Data Science Intensive

Presentation
By Michelle Lia

Aim

To analyse the species dataset to understand more about the biodiversity living on National Parks land.

Insights from species dataset

- There were 15 endangered and 10 threatened species living within the park
- 96% of all species needed no intervention
- 2.72% of all species sighted were considered to be of concern
- .27% were considered to be endangered
- 0.07% were considered endangered

Of the not protected/no intervention group, 76.1% were vascular plants

The category of species that had the most protected species were birds.

Statistical significance of endangered species

Performed chi squared tests to answer the question:

“Are certain types of species more likely to be endangered”?

Tests were performed between birds and mammals, then reptiles and mammals.

Results:

P value (birds and mammals) = 0.69 (insignificant)

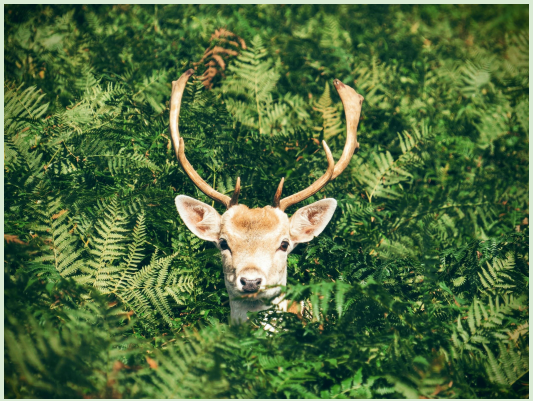
P value (reptiles and mammals) = 0.03 (significant)

This data suggests some categories are more likely to be endangered than others

Recommendations for conservationists

Mammals have the highest % rate of being protected - more could be done to help increase numbers of protected mammal species i.e. introduce endangered mammal species breeding programs

Based on our data certain species such as reptiles are more likely to endangered and conservation efforts would be better directed in towards categories of species more likely to be endangered



Foot and mouth disease study

Scientists wanted to test the number of sheep that had foot and mouth disease in the park had decreased to 10% from 15% with a 90% confidence level. To do this, we needed to work out the appropriate sample size.

To do this we needed to work out the:

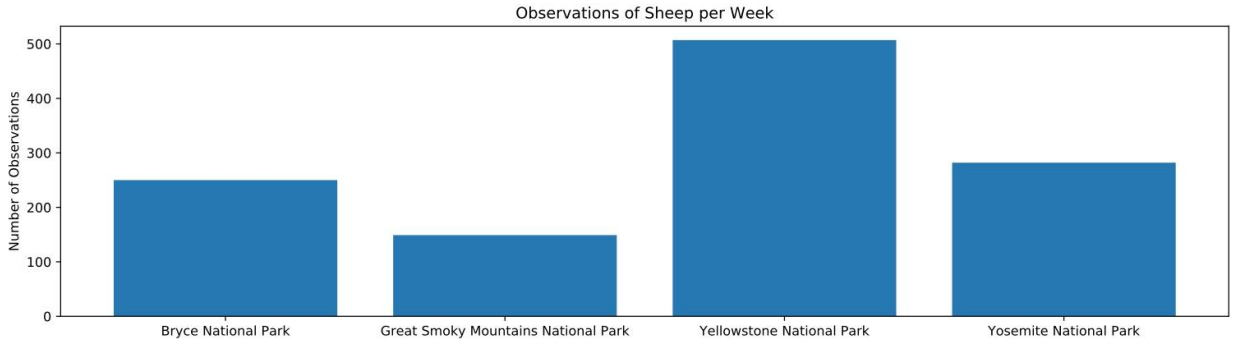
Baseline :15% (we already had this, it was the 15% that was recorded with the disease in the prior year)

Confidence Level: 90% (which we already had)

Minimum detectable effort: 100x5/15 =33.3% (To work this out, calculate the % change as part of the baseline percentage).

Summing this into the calculator, the population size needed at a 90% confidence level was found to be 870.

Appendix: Graphs of sheep sightings in national parks



Yellowstone National Park had the highest sightings of sheep

Conversation status by species