Statistics 600: Project Guidelines

During the last part of the course each student will do an individual project. The project is to conduct a small piece of independent research on a topic which is relevant to this course and to write a paper and give a presentation describing your work. Please devise a project which you would enjoy doing and is consistent with the goals of the course.

1 Project Topics

Your project should be complex enough so that you'll learn a lot from doing it, but not so complex that you can't produce a finished work by the end of the semester. The project should focus on an extension of a topic not fully covered in class. Students can (and should!) meet with me the to discuss their term project during office hours throughout the semester. Here are some suggestions for various approaches to the project:

- Apply a method we study to a novel dataset or problem which requires a bit of adaptation or ingenuity.
- Investigate a method that is an extension of a topic we have covered. Apply the method to data.
- Investigate the performance or sensitivity of a method across a class of problems and suggest/investigate adaptations for cases where performance is degraded.
- Use several methods to solve a challenging problem. Compare/contrast the results and the advantages/disadvantages of the methods.
- Develop the theory for a new method or extend an existing result. Apply the method to a small example.

The list of possible topics is only limited by your creativity. To get you started, here are some suggestions for specific topics:

- MCMC: Develop and apply a Langevin algorithm, or a multiple try Metropolis Hastings algorithm, or a perfect sampling algorithm.
- EM algorithm: we only scratched the surface on this topic in class. Investigate more sophisticated versions of the EM algorithm and apply it to a data set (e.g., ECM, SEM, Bayesian or Monte Carlo EM).
- Bootstrap: compare several approaches to bootstrap inference for a challenging data set.
- Simulation: Investigate bridge or path sampling and apply it to a problem.

2 Format

- Your paper should up to 7 pages of double spaced text ($\geq$ 10 point font). This limit does not include figures, tables and other supporting documentation. Include your final computer code and formatted output in an appendix. If you use any data in the analysis, include the data in the appendix. In the write-up, you should describe the purpose of the project, discuss why your project is important, and provide at least one example worked demonstrating your program. One possible paper format would be: introduction, background, algorithm details, example(s), conclusions.

- Oral presentations should be about 25 minutes long plus time for questions. You should prepare overheads or use the computer projector. Read the article in the coursepack on good presentations (Freeman *et al.*, 1983) before you create your talk.

3 Due dates

- November 18: Submit project proposal for approval. This should be a 1-2 page description of your project. It should include basic summary statistics for the real or simulated data you will use in your project.
- December 9 and 11: presentations
- December 18: Final report due by noon.

4 Grading

Grading of the presentation and report will be based on the following criteria:

- Motivation of problem: Your presentation should include detailed, convincing motivation of the topic. Why is this topic important?
- Statistical quality: Understanding and explanation of the topic. Interpretation of the results of the analysis. Accuracy of presentation and example. Quality of graphs and tables.
- Organization: Logic and clarity of presentation or report.

5 A Bonus!

Every year the Statistical Computing section of the American Statistical Association holds a student paper competition. Those students with quality projects should consider sending in their papers for the competion. You could win a free trip to the Joint Statistical Meetings which will be held in early August in Toronto. At the JSM you will present your paper. It is a big honor to win, so it can't hurt to submit your paper!

6 Some Useful References on Writing

For those of you who will be writing Ph.D. dissertations, the most useful book in the list below is probably that of Higham. I still refer to it for my own writing. For those students for whom English is not your first language, this book also include the chapter "When English is a Foreign Language."

Cleveland, W.S. (1985) *The Elements of Graphing Data*, second edition, Wadsworth, Pacific Grove, California.

Day, R.A. (1988), *How to Write and Publish a Scientific Paper*, third edition, Cambridge University Press: Cambridge.

Ehrenberg, A. (1981), "The Problem of Numeracy", *The American Statistician*, **35**, 67-71.

Ehrenberg, A. (1982), "Writing Technical Papers or Reports", *The American Statistician*, **36**, 326-329.

Gopen, G. and Swan, J. (1990), "The Science of Scientific Writing", *American Scientist*, **78**, 550-558.

Higham, N.J. (1998), *Handbook of Writing for the Mathematical Sciences*, Society for Industrial and Applied Mathematics, Philadelphia.

Strunk Jr., W. and White, E.B. (1979) *The Elements of Style*, third edition, Macmillan, New York.

Tufte, E.R. (1983), *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, Connecticut.