Biodiversity Investigation

Bowen Ya

Section I Endangered Species Study

Data provided: species_info.csv

lists all the species that can be found in the National Parks, and specifically indicates the endangerment level of these species

Sample data:

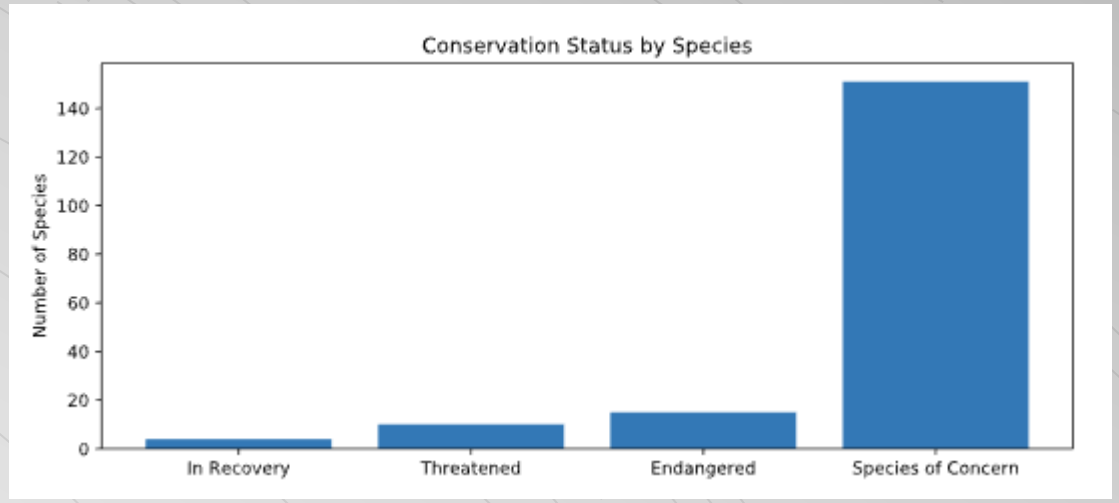
	category	scientific_name	common_names	conservation_status
0	Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	NaN
1	Mammal	Bos bison	American Bison, Bison	NaN
2	Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Dom...	NaN
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	NaN
4	Mammal	Cervus elaphus	Wapiti Or Elk	NaN
5	Mammal	Odocoileus virginianus	White-Tailed Deer	NaN
6	Mammal	Sus scrofa	Feral Hog, Wild Pig	NaN
7	Mammal	Canis latrans	Coyote	Species of Concern
8	Mammal	Canis lupus	Gray Wolf	Endangered
9	Mammal	Canis rufus	Red Wolf	Endangered

Data Interpretation

5 animal categories
- Mammal
- Bird
- Reptile
- Amphibian
- Fish

2 plant categories
- Vascular Plant
- Nonvascular Plant

5541 different species ▶ 180 among them need attention



	conservation_status	scientific_name
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?

	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
5	Reptile	74	5	0.063291
6	Vascular Plant	4424	46	0.010291

We grouped each category by the conservation status of the species, and ran chi-squared tests to different pairs of categories.

Our null hypothesis: The difference between the protection ratios of two categories are due to chance

Chi-Squared Test Results

The p-value of test between Mammal and Bird is `0.687594809666` which means we fail to reject the null, and there are no significant difference between these two categories.

However, the p-value of test between Reptile and Mammal is `0.0383555902297` and we are confident to reject the null hypothesis. Mammals are more likely to be endangered than reptiles!

Recommendation for conservationists:

After running more tests between different pairs of categories of species, we got the conclusion that categories that would significantly more easily be endangered are (in descending order):

Mammal
Bird
Amphibian

Pay closer attention on these categories would be a wise choice!

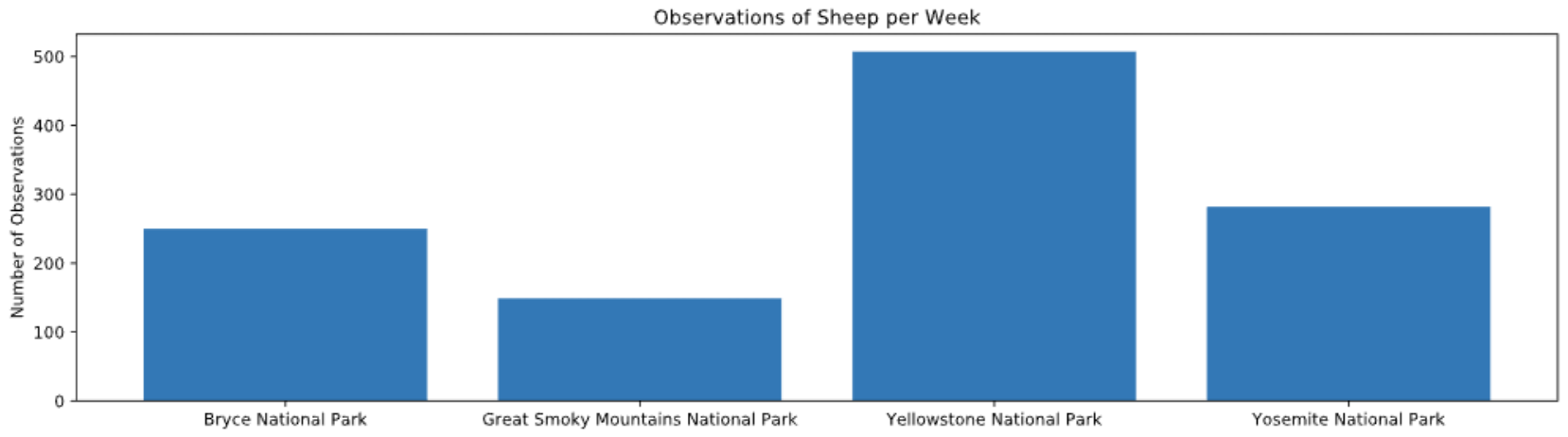
Section II A Study on Foot and Mouth Disease Reduction Effort

Expected to see a 5% Reduction of foot and mouth disease in a certain National Park

In order to do that with confidence, we need a valid sample size

Observations of Sheep in National 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



With our given rate, the sample size calculator shows that a sample size of 870 would give us 90% confidence of our 5% foot and mouth reduction detection.