STA2201H Research Project

Overview

Your final assessment for STA2201 is a research project based on the research question you presented in your proposal and in class. The aim is to apply some of the methods covered in class to gain insight into your question of interest, and write up your analysis in the form of a short research paper.

Tips

- You should be aiming for your submission to read like a short research paper. This means using formal language, spell-checking, writing in complete paragraphs, etc.
- If it is appropriate, refer to and cite previous work.
- For readability, I would suggest hiding your code, and suppressing any warnings and messages in the pdf. All you really want to show are nicely formatted graphs, tables and results.
- Formatting is also important and helps readability.
- Spellcheck, please
- Write all your models using appropriate mathematical notation.
- You should include some type of model diagnostics or validation.
- Be logical. Clearly state the problem you are analyzing, what the data look like, how you are doing it, what the results are, and what could be done in future.
- There is no page limit or page minimum. Note that quantity does not always indicate quality.
- The PDF you submit must be able to be compiled from you .Rmd/.qmd file.
- If your data and/or model is large and it's hard to compile the one document, consider having separate R scripts which run and save output, which is then read into your .qmd file (you would also have to hand in the R scripts).
- A reminder that using generative AI to for written text is prohibited in this course and considered an academic offense.

Below is a suggested outline of your paper. Note that this is meant to be useful as a guide only — you do not have to do this exact outline. To reiterate, your submission should read like a research paper.

Abstract

You may include an abstract that summarizes you aims, data, methods and findings. Note that in RMarkdown/Quarto an abstract can be included in the YAML at the top of the file.

Introduction

Introduce the problem, clearly state your research question and motivation, mention previous work if relevant.

For this section and in general for the whole report, it should be written like a mini research paper. So instead of just stating your research question, try to introduce the topic in a conversational way.

Data

This section should build on what you did for your initial EDA. It should include things like

- A brief sentence explaining the data set you are using
- Whether you subsetted to only include certain respondents
- If values are missing, how are you dealing with this? Are they missing at random?
- The variables in the data set you are using
- Any other variables you created

This section should also include any summary statistics, graphs or tables that

- describe the characteristics of your dataset, and
- illustrate the relationship between variables that are related to your research question

Please don't include pages and pages of graphs here. The goal is to describe the key features of the dataset that motivates the subsequent analysis.

Methods

- Describe the statistical method you are using to answer your research question
- Write down the model and define any notation
- Why is this method appropriate for the research question?
- How are you fitting the model?
- How are you validating the model?

Results

Present your findings, including things like

- Key coefficient estimates and CIs, with interpretation
- Graphs/Visualizations where appropriate
- Diagnostics
- If relevant, discussion of alternative models tried/ changes made / improvement in diagnostics

Discussion

- Discuss results as they relate to your research question
- Were results surprising or expected?
- Brief discussion about potential future work – what could you do differently if you had more data or more time?

References

- Any references you cited