

Biodiversity for the National Parks

Capstone Project
James M Seymour

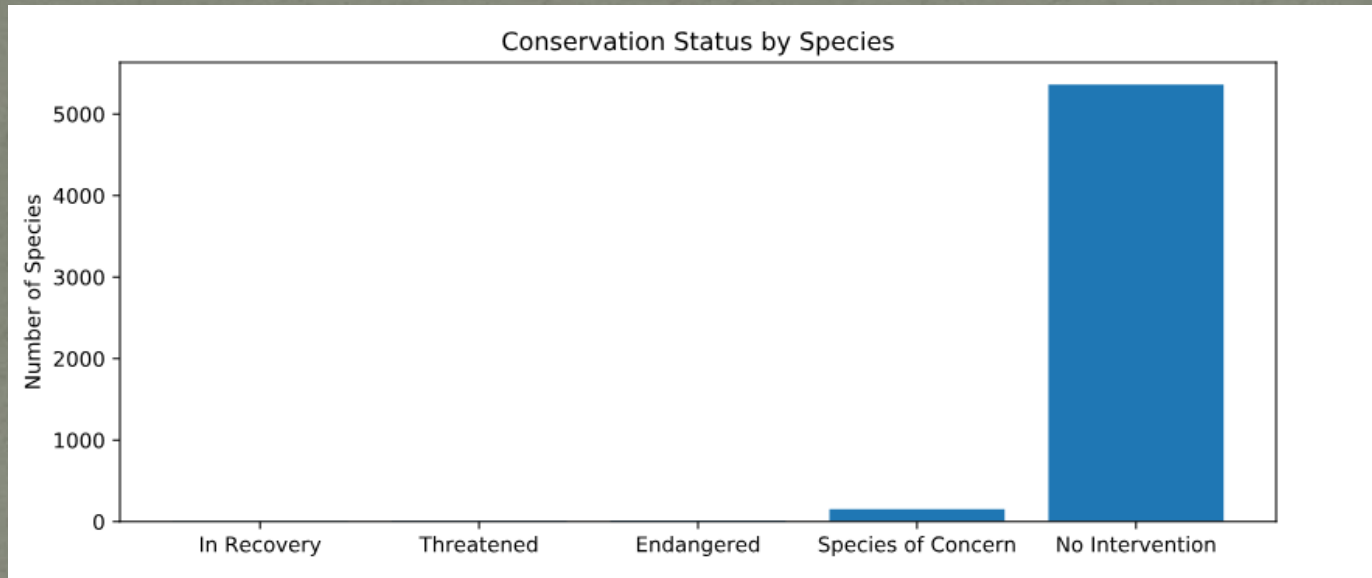
Introduction

- Researchers need help generating and interpreting data involving biodiversity in national parks.
- This presentation will help determine:
- Are some categories of species more likely to need protective intervention than other categories of species?
- Which national park would be the most time efficient place to observe sheep for a disease prevalence study?
- What sample size should be used for the disease prevalence study?

Data in species_info.csv

- A DataFrame was created from species_info.csv so that the data within could be easily manipulated.
- The data set contains records of 5541 unique species.
- 7 categories of species: mammals, birds, amphibians, fishes, reptiles, vascular plants, and nonvascular plants
- Each unique species has 1 out of 5 possible conservation statuses: endangered, threatened, in recovery, species of concern, no intervention needed
- The data set also contains the scientific names and common names of each species.

Species Count by Conservation Status



- Numbers of unique species with each possible conservation status: 4 in recovery, 10 threatened, 15 endangered, 151 species of concern, 5363 no intervention needed
- The large majority of species need no intervention, which can easily be seen from the bar chart above.

Category Counts for Conservation Status

Category	No Intervention	Any Other Status
Amphibian	72	7
Bird	413	75
Fish	115	11
Mammal	146	30
Nonvascular Plant	328	5
Reptile	73	5
Vascular Plant	4216	46

- The table above shows the number of unique species within each category that either do or do not require intervention.

Percentage of Species Protected

Category	Percent Protected
Amphibian	8.86%
Bird	15.37%
Fish	8.73%
Mammal	17.05%
Nonvascular Plant	1.50%
Reptile	6.41%
Vascular Plant	1.08%

- The chart above shows the percentages of unique species within each category that have any conservation status other than “No Intervention”.

Is there a significant difference?

- The percentages are different, but are any of them significantly different? If they are, then that information could be used to support the hypothesis that some categories of species are more likely to need protection than other categories of species.
- A chi-squared test can be used to determine the significance of the difference between these values because a chi-squared test is used for comparing multiple sets of categorical data.

Chi-squared Test Results

- Significance is determined by comparing the p-value calculated by the chi-squared test to 0.05. The difference is significant if the p-value is < 0.05 .
- Mammal-Bird p-value = 0.6876, so the difference between mammals and birds is not significant
- Mammal-Reptile p-value = 0.0384, so the difference between mammals and reptiles is significant. This means that mammals are significantly more likely than reptiles to need protection. This also supports the idea that some categories of species are significantly more likely than others to require protective intervention.

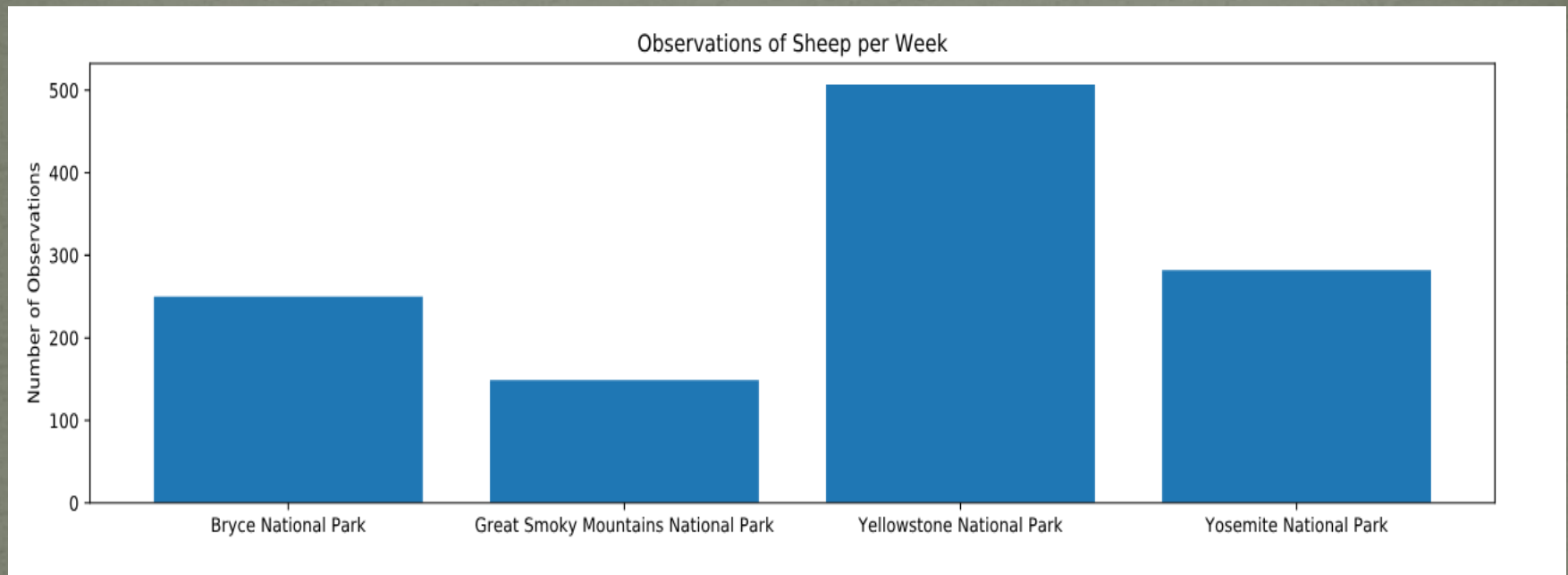
Recommendation for Conservationists

- Based on the results of the chi-squared tests, both mammals and birds are significantly more likely to require intervention than the other categories of species.
- Because of this, it is recommended that conservationists focus more on mammals and birds than other categories of species, as members of these two groups are more likely to require intervention than species from the other groups.

Foot and Mouth Disease in Sheep

- Researchers need help determining the number of sheep with foot and mouth disease in national parks.
- Before any tests can be run, an appropriate sample size for the study must be determined.
- The sample size must be calculated beforehand to avoid biases.
- There are three species of sheep in the parks: domestic sheep (*Ovis aries*), bighorn sheep (*Ovis canadensis*), and Sierra Nevada bighorn sheep (*Ovis canadensis sierrae*)

Sheep Observed by Park per Week



- The chart above shows the number of sheep observed per week in each park: Bryce 250, Great Smoky Mountains 149, Yellowstone 507, Yosemite 282

Sample Size Determination

- An appropriate sample size was determined using a sample size calculator.
- The baseline conversion rate = 15%, the minimum detectable effect = 33%, and the desired statistical significance = 90%
- The sample size calculated using the parameters above = 890
- Number of weeks needed to observe the sample size at each park: Bryce 3.56, Great Smoky Mountains 5.97, Yellowstone 1.76, Yosemite 3.16

Conclusion

- In conclusion:
- Some categories of species are more likely to need protective intervention than other categories.
- Mammals and birds are more likely to need intervention than other categories of species.
- 890 is the appropriate sample size for the sheep foot and mouth disease study.
- Yellowstone would be the most time efficient place to run the sheep foot and mouth disease study.