Capstone Option 2: Biodiversity for the National Parks

By: Shaun Sapp

The Data

- I have been given a CSV file, species_info.csv with data about different species in our National Parks, including:
- The scientific name of each species
- The common names of each species
- The species conservation status
- There are 5541 types of species in the file
- The types of species listed: 'Mammal' 'Bird' 'Reptile' 'Amphibian' 'Fish' 'Vascular Plant' 'Nonvascular Plant'
- The types of conservation statuses listed are: 'Species of Concern' 'Endangered' 'Threatened' 'In Recovery‘ and None
- ‘In Recovery’ is the least populated status while ‘None’ is the most populated status

Significance Calculations

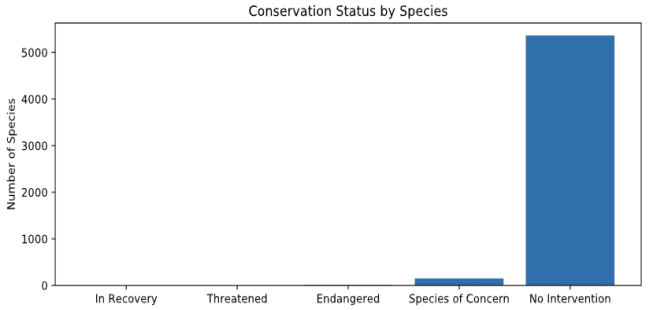
•Reformatted the data to streamline the viewing of the information

•Performed a chi squared test to see the pvalue for the hypothesis “are certain types of species more likely to be endangered?". Our **null hypothesis** here is that this difference was a result of chance

•Some of the answers to that question are that Mammals and Birds but when we ran our chi-squared test, we found a p-value of ~0.688 which indicated they are not significantly different in terms of likelihood of endangerment

•However, Reptiles and Mammals show a significant difference, ie: Mammals are more likely to be endangered than Reptiles. When we compared the percentages of protected reptiles and mammals and ran the same chi-squared test, we calculated a p-value of ~0.038, which **is** significant

•Therefore, we can conclude that certain types of species *are* more likely to be endangered than others



conservation_status	scientific_name
In Recovery	4
Threatened	10
Endangered	15
Species of Concern	151
No Intervention	5363

Recommendation to conservationists: While more attention should be spread to species that have a higher risk you should not ignore the reptiles/amphibians/plants that have lower risks to their lives and species

Foot and Mouth Disease Study

- Park Rangers at Yellowstone National Park have been running a program to reduce the rate of foot and mouth disease at that park. The scientists want to test whether or not this program is working. They want to be able to detect reductions of at least 5 percentage points. For instance, if 10% of sheep in Yellowstone have foot and mouth disease, they'd like to be able to know this, with confidence.
- Using the Sample Size Calculator:
- Given a baseline of a 15% occurrence of foot and mouth disease in sheep at Bryce National Park, I have found that if the scientists wanted to be sure that a >5% drop in observed cases of foot and mouth disease in the sheep at Yellowstone was significant then they would have to observe at least 870 sheep.
- Then, using the observation data I analyzed earlier, I have found that this would take approximately one week of observing in Yellowstone to see that many sheep, or approximately two weeks in Bryce to see that many sheep.

The Remaining Graph