STAT 216 Final Project Guidelines

Due: Friday, December 15, 2023, at 11:59 p.m. EST

Overview

In this final project, you will explore a dataset using the methods you have learned about in STAT 216 this semester. Analyzing a single dataset, you will develop and explore 3 different hypotheses, one from each of the main topic areas:

- Case CQ: Means (t tests, ANOVA, multiple comparisons, non-parametric tests)
- Case CC: Proportions (tests of proportions, Chi-Square, non-parametric tests)
- Case QQ+: Multiple regression models

Four potential datasets to use are on Blackboard named "Baseball.csv", "BreastCancer.csv", "Pregnancy.csv", and "HomelessShelter.csv". You can choose to use one of these datasets, or propose your own to use if you find one that is of interest to you. If you select your own dataset, it must meet the following specifications:

- have at least 2 quantitative variables
- have at least 2 categorical variables
- not be time series
- be able to accommodate multiple hypotheses.
- be small to moderate sized (i.e., less than 300 observations)

You must check your dataset choice with me ahead of time (even if using one of the four listed datasets on Blackboard, please let me know your plan…I want to ensure variety in projects and that people aren't all doing the same analysis).

Components

Each of the three hypothesis that you explore should contain information on why you are exploring it, the analysis, and interpretation of the results.

Your project should have the following structure:

- Part 1: Background information about the dataset
    - This should include a description about the variables that are included in the analysis (not every variable), information that will help the reader to understand the dataset and your overarching questions of interest.
- Part 2-4: Analyses 1-3 for CQ, CC, and QQ+
  Each analysis should contain the following, *all in the context of the problem*:
    - Hypothesis
    - Rationale (why are you exploring this hypothesis)
    - Brief analysis plan
    - R code (and only the relevant R code)
    - Results of the analysis from the above R code
    - Interpretation of the results
- Part 5: Summarize your findings in a discussion of the results.
    - This should tie together the background with all three analyses into a conclusion about the data.

Alternative Approach

If you don't want to do an applied project, I can consider some requests for more theory-based projects. If you want to do some simulations and combine this with reading a journal article, or something like that, we can work that into a project.

Groups

Projects will be completed either as pairs or individually. You may choose your own partner to work with, or if you want a partner but don't have anyone in mind, let me know and I can suggest pairings. If you prefer to work individually, that is fine too, but I do encourage you to consider working with a partner.

You are only allowed to discuss your project with your partner. You cannot discuss the project with any other students. You are not allowed to use LLMs, such as ChatGPT, or other online resources in the completion of this project. You and your partner should apply what you have learned in the course to put together the project.

Please email me regarding your choice of partners.

Delivery

You have a few options in how you would like to present your results/projects:

- Client meeting/office hours talk
    - You can compile all your plots, figures, tables, etc. and present your results to me as if I were a client who came to you with the dataset. This would be an "office hour" type setup where you come with your results and walk me through what you did in a 15-minute discussion.
- Traditional paper
    - Probably a paper in the range of 10-15 pages, including all plots, figures, output, etc., organized in a pseudo-journal article form
- Poster presentation
    - You can put together an electronic poster that you then present to me in a 15-minute time slot
- Pitch your own idea to me
    - If you have a different way in which you would like to present the results of your project, let me know, and I can consider it!

A rubric for grading will be put together soon.