Stat 32: Data Analysis Policy Projects
Spring 2011, Swarthmore College

Class Meetings: Tuesdays and Thursdays, 11:20-12:35, Science Center 145

Prerequisites: {Stat 11, Stat 61, or Econ 31} or {(AP Stat or Stat 1) and permission of the instructor}

Professor:
Lynne Steuerle Schofield
Science Center 148
lschofi1@swarthmore.edu
610-328-7896
Office Hours: Mon 2:30-4:30, Tues 2:30-4:30 or by appointment.

Course description: In this course, students work in teams on a semester-long data analysis problem. A key objective of the course is to expose students to the variety of challenges faced by the data analyst. Students research the scientific background of their problem, consult with area organizations, and communicate their methods and results both in writing and in class presentations. At the end of the semester, each team will present their findings in class to their area organization and write a final report for use by the organization. There will be presentations during the second week of classes describing the projects further.

In Stat 31, we will extend the skills in exploring data and interpreting results that were begun in Stat 11, by building and fitting models, investigating model assumptions, interpreting results from the statistical models, and report writing. We will do this using real data to solve real problems. We will work concurrently on two tracks: lectures on statistical techniques that may be useful in your projects, and regular project meetings within your groups, with me, and with your area organizations.

Textbooks: Chatterjee and Hadi, Regression Analysis by Example, 4ed, Wiley.

Software: We will mainly use the statistical software package JMP (pronounced “jump”) which is available from the Swarthmore Data & Software server. From time to time, we may also R (available free on-line) or SPSS (also available from the Swarthmore Data & Software server). You are welcome to use another computer program if you prefer.

Grading: Final grades will be based on a total score, with homework worth 20%, the midterm worth 15%, the final worth 15%, the final written report worth 20%, oral presentations worth 20%, your group/self/org assessments worth 8% and class participation worth 2%.
**Homework:** Homework will be assigned approximately every two weeks and will be due on Tuesdays by 1pm. **Late homework will not be accepted** with one exception. You may hand in one assignment late once during the semester, in which case it will be due by Friday of that same week by 1pm. All homework should be typed (though equations and other mathematical notation can be written by hand). You may work together with other students, but your write-up must be your own. Note: **I will not accept homework sent by email.** Homework will be graded on statistical content and clarity of exposition and presentation.

**Policy Projects:** The projects should consume the majority of your time. Each group will get work with one organization to analyze their data and attempt to answer their research or evaluation questions. You should plan on meeting at least once a week within your project group, and check in at least once a month with your organization. There will also be two required meetings with me (during the weeks of February 21st and March 21st).

During the semester, each group will make brief presentations to the whole class on the progress of their projects. Each group member must participate in each of these presentations. The complete project work will be presented in an end-of-the-year in-class session with the organizations. Here are some important dates:

<table>
<thead>
<tr>
<th>Milestone</th>
<th>Date</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Meeting with Orgs</td>
<td>Week of Feb 7</td>
<td>All group members participate</td>
</tr>
<tr>
<td>Presentation 1</td>
<td>Mar 1 &amp; Mar 3</td>
<td>10 min. All group members participate</td>
</tr>
<tr>
<td>Presentation 2</td>
<td>Apr 12 &amp; Apr 14</td>
<td>20 min. All group members participate</td>
</tr>
<tr>
<td>1st Draft of Paper</td>
<td>Apr 5</td>
<td>Due at 1PM via email</td>
</tr>
<tr>
<td>Final presentations</td>
<td>Apr 26 &amp; Apr 28</td>
<td>30 min. All group members participate</td>
</tr>
<tr>
<td>Final Paper</td>
<td>Apr 28</td>
<td>Last regular course meeting.</td>
</tr>
</tbody>
</table>

Each group must turn in a formal, written report on the last day of class (Thursday, April 28 2011). A draft of the written report is due Tuesday, April 5 2011. Two times during the semester, each student will be asked to assess the contribution of each group member to the team effort including themselves. The organizations will also be given the opportunity to weigh in on each team members contribution. This will be factored into your overall grade.

**Responsible Conduct in Research Involving Human Subjects:** During the first half of the semester, every student will be required to complete a self-paced, online training program in human subjects research protections. Swarthmore subscribes to the Collaborative Institutional Training Initiative (CITI) which can be found at www.citiprogram.org.

**Midterm:** Thursday March 17th in class  
**Final:** TBA (the final period runs from May 5 - 14). The final will be comprehensive.  
**Final Papers Due:** Thursday, April 28th  
**Final Presentations to Orgs:** Tuesday, April 26th and Thursday, April 28th in class
Stat 32: Some Additional Readings

Useful Statistical Modeling Resources:
Gelman and Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press. Some chapters available on Blackboard and also at Haverford
Cook and Weisberg, An Introduction to Regression Graphics, Wiley. In Cornell
Weisberg, Applied Linear Regression (somewhat more advanced) In Cornell.

References on Statistical Consulting, Scientific Writing, and Professional Ethics: