Project

In this project you will explore a topic of your choosing, to be approved by me, in the realm of advanced data analysis. In particular, you will prepare a 8-15 page (single spaced, 12pt font) paper on your topic as well as an oral report. The oral report will be a 1 hour presentation, with a 15 minute time period being reserved for questions from your fellow classmates; i.e., an entire class period. A general outline for your paper follows:

- An introduction which illustrates why the methodology was originally developed, what it is useful for, the big issues, etc. This section should be supported by no fewer than 5 academic references; i.e., Wikipedia and the such are not appropriate to reference. A bibliography should be provided.

- A full, yet concise methodological development on how the technique is implemented should be provided. For example, I do not expect a full exploration (i.e., a line by line development) of the asymptotic characteristics of your selected methodology, but a statement of the results along with appropriate discussion is expected.

- You should design and carry out a simulation study which illustrates the strength of the proposed method, implementing competing techniques would add to this. Further, the simulation study that you conduct could actually be a replication of an original study conducted by the authors who developed the methodology you are studying.

- **Data application**: A brief discussion of the data that you are going to analyze should be provided, as well as a discussion of the goal of the analysis. You should provide a complete and accurate description of the results, as well as any appropriate interpretations. I would encourage you to choose a data set that is of personal and/or academic interest.

- Fully annotated code, which implements your simulation study and data analysis, should be appended to your paper. Your code does not add to the length of your paper. Note, the use of existing software packages is completely acceptable (and actually recommended). The goal here is that anybody should be able to take your code and alter it for the analysis of another data set, and you should provide the details on how this is done.

- Your paper will be prepared in Latex, and I will provide a general template that you can use to do so.

Guidelines for the presentation:

- Your presentation should succinctly summarize your chosen methodology, from implementation through to your data analysis.

- The target audience for your presentation is your fellow classmates.

- Your presentation should be **professionally** prepared. I would highly suggest the use of beamer in Latex.

The deadlines and grading of your project will be conducted as follows:

- Your paper should be submitted and distributed to your fellow classmates at least 2 weeks prior to your presentation. Your classmates will then review your work and be **very critical**.

- Edits and critiques will be returned to you on the day of your presentation by your fellow classmates.

- You will have until the day of the final exam to return your revised paper to me for final submission.

This process will serve to emulate the referee process, and as such I will also be grading the referees on their performance reviewing you. The details of this process will be explained as time nears.

In what follows, I will provide a list of suggested topics which would be more than appropriate, of course other topics could be suitable.

- The expectation-maximization algorithm for finding MLEs and Louis' method.

- Bootstrap and Jackknife estimation of the sampling distributions, univariate and regression.

- The parametric bootstrap, univariate and regression.

- Measurement error, in the response or independent variables; methods of handling SIMEX, etc.

- Clustering algorithms; e.g., K-means clustering.

- Random effects modeling.

- Single index models, normal errors and within the class of GLMs.

- LASSO in the GLM framework.

- Spatial data analysis, Bayesian or frequentist.

- Comparison and discussion of methods designed to handle the multiple testing problem; Bonferroni, Holm's method, Benjamini and Hochberg FDR, etc.

Important: Do not procrastinate, get started on this as soon as possible. It is really easy to detect work that has been thrown together at the last minute and I expect more from you. If you need any assistance, as always I am willing to help, but much of this should be an independent study.