and plants.
- The information was sorted by species category, scientific and common names and conservation status.
- While scientific_name and common_name were mostly unique, both category and conservation_status could be grouped for further analysis.
- There were 5541 unique species in the data set
- There were 7 different categorizations in the category column 'Mammal', 'Bird', 'Reptile', 'Amphibian', 'Fish', 'Vascular Plant', 'Nonvascular Plant'
- There were 5 different categories in the conservation_status column with counts of nan or not endangered (5633), 'Species of Concern' (161), 'Endangered'(16), 'Threatened'(10), 'In Recovery'(4)

Currently only a small percentage of species included in the data are either endangered, threatened or of concern of the total sample set (3%). However, there are also very few species who are now in recovery after having this designation. The

SIGNIFICANCE CALCULATIONS

3

- After calculating the number of species that are protected by each category type. I found in descending order the percentage of individual species by category from most to least protected were
 - Mammal 17%, Bird 15%, Amphibian and Fish 9%, Reptile 6%, Nonvascular Plant 2%, Vascular Plant 1%
- While the category Bird had the most endangered species on a percentage of the total population it ranked behind mammal
- However, the chi square contingency test returned a result of 68% which tells us that species in the Mammals category are not more likely to be endangered than species in the Bird category.
- The second test comparing species in the Reptile category and Mammal category found that species in the Reptile category are more likely to be endangered as the result of the chi square test was less than 5% (3.8%) which demonstrates significance.

Currently only a small percentage of species included in the data are either endangered, threatened or of concern of the total sample set (3%). However, there are also very few species who are now in recovery after having this designation.

RECO FOR CONSERVATIONISTS

4

While results from the chi square significance test show that Reptiles are more likely to be endangered than mammals they also make up a much smaller proportion of total species. Conservations should continue to focus their efforts on Mammals.

SAMPLE SIZE DETERMINATION

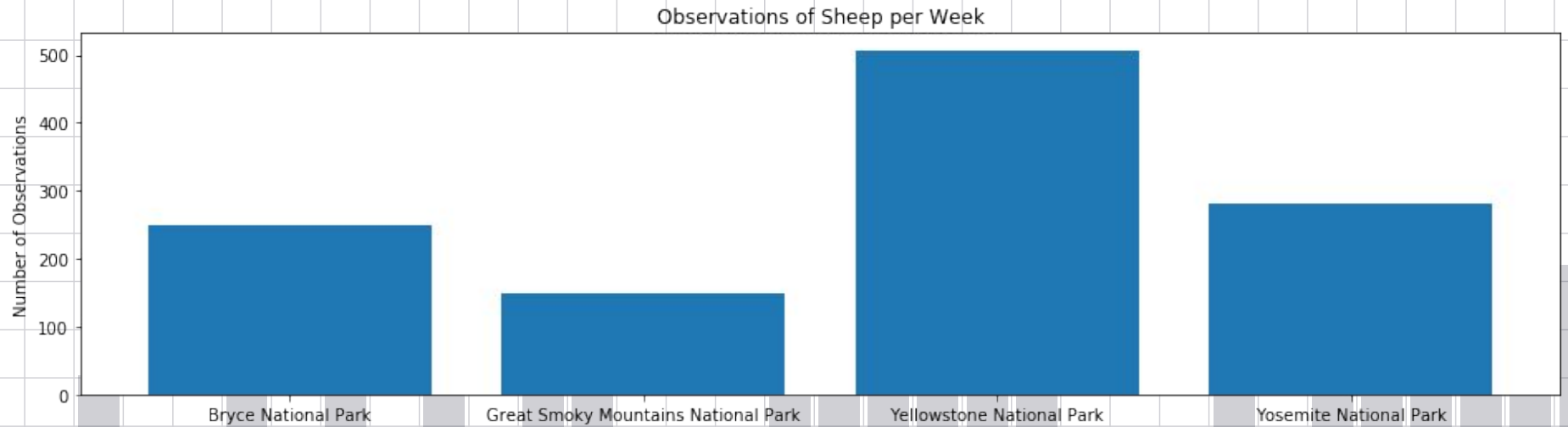
- In order to calculate an adequate sample size to ensure a 90% statistical significance or stated another way, a 10% chance of a false positive result we first need to calculate the Minimum Detectable Effect.
- The minimum detectable effect is 33.3%. This was calculated by dividing the 5% drop in rate of foot and mouth disease in Sheep at Yellowstone we want to see into the 15% rate that has already been calculated from Bryce and then multiplying that calculation by 100.
- Using the baseline rate of 15% from Bryce, the MDE of 33.3% and the statistical significance rate of 90% provided, optimizely returned a required sample size of 520.
- We know that on average 250 sheep are observed each week at Bryce and 507 a week at Yellowstone. By dividing the sample size by the average observed rate at each park I found it would take approximately 2 weeks to observe enough sheep at Bryce and 1 week at Yellowstone to understand if the program is reducing foot and mouth disease by 5%.

GRAPH 1

Conservation Status by Species



GRAPH 2



CREDITS

Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by [SlidesCarnival](#)
- Photographs by [Unsplash](#)