Capstone Project: Biodiversity for the National Parks

Prepared by: Scott Ferry

The Data

The CSV file species_info.csv contains data on a wide range of plant and animal life including:

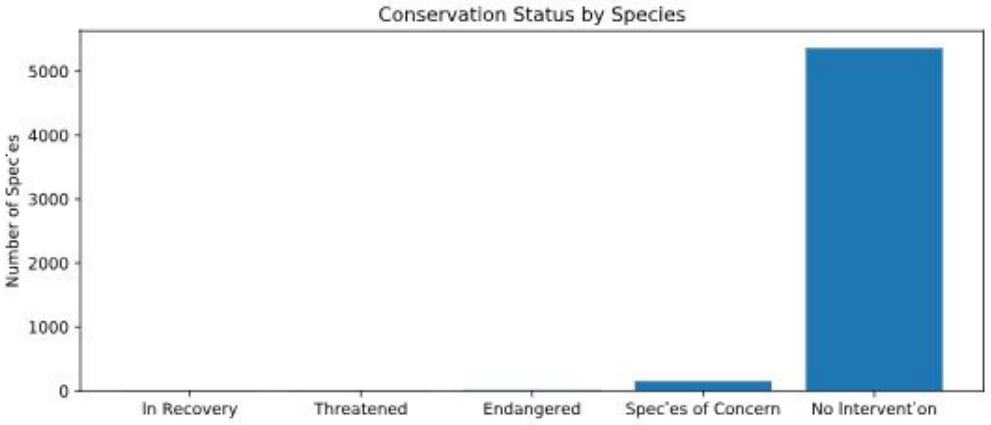
- Category (mammal, bird, reptile, fish, plant)
- Scientific names
- Common names
- Conservation status

From this dataset, we are able to use a wide variety of analysis techniques to gain insights and make inferences about the data provided.

The Data pt2

Once the dust settles and the smoke clears around our data we can easily chart the conservation status by species. That is, how many species fall under the different protection categories.

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



Tests of Significance

After some analysis and data manipulation we were able to calculate what percentage of species fell into the category of protected. We can use this data to answer the question “Are certain types of species more likely to be endangered?”

category	not_protected	protected	percent_protected
Amphibian	73	7	0.087500
Bird	442	79	0.151631
Fish	116	11	0.086614
Mammal	176	38	0.177570
Nonvascular Plant	328	5	0.015015
Reptile	74	5	0.063291
Vascular Plant	4424	46	0.010291

We can compare two species using the Chi Square test. Our null hypothesis is that the differences (in Endangered status’) is due to chance. A P-Value (pval) of less than 0.05 will indicate a significant difference and a rejection of the null hypothesis.

Mammal/Bird Chi Square test pval = 0.688 We can accept the null hypothesis that the difference is due to chance
Reptile/Mammal Chi Square test pval = 0.038 We can reject the null hypothesis that the difference is due to chance

Recommendations

Since we know that certain species are more likely to be endangered than others we can recommend that conservationists focus their efforts on groups that are more “high risk” than others, for instance Mammals and Birds.

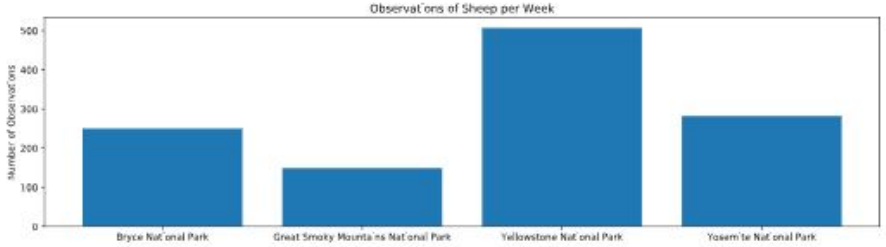
It should be said, however, that the vast majority of animal species fall into the category of “No Intervention Required”. Hopefully our conservationists stay vigilant and hopefully over the years we can see that number rise, not fall.

In Search of Sheep!

In the second part of this project we used a data frame of animal observations across several different National Parks. We sorted through this data go gather information specifically about sheep.

- First we sorted observation data specific to sheeps
- Second we broke the sheep observation data out to which National Parks these observations came from.

	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

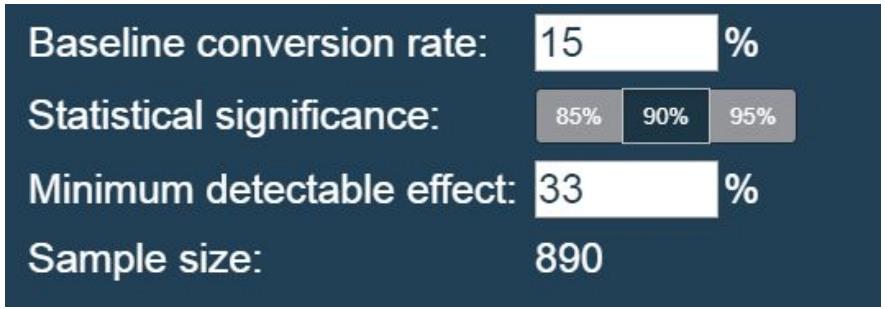


- Lastly, we were tasked with taking this data to figure out a sample size determination for a disease study that is being planned

Foot and Mouth Disease Study

Rangers at Yellowstone park want to see the if a disease reduction program is effective. To make this determination they need to know how many sheep need to be observed in the study to make a inference of the results. Here is what they DO know:

- **15%** of sheep at Bryce National Park have Foot and Mouth disease. This figure will become our **baseline**
- The rangers would like to use the standard 90% statistical significance or confidence interval
- The rangers at Yellowstone want to be able to detect a difference or reduction of at least 5%. Our **Minimum Detectable Effect** would therefore be 0.05/0.15 OR 33%
- Using the sample size calculator and this data the rangers at Yellowstone would need to observe 890 sheep in order to collect enough data (observations) to determine if their program to reduce Foot and Mouth Disease is, in fact, effective.



Foot and Mouth Disease Study pt 2

We can take the previous slide data even further and determine just how long it would take to make the observations required for the study for Yellowstone Park and Bryce National Park

- Yellowstone can make 507 observations per week. At that rate it would take 1.755 weeks to gather enough observations for the study
- Bryce National Park can only make 250 observations per week and therefore would take 3.56 weeks to gather the 890 observations required for the study.