Statistics 4590
Term Project: March 11, 2010
Due no later than: April 9, 2010

There are datasets on the course webpage for you to consider for your project. You can pick one of them (each dataset can be analyze by only one of you, so first-come, first-served), or find your own data. Some of the particular details of what is expected are given with your dataset, but some parts are left for you to discover and implement.

Your written project should be split up into different sections, which contain complete paragraphs and sentences. Your report can be typed or handwritten. I will also expect you to submit any relevant computer output, as described below. This does mean that I will not want to see things R did that you did not want it to do. In particular, please **do not print your entire dataset** and submit it with your report.

As with your assignments, you will be graded on the clarity of presentation, statistical methodology, grammar, spelling-in short, all the things that will be important when you prepare these kinds of reports for a living as a paid statistician.

I will give another handout in a couple of weeks on what I'll expect for the presentation you will give on your project. A reminder that the presentations are scheduled for **Tuesday, April 6**.

A general outline of the format is given below. Note that, depending on your dataset, you may have little or no information to put in a particular section.

Format of Report

In preparing your project, break your report up into the following sections, and use these sections titles to aid in making your project clear and easy to follow:

- **Introduction**: Include description of data, and the questions of interest you will address in your analysis.
- **Exploratory Analysis**: Include things like plots, tables, summary statistics. Not all of these may be needed or applicable. Discuss only the variables that are involved in your questions of interest.
- **Analysis**: Include hypotheses, test statistic values and p-values. Make sure you give conclusions in the context of the question you are addressing. Address any assumptions made or possibly violated. This may include plots to assess model assumptions, and seeing if there are potential outliers in your data. If you can transform the data so that any assumptions are not violated, you should present your analysis based on the transformed data.
- **Conclusions**: Brief summary of findings, including your scope of inference, problems with confounding variables, and so on.
- **Appendix**: Containing any computer output.