Biodiversity for the National Parks

Codecademy – Introduction to Data Analysis – June 2018

jupyter

Data – species_info.csv

	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), Dom...	NaN
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	NaN
4	Mammal	Cervus elaphus	Wapiti Or Elk	NaN
⋮	⋮	⋮	⋮	⋮

Number – 5824 species, 191 with conservation status listed

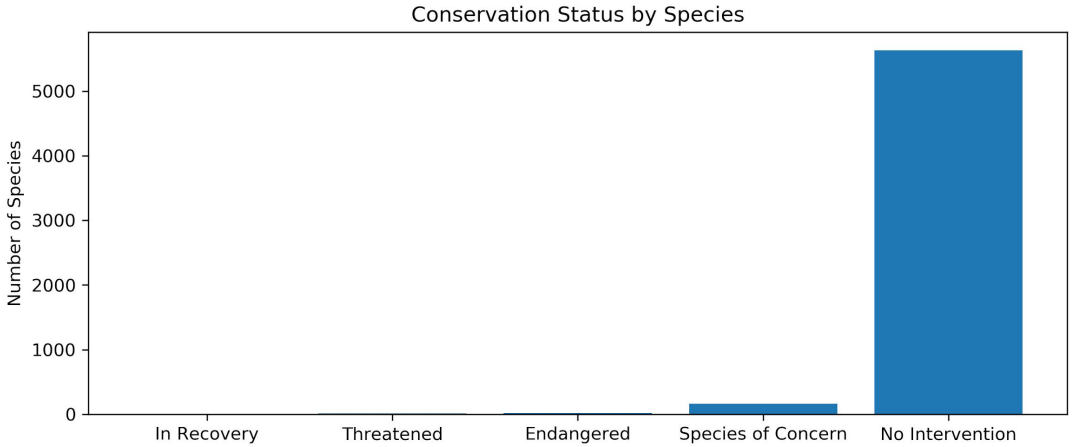
Categories – Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant, Nonvascular Plant

Conservation Statuses – Species of Concern, Endangered, Threatened, In Recovery, Nil (NaN, i.e. no intervention)

Names – Unique scientific names, one or more common names (separated by commas or parentheses)

Conservation status – totals

	conservation_status	scientific_name
1	In Recovery	4
4	Threatened	10
0	Endangered	16
3	Species of Concern	161
2	No Intervention	5633



Conservation Status by Species – table and graph

Conservation status – by category

	category	not_protected	protected	percent_protected
0	Amphibian	73	7	0.087500
1	Bird	442	79	0.151631
2	Fish	116	11	0.086614
3	Mammal	176	38	0.177570
4	Nonvascular Plant	328	5	0.015015
5	Reptile	74	5	0.063291
6	Vascular Plant	4424	46	0.010291

Not protected – Nil conservation status

Protected – Species of Concern, Endangered, Threatened, or In Recovery

Significance calculations

	category	not_protected	protected	percent_protected
0	Amphibian	73	7	0.087500
1	Bird	442	79	0.151631
2	Fish	116	11	0.086614
3	Mammal	176	38	0.177570
4	Nonvascular Plant	328	5	0.015015
5	Reptile	74	5	0.063291
6	Vascular Plant	4424	46	0.010291

Mammals – 17.76% are protected (highest rate)

Birds – 15.16% are protected

Reptiles – 6.33% are protected

Significance calculations

	category	not_protected	protected	percent_protected
0	Amphibian	73	7	0.087500
1	Bird	442	79	0.151631
2	Fish	116	11	0.086614
3	Mammal	176	38	0.177570
4	Nonvascular Plant	328	5	0.015015
5	Reptile	74	5	0.063291
6	Vascular Plant	4424	46	0.010291

Chi squared tests

Using the relevant numbers, we can determine whether there is a significant difference in the likelihood that a species from given categories is protected.

Significance calculations

index	category	not_protected	protected	percent_protected	p-value vs Mammal
3	Mammal	176	38	0.177570	1.000000e+00
1	Bird	442	79	0.151631	4.459017e-01
0	Amphibian	73	7	0.087500	8.416929e-02
2	Fish	116	11	0.086614	3.114526e-02
5	Reptile	74	5	0.063291	2.338465e-02
4	Nonvascular Plant	328	5	0.015015	1.681893e-11
6	Vascular Plant	4424	46	0.010291	1.734911e-70

Using the chi squared test and adding a column with the p-value for each category (vs Mammals) allows us to determine which categories are *significantly* different from Mammals, in terms of the likelihood that a given species from that category is endangered.

The table has been sorted by ‘percent_protected’ in decreasing order.

Significance calculations

index	category	not_protected	protected	percent_protected	p-value vs Mammal
3	Mammal	176	38	0.177570	1.000000e+00
1	Bird	442	79	0.151631	4.459017e-01
0	Amphibian	73	7	0.087500	8.416929e-02
2	Fish	116	11	0.086614	3.114526e-02
5	Reptile	74	5	0.063291	2.338465e-02
4	Nonvascular Plant	328	5	0.015015	1.681893e-11
6	Vascular Plant	4424	46	0.010291	1.734911e-70

With a p-value cut-off of 0.05 for significance, we can show a significant difference between Mammals and:

- Birds or Amphibians **does not exist**
- Fish, Reptiles, Nonvascular Plants or Vascular Plants **does exist**

in terms of likelihood that a given species is endangered.

Recommendations

index	category	not_protected	protected	percent_protected	p-value vs Mammal
3	Mammal	176	38	0.177570	1.000000e+00
1	Bird	442	79	0.151631	4.459017e-01
0	Amphibian	73	7	0.087500	8.416929e-02
2	Fish	116	11	0.086614	3.114526e-02
5	Reptile	74	5	0.063291	2.338465e-02
4	Nonvascular Plant	328	5	0.015015	1.681893e-11
6	Vascular Plant	4424	46	0.010291	1.734911e-70

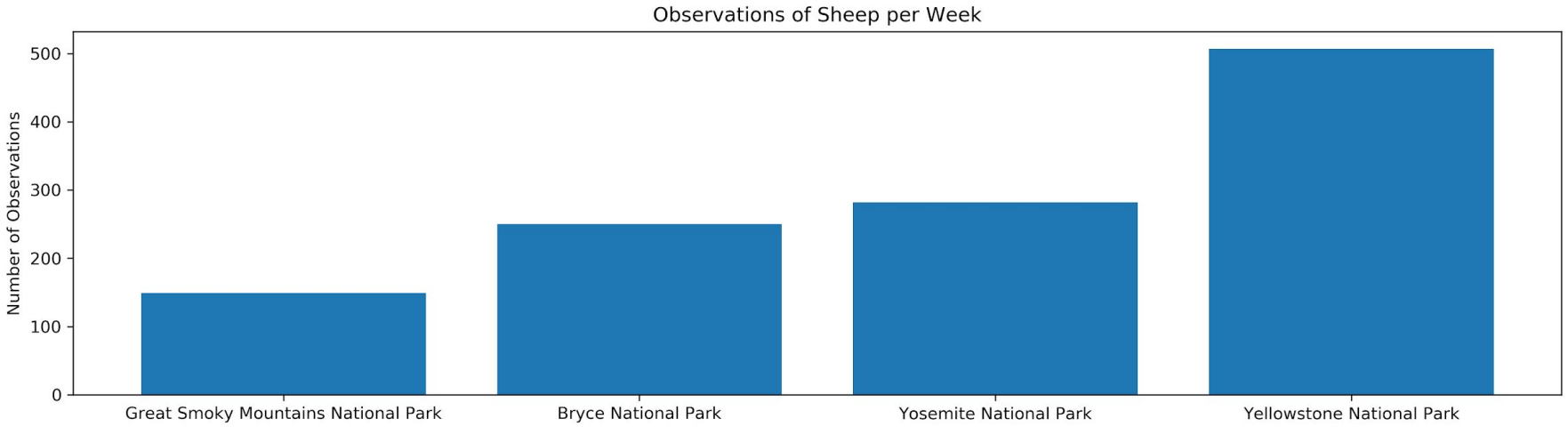
Based on these calculations, it is recommended conservation efforts be focused on **mammals, birds and amphibians**.

Sheep in national parks

	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
5	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True	Bryce National Park	109
6	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True	Yosemite National Park	117
7	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True	Great Smoky Mountains National Park	48
8	Mammal	Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep	Endangered	True	True	Yellowstone National Park	67
9	Mammal	Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep	Endangered	True	True	Yosemite National Park	39
10	Mammal	Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep	Endangered	True	True	Bryce National Park	22
11	Mammal	Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep	Endangered	True	True	Great Smoky Mountains National Park	25

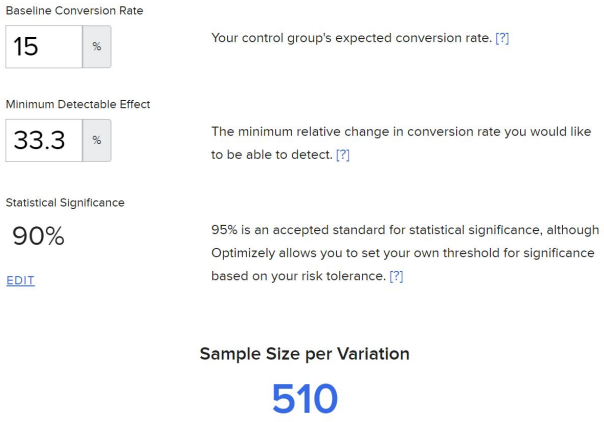
Observations – three sheep species at four national parks over seven days

Sheep in national parks



Observations – three sheep species at four national parks over seven days

Sample size calculations

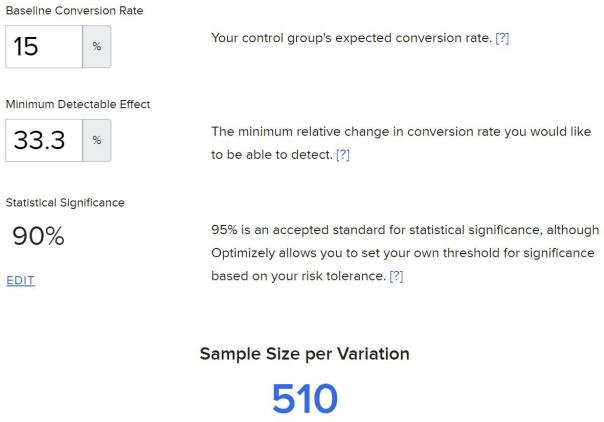


Foot and mouth disease – 15% of sheep (Bryce National Park) affected

Minimum detectable effect – 5% from 15%

Statistical significance – 90%

Sample size calculations



To achieve a sample size of **510**:

Bryce National Park – **two weeks** for approximately 500 observations (rounding up gives **three weeks**)

Yellowstone National Park – **one week** for approximately 500 observations (rounding up gives **two weeks**)