**DATA 557**

**Final Project**

**Winter 2020**

**Instructions**

**1. General Instructions**

This is a group project. Groups should consist of 4-5 students. The requirements for this project are to identify a data set and a set of questions and/or hypotheses to be addressed using the data. The project requires an in-class group presentation, and a written report. The project will be developed through a series of steps (see timeline at end) designed to give you feedback and allow revisions. Your project will be evaluated based on your in-class presentation and final written report; evaluations will consider how well your presentation and final report incorporated feedback received on preliminary submissions. There will be time at the end of each class to work on the projects; however, additional time will need to be spent outside of class to complete the project.

**2. In-class Presentation**

Your group will present your project in class on March 11. The time length for each presentation will be 15 minutes. Your presentation materials can be prepared in any format that can be projected on the classroom desktop, e.g., ppt or pdf. The presentation should briefly describe the aims, including statements of questions and/or hypotheses, the statistical methods, and results. There is no limit on the number of slides; however, it is recommended to use approximately 1 slide per minute.

**3. Written Report**

Your final written report should describe the aims, data set, statistical analysis methods, and results. The written report should be at most 20 pages (single-spaced, at least 11pt font, approx. 1-inch margins), including all tables, graphs, references and appendices. Pages should be numbered. The report should be submitted in pdf format.

Suggested organization of the report (this is optional)

(a) **Abstract**. This should consist of concise statements of the aims, methods, results, and conclusions (eg, 200-300 words).

(b) **Introduction**. Include a clear statement of the aims and questions to be addressed. The goal of the analyses should be clear from the introduction. You may include background material, such as previous research that your analysis builds on, but this should be kept brief.

(c) **Data Set Description**. Describe the data set, including the source of the data, data collection methods, study design (e.g., randomized experiment, observational data, survey, etc.)

(d) **Statistical** **Methods**. Describe the statistical methods used for the analysis, including discussion of underlying assumptions and explanation for why the methodology you used is appropriate.

(e) **Results**. Describe your results clearly and concisely. Use graphical displays and tables to convey descriptive information about the data and the results of the analysis.

(d) **Discussion**. This section should briefly summarize the results and conclusions. Also describe limitations of the analyses, including limitations of the data set as well as of the statistical analyses.

(e) **References**. List books or articles you consulted. (It is not necessary to do a lot of background research, so the reference list should be short.) References for statistical methods used in class (e.g., t-tests, and linear regression) are not required, but references should be given for advanced methods not covered in class.

(f) **Tables and Figures**. These can be included with the text in the body of the report or in a separate section at the end. All should have clear titles/captions.

(g) **Appendices**. You may include more technical aspects of your analysis in an appendix if it will help with the readability of the main body of your report. Appendices are included in the 20-page limit.

**4. Evaluation**

* You will be evaluated separately on your written report and your in-class presentation. The criteria used to evaluate your project are as follows:

(a) **Aims/Questions/Hypotheses**. Are the aims, the specific questions and hypotheses (if any) clearly described? Can the questions be answered using the data set?

(b) **Statistical Methods**. Do the statistical methods used address the aims/questions/hypotheses? Are the methods appropriate? Are the assumptions evaluated?

(c) **Execution of the Analyses**. Are the analyses executed correctly and accurately?

(d) **Interpretation and Presentation of Results**. Are the results interpreted correctly and presented clearly, including both descriptions of results in words as well as with tables and figures?

(e) **Quality of the Oral Presentation**. Do the oral presentation and the presentation slides clearly describe the aims, the data set, and the results? (This will be evaluated by your fellow classmates as well as the instructor and TA.)

(f) **Quality of the Written Report**. Is the written report organized and written clearly and accurately? Are the tables and figures effective in describing the analyses?

**5. Advice**

* Try to keep both your analysis methods and your oral and written presentations as non-technical as possible. A reader with limited understanding of statistics should be able to understand most of your presentation and report.
* You should choose an analysis strategy that addresses the questions posed in a reasonable way and justify your choice of methods. It is recommended that you use methods discussed in class for most of your analyses. It is not necessary to find the optimal statistical method for your problem. If more advanced statistical methods would be helpful you may include some results using these analyses or describe these methods.
* Keep in mind that writing is very time consuming and you should not leave the preparation of oral presentation and written report to the last minute. Divide the writing responsibilities among the members of the group.

**Timeline for Final Project**

Wed January 22: general discussion of project in class

Jan 22-29: during this time identify a group to work with

Wed, January 29 – in-class group discussions to formulate scientific questions and identify data source

Mon, Feb 3 – **submit** preliminary short proposal

Wed, February 5 – in-class group discussions to formulate a short proposal (a 1-page summary of your project including scientific questions and data source)

Monday, February 10 – **submit** short proposal, submissions will be posted on canvas

Wed February 12 – in-class group discussion to revise proposal based on feedback, begin to make analysis plans

Wed February 19 – in-class group discussion on analysis plans

Monday, February 24 – **submit** revised proposal with analysis plans (3-5 pages total)

Wed February 26 – work in class on data analysis

Wed Mar 6 – work in class on analysis, and preparing presentation slides and report

Sunday Mar 8 – **submit** draft presentation slides

Wed Mar 11 – in class presentations

Wed Mar 18 – **submit** final written report