Final Project

Samantha-Jo Caetano

<u>Rough Draft Due Date (2%)</u>: Wednesday December 9, 2020 at 11:59pm ET

<u>Peer Review Due Date (3%)</u>: Monday December 14, at 11:59pm ET

<u>Final Report Due Date (25%)</u>: Monday December 21, 2020 at 11:59pm ET

This Final Project is to be handed in as a report.

This final project should be completed in an R markdown file and should be knit to a pdf document. Your submission will have 3 parts: (i) Output/Final Copy of Report; (ii) R markdown code, .Rmd file; (iii) link to a GitHub repository of your code (this will include your .R scripts for cleaning the code).

Please have all three files available for submission at the due date.

Your Objective

To perform a meaningful statistical analysis on some survey, sample, or observational data.

Note: There is a peer review component to this project.

General Requirements

- As an individual you will select one of the options (a-e), you will perform the appropriate analysis and you will write a report.
- On December 9th you will submit a pdf of a rough draft to be edited by your peers.
- From December 10-14 you will provide feedback on some of your peers' rough drafts.
- On December 21 you will submit a pdf and Rmd of the final report, as well as a GitHub repo link.
- The final report will be a well written and revised document consisting of the following sections (more details in "Report Details"):
  - Title & Authors
  - Keywords
  - Introduction
  - Methodology (Data and Model)
  - Results
  - Discussion
  - References
  - Appendix (Optional)

Options

Working individually, please conduct original research that applies statistics to a question involving surveys, sampling or observational data and then write a paper about it. You have various options for topics (pick one):

    a. Develop a research question that is of interest to you and obtain or create a relevant dataset. This option involves developing your own research question based on your own interests, background, and expertise. I encourage you to take this option, but please discuss your plans

with me. How does one come up with ideas? One way is to be question-driven, where you keep an informal log of small ideas, questions, and puzzles, that you have as you're reading and working. Often, after dwelling on it for a while you can manage to find some questions of interest. Another way is to be data-driven - try to find some interesting dataset and then work backward. Finally, yet another way, is to be methods-driven - let's say that you happen to understand Gaussian processes, then just apply that expertise to an area. (If you select this option it is recommended to incorporate some causal inference of observational data into your report.)

b. (Thanks to Jack Bailey for this idea) Build a MRP model based on the CES and a post-stratification dataset that you obtain, to identify how the 2019 Canadian Federal Election would have been different if 'everyone' had voted. What do we learn about the importance of turnout based on your model and results?

c. Reproduce a paper. Options include:
    -   Angelucci, Charles, and Julia Cagé, 2019, 'Newspapers in times of low advertising revenues', *American Economic Journal: Microeconomics*, please see: https://www.openicpsr.org/openicpsr/project/116438/version/V1/view.
    -   Bailey, Michael A., Daniel J. Hopkins & Todd Rogers, 2016, 'Unresponsive and Unpersuaded: The Unintended Consequences of a Voter Persuasion Effort', *Political Behavior*.
    -   Clark, Sam, 2019, 'A General Age-Specific Mortality Model With an Example Indexed by Child Mortality or Both Child and Adult Mortality', *Demography*, please see: https://github.com/sinafala/svd-comp.
    -   Skinner, Ben, 2019, 'Making the connection: Broadband access and online course enrollment at public open admissions institutions', *Research in Higher Education*, please see: https://github.com/btskinner/oa_online_broadband_rep.
    -   Pons, Vincent, 2018, 'Will a Five-Minute Discussion Change Your Mind? A Countrywide Experiment on Voter Choice in France' *American Economic Review*.
    -   Valencia Caicedo, Felipe, 2019, 'The Mission: Human Capital Transmission, Economic Persistence, and Culture in South America', *The Quarterly Journal of Economics*, please see: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ML1155.
    -   If you have a favourite paper then please let me know by the end of Week 12 so that I can check that it's appropriate.

d. Pretend that you work for Upworthy. Request the Upworthy dataset and then use it to evaluate the result of an A/B test. This request could take a week. Please plan ahead if you choose this option.

e. Critique the following paper: AlShebli, Bedoor, Kinga Makovi & Talal Rahwan, 2020, 'The association between early career informal mentorship in academic collaborations and junior author performance', *Nature Communications*. You should be able to download the data here: https://github.com/bedoor/Mentorship and the paper here: https://www.nature.com/articles/s41467-020-19723-8. For background please see: https://statmodeling.stat.columbia.edu/2020/11/19/are-female-scientists-worse-mentors-this-study-pretends-to-know/ and https://danieleweeks.github.io/Mentorship/#summary.

Process for December 9 11:59pm ET

- As an individual, via Quercus, submit a PDF of your rough draft on Quercus by 11:59pm ET on Wednesday, December 9, 2020.
- At a minimum this must include your title and a fully written Introduction section.
- You will be awarded 2% for completion of the total 30% for the Final Project.
- It is recommended that you also include the (partially completed) reference section here, but this is optional.
- You do not need to include your name in the pdf if you prefer to stay anonymous to your peers.
- The point of this is to get feedback on your work (and to make sure you have at least started thinking about this by December 9th) so you are more than welcome to include other sections that you wish to get feedback on.

Disclaimer: There will be no extensions granted for this submission since the following submission is dependent on this date.

Process for December 14 11:59pm ET

- As an individual, on December 10, you will randomly be assigned a handful of rough drafts to provide feedback. You have until December 14, 2020 11:59pm ET to provide feedback to your peers.
- If you provide feedback to one peer you will receive 1%, if you provide feedback to two peers you will receive 2% if you provide feedback to three (or more) peers you will receive the full 3%.
- Providing feedback is obviously subjective, so we have established a set of minimum requirements:
  - Your feedback must include at least 5 comments (meaningful/useful bullet points).
  - One comment on the appropriateness of the title.
  - One comment on the readability of the writing (I.e. address any edits, grammar, typos, etc.)
  - One comment on how interesting/compelling the writing and potential analysis is.
  - One comment that states whether it is clear which option (a-e) was selected.
  - One comment/suggestion a foreseeable model, weakness, next step, data, reference, etc. (Just give them something useful to work off of.)

Disclaimer: There will be no extensions granted for this submission since the following submission is dependent on this date.

Disclaimer: Please remember that you are providing feedback here. All comments should be professional and kind. It is challenging to receive criticism, and arguably more challenging to provide criticism, and even more challenging to give criticism strictly through text. Please remember that your goal here is to help your peers advance their writing/analysis. Any feedback that is inappropriate will receive a 0 on this section.

Process for December 21 11:59pm ET

- As an individual, via Quercus, submit a PDF of your paper. Again, in your paper you must have a link to the associated GitHub repo.
- Via Quercus you will need to submit the following three files:
  - pdf of your final report.
  - your group .Rmd file.
  - a link to a Github repository with your materials.
- This submission will be graded based off the rubric posted on Quercus and will be worth 25%.