Final Project

Your Name Here

2024-06-07

Final Details

Purpose

The goal of this final assignment is to assess your ability to integrate the many skills you have learned over the semester: filtering and summarizing data, choosing the appropriate data visualizations, and performing and interpreting the appropriate statistical tests.

Task

Write R code which produces the correct data, summaries, plots and analyses. Correctly interpret the results of these plots and analyses.

Criteria for Success

- Code chunks run without errors
- Code produces the correct result

    - Code that produces the correct answer will receive **full** credit
    - Code attempts with logical direction will receive **partial** credit

- Appropriate plot types are used to visualize the data
- Appropriate statistical tests are used to analyze the data
- Written answers address the questions in sufficient detail

Due Date

May 7 at midnight MST

Structure & Guidelines

Interpreting Statistical Results

When I ask you to interpret statistical results, you should roughly follow these guidelines.

- the cut-off for our p-values is always 0.05 ($\alpha = 0.05$)
- report the p-value that we are focused on in that question

- if there are multiple p-values of interest, report all of them
- state whether the p-value indicates a significant difference/relationship
- if applicable, state whether we should or should not reject the null hypothesis

Plotting

All plots should be made using `ggplot2`.

Your options for plot types to choose from are:

- <u>multiple</u> histogram plots
  - use transparency (`alpha`)
  - use `position = "identity"` with multiple groups to see the full distributions
- box-and-whisker plot
  - add points on top of the box plot to show the distribution of the points
- scatter plot
  - add the linear model to every scatter plot

Note: All plots should have modified axis labels and legend labels.

In many cases, this might mean capitalizing the axis label and/or legend label. In other cases, you might want to put units in parentheses after the words (e.g., Body Mass (g)).

Generative AI Guidelines

While I do not recommend using generative AI to complete this assignment (throughout the semester, I've seen it cause more problems than not...), you are allowed to use it on the final.

However, you *must* do so according to the guidelines I've set out for all other assignments this semester.

You can find clear examples of these guidelines on D2L in *Resources > Helpful Info > Assignment Reminders*.

CAUTION!   If you do use generative AI without following the course guidelines, you will receive a 0 for this final and you will not have the option to resubmit. Proceed wisely.

Let's Begin!

Have you read through the "Structure & Guidelines" section?

If so, proceed! If not, scroll back up...

Set-Up (5 points)

1. First, we need to load the `tidyverse` package. (2 points)

```
# Load the tidyverse
library(tidyverse)
```

2. As a reminder, there are 4 datasets that you can use to complete this final assignment:

   - krill data from KrillBase -> `krill.csv`
   - fur seal pups on Maiviken Island -> `fur_seals.csv`
   - Gentoo penguins from Bird Island -> `gentoos_and_temps.csv`
   - Chinstrap penguins from Signy Island -> `chinstraps_on_signy.csv`

Which dataset have you chosen to work with for this final assignment? (1 point)

Answer:

3. Bring your chosen data into R to work with using the `read_csv` function. Remember to save it as an object. (2 points)

Understanding the Data (12 points)

4. Use a function (e.g., `head()` or `str()`) to look at your data. (3 points)

5. In 1-2 sentences, explain what the output from Question 4 is telling you about your dataset. The more detail, the better. (3 points)

Answer:

Data Exploration

Now, spend some time really getting to know your dataset. Information about the datasets can be found below, both in terms of how you should treat each column (quantitative vs. qualitative) as well as at the links to the datasets below (click on the names of the datasets).

You might notice that the data at the links has many more columns than the datasets you have been given or some slightly different data. Focus only on the columns that are in the .csv file provided to you for this assignment.

Krill

- You will want to use the `Krill_per_1m2_log` column as your dependent variable. This is a log-transformed version of the `Krill_per_1m2` column. You don't need to worry about *why* we transformed the data, but this column should be a bit easier to work with.
- The `Date`, `Year`, and `Month` columns should all be treated as *quantitative/continuous* variables.
- While the `Station` column can *technically* be considered a qualitative/categorical variable, please use the `Day_Night` column as your qualitative/categorical variable for this assignment and for your research question.

Fur Seals

- If you choose to use the `Date` or `Year` column, treat them as *quantitative/continuous* variables.
- On the other hand, if you use the `Month` column, treat it as a *qualitative/categorical* variable.

Gentoo Penguins

- If you are using the `Year` column, treat it as a *quantitative/continuous* variable.
- The dataset has a column for the previous year's December temperature. Because counts of chicks took place in January or February (and nests were counted in October/November), the number of chicks is most likely to be related to the temperatures directly after chicks hatched (December of the previous year).

Chinstrap Penguins

- I do not recommend using the `MONTH` or `DAY` column.
- If you use the `YEAR` column, treat it as a *quantitative/continuous* variable.

Back to Assignment Questions

6. Based on your own general knowledge, the information at the links above, and the additional information I've provided, list out the columns (only those in the .csv file) in your chosen dataset and give a brief description (just a phrase is fine) of what the data in each columns represents. (3 points)

Answer:

7. For each of the columns in your dataset, identify whether the data points they contain are qualitative (categorical) or quantitative (numeric). Be sure your answers align with the information I've provided for you above. (3 points)

   Many of your answers to subsequent questions will depend heavily on whether variables are qualitative or quantitative.

   Type out the name of the column and then the descriptor. For example:

   - island_name: qualitative
   - island_size_km: quantitative

Answer:

Building a Research Question (6 points)

In order to write a successful research question, we need to identify our dependent variable and the independent variable that we think affects the dependent variable.

8. First, determine which variable is the *dependent* variable in your dataset (there is usually only one, but there might be two). It must be quantitative. (2 points)

Answer:

9. Next, choose 1 independent variable from your dataset (each dataset has more than one from which to choose). (2 points)

   This should be a variable that you believe might affect the values of the dependent variable. It can be either qualitative or quantitative.

Answer:

10. Now, bring your dependent and independent variables together to form a research question based on the structure presented in class in Module 1 (also see Module 1 Assignment 3). (2 points)

Answer:

Literature Search (6 points)

When we ask a research question, we always want to find out what other people have already found out about that question to give us an idea of what outcome we might expect.

The articles you find don't have to be *exactly* related but should be relevant, as you will be using them at the end of this assignment to help interpret your statistical output.

To find related papers, you will want to search for journal articles that have both your dependent variable and your independent variable included. (e.g., plant height, water).

To determine whether the article is related to your research question, read the title and the abstract.

11. Write the citation for your first article. Then, in your own words, write a 1-2 sentence summary of what the article is about, based on the title and the abstract. (3 points)

Article 1 Citation:

Article 1 Description:

12. Do the same as above for the second article. (3 points)

Article 2 Citation:

Article 2 Description:

Summarizing Our Data (6 points)

Let's get to know our data a little better.

13. Summarize your data by calculating at least 1 measure of central tendency and 1 measure of variability for your your *dependent* variable using the `tidyverse` functions (you might need `na.rm = TRUE`!). (4 points)

Pay close attention here! How you summarize will differ based on whether your *independent* variable is qualitative or quantitative (2 points):

- If your research question has a *qualitative (categorical)* independent variable, find the measure of central tendency and the measure of variability of *each* individual category.
- If your research question has a *quantitative (numeric)* independent variable, find a measure of central tendency and the measure of variability for the independent variable in addition to the dependent variable

Data Visualization (18 points)

To continue our data exploration, we are going to plot our data. Some of the plots might be related to your research question; other may not. That's ok!

At least ONE of the plots should include both your dependent and independent variable. Think critically about which plot that should be based on your variable types.

14. Produce a histogram plot (just one, not multiple on the same plot) of the dependent variable. Write 1-2 sentences describing what information you can gather from plot. (3 points)

Answer:

15. Create a multiple histogram plot of your dependent variable and one other variable (this may or may not be your independent variable). (3 points)

    Make sure we can see each histogram plot (make sure one isn't blocking another by modifying the transparency and position).

16. Create a box-and-whisker plot which includes your dependent variable. Changing the color of the boxes is optional. Make sure to add points to the plot. (3 points)

    Hint: if your plot has NA values that you want to remove, you can remove them from the dataset with the `filter` function and the argument `!is.na(column_name)`.

17. Create a scatter plot that includes your dependent variable. Write 1-2 sentences describing what information you can gather from plot. (3 points)

    Hint: if your plot has NA values that you want to remove, you can remove them from the dataset with the `filter` function and the argument `!is.na(column_name)`.

18. Identify which plot(s) from Questions 15-17 match up with your research question. How you know it matches up correctly (based on qualitative vs. quantitative variables)? (3 point)

Answer:

19. In 2-3 sentences, interpret the plot(s) that you identified in the previous question. What does it tell you about your chosen research question? (3 points)

Answer:

Running Your Statistical Model (15 points)

In this course, we have covered 3 different statistical analyses: t-tests, ANOVAs, and linear regression.

It is now time to decide how to apply these statistical analyses to your own research question!

20. Based on everything you know about your independent and dependent variables and the requirements for each of the statistical analyses we covered in the course, determine which statistical analysis you should perform to answer your question. In 1-2 sentence, explain why. (3 points)

Answer:

21. Perform the appropriate statistical analysis below. If necessary, add an additional line of code to print out the results. (3 points)

    Hint: if your dataset has NA values that you want to remove, you can remove them from the dataset with the `filter` function and the argument `!is.na(column_name)`.

22. If necessary, run a pairwise comparison. If a pairwise comparison is unnecessary for your statistical analysis, explain why. (3 points)

Answer:

```
# Space for code to run a pairwise comparison if needed
```

23. Interpret the results of your statistical analysis (including the pairwise comparison, if relevant). In your answer, include the p-value and if the result of your analysis is significant. (3 points)

*Answer***:**

24. What do these results mean in the context of your research question? (3 points)

*Answer***:**

Expanding Our Horizons (6 points)

25. Spend some time thinking about the outcome of your research question. Is the outcome what you were expecting or not? Why or why not? What additional information would you like to have to add context to your results? Your answer should be 4-5 sentences. (3 points)

Answer:

26. Think about your research subjects in the context of ecosystem services. What type of ecosystem services might the study of your research question fall under? Explain your rationale. (3 points)

*Answer***:**

Communicating Your Results (6 points)

When researchers have come to some type of conclusion about a research question they have asked, we usually write up our findings in a scientific article to share our results with the scientific community and beyond.

Just like the articles you found in the Literature Search section, most articles have an abstract that gives any potential reader some information about the topic of the research paper and the conclusions the researchers have drawn.

Now that you have completed the process of coming up with a research question and finding some results, we want to communicate these results through writing an abstract. Feel free to look at your scientific articles for inspiration.

27. Your abstract should be structured in the following manner, with at least one sentence for each topic (6 sentence minimum):

- General information about your dataset
- The research question your have chosen to ask
- What your data summarization told you
- What your data visualization shows you
- What statistical analysis you performed and its outcome
- Your final interpretation of your research question: does your independent variable impact your dependent variable and, if so, in what way?

Be sure you include information from your articles for both of your articles from your literature search somewhere in your abstract! Use in-text citations (e.g., Golding & Anderson, 2024; Bledsoe et al., 2024) to show where you do so.

*Abstract***:**

The End!

Take a look back at everything you just did! You've come a *very* long way since the beginning of the semester and should be quite proud of that.

Great work, and thanks for a nice semester!