Project Guidelines

The course project for Math 570 assumes the role of final exam. The project is done through a team effort, and consists of:

1. finding one or more multivariate data set or data sets;
2. choosing scientific questions of interest;
3. answering the questions by using appropriate multivariate analysis techniques;
4. conveying your findings of the analysis to broader audience by writing reports.

Team Members

Each team is composed of three to four team members. As a project may involve different skill sets, and different team members may contribute to the team in different ways, the diversity among the team is appreciated. Hence, the following restrictions are imposed.

1. A team cannot have all members from the same gender.
2. A team cannot have all members from the same year of graduate study (e.g. there has to be at least one first-year student and at least one second-year student in a team).
3. A team cannot have more than two members who are receiving financial support from the mathematical sciences department. In the case that a team has two such members, then these two members may not be from the same final project team in Math 532 (Regression II). A list of team composition for Math 532 (Regression II) has been acquired from Prof. Ganggang Xu.
4. A team cannot have all its three or four members from the same final project team in Math 532 (Regression II).

Please coordinate among teams to ensure that the formation of all teams complies to the restrictions above.

Important Timelines

1. Team formation. Each team should name a team leader. The team leader is responsible for communicating with the instructor while keeping the others in the loop. By **March 25** (before the Spring break), the team leader should inform the instructor
   I. the name of the team,
   II. the composition of the team,
   III. the main responsibility of each team member,
   IV. and the nature of the data set that the team plans to use.
2. Preliminary report and reading sessions. **By as early as April 25**, each team should have a preliminary report ready. The preliminary report should be at least 2 pages but not exceeding 4 pages (not including the title page.) The instructor will arrange for the preliminary reports be read in class sessions. This report reading process is designed to help people improve report writing, as it is one important quality that a professional statistician should have. Hence, each team should try to secure one person who is trustworthy on writing. The instructor will announce the scheduling of the report reading at a later time.
3. Final report. The final report is due on **May 15**. The final report is written by expanding the preliminary report but is not to exceed 6 pages (including references and not including the title page.)

Reports

Both the preliminary report and the final report have the following components.

1. Abstract: 1-page long summary of the project, explaining the data of interest, scientific questions involved, the methods used, and the findings, in non-technical terms.
2. Introduction: an introduction to the underlying problem which briefly summarizes the motivation and goal of the work.
3. A section explaining the data: how are the data obtained and processed; when applicable, some summary statistics for the data; the covariates of the data; the implication of certain variables.
4. A section detailing the steps of the analysis and the findings. Figures and tables are to be used when necessary. For a report of this length. Two to five figures and tables in total should be appropriate. More figures should be left to the supplementary material.
5. A section discussing alternative plans and secondary findings if any.
6. A conclusion section.

Computer programming code is not a part of and should not be included in a report.

Writing tips and style requirement

Students are advised to read the following book and article before writing the report.

1. The Elements of Style by Strunk and White - only 7 dollars on Amazon and multiple copies are available in the library as well.
2. Effective writing in mathematical statistics by Marron.

The following style requirements are to be fulfilled for the reports.

1. Be written and complied using Latex.
2. Font size 11 or 12.
3. Left, right, top and bottom margins of 1 inch each.
4. Single-spacing (not double-spacing).
5. Uses `asa.bst` bibliography style to cite reference.

A .tex template file will be sent to you for your reference.

Data Sets

There is no good data set; there is no bad data set. There are only data sets that are *interesting*. An interesting data set should have roots in real applications, and should have a multi-dimensional or high-dimensional nature. Preference are given to those which were collected from lab science, social science, and engineering. Projects are not necessarily evaluated on how many fancy methods have been used, but instead, how well the team can utilize different approaches learned in the course to undertake the statistical tasks they desire.

Below are some resources for obtaining data sets:

1. http://www.lib.jmu.edu/resources/statistics_datasets.aspx

2. http://www2.stetson.edu/~jrasp/data.htm
3. http://www.statsci.org/datasets.html
4. https://www.kaggle.com/