

Capstone Option 2: Biodiversity for the National Parks

BY: JENIXON LIM

Source information

The source file is 'species_info.csv'.

The file contains data regarding different species in several National Parks.

The file records four different parameters for each species (see figure below)

1. Category – i.e. species type such as 'mammal' or 'reptile'
2. Scientific name
3. Common name
4. Conservation status

| | 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), Domesticated Cattle | nan |
| 3 | Mammal | Ovis aries | Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral) | nan |
| 4 | Mammal | Cervus elaphus | Wapiti Or Elk | nan |

Source information

There are 5541 different species in total.

There are seven unique species types (category):

1. Mammal
2. Bird
3. Reptile
4. Amphibian
5. Fish
6. Vascular plant
7. Nonvascular plant

There are five unique values of conservation status:

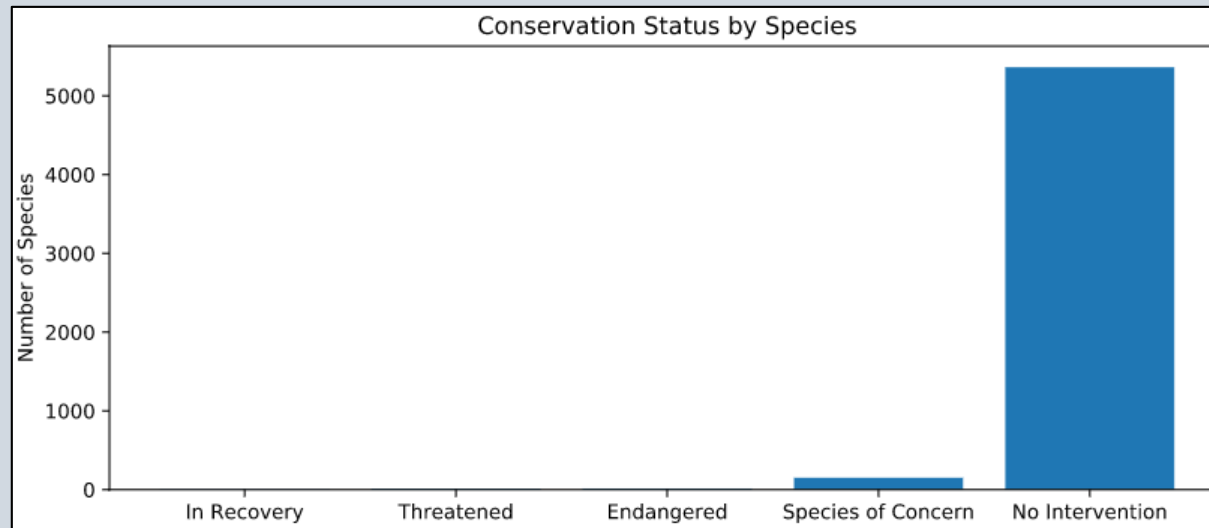
1. Species of concern
2. Endangered
3. Threatened
4. In recovery
5. 'Nan' or 'Null' – These null values indicate that the species is not under any protection status, thus, null values were renamed as 'No Intervention'

Species count by conservation status

A count of the different species was done according to their conservation status.

The majority of species (5363, 96.8%) are not under any form of protection (see table and graph below).

| Conservation Status | Number of Species |
|---------------------|-------------------|
| Endangered | 15 |
| In Recovery | 4 |
| No Intervention | 5363 |
| Species of Concern | 151 |
| Threatened | 10 |



Protection status by category (species type)

For each species type, a count was undertaken to determine the number of protected and unprotected species for each category.

The two categories with the highest percentage of protected species were mammals (17.05%) and birds (15.37%) indicating these two categories are the most likely to be endangered.

Both vascular and nonvascular plants had the lowest percentage of protected species (1.08% and 1.50% respectively) indicating that these were the least likely to be endangered.

| Category | Not Protected | Protected | Percent Protected |
|-------------------|---------------|-----------|-------------------|
| Amphibian | 72 | 7 | 8.86% |
| Bird | 413 | 75 | 15.37% |
| Fish | 115 | 11 | 8.73% |
| Mammal | 146 | 30 | 17.05% |
| Nonvascular Plant | 328 | 5 | 1.50% |
| Reptile | 73 | 5 | 6.41% |
| Vascular Plant | 4216 | 46 | 1.08% |

Chi-square test – Mammals and Birds

A chi-square test was done to determine if the differences in the likelihood of mammals (17.05%) and birds (15.37%) being endangered is significantly different.

The chi-square test is appropriate as the data is categorical and we are comparing two data sets.

The null hypothesis was that there is no difference between the likelihood of mammals and birds becoming endangered.

A contingency table (below) was used for the chisquare test.

| | Protected | Not Protected |
|--------|-----------|---------------|
| Mammal | 30 | 146 |
| Bird | 75 | 413 |

The resulting p-value was 0.6876, this is well above a 95% confidence level p-value of 0.05.

Therefore, mammals are no more likely to be endangered than birds.

Chi-square test – Mammals and Reptiles

The chi-square test was repeated to determine if there is a significant difference between the likelihoods of mammals (17.05%) and reptiles (6.41%) being endangered.

The null hypothesis was that there is no difference in the likelihood of mammals and reptiles becoming endangered.

The contingency table is as below:

| | Protected | Not Protected |
|---------|-----------|---------------|
| Mammal | 30 | 146 |
| Reptile | 5 | 73 |

The resulting p-value was 0.038. This is well within the 95% significance p-value of 0.05.

Therefore, mammals are more likely to be endangered than reptiles.

Recommendations

Based upon the chi-square tests of significance, it can be concluded that some species types are more likely to be endangered than others.

Therefore, conservationists should prioritise their efforts according to the likelihood of a species type being endangered.

The table below ranks from high priority to low priority species types according to their likelihood of being endangered.

| Priority | Category |
|----------|-----------------------------------|
| High | Mammal, bird |
| Medium | Amphibian, fish, reptile |
| Low | Nonvascular plant, vascular plant |

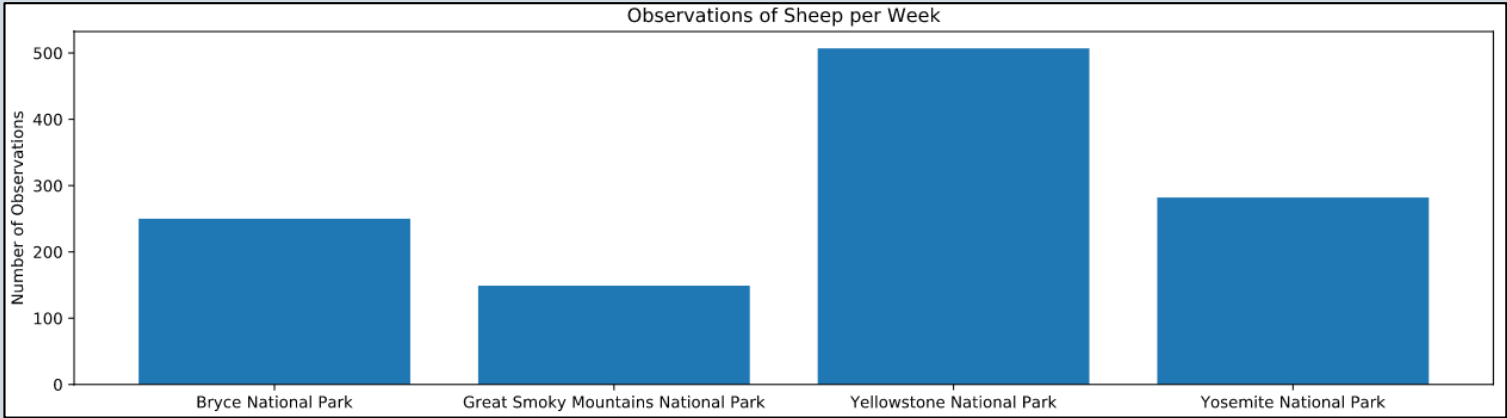
Sheep observations

There are three different species of sheep:

- 1. Ovis aries
- 2. Ovis canadensis
- 3. Ovis canadensis sierra

The total number of sheep observed each week in each national park is shown in the below table and graph.

| Park Name | Observations |
|-------------------------------------|--------------|
| Bryce National Park | 250 |
| Great Smoky Mountains National Park | 149 |
| Yellowstone National Park | 507 |
| Yosemite National Park | 282 |



Sample size determination – Foot and mouth disease study

The baseline conversion rate was 15%. This is the only historical information available, it is the percentage of sheep at Bryce National Park that had the foot and mouth disease.

As the rangers want to detect reductions of at least 5%, the minimum detectable effect would be $100 \times 5/15 = 33\%$.

The above numbers were fed into Optimizely's A/B Test Sample Calculator:
'<https://www.optimizely.com/sample-size-calculator/>'

The resulting sample size required was 520 samples.

Therefore, 520 was divided by the number of observations at each park to attain the required sample periods for each park. This is shown in the table below.

| Park Name | Observations | Sample Periods Required (Weeks) |
|-------------------------------------|--------------|------------------------------------|
| Bryce National Park | 250 | 2.08 |
| Great Smoky Mountains National Park | 149 | 3.49 |
| Yellowstone National Park | 507 | 1.03 |
| Yosemite National Park | 282 | 1.84 |