Introduction to Data Analysis

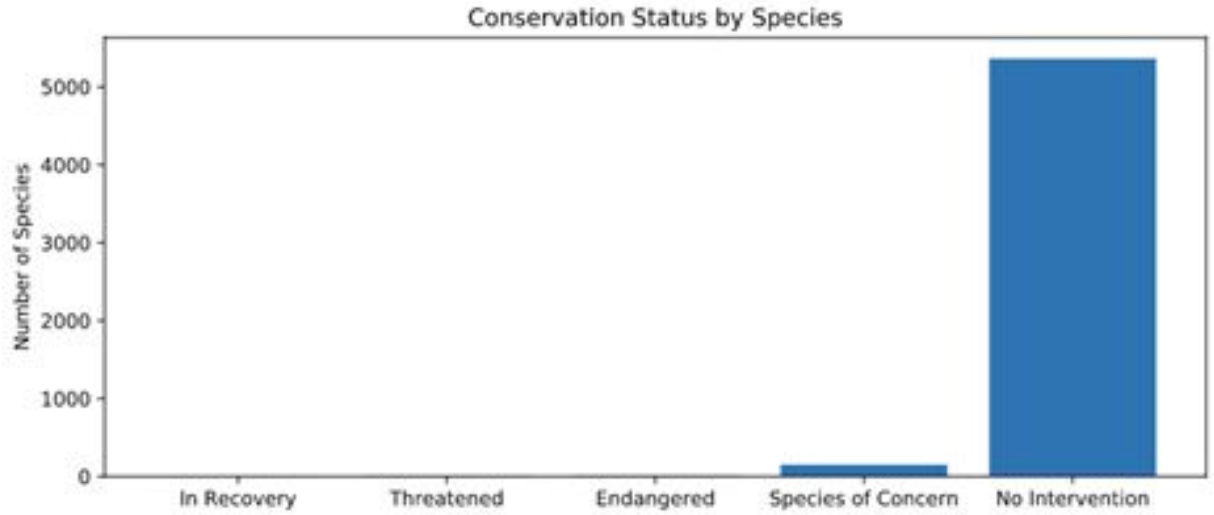
Capstone project option 2: Biodiversity for the National Parks

Katherine Duerden

Description of species data

There are 5541 different species identified, which fall into 7 types; Mammal, Bird, Reptile, Amphibian, Fish, Vascular and Nonvascular Plants. Their conservation status ranges from ‘Species of Concern', 'Endangered', 'Threatened', 'In Recovery' and those without any conservation status.

As shown in the graph below the majority of species (5363 of 5541) do not require an intervention.



Significance calculations

I found the percentage of each species type which were endangered was highest for mammals (17%) and birds (15%), and lowest for Vascular (1%) and Nonvascular plants (1.5%).

I carried out further analysis to test whether the difference in endangered status between the different species types was due to chance or significant, using a chi-squared test.

14 comparisons of species types were carried out, which involved putting the two variables into a contingency table, and calculating the p-value.

I found p-value of less than 0.05, which indicate a significant difference, in 2 cases.

Recommendation

My analysis of the data showed that certain types of species more likely to be endangered.

I set a p-value of 0.05 to determine if a difference was significant.

- Comparing the endangered status of Mammals and Reptiles, the p-value was ~0.038356
- Comparing Amphibians and Nonvascular plants, the p-value was ~0.001779

In 12 chi squared tests I carried out the p-value was above ~0.05 and so these differences can be ascribed to chance.

I recommend that the conservation of mammals and amphibians should be prioritised, as my analysis shows these species types are more likely to be endangered.

Sample size determination

Observations from the previous year showed foot and mouth disease in 15% of sheep at Yellowstone park. Using this as the baseline percentage, I found that a sample size of **510** would be needed to observe a 5% reduction in the disease.

Using the data on observations of sheep per week in each park, as shown in the graph below, I found that 1 week of observations are needed at Yellowstone Park to detect an effect. It will take 2 weeks to detect an effect at Bryce Park.