STAT 3690 Final Project Guideline

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Due at 11:59 pm of Apr 25, 2022

The final project serves as an opportunity to apply what you learn in class to real datasets. Also, it is a practice of using `R` as well as polishing writing skills. Each of you should submit one individual written report for final assessment.

Finding a dataset

Feel free to pick up a dataset interesting to you. Meanwhile, it is expected to 1) *be publicly accessible* AND 2) *include some data points collected in the recent ten years* (i.e., NOT the entire dataset was created before/in 2012). Here are some repositories of datasets for your reference, but it is not mandatory to stick to them.

- Data And Story Library (DASL, dasl.datadescription.com)
- Inter-university Consortium for Political and Social Research (ICPSR, www.icpsr.umich.edu/icpsrweb/ICPSR)
- Kaggle (www.kaggle.com/datasets)
- The Cancer Genome Atlas (TCGA, www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga)
- UCI Machine Learning Repository (archive.ics.uci.edu/ml/index.php)

Data analysis

Typically, a data analysis begins with a description of the dataset. What follows is an exploration involving summary statistics and/or data visualization. A scientific hypothesis/question/concern is further induced. In verifying the hypothesis, you ought to choose appropriate statistical methods and justify your choice. Specifically, pay attention to the prerequisite/assumptions of each method and explore the extent to which they are met by your data.

Although there is no strict restriction on methods involved, the analysis should be centered around the application of methods introduced in our course.

In addition, the analysis should be numerically carried out by using `R`, if possible.

Written report

A written report of **no more than 8 pages** (excluding tables, figures, appendices and bibliography) must be prepared and submitted electronically in the PDF format. It should cover the following parts.

1. *Abstract*: a short description of your dataset as well as the problem you are trying to solve; the purpose of your research; the methods used to find the solution; the results and implications of your findings.
2. *Introduction*: data description; data source (i.e., how to access the data); summary and exploratory analysis of data; relevant literature review; clear scientific hypothesis/research question.
3. *Methods*: role of each method clearly stated; methods and assumptions described accurately, including the necessary justification.
4. *Results*: results accurately illustrated; research question adequately answered.
5. *Discussion/Conclusion*: results clearly and completely summarized; limitations and/or concerns stated.

6. *Appendix*: `R` code trunks enclosed.
7. *Bibliography* (if needed).

More appreciation will go to reports with merits including but not limited to: ideas presented in a logical order, sections well-organized, no grammatical/spelling/punctuation error, AND appropriate level of details.