Introduction to Data Analysis

Capstone Project: Biodiversity for the National Parks

Prepared by: Meg O'Neill

Species Data

Working with species_info.csv

- CSV contained a large amount of species data
- One could use the data to answer a lot of different questions relating to species' status in our National Parks
- Data included plant and animal species
- Majority of species included require 'No Intervention'
- Only 4 species 'In Recovery'

Endangered Status Significance

OBSERVATIONS

- Mammals (17.8%) and Birds (15.1%) had the highest percentages of protected species
- Vascular Plants had the lowest (1%)
- Amphibians and Fish had almost equal protected percentages (8.8% and 8.7% respectively)
- Overall, plants species (Vascular and Nonvascular) had far lower percentages of protected species (1% and 1.5% respectively)

SIGNIFICANCE FINDINGS

- To determine whether the difference between the **Mammal and Bird** percentages was significant, I performed a Chi-Squared Test using the numbers of 'protected' and 'not-protected' species from each category and found that the **difference was *not* significant** (with a pval of 68.8%)
- However, when comparing the difference between **Mammals (17.8%) and Reptiles (6.3%)** via another Chi-Squared Test, the returned pval of 3.8% indicated that the **difference between those categories of species *was* significant**

Significance calculations confirm that **certain types of species are more likely to be endangered.**

Given these findings, it is recommended that conservation efforts be focused on species in the **more vulnerable Mammal and Bird categories** when possible.

870

To determine the sample size of sheep that would need to observed from each park to ensure the foot and mouth percentages are significant, I first took the baseline of 15% that was observed in Bryce National Park and, in conjunction with the scientists' desire to see reductions of at least 5%, I was able to calculate the Minimum Detectable Effect (33%).

With that information at hand, and using the Sample Size Calculator, I was then able to determine the appropriate sample size of sheep (870) that would need to be observed at each park in order to make sure the foot and mouth percentages are significant.

Related Graphs

Conservation Status by Species

Observations of Sheep per Week