codecademy PRO / Introduction to Data Analysis

Biodiversity Capstone Project

by Werner Wasmayr

1. Data in „*species_info.csv*“

The data in the file '*species_info.csv* contains **5541** species.

There are seven types of species: 'Mammal', 'Bird', 'Reptile', 'Amphibian', 'Fish', 'Vascular Plant' and 'Nonvascular Plant'

In the conservation status there are four values present when beginning tthe analysis: 'Endangered', 'In Recovery', 'Species of Concern' and 'Threatened'

2. Significane calculations for endangered status between different categories of species

Category	Not Protected	Protected	Percent Protected
Amphibian	72	7	0,088608
Bird	413	75	0,153689
Fish	115	11	0,0873
Mammal	146	30	0,170455
Nonvascular Plant	328	5	0,015015
Reptile	73	5	0,064103
Vascular Plant	4216	46	0,010793

3. Recommendation for conservationists concerned about endangered species

The question was: "are certain types of species more likely to be endangered?".

According our chi-squared test, the answer is **yes**! We calculated a p-value of ~0.038, which **is** significant.

That means, that certain types of species **are** more likely to be endangered than others.

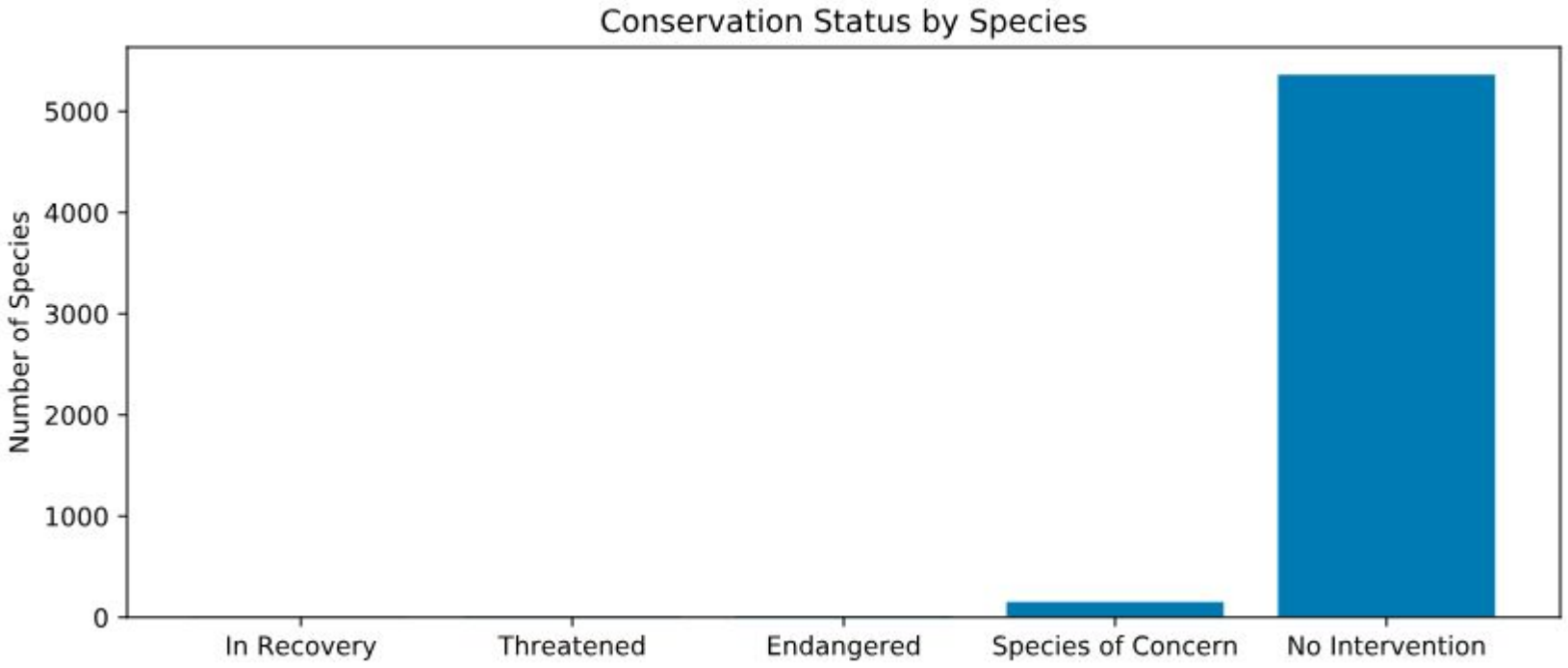
4. Sample size determination for the foot and mouth disease study

The sample size was calculated with the following parameters:

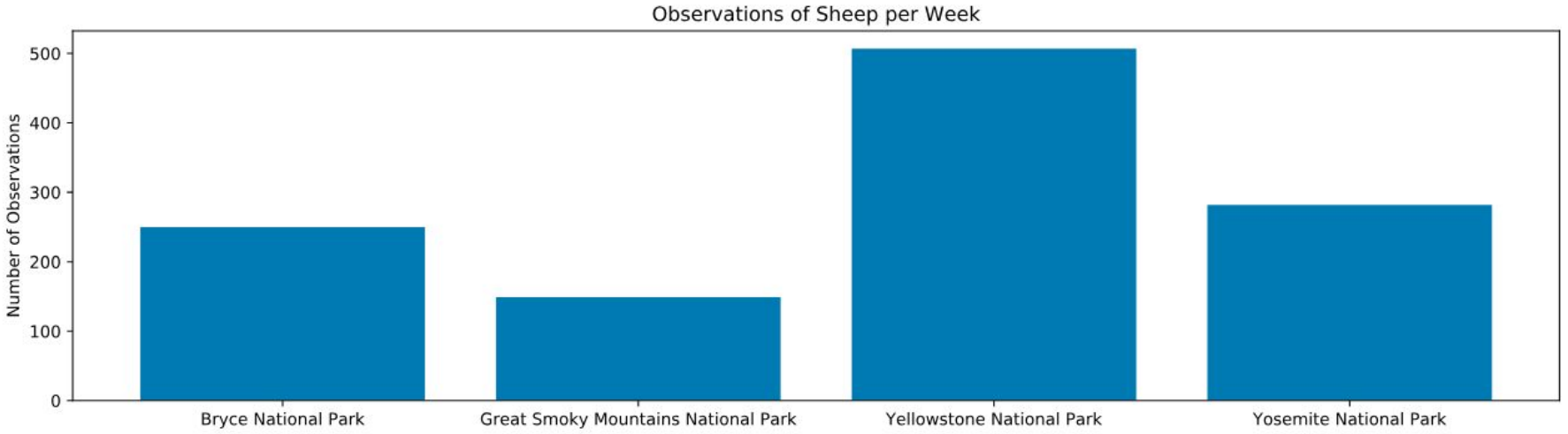
- Baseline conversion rate: 15%
- Statistical significance: 90%
- Minimum detectable effect: 33%

The result is a sample size of: **890**

5. Graphs


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

5. Graphs


Observations of Sheep per Week
Number of Observations
0
100
200
300
400
500
Bryce National Park
Great Smoky Mountains National Park
Yellowstone National Park
Yosemite National Park