STAC67 Case Study Guidelines

Introduction

- ☐ Describe your research question/hypothesis/study aims
- ☐ Brief background about the topic
    - Optional: Include references to research already done in this area (APA format)
- ☐ Describe how your data was cleaned
- ☐ Brief description of what analyses you will conduct in the paper
- ☐ Explain why some variables will be left out of the model (if you decide not to use all the variables listed)

Description of Dataset

- ☐ Include descriptive statistics for each one of the variables that you will be analyzing
- ☐ Add graphs to help illustrate distributions of data
- ☐ Check the distribution of the response variable and explore the relationships between response and explanatory variables (also between explanatory variables themselves)

Building Model and Model Validation

- ☐ Build the best multiple regressions model that you can find
    - Note: There is no one "correct" model. We are leaving it to you to play around with the data and find a model that you believe best describes the data.
- ☐ Interpretation of the model is the most important part of this section.
- ☐ Clearly indicate your final (selected) **regression equation** based on your output from R.
    - Co-efficient interpretations are required
- ☐ Validate your final selected model.

Model Diagnostics

- ☐ Check if regression assumptions are valid using residual diagnostics
- ☐ Again, it's not enough to just insert the plots in the paper. Make sure to interpret what they mean in relation to your model.
- ☐ Check if there any outlying and influential points.

Conclusion

- ☐ Summarize findings
- ☐ Address limitations of your study
- ☐ Suggest future directions/extensions

Reference

- APA citation of references that you used for this project (if you used information from external sources

Optional

- Add something to this paper that was not taught in the course, or is an extension to something that was taught

Something to think about for your paper and presentation

Assume that the TAs marking your projects have no background knowledge about the dataset. You should try to convey the information with sufficient detail, so we are able to understand what your project is on, but find a balance (don't go overboard with the details)!

Note: This is not an exhaustive list of everything that can be added to your report. It should serve as a template to guide your writing, but feel free to add sections if you think it is necessary for your own report!

Tips

This case study may seem overwhelming at first if you have no prior exposure to raw data analysis. Here's a few things that help:

1) Get on a group call and figure out everyone's schedules for the next few weeks
    a. Divide up the work accordingly – ensure that everyone has something to do to contribute to the project
2) Share a google doc with your group and come up with a weekly schedule
    a. Set deadlines every few days and ensure that they are met
3) Schedule de-briefing sessions at the end of the week so that everyone is on track with the project progression
4) When doing the analyses, it's often helpful to have someone look over your work
    a. Try jumping on a zoom call and sharing your screen while you're doing the analysis so you can discuss and look over each other's work in real-time
5) Have fun!
    a. This could be the first time you are actually applying the statistical techniques you've been taught in your statistics classes to a real-world dataset. It's the skills you learn from projects like these that will stick with you in the future and help you land a job, get a research position or even ace your final exam‼