

Exam 1

CS3300 Data Science

Overview

In the first half of this course, you learned how to clean and explore data using a variety of techniques. In this exam (which is structured similarly to a lab), you are going to demonstrate your knowledge of these techniques to perform an analysis of a provided data set.

Background

Conservation ecologists need to track population levels of various species to determine the health of those populations. For example, dips in population levels might indicate new dangers to populations, while rises in population levels may indicate that conservation efforts are effective.

Conservation ecologists have multiple approaches for tracking population levels. For example, in mark and recapture methods, ecologists capture animals, tag them, release them, and recapture the animals. Alternatively, ecologists may set up cameras and count how many times each individual is seen.

Dr. Rachel Reid is focused on a third approach – using "leftovers" such as scat and animal carcasses used from food. In a 2015 paper, Dr. Reid reported on an effort to determine if it is possible to distinguish between bobcat, coyote, and grey wolf scats from morphological (e.g., length, volume, etc.), biogeochemical (e.g., chemical composition), and contextual traits (e.g., location). Morphological traits are the least costly properties to evaluate and can be done in the field, while biogeochemical traits may require expensive equipment and need to be analyzed in a lab.

We have provided Dr. Reid's data set. Your job is to apply your data science knowledge to analyze the data and answer the research questions posed by Dr. Reid. You will also have to do some background research on the three species to interpret and explain the differences. Using this data set, you need to answer two questions:

1. Which (if any) morphological and biogeochemical traits distinguish between originating species of the scat samples?
2. Why do you think those traits differ across species?

Instructions

You will organize your notebook as a paper with four sections. You will be evaluated on choosing the appropriate techniques for analyses, accuracy of the analyses, effort and completeness in interpreting and discussion the data, and presentation and formatting of plots (e.g., axis labels). You are not allowed to work with or discuss the exam with anyone. When you are done, save your Jupyter notebook as a PDF and upload that file through Canvas.

Part I: Introduction and Background

1. Read the provided article by Dr. Reid.
2. Search the internet to find a descriptive article or web pages on the biology of each species (coyotes, bobcats, and grey wolves). Encyclopedia, magazine, etc. type articles are sufficient. The article should answer questions such as what they eat, their size, and where they are found. Include a bibliography at the end of the notebook with your sources.
3. Find an article reviewing population estimation methods for (conservation) ecology and their respective strengths and weaknesses (these two articles might be helpful: [article1](#) and [article2](#) (you can just read the paragraph on “Population Research Methods”)).
4. Write an introduction. Summarize what you've learned about the three species and provide some motivation for this analysis. For example, explain why anyone might be interested in tracking coyote, bobcat, and grey wolf species. Are any of these species endangered? Are they predators that represent risks to human society, agriculture, or domestic animals? Explain the motivation for estimating population sizes from scat samples. Restate the given research questions, and explain why it might be possible to observe morphological and biogeochemical traits differences between originating species.

Part II: Analysis

5. Load and clean the data set. All columns must be converted to the correct types. Identify any outliers or anomalies in the data. If outliers or anomalies are identified, describe the rationale behind categorizing them as such.
6. Create a table that categorizes each variable in the data set as morphological, biogeochemical, contextual, or not a trait (e.g., species).
7. Refine the research questions into more specific questions or hypotheses that test the relationship between the morphological and biogeochemical traits and species. Explore these questions using the visualization and statistical testing techniques you've learned. (Do not analyze contextual traits.) Determine which traits can be used to distinguish at least one species from the rest.

Part III: Discussion and Interpretation

9. Make a table of all traits that separate at least one species from the rest. Use descriptive statistics (e.g., mean, mode, etc.) to characterize the values for each species.
10. Using what you learned from your research, explain how the traits with differences relate to differences in the biology of the three species.

Part IV: Conclusion

11. Write a conclusion. The conclusion should summarize your key findings and describe what someone could do to continue or expand on this work.