Statistics: Continuous Methods
STAT 652, Spring 2011

Data Analysis Project

The data analysis project is to be completed by all students in STAT652. The project will result in a project report and presentation given at the end of semester (see the timeline below)

Project Goals:

- Learn theoretical methods and develop practical skills of statistical data analysis;
- Practice in formulating realistic research questions for a given data set;
- Get experience with exploring and analyzing real-life data sets;
- Learn how to interpret results of data analysis in terms of the applied process;
- Learn how to prepare a scientific data analysis report;
- Practice oral presentation skills.

Methods to use:

The project should combine theoretical and numerical methods studies in the class, which can be divided into the three broad categories:

- Exploratory data analysis and descriptive statistics
- Distribution fits, goodness-of-fit tests
- Association between variables: correlation measures and regression

A project would necessarily include exploratory data analysis and data description, as well as method from at least one other category. It might happen that you need (want) to apply methods outside of the scope of this class; this is highly encouraged and should be discussed with instructor.

Remark: ANOVA methods will be studied at the end of the class, so they are not included in the topics list above. This should not discourage you from using ANOVA in your analysis, if needed.

Project report:

A project report should be written according to scientific standards; the project format and requirements will be discussed later in class.
## Data Analysis Project Timeline

<table>
<thead>
<tr>
<th>To do item</th>
<th>Pts</th>
<th>Deadline</th>
<th>Comments</th>
</tr>
</thead>
</table>
| Think what data sets to use. In this class we mainly study methods to detect association between random variables. Hence, stick to:  
  • Several sets that might be related (whether they are actually related might be the question of your research)  
  • Continuous data  
  • Large number of observations (>100)  
  • Something you are interested in! |     | Feb 10   |          |
| Meet w/instructor to discuss the data, possible research questions, and the methods to apply. | 10  | Feb 17   |          |
| Finalize the data set for analysis, Formulate research questions (you might expand/refine/update them later), Start working on data analysis |     | Feb 24   |          |
| Data analysis is mostly completed, You know how to how to interpret the data and answer all the research questions, You might still need to do some analysis to improve results/figures, etc., Start preparing final report |     | April 1  |          |
| Project report draft submitted                                            | 10  | April 21 |          |
| Final project report submitted, Presentation ready                        | 10  | April 28 |          |
| Project presentation                                                      | 30  | May 3    |          |
| Project report evaluated                                                  | 40  |          |          |
| **TOTAL points**                                                          | 100 |          |          |