Capstone Project

Biodiversity for the National Parks

Codecademy Introduction to Data Analysis

James Horne

Provided Data

Data Supplied in species_info.csv

Data supplied included column headings

- Category
  - amphibian, bird, fish, Nonvascular plant, vascular plant and reptile
- Scientific Name
  - Scientific names of species in each category
- Common names
  - Linked common names to each scientific name for the species
- Conservation status
  - Endangered, In Recovery, Species of concern, Threatened and Null

Interesting Points

- Majority of species did not have an assigned conservation status
- For all species categories, <20% of the species were protected - mammals being the highest (17%)

Significance Calcs

Chi-Squared Test

- Chosen method for significance calculation due to multiple sets of numerical data
- Null Hypothesis ‘The difference between percentages of protected species is a result of chance’
- Results show that it is not always a result of chance
- We can conclude some species are more likely to be endangered than others
- P-value of reptiles to mammals ~0.038

Recommendations

For conservationists who are concerned about endangered species, it is statistically more likely for some species to be endangered than others.

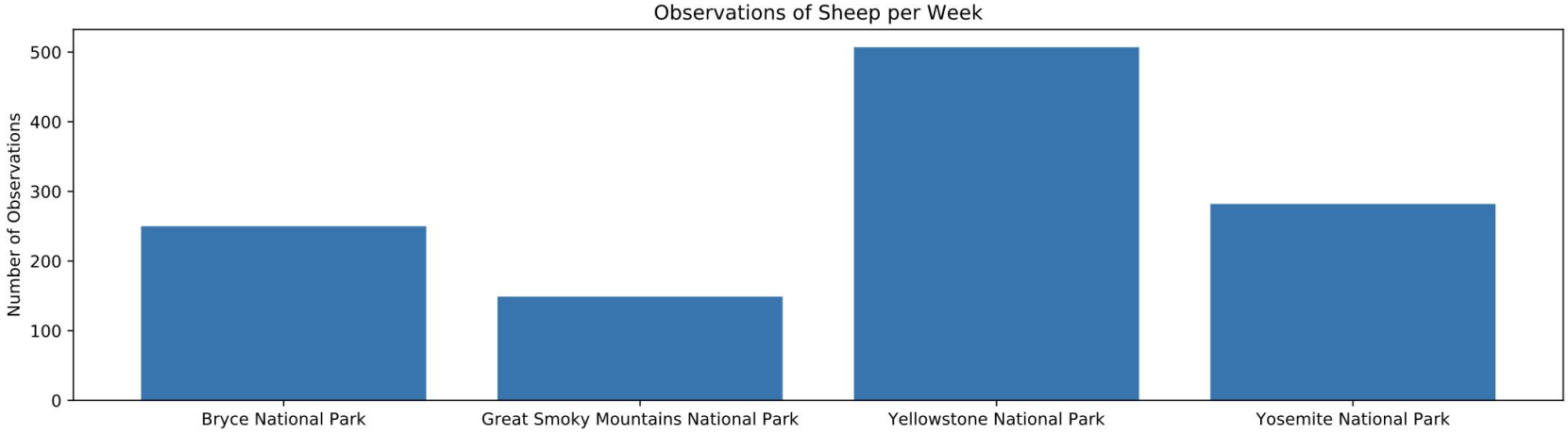
Before creating plans to help protect species, it would be beneficial to run more analyses to check which ones should be focused on first.

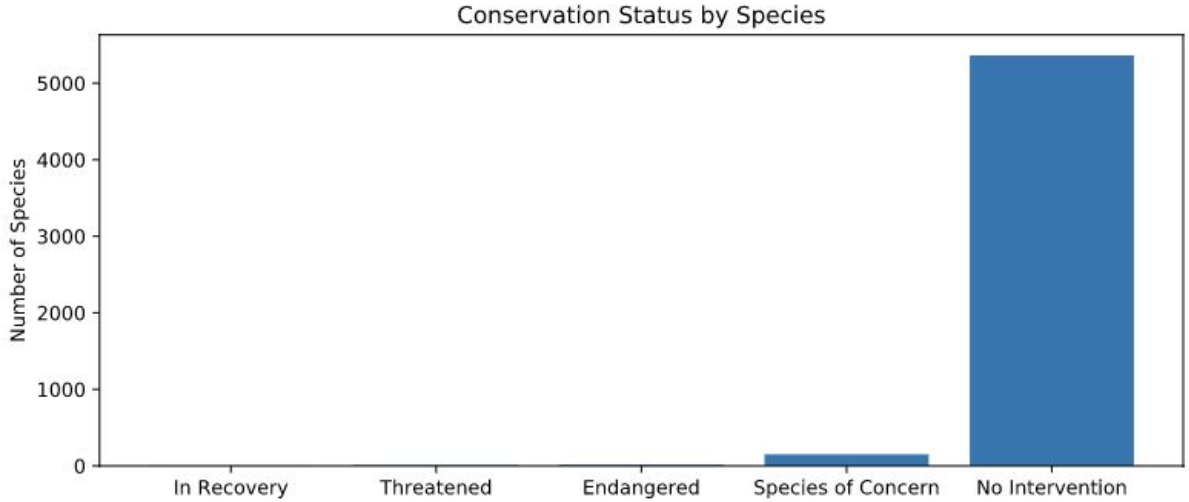
Sample Size Determination

Calculations used for Sample Size

- Baseline percentage from prior studies = 15%
- We want to be able to detect a minimum change of 5 percentage points
- Using the formula provided ‘Minimum Detectable Effect’ = 33.3%
- Finally using the provided calculator the amount of time needed to observe the required sample size of 870:
  - Yellowstone National Park = 1.72 weeks
  - Bryce National Park = 3.48 weeks

Graphs


Observations of Sheep per Week
Number of Observations
0
100
200
300
400
500
Bryce National Park
Great Smoky Mountains National Park
Yellowstone National Park
Yosemite National Park


Conservation Status by Species
Number of Species
0
1000
2000
3000
4000
5000
In Recovery
Threatened
Endangered
Species of Concern
No Intervention