CS 7280 **Project** Spring 2015

The goal of the project is to gain hands-on experience with analysis of an experimental dataset. You should select a real-world problem, and a dataset, that can be analyzed using methods covered in class.

The project should be conducted in groups of 2-3 people. Each group should work independently, but you are welcome to discuss technical issues on the mailing list. Completion of the project will include a project proposal, a report, and a review of the project by another team.

Group members: due Thursday February 26 in class.

Please form groups of 2-3 class member who will work jointly on a project.

Project proposal: due Thursday March 19 by email before noon.

The proposal should contain a short description of the project. I encourage you to talk to me about the topic before submitting the proposal. The 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 send Word documents.
2. **Authors:** List all the group members.
3. **Description of the problem:** 1-2 paragraphs describing the problem, and the scientific question that you'd like to address with this dataset. You are welcome to use data from your research, but you are expected to make an additional effort for the class. You can also use a dataset from the KNNL Appendix C, provided that this dataset has not been discussed in class. Please contact me if you have difficulties finding a dataset.
4. **Summary of the data:** 1-2 paragraphs describing the variables that you will consider in the analysis, issues such as violations of independence, missing data and outliers.
5. **Methods:** 1-2 paragraphs describing analysis methods that you will consider, and potential methodological difficulties. The methods should be more elaborate than a simple t-test, and not 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. Please briefly describe the available preliminary results if you have any.
6. **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.

Project report: due Thursday April 16 by email before noon.

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 send Word documents.
2. **Authors:** To ensure blinded review, please do not list the authors of the report, but only the number of the group. I'll send you the number of the group as part of the feedback on the project proposal.
3. **Introduction:** Provide a short background of the project (e.g., what kind of scientific question is to be answered with this dataset). Provide a short non-technical summary of your analysis.

4. **Methods:** Summarize the statistical methodology used, and use mathematical formulae and notation when appropriate.

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. **Statement of contributions:** Please state explicitly how each member of the group contributed to the project

9. **Appendix:** Add computer code, plots, and other relevant technical details that will help me evaluating your work.

10. **Statement of contributions:** In the **last separate page**, please list the names of the authors, and state explicitly how each member of the group contributed to the project. I'll remove this page prior to forwarding the report for evaluation.

Project evaluation report: due Friday April 24 by email before noon.

To gain expertise in evaluating statistical analyses performed by your collaborators, each class member will evaluate a project completed by another group. Project evaluation report should not exceed 1 page, 11 points font single-spaced, and should contain the following:

(a) **Format:** Submit the document as a pdf file. Please do not send Word documents.

(b) **Author:** To ensure blind review, give your personal id that I'll send you together with the report. Please avoid using real names.

(c) **Description of the problem:** 1-2 paragraphs describing the scientific question addressed by the project, statistical methods used, and major conclusions.

(d) **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.

(e) **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.

(f) **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 dataset, or use alternative statistical methods to improve the inference?