CS7200 **Project guidelines** Spring 2020

The goal of the project is to either gain hands-on experience with a real-life statistical problem of your choice, or explore the properties of a statistical method.

Completion of the project includes a written and oral project proposal, a written and oral project report, and a written review of the project by another team. Each group should work independently, but you are welcome to discuss technical issues on Piazza.

Projects having any degree of similarity with work by any other group, or with any other document (e.g., found online) is considered plagiarism, and will not be accepted. The minimal consequence is that all the group members will receive the project score of 0, and the best possible overall course grade will be C. Additional consequences are described at
`http://www.northeastern.edu/osccr/pdfs/Resources/Faculty_Guide_to_Academic_Integrity.pdf`

Group members: due Tuesday, February 11.

Please form groups of 1-3 class members who will work jointly on a project.

On the due date each group should submit a document stating (1) name, (2) department and (3) year of study of each group member. Please talk to me if you can not find a group. I may need to change group memberships to ensure a balance of skills within and between groups.

Project proposal: due Friday, February 27.

Each group will prepare a written proposal, which should not exceed 1 page, 11 points font single-spaced. Appendices are allowed, but will be reviewed at the instructor's discretion. I encourage you to talk to me about the topic before submitting the proposal. The report should contain the following:

1. **Format:** Submit the document as a pdf file. Please do not include Word documents.
2. **Authors:** List all the group members, the expected degree and the year in the program.
3. **Description of the problem:** 1-2 paragraphs describing the problem, and the specific question that you'd like to address. Please contact me if you have difficulties finding a problem.
4. **Summary of the data:** 1-2 paragraphs describing the variables that you will consider in the analysis, and issues such as correlations, violations of the model assumptions, missing data, outliers.
5. **Methods:** 1-2 paragraphs describing analysis methods that you will consider, and potential methodological difficulties. The methods should not be too different from the topics covered in class. Please talk to me before the deadline if you are not sure of what method to use.
6. **Preliminary results:** Some preliminary results are expected. Please briefly describe the available preliminary results if you have any.
7. **References:** Please only add references that are explicitly used in the text. Make sure that you cite the sources of data, and the associated publications. Use consistent format and numbering scheme.

Each group will make a short in-class presentation of the proposed work, with slides following the same format.

Project report: due Monday, April 20.

The report will consist of an oral and a written part. Each group will prepare a written report. The text should be at most 8 pages, 11 points font single-spaced, including figures, and excluding references and appendices. The report should be presented in a format of a scientific paper, and points will be taken off if the report does not follow the required format. The report should contain the following sections:

1. **Format:** Submit the document as a pdf file. Please do not submit Word documents.
2. **Authors:** Names of the authors.
3. **Introduction:** Provide a short background of the project (e.g., what kind of question is to be answered, and why this is of interest). Provide a short non-technical summary of your analysis.
4. **Methods:** Summarize the methodology used, and use mathematical formulae and notation.
5. **Results:** Show the results of your analysis. Summarize the results with a small number of most important figures or tables, and keep the description short.
6. **Discussion:** What did you learn from this analysis? What additional steps could be potentially performed to improve your analysis?
7. **References:** Please only add references that are explicitly used in the text. Make sure that you cite the sources of data, and the associated publications. Use consistent format and numbering scheme.
8. **Appendix:** Add links to code repository, plots, and other relevant technical details that will help evaluate your work.
9. **Statement of contributions:** Please list the names of the authors, and state explicitly how each member of the group contributed to the project.

Each group will make a short in-class presentation of the work, with slides following the same format.

Project evaluation report: due Wednesday, Apr 22.

To gain expertise in evaluating work by your colleagues, each class member will provide an anonymous written evaluation of a project by another group. Project evaluation report should not exceed 1 page, 11 points font single-spaced, and should contain the following:

1. **Format:** Submit the document as a pdf file. Please do not submit Word documents.
2. **Title:** The title should contain the id of the group being reviewed.
3. **Author:** To ensure blind review, give your personal id that we will send you together with the report. Please avoid using real names.
4. **Description of the problem:** 1-2 paragraphs describing the question addressed by the project, statistical methods used, and major conclusions.
5. **Positive aspects of the analysis:** 1-2 paragraphs describing aspects of the analysis that you think are interesting and correct. If some aspect of the work was particularly innovative, was not discussed in class or required a major effort, please point this out.
6. **Negative aspects of the analysis:** 1-2 paragraphs describing aspects of the analysis that you think were incomplete, sub-optimal or incorrect. Avoid harsh and unjustified criticism, and be constructive. Do not write comments that you'd not want to receive for your own work.
7. **Possible extensions:** If given an opportunity and time, how would you improve/extend this work? Would you be able to answer additional questions with this project, or use alternative methods to improve the analysis?