Biodiversity for the National Parks

Capstone project

Data: Species in National Parks

Introduction

This section will focus on the analysis of the data, describing the diversity of the species in the National parks. The data contains information about the scientific and common name of each species, the category they fall into and their conservation status.

First look at the data

The data frame consists of 5541 different species.

They are divided in 7 distinct categories : Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant and Nonvascular Plant.

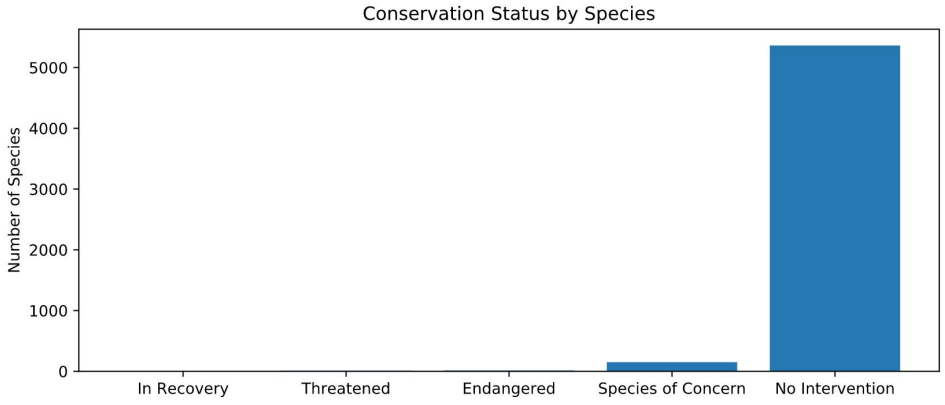
All the species are assigned one of 5 conservation statuses: ‘Species of Concern' 'Endangered' 'Threatened' 'In Recovery’ and ‘nan’, where ‘nan’ refers to all species that are of no concern. For better result, the “nan” value was replaced with more understandable “No Intervention”

The table below shows the number of species that fall under each status category.

conservation status	number of species
No Intervention	5363
Endangered	15
In Recovery	4
Species of Concern	151
Threatened	10

Graph: Conservation status by Species

The image on the right shows more visual representation of the data from the previous slide, showing the number of species that falls under each of the conservation statuses.



Investigating the Endangered Species

Introduction

Going through the endangered species DataFrame raises an interesting question. **Are certain types of species more likely to be endangered?** This section aims to answer that question

Determining which types are more likely to be endangered - 1

In order to answer the given question, several steps need to be taken.

If we assume that all species with conservation status 'No Intervention' are not protected and all others are, we can calculate how many species in certain category are protected and how many aren't. The result is shown on the table on the right.

category	number of not protected species	number of protected species
Amphibian	72	7
Bird	413	75
Fish	115	11
Mammal	146	30
Nonvascular Plant	328	5
Reptile	73	5
Vascular Plant	4216	46

Determining which types are more likely to be endangered - 2

Based on the results of the calculations of the previous slide, we can calculate what percentage of the species in each category are consider of concern or, to put it simply, are protected. The result can be seen on the table on the right.

category	number of not protected species	number of protected species	percentage of protected species
Amphibian	72	7	0.088608
Bird	413	75	0.153689
Fish	115	11	0.087302
Mammal	146	30	0.170455
Nonvascular Plant	328	5	0.015015
Reptile	73	5	0.064103
Vascular Plant	4216	46	0.010793

Analysis of the results

Analyzing the results from the previous slide, it would seem that ‘Mammal’ and ‘Birds” are more likely to be endangered than all the other species, with ‘Mammal’ having a slightly higher chances to be endangered.

This raises new questions: **is this difference significant?** In order to answer that question we need to create a hypothesis and test it.

Chi Square Test and results

Since our data is divided into categories, it is more appropriate to choose a test such as Chi Square Test to test the hypothesis. The test will be performed with the following set up:

- The null hypothesis we state that this difference is due to chance.
- The test will use two sets of data - the information about the number of protected and not protected species in the categories ‘Mammal’ and ‘Bird’

Results: The Chi Square Test calculated a pvalue of 0.687594809666, which is marginally larger than 0.05 so the null hypothesis can not rejected. Which means that the difference in the results is due to chance and therefore not significant

In comparison, if we run the Chi Square Test with the same hypothesis but for categories ‘Mammal’ and ‘Reptile’, we get pvalue = 0.0383555902297, which is less than 0.05 so we reject the null hypothesis which means that any difference between those two categories is significant

Conclusion: Some species are more likely to be endangered than others

Recommendations

Since species in categories such as ‘Mammal’ and ‘Birds’ are more likely to be endangered than others, conservationists should focus more attention on preserving those species and their natural habitat. In addition, other species in these categories, which are currently considered not endangered, should be kept under observation in case their numbers drop so that actions can be taken before they even go in ‘Species of Concern' status.

Observation Results Analysis

Introduction

This section will focus on the analysis of the data, gathered as a result of recording of sighting of different species in several national parks, for a period of 7 days. The focus on the research is the evaluation of a program concerning sheep species in the 4 different national parks.

First look at the data

The focus on this research is different types of sheep, observed in several national parks.

Going back to part one of this presentation and using the data from the species that can be found in national parks, we can find out that all of the observed sheep fall into three main scientific categories: Ovis aries, Ovis canadensis and Ovis canadensis sierrae.

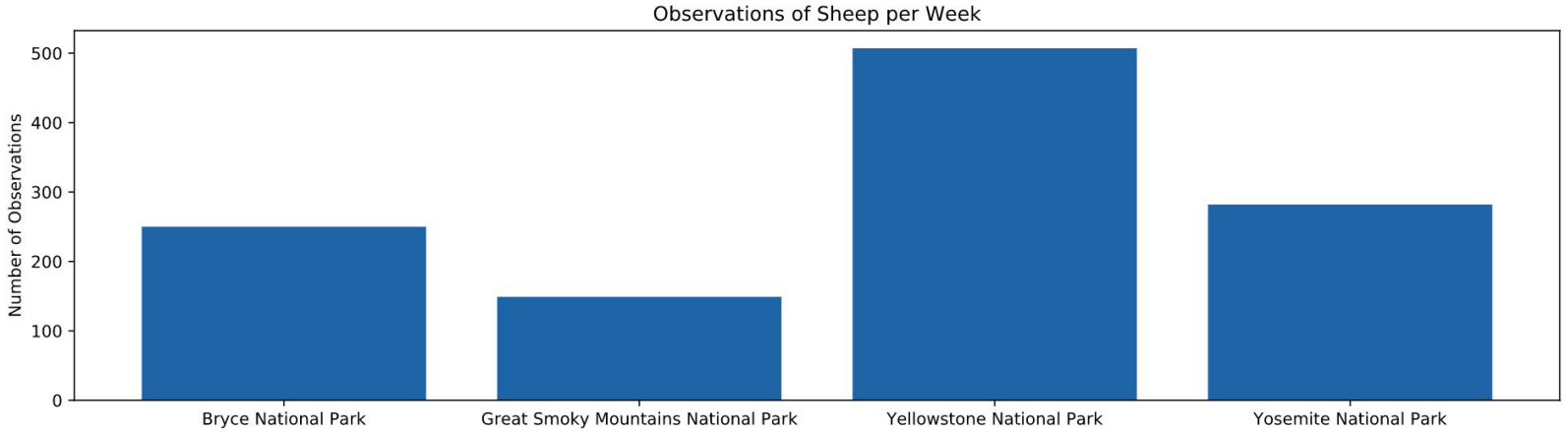
Using that data and the data from the observation, we can determine how many total sheep sightings, across all three species, were made at each national park

The table below shows the result of that calculations

park	observed sheep
Bryce National Park	259
Great Smoky Mountains National Park	149
Yellowstone National Park	507
Yosemite National Park	282

Graph: Observations of Sheep per Week

The image below shows more visual representation of the data from the previous slide, showing the number of sheep observed in the different parks in the course of 1 week.



Foot and Mouth Reduction Effort - Sample Size Determination

We want to evaluate a program to reduce the rate of foot and mouth disease at Yellowstone National Park. In order to determine if it is working, we need to calculate how many week scientist need to spent in the park, observing the sheep.

First, we need to calculate the sample size. For baseline, we are going to use last year data from Bryce National Park, where 15% of the sheep had foot and mouth disease: **baseline = 15%**. **Minimum Detectable Effect = = 33%,** and the **statistical significance = 90%.** With this we determine that we need a sample size of **520** sheep.

Looking at the data from previous slides, the scientists in Yellowstone National Park will need 520/507 or approximately **1 week** to determine if the program is working.

On the other hand, the scientists from Bryce National Park will need 520/250 or approximately 2 **weeks** to finish the study.

Thanks!

Contact :

Borislava Spasova

b.spasova@gmail.com