Codeacademy Capstone Project Option2: Biodiversity

Submitted by Gahgene Gweon

species_info.csv description

- Each row of the species_info file has 5 columns as follows;
    - Category: type of species, e.g. mammal
    - scientific_name: scientific name of each species
    - common_names: common name of each species
    - conservation_status columns: species conservation status
- Species_count = 5541
- Species_type = ['Mammal' 'Bird' 'Reptile' 'Amphibian' 'Fish' 'Vascular Plant' 'Nonvascular Plant']
- Conservation_statuses = [nan 'Species of Concern' 'Endangered' 'Threatened' 'In Recovery']
- conservation_status =
    - Endangered 15, In Recovery 4, no intervention 5363, Species of Concern 151, Threatened 10

Endangered status between different categories of species

```
        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
```

- Majority (~99%) of plants (nonvascular & vascular) are not being protected.
- Higher percentage of reptiles (6%), amphibian (8%) & fish (8%) are protected than plants.
- Birds (15%) & mammals (17%) are most protected species.

Recommendation for conservationists about endangered species

- Birds & Mammals are most protected at 15% and 17%, respectively (non-significant difference using chi-squared test, p=0.687)

- More mammals (17%) are protected compared to reptiles (6%) (chi-squared test, p=0.03)

- Our analysis shows that mammals and birds are the most endangered species, thus we recommend conservationists to focus on saving mammals & birds if they have to put priorities on the categories of endangered species due to economic/ time constraints.

Sample size determination for foot & mouth disease study

- baseline = 15
- minimum_detectable_effect = 100*5/15 =33.33%
- sample_size_per_variant = 870
- Therefore, scientists would have to spend approximately 2 weeks at yellowstone (yellowstone_weeks_observing = sample_size_per_variant/507) & 4 weeks at bryce (bryce_weeks_observing = sample_size_per_variant/250) to observe enough sheeps.

Graphs

Conservation Status by Species

Observations of Sheep per Week