# Introduction to Data Analysis
# Jesse Fiorito - Spring 2018

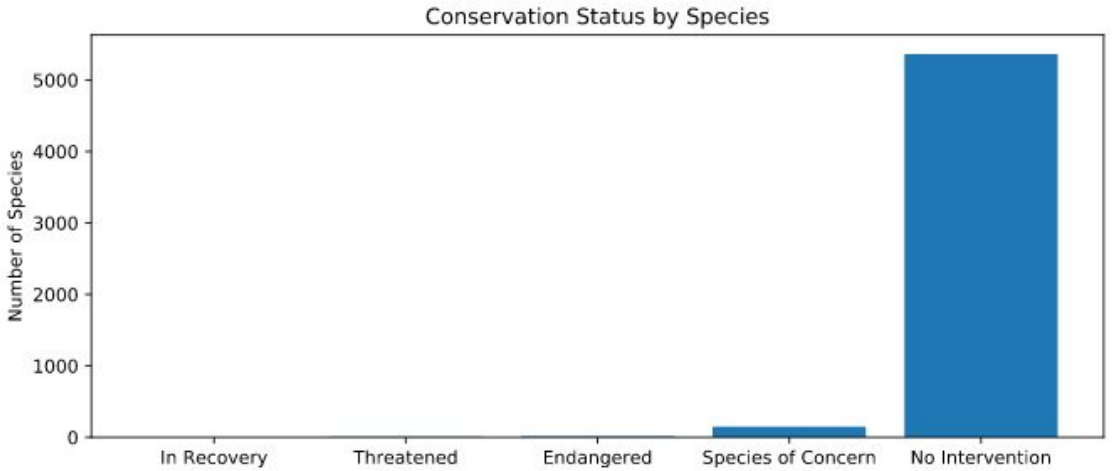
## Capstone Option 2: Biodiversity for the National Parks

# Species_info.csv. Data Description

I noticed the following while working through the notebook - data about different species in the National Parks, including:

- The scientific name of each species
- The common names of each species
- The species conservation status

| Category | Not Protected | Protected |
|---|---|---|
| Amphibian | 72 | 7 |
| Bird | 413 | 75 |
| Fish | 115 | 11 |
| Mammal | 146 | 30 |
| Nonvascular Plant | 328 | 5 |
| Reptile | 73 | 5 |
| Vascular Plant | 4216 | 46 |



# Significance calculations for endangered status between different categories of species.

Based on our significance calculations that we did for endangered status between different categories of species:

- **Chi-Squared Test for Significance**
  - Contingency (protected birds and mammals)
    - pval = 0.687594809666 (~0.688)
    - The difference between the percentages is not significant because pval > 0.05
  - Protected Reptiles and Mammals
    - pval_reptile_mamma = 0.0383555902297 (~0.038)
    - The difference between the percentages is significant because pval_reptile_mammal < 0.05

| Conservation Status | Count |
|---|---|
| Endangered | 15 |
| In Recovery | 4 |
| Species of Concern | 151 |
| Threatened | 10 |

# A recommendation for conservationists concerned about endangered species

- **Conservation Status**
  - Species of Concern: declining population or appears to be in need of conservation
  - Threatened: vulnerable to endangerment in the near future
  - Endangered: seriously at risk of extinction
  - In Recovery: formerly Endangered, but currently not in danger of extinction throughout all or a significant portion of its inhabitable range

Are certain types of species more likely to be endangered? Based on our significance calculations:

- There was a slight difference in the percentages of birds and mammals that fall into a protected category
- Null hypothesis: This difference was a result of chance
- Conclusion: Certain types of species are more likely to be endangered than others
- Recommendation: Types of Species should be prioritized for protection based on conservation status

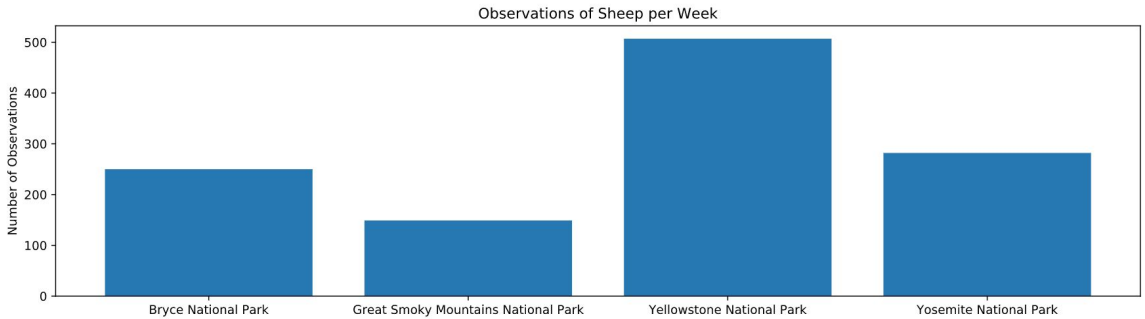
# Foot and Mouth Disease Study Sample Size Determination

Based on our calculations the foot and mouth disease study (use the default level of significance of 90%):

- Baseline conversion rate: 15% (*Recorded by Bryce National Park scientists)
- Statistical significance: 90%
- Minimum Detectable Effect: 10%
- Sample size: 12000

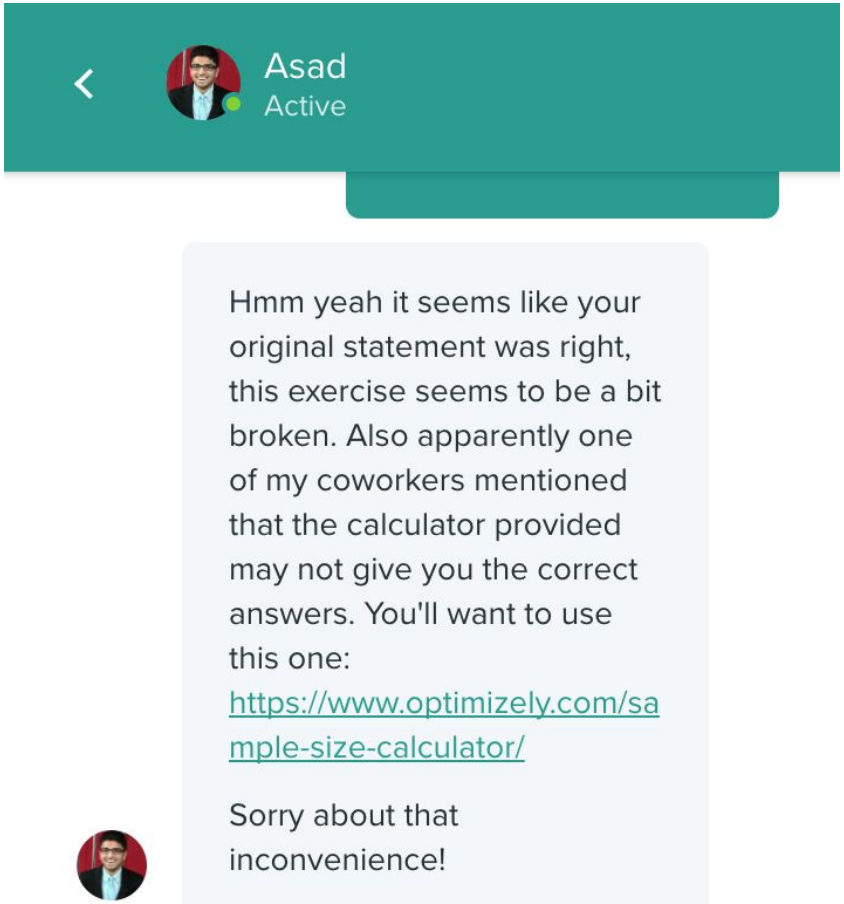
If the scientists wanted to be sure that a >5% drop in observed cases of foot and mouth disease in the sheep at Yellowstone was significant they would have to observe at least 510 sheep.

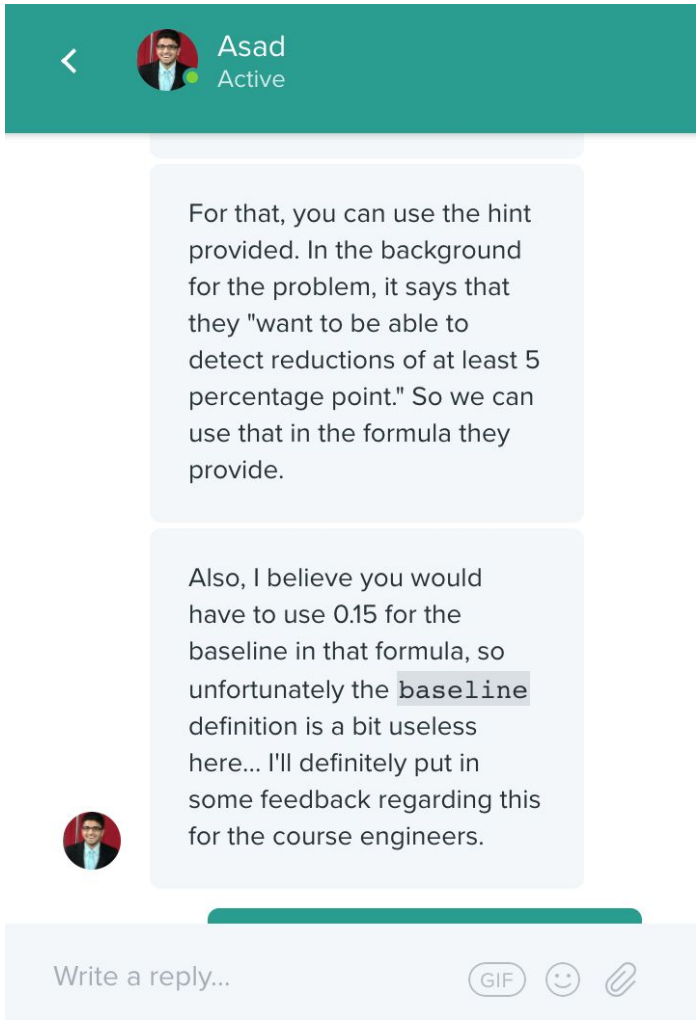
- Weeks required to see 510 sheep to test for >5% drop in cases of foot and mouth disease:
  - One week of observing in Yellowstone National Park
  - Approximately two weeks in Bryce National Park



| Park Name | Observations |
|---|---|
| Bryce National Park | 250 |
| Great Smoky Mountains National Park | 149 |
| Yellowstone National Park | 507 |
| Yosemite National Park | 282 |

| Scientific Name | Park Name | Observations | Category | Common Names | Conservation Status | Is Protected? | Is Sheep? |
|---|---|---|---|---|---|---|---|
| Ovis canadensis | Yellowstone National Park | 219 | Mammal | Bighorn Sheep, Bighorn Sheep | Species of Concern | TRUE | TRUE |
| Ovis canadensis | Bryce National Park | 109 | Mammal | Bighorn Sheep, Bighorn Sheep | Species of Concern | TRUE | TRUE |
| Ovis canadensis | Yosemite National Park | 117 | Mammal | Bighorn Sheep, Bighorn Sheep | Species of Concern | TRUE | TRUE |
| Ovis canadensis | Great Smoky Mountains National Park | 48 | Mammal | Bighorn Sheep, Bighorn Sheep | Species of Concern | TRUE | TRUE |
| Ovis canadensis sierrae | Yellowstone National Park | 67 | Mammal | Sierra Nevada Bighorn Sheep | Endangered | TRUE | TRUE |


Asad
Active
Hmm yeah it seems like your original statement was right, this exercise seems to be a bit broken. Also apparently one of my coworkers mentioned that the calculator provided may not give you the correct answers. You'll want to use this one: https://www.optimizely.com/sample-size-calculator/
Sorry about that inconvenience!
Just now.


Asad
Active
For that, you can use the hint provided. In the background for the problem, it says that they "want to be able to detect reductions of at least 5 percentage point." So we can use that in the formula they provide.
Also, I believe you would have to use 0.15 for the baseline in that formula, so unfortunately the baseline definition is a bit useless here... I'll definitely put in some feedback regarding this for the course engineers.
Write a reply...
GIF

https://localhost/

| | | | | | | |
|---|---|---|---|---|---|---|
| 2233 | Vascular Plant | Festuca filiformis | Fineleaf Sheep Fescue | No Intervention | False | True |
| 3014 | Mammal | Ovis canadensis | Bighorn Sheep, Bighorn Sheep | Species of Concern | True | True |
| 3758 | Vascular Plant | Rumex acetosella | Common Sheep Sorrel, Field Sorrel, Red Sorrel, Sheep Sorrel | No Intervention | False | True |
| 3761 | Vascular Plant | Rumex paucifolius | Alpine Sheep Sorrel, Fewleaved Dock, Meadow Dock | No Intervention | False | True |
| 4091 | Vascular Plant | Carex illota | Sheep Sedge, Smallhead Sedge | No Intervention | False | True |
| 4383 | Vascular Plant | Potentilla ovina var. ovina | Sheep Cinquefoil | No Intervention | False | True |
| 4446 | Mammal | Ovis canadensis sierrae | Sierra Nevada Bighorn Sheep | Endangered | True | True |

| | category | scientific_name | common_names | conservation_status | is_protected | is_sheep |
|---|---|---|---|---|---|---|
| 3 | Mammal | Ovis aries | Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral) | No Intervention | False | True |
| 3014 | Mammal | Ovis canadensis | Bighorn Sheep, Bighorn Sheep | Species of Concern | True | True |
| 4446 | Mammal | Ovis canadensis sierrae | Sierra Nevada Bighorn Sheep | Endangered | True | True |

| | 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 |

| | scientific_name | park_name | observations |
|---|---|---|---|
| 0 | Vicia benghalensis | Great Smoky Mountains National Park | 68 |
| 1 | Neovison vison | Great Smoky Mountains National Park | 77 |
| 2 | Prunus subcordata | Yosemite National Park | 138 |
| 3 | Abutilon theophrasti | Bryce National Park | 84 |
| 4 | Githopsis specularioides | Great Smoky Mountains National Park | 85 |
| 5 | Elymus virginicus var. virginicus | Yosemite National Park | 112 |