

Statistical Inference Project Overview

Every student in Statistical Inference 1 is required to complete a semester project analyzing real data. The subject area, goal, and topic are entirely up to you. To help make sure your project plan is viable and within the scope of the class, you should plan to meet with the instructor prior to turning in your data analysis and presentation.

Important due dates:

Task	Due Date
Project proposal and group formation	November 1
Submit your data (or a sample of your data if the data set is too large)	November 17
Rough draft of data analysis	December 1
Group presentations (we may have to stay slightly beyond our scheduled class time)	December 12
Division of labor write up—write a summary of which group members completed which parts of the project. This will help me assign individual grades.	December 12

Important notes:

- You ARE required to work in groups for the project, but you can divide the work in many different ways. Just keep track of who does what, so that everyone can get appropriate credit. Groups can be 3-5 people, depending on the size of the project.
- In general, I suggest choosing a research question that can be broken down into smaller statistical question. Each person in the group should take responsibility for analyzing at least one statistical question.
- Alternately, one person can focus on accessing/scraping the data, one person can focus on cleaning/managing the data, one person can focus on summarizing and visualizing the data, etc. It depends on how intensive these tasks are in your project. Keep in mind, however, that the focus of this class is **statistics**, so it's a good idea for each person to demonstrate their understanding of some statistical concepts.
- There is no single correct way to divide the workload for your project. Check with me if you aren't sure, and make sure that you document which group member completes which tasks.

Your final presentation should be submitted in R Markdown (you can choose whatever output format you prefer; it doesn't have to be beamer). Presentations will take place on December 12, and should be about 15 minutes per group. Do not go over your time limit! If you have trouble fitting everything in to your time limit, come see me and we can discuss cutting down parts of your project for the presentation.

Statistical Inference Project Overview

Presentation Components:

1. Introduction
Give some background on your topic. Explain any relevant ideas or concepts. Identify any populations you will be working with. Why is there a need to research this topic? State the general form of your research question.
2. Literature Review/ Background and Methods
Conduct either a formal literature review or an informal background research overview to figure out how to analyze your topic. What do experts say about how to study your topic? How have other studies measured these variables? What data is available? What data is it possible to measure? Keep in mind your ultimate goal.
3. Research Question(s)
Based on your literature review, formulate your specific research question(s) and any sub-questions. Use a statistical perspective in framing your question(s); in other words, refer to the parameters of interest and the types of conclusions you will draw.
4. Data Collection & Experimental Design
Analyze the target population and sampling methods for the data you accessed. Identify any potential sources of systematic or random errors. If appropriate, discuss alternate forms of data collection that you might use (in an ideal world). Use this discussion to demonstrate your understanding of the topics we covered in experimental design at the beginning of the semester.
5. Data Analysis (graphs, tables, etc.)
Explore and summarize your data with numbers and graphs as appropriate. Are there any problems with your data (e.g. missing values)? Are there any surprises in your data (e.g. outliers)? Visualize your data using ggplot or other R tools. Conduct inference (hypothesis tests, confidence intervals, etc.) to address these questions as appropriate. Explain your results in writing. Possible analysis includes:
 - a. Hypothesis test for a mean or proportion (to test a claim based on previous research, for example)
 - b. Hypothesis test for the difference of two means or two proportions (to compare two groups)
 - c. Confidence intervals to estimate parameters
 - d. Regression analysis and inference to study the relationship between your variables; include F-tests, t-tests, ANOVA.
 - e. Model selection and regression diagnostics, where appropriate.
 - f. Transformations of data, if appropriate
6. Conclusions

Statistical Inference Project Overview

What is the main idea people should take away from your project? What do the results of your analyses show? What recommendations can you make? What limitations are there to your study? What would you like to follow up on in a later study?

7. Works Cited

Include a list of all sources you used for this project. You may choose whichever citation method you prefer (MLA, APA, etc.), but each source should be clearly identified.

8. Appendix

Include any details about the statistical analyses performed that you did not include in the data analysis section: calculations, simulations, formulas, etc.