BIODIVERSITY

analysing data on endangered species from national parks

stephan zobl

codeacademy • introduction to data analysis intensive • final capstone project

OVERVIEW

Mammal
Bird
Reptile
Amphibian
Fish
Vascular Plant
Nonvascular Plant

species category

data for

5541

species

Species of Concern
Endangered
Threatened
In Recovery

conservation status

CONSERVATION STATUS: A CLOSER LOOK

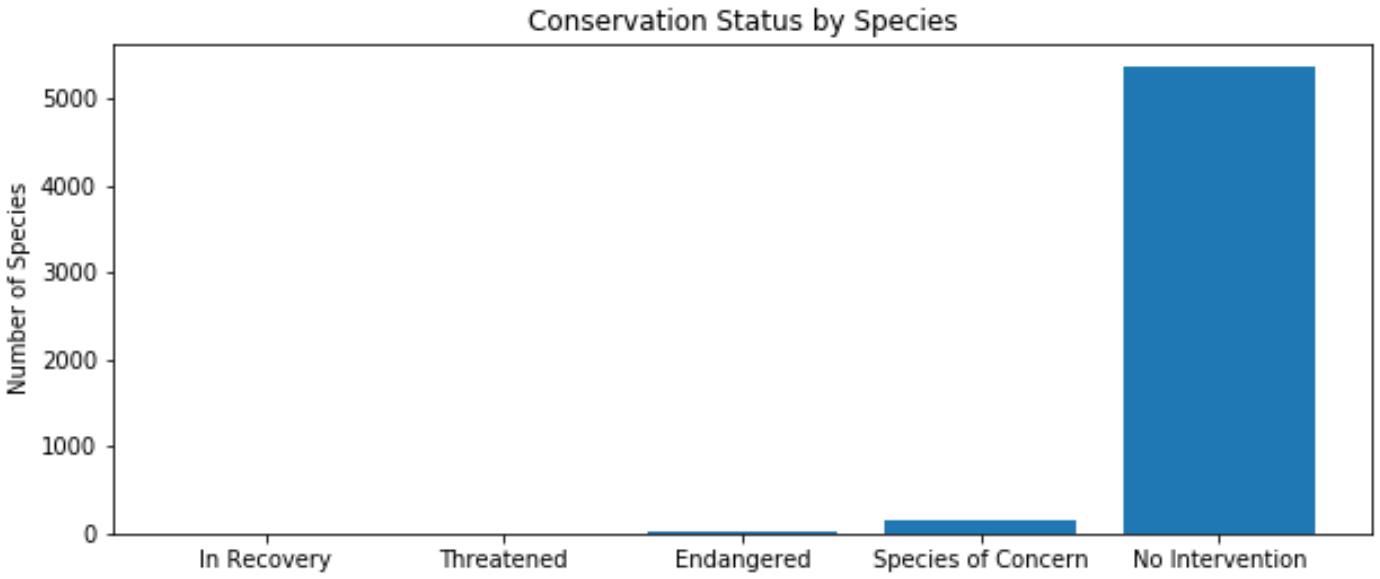
	conservation_status	scientific_name
0	In Recovery	4
1	Threatened	10
2	Endangered	15
3	Species of Concern	151
4	No Intervention	5363

species counts

No Intervention
Species of Concern
Endangered
Threatened
In Recovery

conservation status

CONSERVATION STATUS: A CLOSER LOOK



species counts

No Intervention
Species of Concern
Endangered
Threatened
In Recovery

conservation status

CONSERVATION STATUS: A CLOSER LOOK

not_protected
protected

protection status

No Intervention
Species of Concern
Endangered
Threatened
In Recovery

conservation status

CONSERVATION STATUS: A CLOSER LOOK

`not_protected`
`protected`

protection 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

protection status by species

QUESTION

are species in the **Mammal** category
more likely to become endangered than those
in the **Bird** category?

ENDANGERED STATUS: MAMMALS VS BIRDS

chi-square test

	protected	not_protected
Mammal	30	146
Bird	75	413

The difference between these values is random and not statistically significant; based on this data, mammals are not more likely to be endangered than birds.

null hypothesis

0.05
threshold

0.69
test result

p-value

The difference between these values is statistically significant; based on this data, mammals are more likely to be endangered than birds.

alternative hypothesis

QUESTION

are species in the **Mammal** category
more likely to become endangered than those
in the **Bird** category?

based on the data, this is unlikely to be true.

CONSERVATION STATUS: A CLOSER LOOK

`not_protected`
`protected`

protection 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

protection status by species

QUESTION

are species in the **Mammal** category
more likely to become endangered than those
in the **Reptile** category?

ENDANGERED STATUS: MAMMALS VS REPTILES

chi-square test

	protected	not_protected
Mammal	30	146
Reptile	5	73

The observed difference is random and not statistically significant; based on this data, mammals are not more likely to be endangered than reptiles.

null hypothesis

0.05
threshold

0.04
test result

p-value

The difference between these values is statistically significant; based on this data, mammals are more likely to be endangered than reptiles.

alternative hypothesis

QUESTION

are species in the **Mammal** category
more likely to become endangered than those
in the **Reptile** category?

based on the data, this is likely to be true.

CASE STUDY

program to reduce occurrence of foot and mouth disease in sheep at Bryce National Park and Yellowstone National Park

QUESTION

is this program working?

how many sheep do we need to test in order to confidently make predictions about the population as a whole?

how long will this take?

FOOT AND MOUTH DISEASE PROGRAM

duration

confidence

15%
with disease

current

sample size

5%
reduction

target

A/B Test Sample Size Calculator

Powered by Optimizely's Stats Engine

Baseline Conversion Rate

15 %

Your control group's expected conversion rate. [?]

Minimum Detectable Effect

33.3 %

The minimum relative change in conversion rate you would like to be able to detect. [?]

Statistical Significance

90%

EDIT

95% is an accepted standard for statistical significance, although Optimizely allows you to set your own threshold for significance based on your risk tolerance. [?]

Sample Size per Variation

510

Source: https://www.optimizely.com/sample-size-calculator/

FOOT AND MOUTH DISEASE PROGRAM

duration

90%
confidence

15%
with disease

current

510
observations

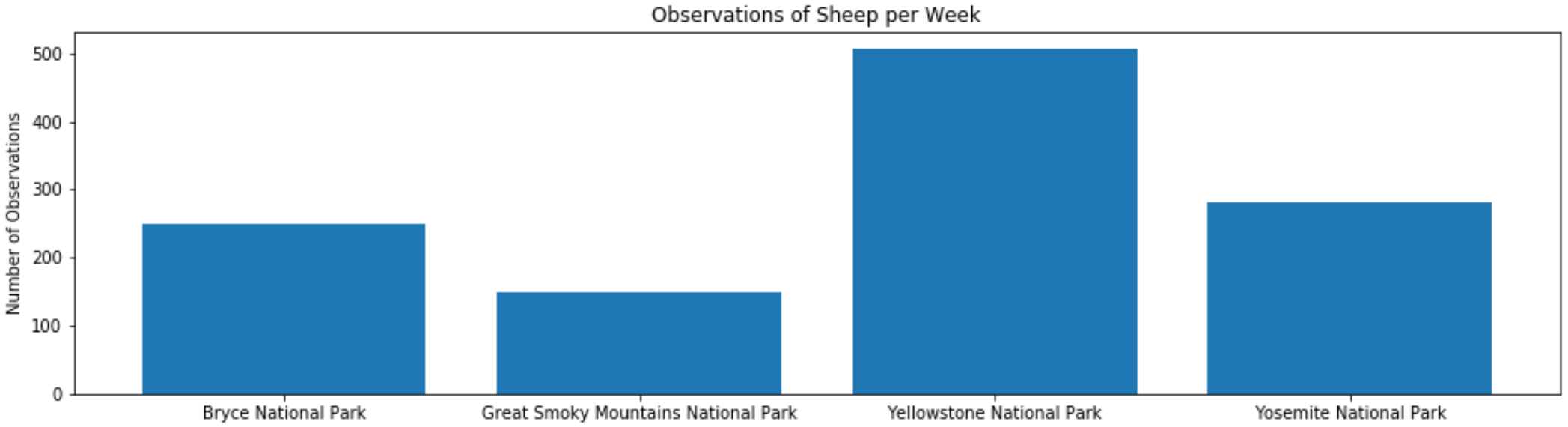
sample size

5%
reduction

target

FOOT AND MOUTH DISEASE PROGRAM

duration



FOOT AND MOUTH DISEASE PROGRAM

bryce national park:

2 weeks

of observation

duration

yellowstone national park:

1 week

of observation

	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

15%

with disease

current

510

observations

sample size

5%

reduction

target

QUESTION

is this program working?

in order to determine this, we need to sample at least 510 sheep

which, based on observation data, will take

at least 2 weeks at Bryce National Park

and at least 1 week at Yellowstone National Park.

FINDINGS & RECOMMENDATIONS

species in the **Mammal** category
are not more likely to become endangered
than those in the **Bird** category

species in the **Mammal** category
are more likely to become endangered
than those in the **Reptile** category

in order to determine this whether the program to reduce foot and mouth disease in sheep at Bryce National Park and Yellowstone National Park is working, we need to sample at least **510** sheep which, based on observation data, will take at least **2 weeks** at Bryce National Park and at least **1 week** at Yellowstone National Park

stephan zobl

stephanzobl@gmail.com

helloimstephan