Biodiversity for the National Parks

Capstone Option 2

Dan Earley

Data in species_info.csv

- The data within species_info.csv consisted of:
  - Category – 7 different categories of species
  - Scientific/Common Name – thousands of names identified, with a mix of duplicate entries for both columns. Unique counts were used to de-duplicate these lists.
  - Conservation Status – 4 pre-determined status types (endangered, in recovery, species of concern, threatened). Many species did not fit into any of these statuses, so we created a 5th type – no intervention – for those.
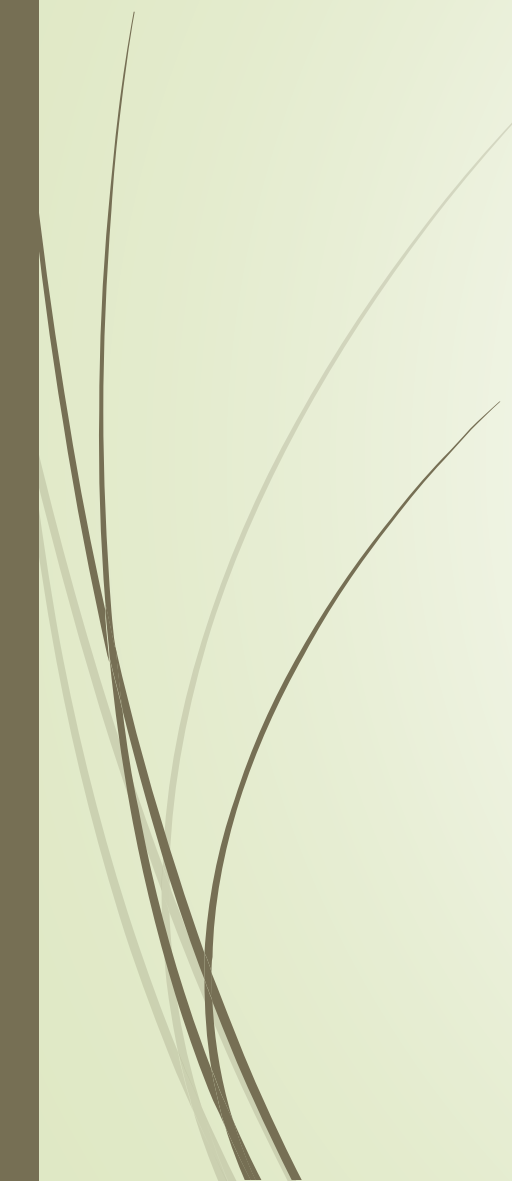
Significance calculations & recommendations

- At first glance, there seemed to be a pretty obvious trend that birds (15.37%) and mammals (17.05%) were more likely to be endangered than all other categories.

```
is_protected          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
```

- After running the chi-squared test, however, we found that the higher percentage was a result of chance (p-value of .688). This didn't apply to all species, however. When we tested the difference between reptiles and mammals, we found that the result was significant (p-value of .038), so some species types are more likely to be endangered than others.

- Recommendation would be to run significance tests on each pair of categories to determine the likelihood of endangerment, based on lowest p-value.

Sample size determination for Foot and Mouth Disease study

- Using existing data from Conservationist recorded sightings of species, in combination with observed results of F&M Disease from Park Rangers, we were able to determine the sample size needed to detect a 5% reduction in the disease at any given park.

- At 90% significance and a 1/3 observable effect from the baseline, we would need to observe 870 sheep from each park to test the results of the new program.

Graphs

	conservation_status	scientific_name
0	Endangered	15
1	In Recovery	4
2	No Intervention	5363
3	Species of Concern	151
4	Threatened	10

```
5541

['Mammal' 'Bird' 'Reptile' 'Amphibian' 'Fish' 'Vascular Plant'
 'Nonvascular Plant']

[nan 'Species of Concern' 'Endangered' 'Threatened' 'In Recovery']

   conservation_status  scientific_name
0            Endangered               15
1           In Recovery                4
2       No Intervention             5363
3    Species of Concern              151
4            Threatened               10
```
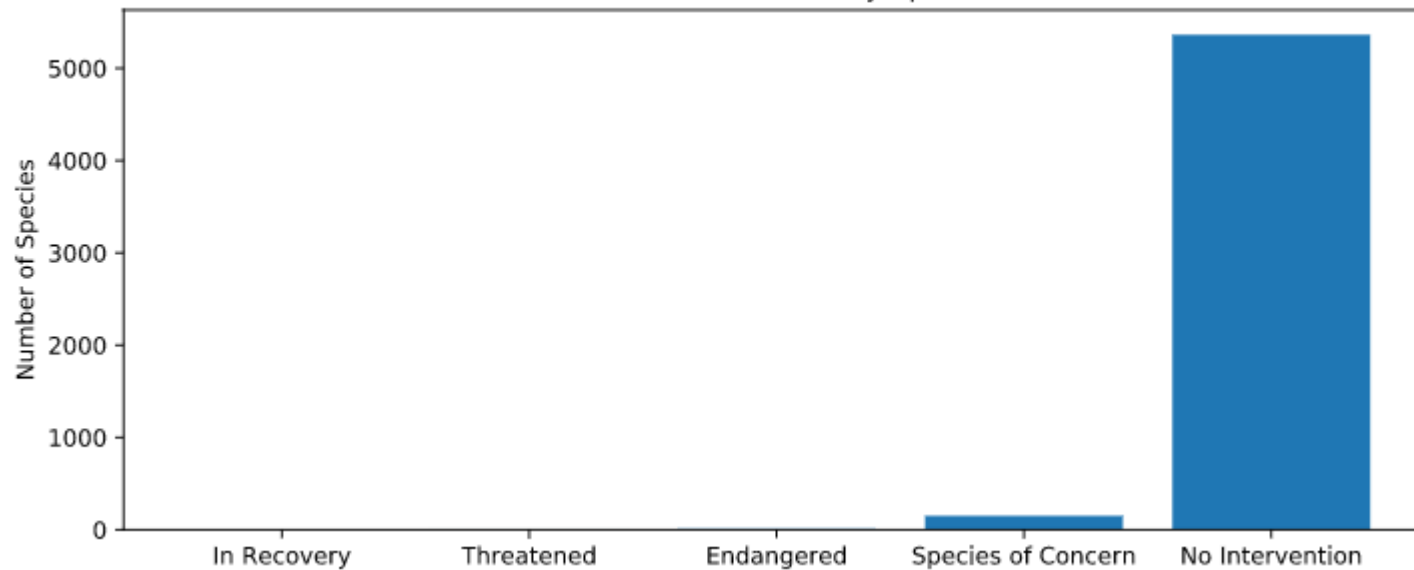

Conservation Status by Species

Graphs cont'd

is_protected	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

	category	scientific_name	common_names	conservation_status	is_protected	is_sheep	park_name	observations
0	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True	Yosemite National Park	126
1	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True	Great Smoky Mountains National Park	76
2	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True	Bryce National Park	119
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True	Yellowstone National Park	221
4	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True	Yellowstone National Park	219

	park_name	observations
0	Bryce National Park	250
1	Great Smoky Mountains National Park	149
2	Yellowstone National Park	507
3	Yosemite National Park	282

Graphs cont'd

Observations of Sheep per Week
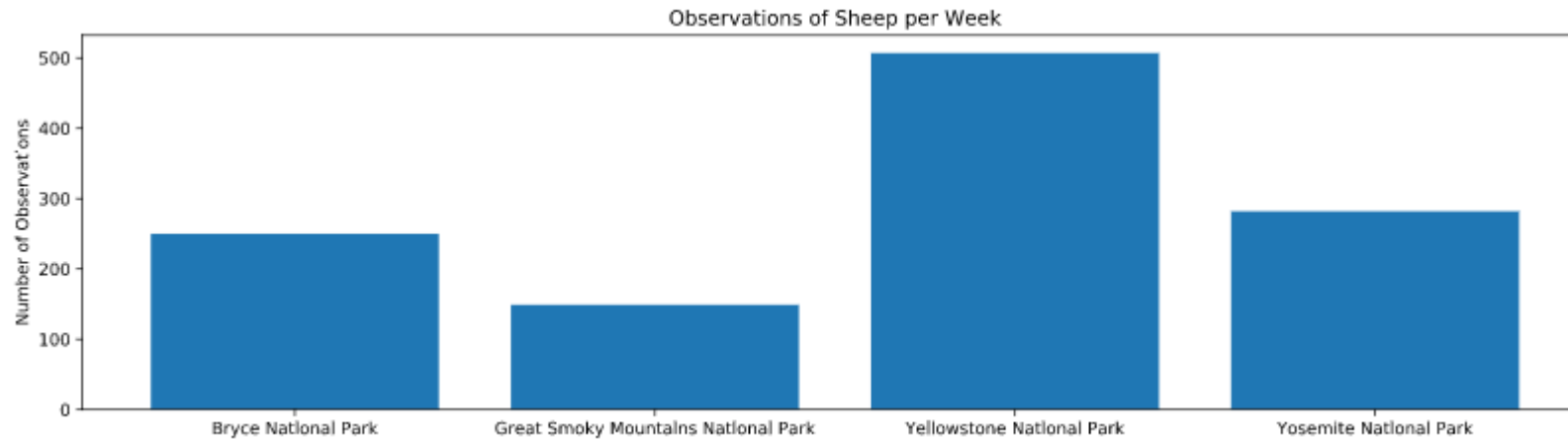
Baseline conversion rate: 15 %

Statistical significance: 85% 90% 95%

Minimum detectable effect: 33.33 %

Sample size: 870