Introduction to Data Analysis

Capstone Option 2: Biodiversity for the National Parks

Data in species_info.csv

- 5541 types of species
- 5 types of species: mammal, bird, reptile, amphibian, fish, vascular plant, nonvascular plant
- Each species assigned a conservation status: Endangered, In Recovery, No Intervention, Species of Concern, Threatened

Conservation status	Number of species
Endangered	15
In Recovery	4
No Intervention	5363
Species of Concern	151
Threatened	10

Significance calculations for endangered status between different categories of species

- Since data is categorical and 2 pieces of data are being compared, we use a chi-squared test to determine significance

Significance calculations for endangered status between mammals and birds

- Contingency table:

	Protected	Not protected
Mammal	30	146
Bird	75	413

- Using chi2_contingency from scipy.stats: p-value = 0.68759 > 0.05
- No significant difference

Significance calculations for endangered status between mammals and reptiles

- Contingency table:

	Protected	Not protected
Mammal	30	146
Bird	5	73

- Using chi2_contingency from scipy.stats: p-value = 0.03836 < 0.05
- Difference is significant!

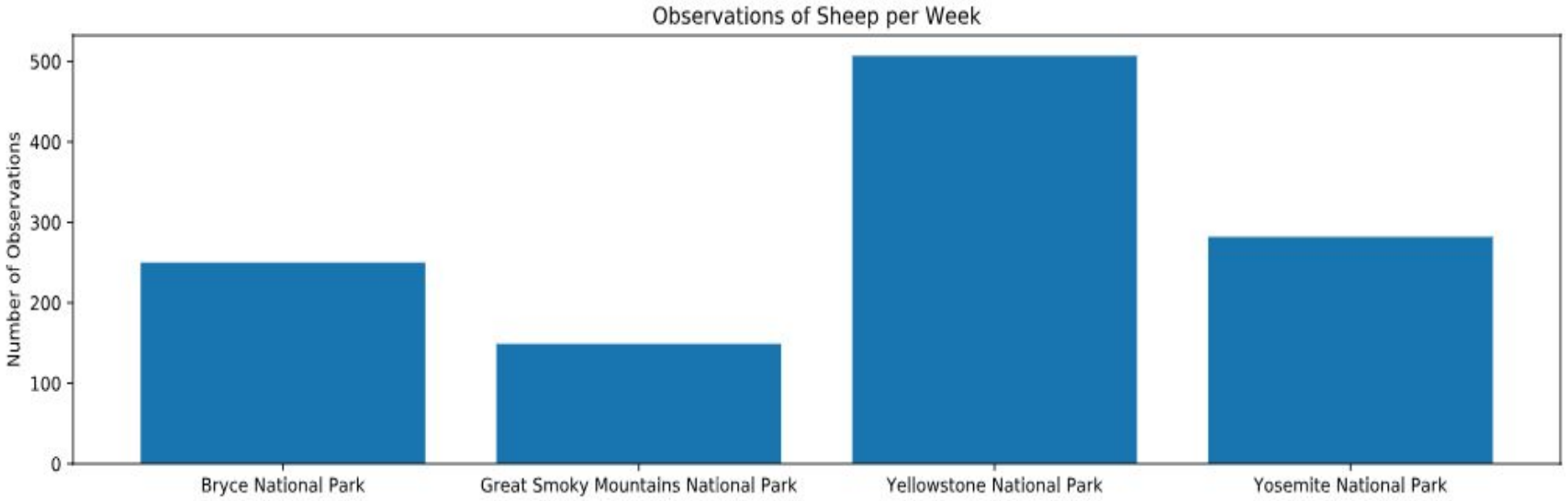
Recommendation for conservationists concerned about endangered species

- Since difference between mammals and reptiles is significant, we can conclude that certain species are more likely to be endangered than others.
- Conservationists should give more attention towards species that are more likely to be endangered.

Sample size determination for food and mouth disease study

- Since last year it was recorded that 15% of sheep at Bryce National Park have foot and mouth disease, baseline conversion rate = 15%
- Since we want to be able to detect reductions of at least 5%, minimum detectable effect = 5/15*100% = 33.3%
- Using sample size calculator, sample size per variant = 870

Graph: Observations of Sheep per Week



Graph: Observations of Sheep per Week