Biodiversity Investigation for National Park Service

Gulcin Altinok

Understanding the data from species_info.csv

4 different sets of data is stored in this file:

1. **Category:** 7 different categories (Mammal, bird, reptile, amphibian, fish, vascular plant, nonvascular plant)
2. **Scientific name:** 5541 different species of animals
3. **Common names:** 5504 different common names of animals
4. **Conservation status:** 5 different conservation status (nan, species of concern, endangered, threatened, in recovery)

***Conservation status: Only a small portion of the whole species need some sort of a protection. 180/5541 -> (3,24%). Most of the species have no intervention (5363). This means that most of the species are not species of concern, endangered, threatened or in recovery.

Significance calculations to determine whether certain types of species more likely to be endangered?

To be able to check whether certain type of species have significantly different protection rates from each other therefore different chances to be endangered, we can perform significance tests on the protection values of two different categories.

is_protected	category	not_protected	protected	percent_protected
0	Amphibian	73	7	0.087500
1	Bird	442	79	0.151631
2	Fish	116	11	0.086614
3	Mammal	176	38	0.177570
4	Nonvascular Plant	328	5	0.015015

Ho:There is no significant difference between the percentage protection values of birds and mammals

Birds and mammals seem to be more vulnerable as the protection_percentage seems to be higher for these two categories, compared to others. These two categories can be an interesting start to compare the protection rates of different categories.

Test Choice: Chi 2 Test

- Data is categorical, not numerical
- We are comparing two pieces of data

Result: p-value=0.68 > 0.05 -> we cannot reject Ho. Therefore we conclude that the difference between the protection rates of mammals and birds are due to chance.

Recommendation for conservationists.

Mammals vs. reptiles (2nd comparison category): Same Ho is applied to mammals and reptiles, p-value=0.03 < 0.05, we now reject the Ho. This means that mammals have a significantly higher protection rate than reptiles, and hence more likely to be endangered.

Therefore we know that certain type of species are more likely to be endangered than others. Conservationists keep this in mind while determining where to focus their studies and effort more. E.g. they should give more importance to mammals and species needing protection there compared to reptile category.

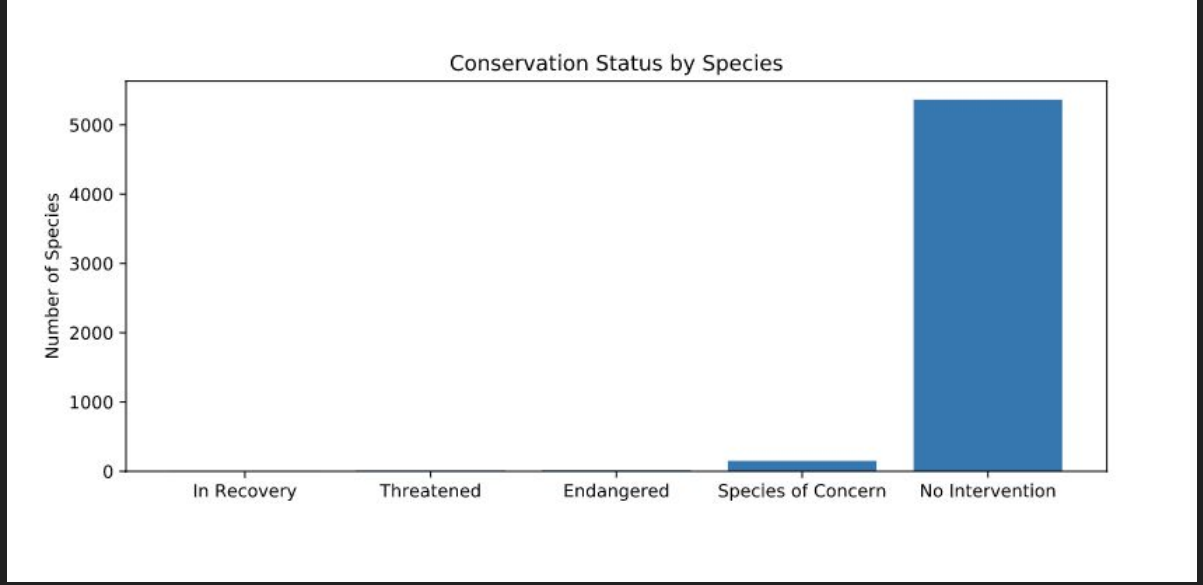
****To track for example a mammal species, such as sheep, to observe their overall counts and health, they can observe Yellowstone National Park (provides the greatest number of observations/week), Yosemite, Bryce and and Great Smokey Mountain's national parks respectively.*

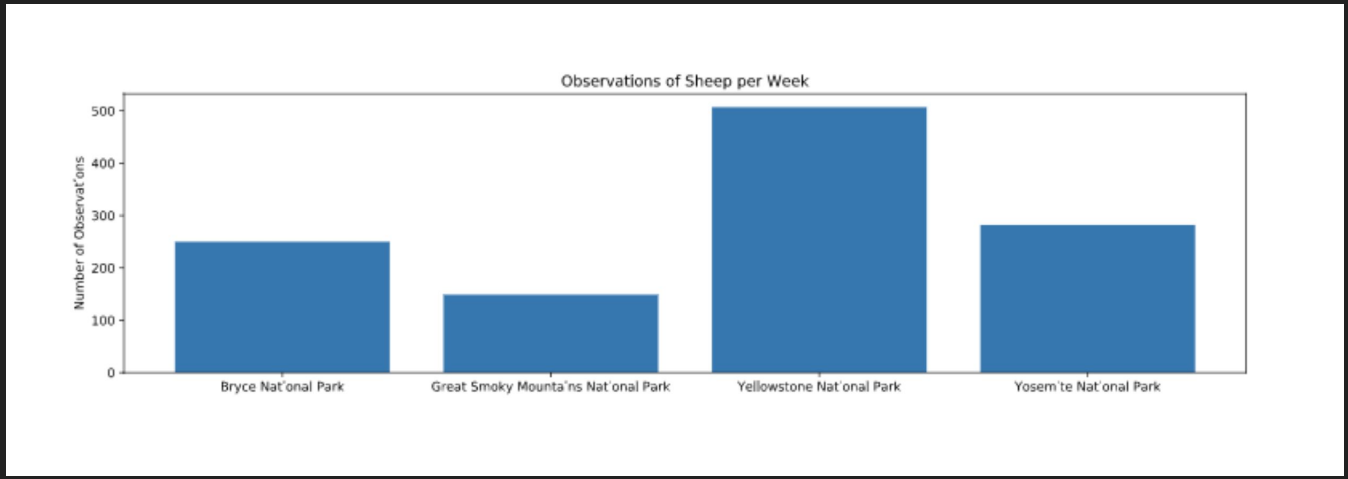
Recommendation Sample Size Determination for Foot and Mouth Disease

In order to determine the sample size for foot and mouth disease study, we need to determine what the baseline conversion, statistical significance we're looking for in the test results and minimum detectable effect of the study (lift) are.

- **Baseline conversion:** 15%, as we know 15% of sheep at Bryce National Park had foot and mouth disease in last year's recordings
- **Statistical significance** set as default 90%
- **Minimum detectable effect:** Scientists want 5 percent point reduction from the current rate of animals with the disease. (5/15)*100=33.3%
- **Sample size:** Depending on these variable we concluded that the amount of sample needed for this study to show 90% significant results of 5% reduction in the disease rate, we need to have 870 samples.

This number can be reached in approximately (870/507) 1,5 week in Yellowstone park, whereas in Bryce Park we need to wait (870/250) around 3,5 weeks


Conservation Status by Species
Number of Species
0
1000
2000
3000
4000
5000
In Recovery
Threatened
Endangered
Species of Concern
No Intervention


Observations of Sheep per Week
Number of Observat'ons
0
100
200
300
400
500
Bryce Nat'onal Park
Great Smoky Mounta'ns Nat'onal Park
Yellowstone Nat'onal Park
Yosem'te Nat'onal Park