Biodiversity for the National Parks

Simon Koudelka

Dataset: Description

I have analyzed a dataset that contained information on endangered species, in particular:

- Category of the species
- Scientific name of the species
- Common names of the species
- Conservation status

Boiling the dataset down to find unique values gave us these **7 categories of the species**:

- Mammal
- Bird
- Reptile
- Amphibian
- Fish
- Vascular Plant
- Nonvascular Plant

Analysis of the **Conservation status** provided us with the following overview of the species. While most species have no intervention, only 4 were formerly Endangered, but recovered.

Endangered	15
In Recovery	4
No Intervention	5363
Species of Concern	151
Threatened	10
Total species in the dataset	5541

Again, this bar chart clearly shows that most of the species in the dataset require no intervention.

If we wanted to do a graphical analysis of the species that do require intervention, it would be best to exclude the 'No Intervention' category so that the differences between other categories would be more apparent.

Conservation Status by Species
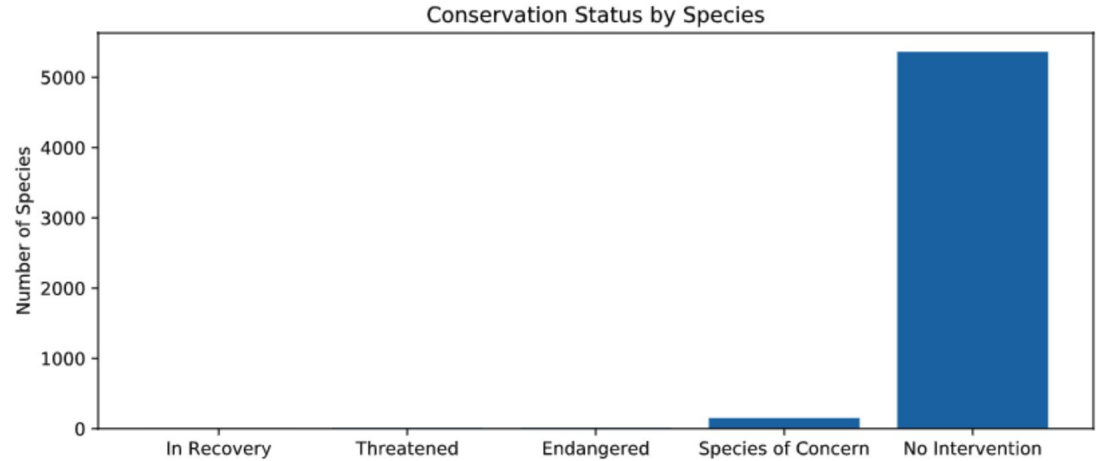
Significance Calculations

The following pilot table allows us to easily compare the percentage share of each species category that is protected.

We can observe that Vascular Plants are the least endangered, as only about 1% of them are under some form of protection. On the other hand, Mammals exhibit 17% protection rate, making them the most vulnerable group. Therefore, I would advise the park rangers to allocate the most resources to protect Mammals. However, in order to have certainty that this claim is correct, we would need to test significance (see next slide).

Alternatively, it could be the case that authorities of the National Parks have just paid more attention to Mammals at the expense of other species and thus making them the most protected category. In such case I would advise rangers to pay more attention to other species, too.

```
          category  not_protected  protected  percent_protected
         Amphibian             72          7           0.088608
              Bird            413         75           0.153689
              Fish            115         11           0.087302
            Mammal            146         30           0.170455
  Nonvascular Plant            328          5           0.015015
           Reptile             73          5           0.064103
     Vascular Plant    |      4216         46           0.010793
```

From the previous table, it seemed that e.g. Mammals were more endangered than Birds. To make sure this claim is true, I have tested for significance with a Chi-square test (H0=difference is a matter of chance)

I have compared these categories of species with respect to the numbers of protected and unprotected observations:

- Bird and Mammals -> p-value of 0.687594809666 -> we cannot reject H0 -> difference is not significant

and

- Reptiles and Mammals -> p-value of 0.0383555902297 -> we can reject H0 -> difference is significant

Therefore, we can claim that Mammals are indeed more endangered species than Reptiles. Hence authorities should pay more attention to Mammals than to Reptiles. On the other hand, we cannot claim that Mammals are more endangered than Birds due to low significance.

Sample Size Determination

Rangers are trying to decrease foot and mouth disease. To see whether the program is working, we want to see a change of at least 5 percentage points. We knew that last year 15% of sheep at Bryce National Park had foot and mouth disease, which determines our baseline. We use these two values to calculate the Minimum Detectable Effect = 100 * 0.05 / 0.15 = 33.3%

With the significance level at 90%, we had all the values to calculate the sample size that is required to tell whether the program has worked: 870 observations per variant. Then I calculated how long the test would need to run to collect the required observations based on the number of observations from the past (see graph below): in Yellowstone it would be about 1.7 weeks, in Bryce Park about 3.5weeks.

Observations of Sheep per Week