BST 210 **HW 8: 2nd Project Check-In**

Due 11:59pm, Sunday, November 8, 2020

The goal of this project check-in is to flexibly promote further progress and thinking in your group project work, with relevance to the methods we've learned in class thus far. It is important to contemplate your research questions in the context of various method frameworks, even if ultimately you decide to utilize results from only a subset of those. You will also begin draft writing certain sections, which will serve to facilitate organization of thoughts and plans, possible adjustments, and a feedback loop for informing your actual analyses and progress. Your final reports will include more than a journal submission would require, as this is a semester-long project for a statistics course. Throughout the process of completing this project however, you should be thinking in terms of including sections, arguments, figures and following the general approach of, the preparation of a manuscript (with the presumption that it will flexibly include more than an article would).

1. **Group Identification**: Print your Group Number, Group Name if you have one, and the names of each group member, clearly at the top of your 2nd Project Check-In submission. You will submit this assignment as a Group (not individually). (The portal is set up to accept this assignment in grouped format.)

2. **Background knowledge, feedback, and eyes on the field**:
    a. Have you done a literature review of similar problems, data and questions in order to inform your own? Have similar questions been posed/answered before? What analysis methods were used? Do you plan to incorporate anything gained from a literature review in your project, and if so, what?
    b. Have you reviewed your peers' and Teaching Team's review(s) of your 1st Project Check-In?
        i. Is there any insight or advice that was shared with your group that you feel is viable for incorporation into the direction of your project? (yes or no, and if no, please explain)
        ii. If yes, how will you implement these ideas, and why?

    ** Have you contacted the (non-teaching team) domain experts you listed in the 1st Project Check-in, as needed?

3. **Analysis Plan**: Clearly update your analysis plan at this stage, or specifically outline/delineate such if you did not earlier. This may simply involve a few updates and additions, or it may require more

substantial changes, given you've progressed further into the course and thus have been exposed to additional analysis methods.

4. **Missing Data**: Explain your assessment of the presence of missing data in your project.
    a. Roughly what type of missing data do you think exists? How did you arrive at this? Explain the steps you took to discover whatever you have learned about missing data in your project, and whether these missing data require attention of some sort.
    b. What steps might you take to accommodate missing data in your project? Please explain the specific steps, and if you cannot, at least explain the thought process – not everyone is able to viably manage missing data in a study in a strictly formal sense, but you should be able to explain what you will do to address what you believe to be the situation, and why you would take the steps you intend to.

5. **Modeling**:
    a. Linear, flexible/additive or other methods (LASSO, ridge) from this topic:
        i. If your analysis plan includes any of these approaches, please begin fitting such models, performing diagnostics and evaluation, model selection, interpretations, etc. Please give some summary bullet points including model statements, whether assumptions are met, any transformations, fitted model interpretations, plots which might be important, etc.
        ii. If your analysis plan does not involve these methods, please consider a way to formulate any variables in your data to which you can then apply these models. Often these can be very interesting secondary analyses, or may lead to primary analyses of interest. Sometimes these methods can simply be parts of an exploratory data analysis step.
        iii. If your data and project are truly not supportive of any linear, flexible or other approach in this topic, please answer No to this problem and explain why not.
    b. Logistic, multinomial, ordinal:
        i. If your analysis plan includes any of these approaches, please begin fitting such models, performing diagnostics and evaluation, model selection, GOF, interpretations, etc. Please give some summary bullet points including model statements, whether assumptions are met, fitted model interpretations, plots which might be important, etc.

ii. If your analysis plan does not involve these methods, please consider a way to formulate any variables in your data to which you can then apply these models (most data can be categorized in some meaningful form). State clearly how you created such new variables, and your reasoning behind cutoff points and groupings. Often these can be very interesting secondary analyses, or may lead to primary analyses of interest.

iii. If your data and project are truly not supportive of any logistic, multinomial, or ordinal models, please answer No to this problem and explain why not. (Most data can be categorized in some meaningful way--choosing this option simply because you don't want to spend time attempting these modeling methods, is not sufficient.)

c. Poisson:

i. Please consider ways to evaluate your data involving a Poisson Regression approach, even if your intention was not initially to include such (perhaps this could be a sub-study, or used to answer secondary questions). Outline how you would set up this analysis, and then fit one or more Poisson models and interpret the results. If your data were not collected or arranged in a format that would facilitate a Poisson approach, please consider whether there are ways to reformulate a variable(s) of interest in such a way that would warrant Poisson regression.

ii. If your data and project are truly not supportive of any Poisson modeling, please answer No to this problem and explain why not. (Choosing this option simply because you don't want to spend time attempting any Poisson modeling, is not sufficient.)

d. Survival Analysis:

** This question is mainly for those who are already considering survival analysis as part of the project – others can wait until we get to this part of the course to decide, or you can begin thinking about it a bit now too.

i. Are you planning to incorporate survival analysis into your project (Definitely yes, Not sure yet, or Definitely No as our data don't warrant it)?

ii. If yes, please delineate here what you are initially planning (write the outcome, predictors, and state the general question you seek to answer)– this can be just a few sentences. If not sure yet, that is fine, and you may consider such once we reach that Topic in the course. (We will cover this topic in upcoming weeks, so you are not expected to understand or even be that familiar with these methods, but if you are even contemplating survival techniques for your project per, please share whether

this is the case, here.) If definitely no, please explain why not (time-to-event data is specifically collected as such, though depending on your data and project, there may be a way to restructure data into this format, given what you currently have to work with).

6. **Writing:**
    a. Please complete a solid draft of a project Abstract (you won't have Results or Conclusion yet presumably, but you can begin to get into place any background, aims, methods, keywords, and more), just as one would see in a strong journal article. You may even want to consider setting up tables that you will populate later, and outline where your sections will be placed, or what types of figures you may want to include.
    b. Please complete a solid draft of the Introduction/Background, Data description and motivation.
    c. Please create an outline of your Materials and Methods. (You may also outline any Results or Discussion sections, if you feel equipped to do so.).

    ** The point is to continue with planning, EDA and data analysis, but also to get writing where you can! This should all be balanced with the caveat that you are still exploring with analyses and direction, and will finalize which methods and results to include, discuss, and report, down the road in the final report. Any writing that you begin to draft now, can be updated. Some students may wish to create an overall outline of their final report, and begin to write and populate a few sections now, to help spread the work out over a greater interval of time, rather than needing to do it at the end of the term. Groups should direct their own work and feel free to meet, communicate, and update the project, throughout the course.

7. **Appendix**: Include all code and output (at this stage; clearly this will be ongoing) in an appendix for the teaching team (<u>only</u>) to reference.

8. **Intentions**: (Preliminary) Is this group's intention to try to reach publication (e.g. abstract, poster, paper) involving work/results from this project? (Publication is by NO means a requirement, and you may not know for sure at this stage what your group may ultimately decide. There is no right or wrong answer to this question.)