Biodiversity Capstone Project Investigating Endangered Species

Annabel Neff - September 2018

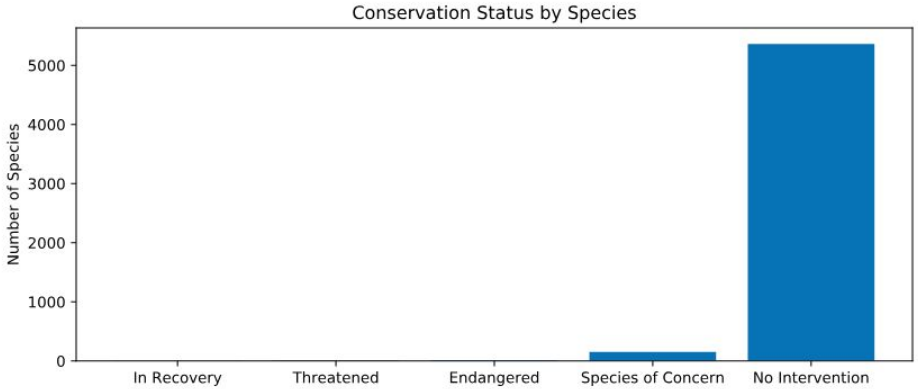
Inspecting Species DataFrame

Species DataFrame Information

- Table species contains the following rows:
  - Index
  - Animal category
  - Scientific name of animal
  - Common name of animal
  - Conservancy status
- Species_info.csv is a comma separated value file
- There are 5,541 unique species in the file
- The different conservation statuses define the level of endangerment for each species
  - **Threatened**: vulnerable to endangerment in the near future.
  - **Endangered**: seriously at risk of extinction.
  - **In Recovery**: formerly Endangered, but currently not in danger of extinction throughout all or a significant portion of its inhabitable range.
  - **No Intervention:** No data available

Species by Conservation Status

- This chart shows that we do not have data regarding conservation status for 5,363 of the species accounted for, as shown in the bar with conservation status “No Intervention”
- Outside of species with no intervention status, the largest subset of conservation status is “Species of Concern”, with 151 species accounted for
- The smallest subset of conservation status is “In Recovery” with 4 species accounted for



Protected vs. Non-protected Categories

- Mammals are the most likely to become endangered because 17% of their species are not being intervened upon
- Birds are the second most likely to become endangered with 15% with no intervention
- Vascular Plants are the least likely to become endangered because only 1% of their species have not been intervened upon

index	category	not_protected	protected	percent_protected
0	Amphibian	72	7	8.86%
1	Bird	413	75	15.37%
2	Fish	115	11	8.73%
3	Mammal	146	30	17.05%
4	Nonvascular Plant	328	5	1.50%
5	Reptile	73	5	6.41%
6	Vascular Plant	4216	46	1.08%

Hypothetical Testing: Are some species more likely to be endangered than others?

Are certain types of species more likely to be endangered?

- To test whether one type of species is more likely to be endangered than another, a hypothesis test is required
- The information being tested are two types of categorical data
- The most appropriate type of test is a Chi-Squared test

Are Mammals more likely to be endangered than Birds?

- To the right is a contingency table comparing Mammals and Birds organized by whether or not they are protected
- Using the Chi-Squared Test this data can be compared to determine a p-value
- If the resulting P-value is greater than 0.05, there is no significant statistical difference and the null hypothesis is true

	Protected	**Not Protected**
Mammal	30	146
Bird	75	413

0.69

The P-value is greater than 0.05 which means the difference between a Mammal and a Bird being endangered is due to chance.

Comparison between Bird and Reptile Endangerment

Further Chi-Squared Tests can determine if additional combinations of animals are statistically related

Based on the contingency table below and the results of a Chi Squared Test, Mammals are statistically more likely to be endangered than Mammals.

The P-value of this test is 0.038, which is lower than 0.05 and therefore significant.

	Protected	**Not Protected**
Reptile	30	146
Mammal	75	413

Chi-Squared Test on all available Species Data

When comparing all data, is there a significant likelihood that a species will be endangered and not another?

Below is a contingency table for all data available on our species.

The P-value of a Chi-Squared Test performed on this data is 5.51082804731e-89, which is lower than 0.05 and therefore significant.

	Protected	Not Protected
Amphibian	7	72
Bird	75	413
Fish	11	115
Mammal	30	146
Nonvascular Plant	5	328
Reptile	5	73
Vascular Plant	46	4216

Are certain types of species more likely to be endangered?

Based on the Chi-Squared Tests I performed on the data available, yes some animals are more likely to be endangered than others.

Recommendations for Endangered Species

- Identify all species that are statistically likely to become endangered using Chi-Squared Test
- Focus on reducing number of endangered species for those more statistically likely to become endangered
- Lower the number of Non Intervention Species, starting with animal categories with the largest amount of non-intervened upon species

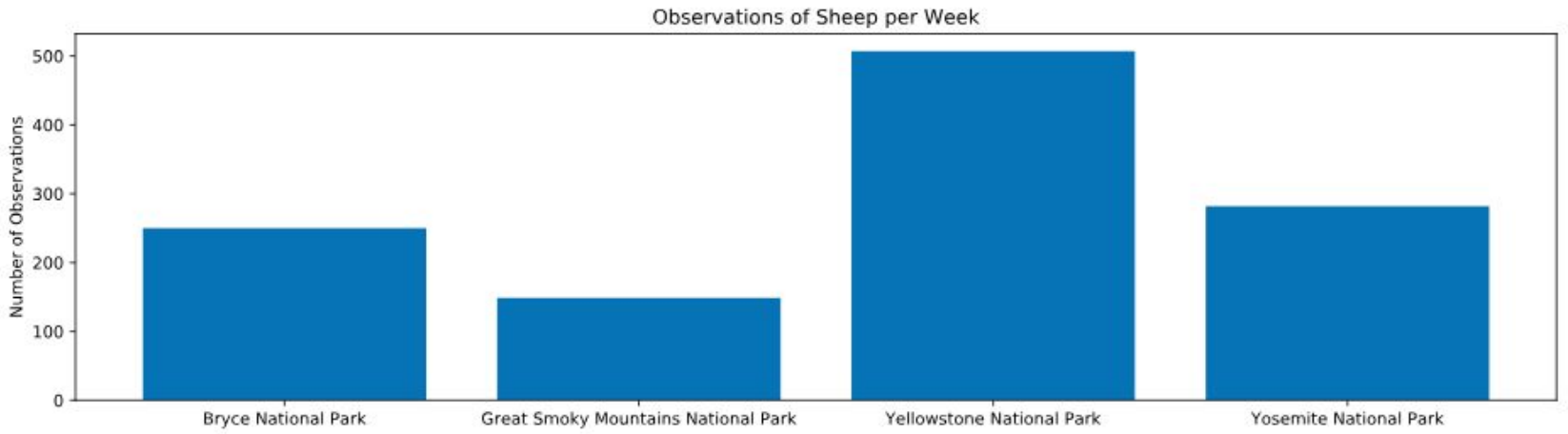
Foot and Mouth Disease Study: Sample Size Determination

Observations and Species DataFrames

- We received data on species observations throughout several National Parks over the past 7 days
- To further investigate Foot and Mouth Disease, I narrowed the data down to Sheep sightings

index	scientific_name	park_name	observations	category	common_names	conservation_status	is_protected	is_sheep
0	Ovis canadensis	Yellowstone National Park	219	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	TRUE	TRUE
1	Ovis canadensis	Bryce National Park	109	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	TRUE	TRUE
2	Ovis canadensis	Yosemite National Park	117	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	TRUE	TRUE
3	Ovis canadensis	Great Smoky Mountains National Park	48	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	TRUE	TRUE
4	Ovis canadensis sierrae	Yellowstone National Park	67	Mammal	Sierra Nevada Bighorn Sheep	Endangered	TRUE	TRUE

Sheep Sightings Grouped by National Park



Foot and Mouth Reduction Effort Data

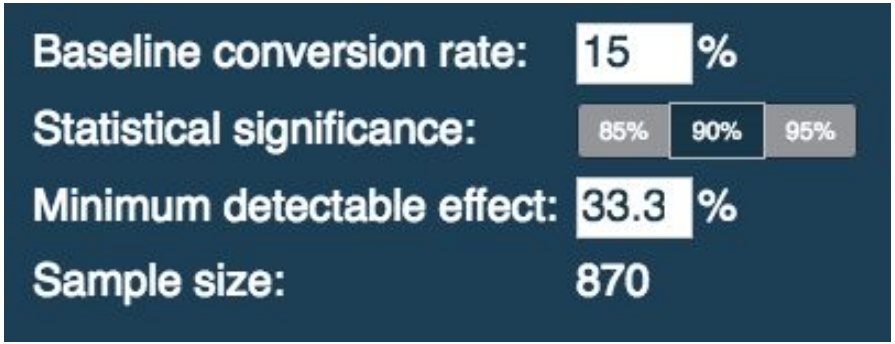
Park Rangers at Yellowstone National Park would like to test the success of their Foot and Mouth Reduction Effort. To do so, I will use the information available to me to create a Sample Size Determination.

What we know: 15% of sheep at Bryce National Park have foot and mouth disease.

	park_name	observations
0	Bryce National Park	250
1	Great Smoky Mountains National Park	149
2	Yellowstone National Park	507
3	Yosemite National Park	282

Determining the Sample Size

Using a sample size calculator (pictured below) and the data available, I determined the sample size should be 870 sheep.



Final Conclusions for Sample Sizes

Timeline for testing success rates

- It will take about 1.7 weeks to reach the sample size and determine success in the Yellowstone National Park.
- It will take about 3.5 weeks to reach the sample size and determine success in the Bryce National Park.

Thank you!