

Final Project

Statistics 679 — Spring 2018

Description

For the final project, pick a group of from 2–4 students. Obtain a data set on a topic of your choice, identify questions of interest, carry out a data analysis using multilevel modelling methods presented in the course, and write a summary of what you learn about the questions of interest. The final report should be between 8 and 12 pages, including graphs. Prepare the report using RStudio and knit into a PDF document. Include at the end of the written report, a reflection on what you have learned by doing the final report. You also need to turn in data and code adequate to replicate your analysis.

Due Dates

- *April 27, 11:59pm*: Let me know by email (from all group members) who is in your group and what data you plan to analyze.
- *May 8, 11:59pm*: Upload:
 - a PDF report on Canvas;
 - any R and Stan code needed to replicate your analysis;
 - a file with your data.
- *May 8, 11:59pm*: Answer a brief questionnaire about contributions made by group members.

I will set up Canvas so a single upload suffices for the group once I know your groups.

Data

There are very few limitations on the data set you use, as long as it obtained legally and ethically. - You must report the source of the data and include a copy of it with your report. - Your analysis must differ substantially from any other analyses publically available for this data. (Don't find a Stan case study and just replicate it.) - Choose data where there is an interesting question to you that can be addressed. - Your data may come from an R package or a data repository — but **bonus points** if you gather a data set that has not been previously distributed.

Grading Rubric

I will evaluate your projects by scoring you from 0 to 5 for each category according to this grading rubric.

- 5: **Excellent** — greatly exceeded expectation; a model for others to follow.
- 4: **Very good** — very good job deserving of a solid A grade.
- 3: **Adequate** — passing, but some minor deficiencies.
- 2: **Substandard** — at least one major deficiency; below expectations.
- 1: **Poor** — several major deficiencies; well below expectations.
- 0: **No Credit** — this element of the report is either completely missing or only superficially included.
- Data:
 - Is the data set appropriate for the course?
 - Are there interesting questions to address with the data?
 - Are there complexities in the data that make the subsequent analysis a good learning experience?

- How much effort did you put forth to collect the data?
- Do you describe what each variable means?
- Do you describe how the data was collected?
- Do you report the source of the data?
- Does your report include a brief table with a descriptive summary of the data (including key numerical summaries of key variables, samples sizes, and so on)?
- Graphics:
 - Do plots show key features of the data well?
 - Are there good labels?
 - Do you make effective use of color/shape/symbols?
 - Do your plots provide information about the questions of interest?
 - Are plots an appropriate size for their relative importance?
 - Are there any plots that ought to be included, but are not?
 - Are there plots that serve very little purpose?
 - Are there plots that help to assess model fit?
 - Are there plots that help to make inferences about questions of interest?
- Analysis:
 - Have you selected an appropriate model for your problem?
 - Can you justify well your choice of analysis?
 - Does your analysis address the questions of interest?
 - Does your analysis match the structure of the data (such as the design of the experiment or the process of data collection)?
 - Are you making good choices when presenting numerical results of your model?
 - Can the reader understand your model description?
 - Are summaries of fitted models well done?
 - Does your report include a comparison between a model that uses a multi-level approach and one that does not?
 - Does your report explain the process of how you built from a preliminary to a final model?
- Communication:
 - Is your report well organized?
 - Is it easy for me to understand the questions you wish to address?
 - Do you clearly describe your analysis approach?
 - Can I understand your results and how they relate to the questions of interest?
 - Is your language usage good (with understanding on my part that groups without native English speakers may have more challenges in excelling)?
 - Are there clear section titles?
 - Does the report have a clear and effective format?
 - Does each table or figure include a caption or paragraph that helps to interpret the information contained within?
 - Does your report effectively interpret results and how they address the questions of interest?
 - Does your report contain an effective summary?
- Code:
 - Can I easily see what your code does?
 - Can I read in your code and run it on your data with little effort?
 - Does your code include some helpful comments?
 - Does your code correctly implement the analysis described in the report?
- Reflection:
 - Does your report contain a section that reflects on your learning?