Biodiversity

by Aaron Wolf

species_info.csv

This was an interesting dataset that looked into the over 5500 species of wildlife ranging from a variety of categories. These species also varied in their conservation status. Just over 3% of the species in the study were *at least* of Species of Concern. (Under Step 4 there seems to be an anomaly where we're missing 2 species under "no intervention").

After the pivot table it was easy to determine that birds dominated the need for protection in raw numbers. In terms of percentage, however, most need was with mammals. However, the chi squared test determined that the difference isn't significant.

species_info.csv (continued)

In Step 6 we looked at Sheep. The findings were of sheep sightings in the past 7 days. We looked at protected species of sheep versus non-protected species of sheep, among other statistics. Most of the sightings were done in Yellowstone National Park, but there were significant sightings in other parks as well.

In the last exercise we figured out the minimum detectable effect as park rangers tried to decrease the population of sheep with Foot and Mouth Disease.

Significance Calculations

The vast majority of species didn't need any sort of intervention.
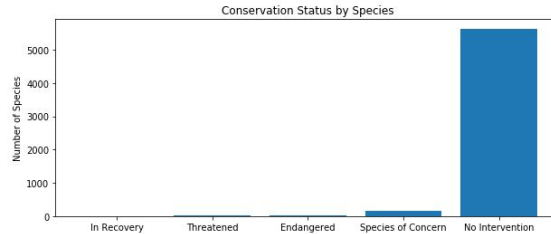
However, those in need of intervention had different spreads. This table breaks down the type of wildlife and their protected status.

	category	not_protected	protected
0	Amphibian	72	7
1	Bird	413	75
2	Fish	115	11
3	Mammal	146	30
4	Nonvascular Plant	328	5

Significance Calculations (cont.)

From table this we can calculate the percent of each category that is a protected status.

	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

Using the chi square test there is a significance between a few of these species. Namely, reptile and mammal, but not so much between bird and mammal.

```
#Chi Test between Mammal and Bird
contingency = [[146, 30],
               [413, 75]]

chi2, pval, dof, expected = chi2_contingency(contingency)
print(pval)

0.687594809666
```

Let's test another. Is the difference between `Reptile` and `Mammal` significant?

```
contingency2 = [[73,5],
                [146,30]]
chi2, pval, dof, expected = chi2_contingency(contingency2)
print(pval)

0.0383555902297
```

Conservation Recommendation

Based on these significance tests, it would be my opinion that the focus of conservation efforts be geared towards mammals and birds. They have the most need in terms of raw numbers and percentages. Moreover, the chi square test shows that there is no difference between them.

Reptiles are not in need of conservation as much based on all of the criteria discussed above.

Step 6: Sheep Sample Size

Sightings of sheep were recorded across 4 different national parks.

Step 1 was to find out which species were actually sheep, not just their names. We did this by using a lambda function to remove any rows in the dataframe that didn't include the word 'sheep' in the name, and made sure it was mammal.

Step 2 was to use the groupby function to break the different species of sheep into the different parks where they were seen.

Step 3 was to make a bar graph to compare the different observations in different places.

Step 4 was to determine the amount of time needed to observe the number of sheep in Yellowstone for foot and mouth disease. It's just over 1 week.

Graphs

Conservation Status by Species

Observations of Sheep per Week