STOR 664: Project Assignment

Instructor: Dan Kessler

Fall 2024

The course project is an opportunity to put into practice what we have learned in class: you will apply statistical methods in order to shed light on a question you find interesting. You will need to choose your question, identify a dataset that you can use, develop an analysis plan, execute your analysis plan, and then interpret your results. Your analysis must involve multiple regression analysis in some way, but beyond that you have broad flexibility. You may work in groups of up to three students; students in the same group will receive the same grade. If you prefer to work in a group of two or by yourself, that is fine, too). While all students in a group are jointly responsible for the work, it is acceptable to divide the workload in a way that all group members feel is fair, e.g., having one member of the group focus on code/results and another member focus on writing.

In order to keep things on track, we will break the project down into a set of four smaller assignments. Each component will have its own deadline as given below. In essence, you will be writing your report section-by-section. The idea is that when each of the components are concatenated, you will have written a (brief) scientific paper that answers a question that you find interesting. As you proceed, you may choose to revise earlier sections to accord with what you do later, but I encourage you to reflect on how this process accords (or doesn't) with the way statisticians generally understand the scientific method.

Although there is some variability, scientific papers generally include four primary sections. You will tackle each of these sections in a separate assignment (with deadlines noted below). Please prepare your report as a single document; at each deadline you will submit the entirety of your report (thus far) as a PDF, although you'll only be evaluated on the most recent additions. The sections, their suggested content, and deadlines (all 11:59 PM ET) are as follows. We will endeavor to give you prompt feedback on your project after each submission so that you know if you are on the right track.

1. Introduction (Due *[2024-11-04 Mon]*)

   - What question do you want to answer?
   - Why is it important?
   - Is there any existing relevant literature?
   - What data will you use?
   - There should be no (or minimal) math in this section

2. Methods (Due *[2024-11-18 Mon]*

   - How will you answer your question?
   - What models will you fit?
   - What assumptions are you making, and will you check them?
   - What are your null hypothesis(es)?
   - It's ok to use equations and math here, but they should be well-explained
   - Will your analysis be adaptive (e.g., involve model selection)?
   - How will you account for multiple comparisons?

3. Results (Due *[2024-11-25 Mon]*

   - What were the results of applying your methods?
   - Include figures and/or tables as appropriate
   - Reserve interpretation of findings for Discussion

4. Discussion (Due *[2024-12-04 Wed]*

   - What have you learned about the question that you set out to answer?
   - Was your answer conclusive?
     - If not, what could be done to make the findings conclusive? More data? Better methods?
   - What limitations did your data and/or analyses have?
   - Were your assumptions reasonable?
   - What directions for future work do your findings suggest?
   - Final step: add an abstract to the beginning of your report

Grading The project will count for 25% of your final grade in the class. It is worth 25 points: each section will be scored on a 5 point scale when it is submitted, and after final submission the entire report will be evaluated holistically and scored on a 5 point scale. Note: In order to make this work on Canvas, the "Discussion" section assignment carries 10 points (5 for Discussion plus 5 for holistic evaluation of the entire report).

Your report should include in-text citations (and accompanying bibliography) when you rely on external resources (e.g., if your dataset is taken from a published paper, you should cite it), but in general when you rely on things you learned in class you do not need to cite the course textbook or lecture. You can adhere to whatever standards you prefer (e.g., APA). I suggest you prepare your report in LaTeX, but you can use whatever tool you prefer so long as it produces a PDF. The length of the report will vary substantially depending on how you typeset it and how many figures/tables you include, but each of your sections should contain no fewer than three paragraphs of written text (although they may be appreciably longer than that).

Your project should reflect new work, i.e., it should not be recycled from something that you have already done for another class or in the course of a research project (although you are free to use data with which you already familiar). Of course, if you find your project interesting and wish to continue working on it after the class (with an eye to eventual publication), you are free to do so. Indeed, if you are engaged in research, it may be wise to choose a project that aligns with your research (e.g., this is an opportunity to become familiar with a dataset or scientific area that you anticipate using in the course of your other work).

There should be an area for acknowledgments at the end of your report in line with the course requirements for all submitted work. In addition to the usual reporting, it should also address how each group member contributed to the project (e.g., "Shaleni wrote the majority of the report, but Dan conducted the analysis and made the figures"). It should be included at each submission deadline and should get progressively more detailed as you work through the project.