

Statistics 3220
Spring 2019
Final Project (25% of Final Grade)

Introduction

Purpose

Often times in "the real world" when you are given a data set or are trying to answer a research question, you may not be equipped with the analysis skills to answer the question. You will have to research and learn a technique on your own. Additionally, you will need to communicate the understanding of the new technique and your analyses to any audience. Throughout the semester, I hope that I have provided you with opportunities to develop the skills to be able to do just that. The purpose of the final project is to demonstrate your ability to learn a new analysis technique and present an analysis.

Groups

You will work in teams of 3-4 students. You will fill out a [Google Survey](#) with your group members and topic preferences by Monday, April 8th at noon (the sooner the better). If you would like to change groups, but need help finding a new group see/email me before that time. It is your responsibility to communicate your concerns directly to your group membership and your anticipated workload. Each member will contribute equally to the group. If you are having difficulties with a group member, please inform the instructor immediately to help in the communication. At the mid-point and conclusion of the project, you will submit a team evaluation.

Important Dates

- Monday, April 8th: Class time is used for new material from Ch 11 & Group work
- Friday, April 19th @ 9pm: Final Project Proposal is due to Collab (including individual "HW Problems")
- Wednesday, April 24th: Class time is used for the project (Required)
- Monday, April 30th: Class time is used for the project (Required)
- TBD: Project Workshops and Individual Meetings with me
- Thursday, May 2nd @ NOON: Final Report and Presentation Slides Due to Collab
- Thursday, May 2nd @ 2-5: Section 001 Presentations
- Saturday, May 4th @ 2-5: Section 002 Presentations

Proposal: Due April 19th @ 9pm (100 points)

The purpose of the proposal is to be sure that you understand your technique and have selected appropriate data. The proposal has an individual component and a group component. The individual component is listed below and consists of mid-project evaluation and textbook problems related to your topic. The group component will be submitted as a written report to Collab as a .pdf document. There is a page limit of 3 pages (excluding the title page). The proposal should be readable in terms of the written work and formatting. Reports submitted in a different format or late will not be accepted. The proposal should include the following sections: You will have the opportunity to get feedback while you complete your proposal if you would like. Then you will receive complete feedback before you go on to complete the project. The following components are to be included in your group proposal:

Title Page

Your title page should include the names of your group members, the title of your project (something related to the research question, not just "STAT 3220 Final Project: Proposal"), the course number and section, and the semester.

Methodology Introduction

Here you will explain at least two advantages and two disadvantages of your technique. You do not need to explain the method in detail (that will be done in the final report). Consider example of when this method is useful and when it is not (type of data or other situations), the computation power required, the ability to interpret results practically, etc. This should be about 2 paragraphs.

Research Question

This section is brief summary of the research question you intend to answer. This section should include motivation behind your research question (why it is interesting and relevant) and what your question is.

Data Summary

Data Description

This should include two aspects: information about how the data were collect (what study it was used for) and a data dictionary. For the former, you should include the information you have about the study (who, what, when, where, why, etc). If there is a link to a paper or site, you should include that as well, but you must use your own words to describe the study. The second part is to present a data dictionary (code book) of the variables you will use in your analysis. The code book should include the names of the variables, the description of the variables, and whether the variable was quantitative or categorical. If quantitative, include

the units; if qualitative include the levels and how you will define the dummy variables. This section should clearly state your response variable. (**NOTE: Depending on your topic, you will have different guidelines for the number/type of variables.**)

Why the data is appropriate for your technique

You should explain how you know this data is appropriate for your technique. (EX: If your technique was MLR, you would say that your response variable is continuous and you intend to use the model for prediction. etc.) This section might also include some EDA if necessary to show that you have met the conditions for your data.

Copy of the Data

You are to submit a .csv or .txt of your actual data. This data file should include any transformations you decided to do, such as recoding of variables. If you have a data set of more than 200, you might consider taking a sample (depending on your technique). If you need help with taking a sample of your data, please see me.

Individual Component- Submitted Separately

To demonstrate your understand of the technique, you will submit responses to "textbook questions". You and your group (and others with the same topic) will have the same assignment, so like a mini homework, you may work together, but you will submit your own work and answers. Additionally, you will submit a evaluation of your group and your contribution to the proposal. Both will be submitted as one "Test & Quiz" in Collab.

Final Report: Due May 2 @ NOON (200 points)

To demonstrate your ability to communicate your findings formally in writing, you are required to submit a report. The final project will be submitted as a formal statistical report to Collab as a .pdf document. Reports submitted in a different format or late will not be accepted. There is a page limit of 10 pages-single spaced, 12 point font, 1 inch margins, Times New Roman- excluding the appendix and SAS code. The report should be readable in terms of the written work and formatting (page numbers, section titles, complete sentences). Follow the statistical report writing document and the Project 1 workshops for what should be included in each section of the final report. The final report should include the following sections:

Title Page

Your title page should include the names of your group members, the title of your project (something related to the research question, not just "STAT 3220 Final Project"), the course number and section, and the semester.

Introduction

This section will be similar to your "research question" section from the proposal, but should also include a brief summary of the analysis you used.

Data Summary

This section will include the data summary from your proposal unless there was a change to the data, research question, or major/minor improvements were recommended. Additionally, you will also need to add a complete exploratory data analysis (unless you submitted that in the proposal).

Methodology

This is something that is different from project 1. You will provide a thorough, statistical description of the technique. Include: What are the assumptions of this procedure; Describe the exact procedure of method; When is this method most useful; What are the drawbacks to this procedure (you may copy or elaborate on the section "Methodology Introduction" section from the proposal). This section should include about one page and does not need to include any proofs that we would not perform in our class. To complete this section you are to include two additional relevant sources (can use specified section from the textbook if applicable, resources at the end of the chapter, another textbook, etc). This section will have no context to your research question.

Analysis

Now you will perform the analysis on your data. You should be thorough (as you were with Project 1). Include any model building process, testing procedures, checking for assumptions, assessing the model, etc. This section will end with a final "best" model.

Conclusion

Your conclusion should consist of 2 parts: interpreting your results of the analyses in context of the problem and commenting on areas of future improvements. Then you would comment on a limitations of this analysis and your data and how you would continue or improve on this research in the future. (Again, similar expectations of Project 1)

SAS Code

You are to include the code used in your analysis. It should be annotated so it is clear what each step was used to do.

Optional Appendix

You may put any output or graphics that were not needed. Reminder this section will not necessarily be read. Do not refer to anything in the appendix that is vital to the understanding of your project.

Peer Evaluation- Submitted individually

You will submit final evaluation of your group and your contribution to the project as a "Test & Quiz" in Collab by 5pm on Sunday, May 5th. (This gives you time to write about presentations as well.)

Presentation: Slides due May 2 @ NOON, present during final exam period (50 points)

To demonstrate your ability to communicate your findings orally, you are to give an 5-10 minute oral presentation to your peers (and me!). You will submit the slides for your presentation by May 2 @ NOON as a pdf to Collab. You will then give your presentation during the final exam period. For each class presentations will be split into two parts: half of the class will come to the first half (2-3:30) and half of the class will come to the second half (3:30-5). Information on presentation order will come at a later date. Your presentation will be graded on the following:

- Visual Aesthetic
- Includes all components
- Clearly Presented: Written slides are clear, Verbal presenting is clear, All analyses are correct
- Equal presenting from entire group
- Peer Evaluation of other groups: You will submit an evaluation of 2 other groups on paper during the exam period.

OPTIONAL: Submit to a Competition

There is an opportunity to submit your project (for any class project) to **Undergraduate Statistics Project Competition and Electronic Undergraduate Statistics Research Conference**. More details can be found here: <https://www.causeweb.org/usproc/>. If your group is interested, let me know and I will meet with your group when the project is complete to help meet the guidelines for submission.

Topics

You will be assigned one of the following topics. Depending on the number of groups there will be a maximum of 2 groups in each class assigned the same analysis. I will provide at least one resource if the topic is not described in detail in our textbook. Remember that you will still need to find at least two additional scholarly resources beyond what I provide to write your methodology section. These can come from references listed at the end of each chapter, another textbook, a journal entry, etc.

1. Model Validation (Ch 5.11 provides a basic overview of some procedures, you may use those or other techniques)
2. Weighted Least Squares (Ch 9.4 provides a basic overview, I would expect more advanced techniques)
3. Logistic Regression (You will need to consult other sources for assessing the model, assumptions, other techniques, etc.)
4. Ridge Regression (Ch 9.7 provides a basic overview, but you will need to consult other sources for how to apply it)
5. Lasso Regression (Not in our textbook)
6. Robust Regression (Ch 9.8 provides a basic overview, but you will need to consult other sources for how to apply it. Note there are several techniques.)
7. Time Series Regression (Ch 10.4-10.10 lists several options of how to compose an appropriate model)
8. Two Factor ANOVA with a comparison of different multiple comparison tests (Ch 12.8 provides information behind some of these tests)
9. Three factor ANOVA (Ch 12.6 provides a good overview)
10. Poisson Regression (*Practicing Statistics* Ch 8)
11. Survival Analysis (*Practicing Statistics* Ch 9)
12. Select another technique that is not listed above (two examples are Robust Regression and Box-Jenkins). The technique need to be related to modeling for prediction/estimation.