



BIODIVERSITY FOR THE NATIONAL PARKS

Capstone assignment by Sanne Deurloo

Details about the data to be analyzed from species_info.csv

- The National Parks Service liked to have some data analysis done on the conservation statuses of the species and to investigate if there are any patterns or themes to the types of species that become endangered. Available data includes:
 - The scientific name of each species
 - The common names of each species
 - The species conservation status
- Amount of species: 5541
- The different categories present in this dataset: Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant and Nonvascular Plant
- The different conservation statuses present in this dataset: Species of Concern, Endangered, Threatened, In Recovery and Null values which were replaced by No intervention

Conservation statutes explained and division

- Explanation of the different conservation statuses:
 - Species of Concern: declining population or appears to be in need of conservation
 - 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
- Division of the 5541 species in the dataset:

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

Are certain types of species more likely to be endangered?

- As one can see from the table below, some categories have a higher percentage of protected species than others, the mammal stands out with 17% of its species being protected, followed by the bird with 15%.
- The vascular plant has the lowest percentage of protected species, only 1%.

	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

But is there a significant difference?

- One can see that it looks like mammals are more likely to be endangered than birds
- What we want to test is the following: is this difference only due to chance? (the 'null hypothesis')
- We can test this with a chi-squared test, because we want to compare two categories: mammals and birds. For this we have to create a contingency table:

	Protected	Not-protected
Mammal	30	146
Bird	75	413

- Running the test the p-value is 0.69, which is not statistically significant

Continued

- What about the difference between mammals and reptiles?

	Protected	Not-protected
Mammal	30	146
Reptile	5	73

- There we see a p-value of 0.04, which is statistically significant, meaning that reptiles are more endangered compared to mammals, which is not only due by chance

Conclusion

- We can conclude that certain types of species are more likely to be endangered than others.
- This means that it is important for conservationists concerned about endangered species, to notice this statistically significant result.
- Apparently not all species get the same treatment and this should be changed to preserve these wonderful species.
- As you can see this little guy does not agree with this approach:

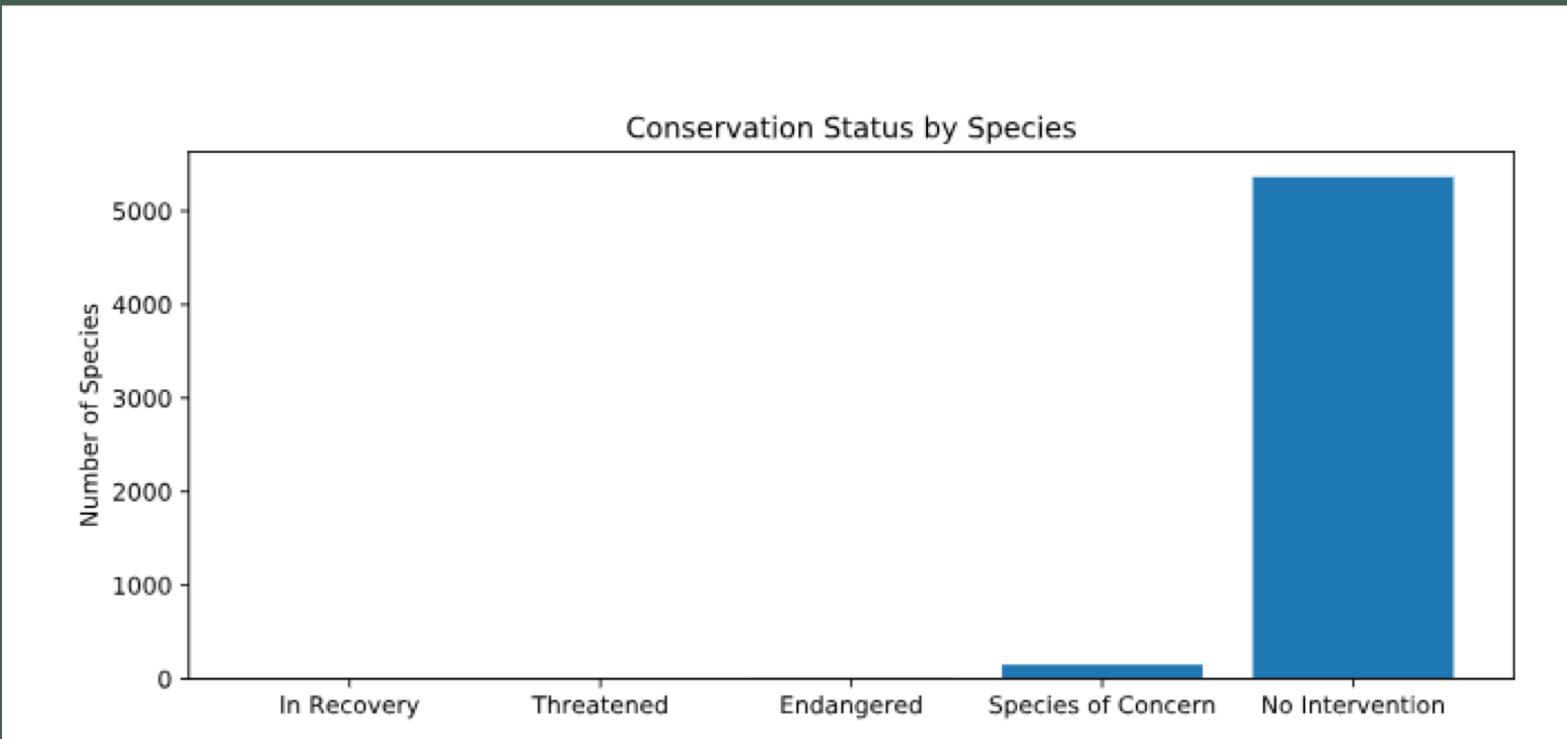


Sample size determination - foot and mouth disease study

- Baseline percentage = 15% (last year it was recorded that 15% of sheep at Bryce National Park have foot and mouth disease)
- Minimum detectable effect = 33.3 ($100 * 5 / 15$)
- Sample size per variant = 870 (calculated in the sample size calculator)
- Yellowstone weeks observing = 1.7 ($870 / 507$)
- Bryce weeks observing = 3.5 ($870 / 250$)

Appendix

Conservation status by species



Observations of sheep per week

