Biodiversity for the National Parks

By Nagasuryarama Vegesna

Species Information

- The data frame contained information from the National Parks Service
- It the data was arranged in columns of Category, Scientific name, common name and, conservations status
- Unique category = Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant and, Nonvascular Plant
- Unique Scientific Names = Endangered ,In Recovery, Species Of Concern, Threatened and, No Intervention
- Data was collected for 5541 unique species
- In the category group, some of the mammals were repeated with different conservation status which didn’t make sense
- When the observation and species data frames were combine it made sense as data was collected form different parks.

Significant Calculations for species

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

	category	not_protected	protected	percent_protected
0	Amphibian	72	7	0.088608
1	Bird	413	75	0.153689
2	Fish	115	11	0.087302
3	Mammal	146	30	0.170455
4	Nonvascular Plant	328	5	0.015015
5	Reptile	73	5	0.064103
6	Vascular Plant	4216	46	0.010793

Percent Protected	Category
17 %	Mammal
15 %	Bird
8.8%	Amphibian
8.7%	Fish
6.4%	Reptile
1.5%	Nonvascular Plant
1.0%	Vascular Plant

- The chart to the left shows us species categorized based on their conservation status.
- No Intervention means that they are not at risk and don’t need protection
- While Endangered, In recovery, Species of Concern and, Threatened need protection

- The chart in the middle show us the different category of species and what percent of them need protection
- The table to the right ranks the species category based on percent protected

Chi-Square Test

- We tested our null hypothesis that the difference between percent protected was due to chance.
- Using the Chi-Square test we found that the P-value for birds and mammals was 0.68 which is not significant and was a result of chance
- The same for reptiles and mammals resulted in a P-value of 0.038 which significant and show that mammals are more likely to be endangered than others.
- For Amphibians and Fish , the p-value was 0.824 which shows that is due to chance.

Calculations for Sheep Observed

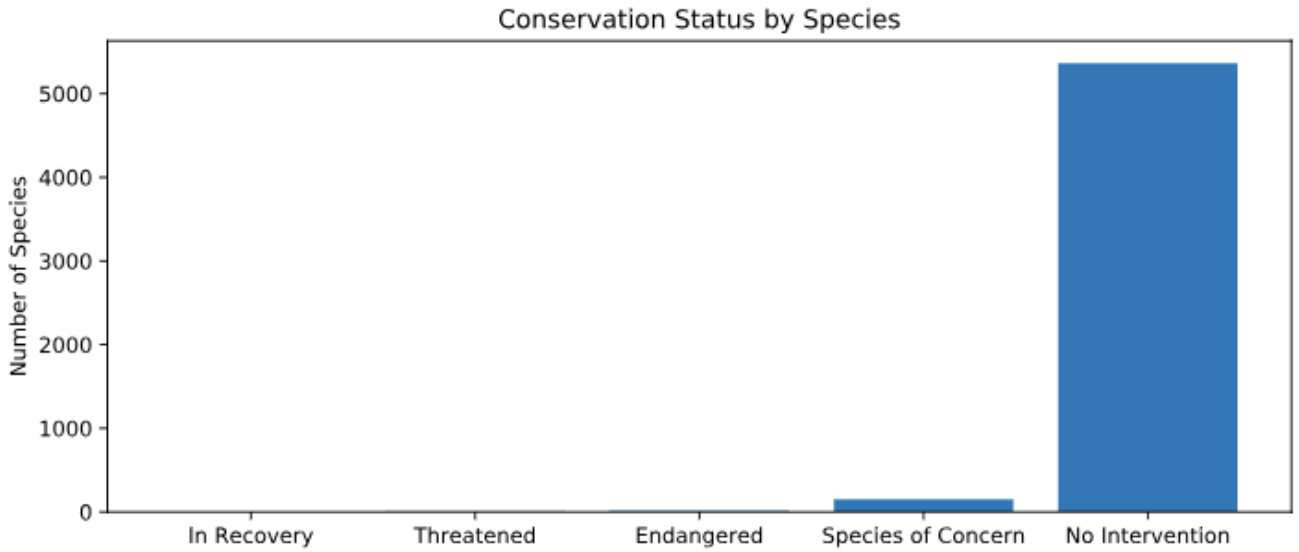
	park_name	observations
0	Bryce National Park	250
1	Great Smoky Mountains National Park	149
2	Yellowstone National Park	507
3	Yosemite National Park	282

- The data in the table show the number of observations for different mammal sheep species that have been observed for a week

Sample Size for Foot and Mouth Disease

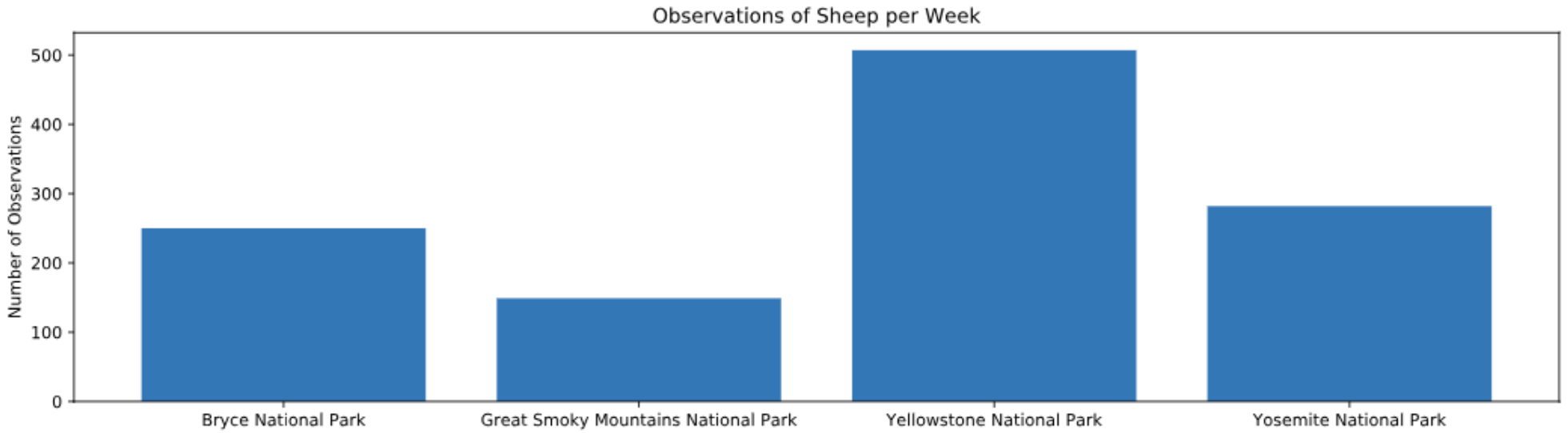
- The baseline for the for conversion rate was 0.15 as that was the percent of sheep with foot and mouth disease last year.
- We wanted the find a reduction of 5 percent point this year which would give us 0.33 minimum detection rate
- We are looking for a 90 % significance which would result in a sample size of 890
- For this sample size, it would take us about 1.75 weeks to observe sheep in Yellowstone and 3.56 weeks at Bryce National Park

Species in grouped by Conservation status



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

Sheep observed per Week in different parks



	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