Biodiversity Capstone Project

Investigating Protected Species

Rehan Choudhary

Introduction

The National Park Service (NPS) has requested the following two analyses:

1. Analyze data on the conservation status of various species and determine whether certain types of species are more likely to be endangered.
2. Determine the sample size needed to detect a 5% point reduction in foot and mouth disease among sheep with 90% statistical significance. And based on observation data, estimate how long it will take to sample the required number of sheep (e.g., the required sample size).

Dataset on Conservation Status of Species at National Parks

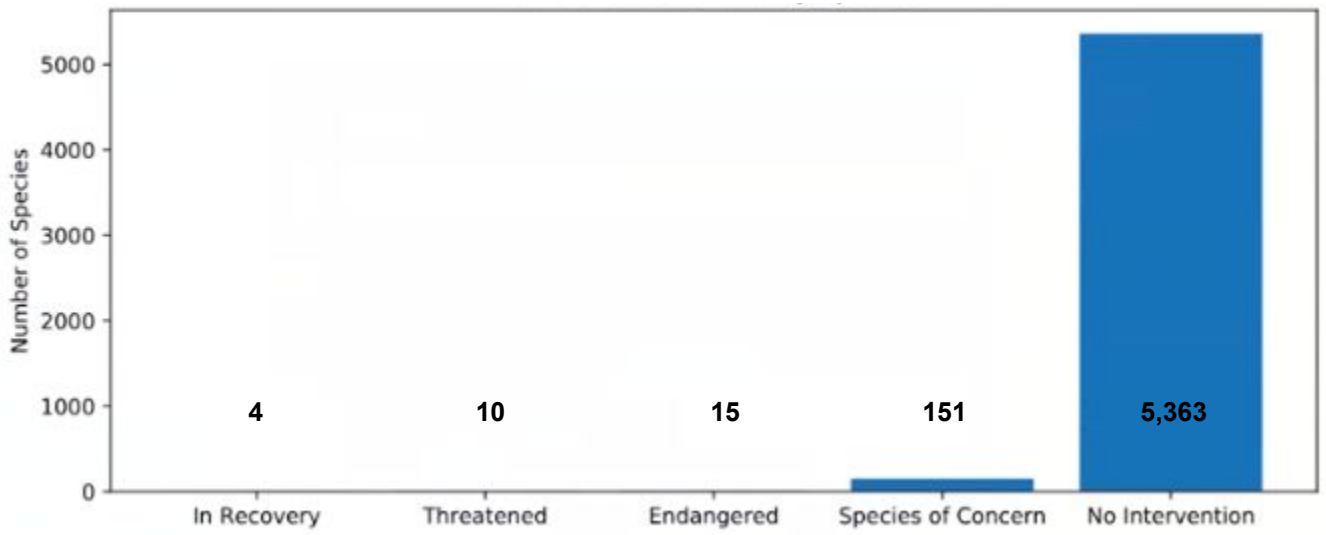
NPS provided a CSV file (species_info.csv) with the following data about different species in our National Parks:

1. Scientific name of each species
2. Common names of each species
3. Species conservation status (Endangered, In Recovery, No Intervention, Species of Concern, Threatened)

Upon further inspection, the file contained data for a total of 5,541 different species and seven different types of species (mammal, bird, reptile, amphibian, fish, vascular plant, nonvascular plant).

Conservation Status by Species Type

Approximately 97% of species are not endangered
and thus require no intervention.



Are Certain Types of Species More Likely to be Endangered?

Since we are dealing with categorical data, the chi-squared test can be used to determine whether the difference between two species is statistically significant.

Species Type	Not Protected	Protected	Percent Protected
Amphibian	72	7	8.86%
Bird	413	75	15.37%
Fish	115	11	8.73%
Mammal	146	30	17.05%
Vascular Plant	328	5	1.50%
Reptile	73	5	6.41%
Nonvascular Plant	4,216	46	1.08%

Is the difference between Mammals and Birds statistically significant?

Running the chi-squared test, we found a p-value of about 0.688, so we conclude that the difference between the percentages of protected birds and mammals is not significant and is a result of chance.

Is the difference between Mammals and Reptiles statistically significant?

Comparing the percentages of protected reptiles and mammals and running the same chi-squared test, we found a p-value of about 0.038, which is significant.

Recommendations for Conservationists

Conservationists can use chi-squared testing to understand which species are more likely to be endangered and focus limited resources on addressing the endangered status of those species.

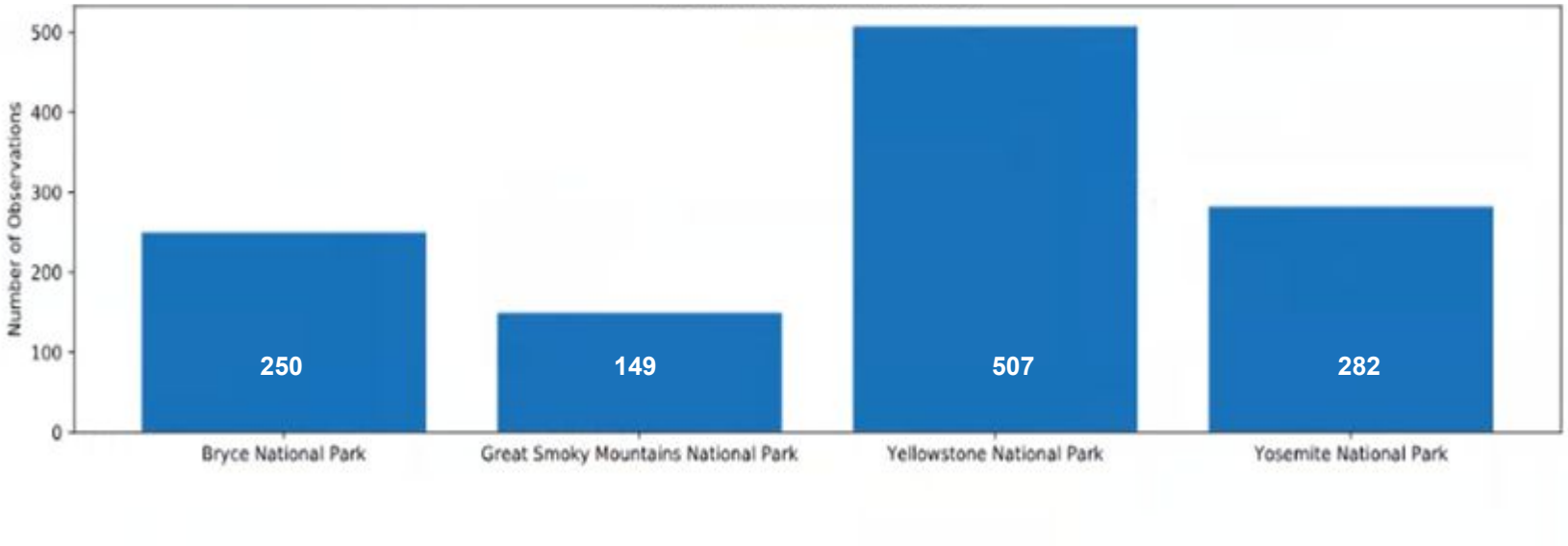
Dataset on Observations at National Parks

NPS provided a CSV file (observations.csv) with the following data, reported by Conservationists, about sightings of different species in our National Parks for the past 7 days:

1. Scientific name of each observed species
2. National park where species was observed
3. Total number of sightings of each species

Observations of Sheep per Week by National Park

Sheep sightings by park were identified by merging observation data (observations.csv) with species data (species_info.csv).



Number of Sheep that Need to be Observed?

Sample Size

Using a sample size calculator and given a baseline conversion rate of 15% (occurrence of foot and mouth disease in sheep), a minimum detectable effect of 33.3% can be detected with 90% statistical confidence by observing at least a total of 870 sheep.

Yellowstone National Park

It will take 1.7 weeks or almost 2 weeks of observation to observe at least a total of 870 sheep at Yellowstone National Park.

Bryce National Park

It will take 3.5 weeks or almost 4 weeks of observation to observe at least a total of 870 sheep at Bryce National Park.

Thank you