**Department of Sociology and Anthropology**

**Carleton College**

Explorations in Social Data Analysis Peter D. Brandon

Soan 239 Leighton Hall 229

Spring Term 2011 Phone: 222-7199

Email: [pbrandon@carleton.edu](mailto:pbrandon@carleton.edu)

Class Meets: Mon., & Wed., (1:50 PM – 3:00 PM) & Fri. (2:20 PM – 3:20 PM).

Office Hours: Mon., & Wed., 11:30 PM – 1:00 PM or by appointment on Mon., Wed., & Fri.

# *I. Course Description*

This course introduces students to basic statistical methods for exploring, analyzing, and summarizing social data. Students learn how to investigate social data, generate results, and make inferences about their social world. The first half of the course focuses on why we have social survey data and the techniques for describing and summarizing social data. Topics include the value of social data, generating and visualizing social data, descriptive statistics, and linking social data to what we know about probability distributions. The course’s second half concentrates on making inferences from data on the social world and detecting differences between and among social groups. Topics include confidence intervals and significance tests, two- and multi-group comparisons, and associations between socially constructed categories. The schedule details weekly topics. Learning experiences will involve team work while others require meeting with members of a course resource support team. This course provides the necessary prerequisites for SOAN 240 and SOAN 280.

# *I.a. Aims of the course*

The course aims to equip students with basic statistical tools for analyzing social data and to provide them with opportunities to apply those tools to social issues and problems. The course seeks to promote students creative explorations of social data, enhance their understanding of how and when to use statistical tools, and offer them a quantitative reasoning learning experience. These goals combined with an aim to help students visualize social data will deepen students’ appreciation of the usefulness of statistics for investigating the social world.

Specifically, the course aims to build students’ knowledge and skills in the following areas:

* assessing survey data for the study of the social world;
* choosing, applying, and interpreting basic statistical methods for analyzing survey data;
* using statistical software to generate findings about the social world;
* constructing compelling arguments based upon quantitative analyses;
* reading social science studies with a deeper appreciation of the basic statistical methods used;
* writing clearly about those inferences drawn from their analyses.

# *I.b. Details about the course*

*I.b.i.:* Team work

Much effective learning can take place through team work. Every student must belong to a small group with two or three; team membership is chosen by me. Your final grade in the class will depend partly, though not exclusively, on your team’s performance. The final quantitative writing assignment (based on your team’s data analyses) is a team-based exercise. Members of the team with the highest score at the end of the term will each receive 5 bonus points toward their final grade and dinner out with me.

*I.b.ii.* STATA: The Statistical Software Package

An essential part of this course is using the statistical software package, STATA. You will use STATA to manage and analyze survey data about the social world and generate empirical evidence to support claims you make in your final team projects. Students will need thumb drives to save their STATA do files, which should correspond to weekly assignments, examinations, and the final team paper. Note that there is no lab component of this course to teach you STATA. You must commit to learning STATA, meeting with the course assistant, and coming to see me during my office hours to help you learn STATA. Also, your team can serve as a learning resource. Please see me and I will help you. Learning STATA will accelerate your overall learning experience of basic statistics and with quantitative reasoning.

*I.b.iii.* Data sets

I will provide you with nine different STATA data sets for weekly assignments, weekly examinations, and final team projects. Please keep copies of these data sets on a thumb drive. Please see the attached handout entitled *“Data Set Descriptions”* that describes the data sets, gives some information about which social surveys produced these data sets, defines variables in each data set, and offers hints to the sorts of topics your team could analyze with them. The data sets cover topics such as crime, gender and women’s issues, AIDS, unions, indigenous issues, and child care. You must see the librarian during the term at least once to ask for assistance in finding links to more detailed descriptions for these data sets or background information. By Wednesday of Week five, your team must advise me of when exactly you met with the librarian.

I picked the nine data sets for good reasons. First, finding an appropriate data set to answer a social science question takes time and resources, skill, and experience. (We focus on locating data for social science research in some of my other classes. See course descriptions for SOAN 240, 215, and 229.) We put this aspect of social science research aside so you can focus on developing your quantitative analysis and reasoning skills. You would not have time to create the latter if you also had to find data sets too. So, I provide the data sets that have been drawn from some of the finest and richest social science surveys available. With ready-to-go, STATA-formatted data sets, you can concentrate on learning basic statistical methods, including data visualizations.

*I.b.iv.* Quantitative Writing

A key component of this course is providing teams of students with an opportunity to write effectively about empirical facts they have explored using social survey data and have assessed using statistics. The course, therefore, will give student teams the opportunity to write about the empirical findings they have discovered. The course teaches the fundamentals of statistics, but in contrast to traditional ways of teaching statistics, this course requires student teams to write clearly and simply about their empirical discoveries. Students must mold their graphical, tabulated, and statistical depictions of the social world into well-crafted, short, and persuasive pieces of writing. Up to 15-page final team papers will provide students with a venue for practicing their quantitative writing and reasoning. I urge all student teams to meet with Carleton College writing specialists.

# *II. Requirements*

## Completion of ten weekly examinations

## Capstone final paper done in teams

## Team participation grade

# *III. Grading*

|  |  |
| --- | --- |
| Completion of ten weekly examinations | **60%** |
| Capstone final paper done in teams | **35%** |
| Intra-team participation rating score | **5%** |

# *IV. Assessment*

*IV.i.*  Weekly Assignments

At the end of each Friday class, I will hand out to each student the next week’s assignment. You are expected to work on the assignment over that week. All assignments are ungraded and thus never handed in to me. However, questions from the week’s assignment will definitely appear on the following Friday’s examination. So, some of the questions on an assignment will be on the following Friday’s examination. In other words, the weekly assignment prepares you for the following week’s examination. If you complete the assignments and studied the assignment contents you should be prepared for Friday in-class examinations. Learning statistics and applying statistical methods to the social world requires you to do your homework. You are not expected to complete the assignments in teams, but your team might aid your learning of the material and checking assignment answers. Each assignment builds upon the previous one.

*IV.ii.*  Weekly examinations

Each Friday you will take an in-class examination at the computer in the CMC lab. Each examination is based on questions that were included in the previous week's assignment. If you have done the previous week's assignment questions you will have prepared well. Your having completed all assignment questions, taken notes during class, and reviewed the textbook should result in you receiving a high examination score. The final in-class examination, Exam 10, will occur during finals week again in the CMC lab. The time you will need for the last exam is the same as the previous ones, no longer. You cannot self-schedule this last exam. Note also that because you will have weekly exams, there is no final exam, just the exam for week 10.

More specifically, I will take one or more questions from the previous week's assignment and ask you to again answer those questions in the Friday class exam. You will not be permitted to have notes available, books open, or any other electronic media open on the computer. You will just have STATA opened on the computer and you will access the data for the examination from the computer. I, of course, will supervise each Friday's exam. At the end of the examination, you will print out your STATA log containing your answers and pick it up, make sure your name is on your log, assemble the STATA log output in the correct order, staple the pages of your log together, and then hand it to me. The short weekly examination will not be onerous, always be based on the previous assignment, and intended to test your integration of the material and your statistical analysis skills using STATA. Examinations, like assignments, build upon the previous examination.

*IV.iii.* Final team paper

The date for the final team paper is the last day of class. I will provide you with a handout by week 7 on my expectations for the final paper and formatting standards. However, by the end of week 4 I will ask your team to write me which of the available data sets you will use for your final team paper. Please note that you have to also hand in the STATA “do” program that you used for your team paper. This requirement is mandatory so that I can replicate your results. **The data set your team uses for your final paper must be one of the ones I have listed**.

*IV.iv.* Team participation and self grading

Your contribution to your team is expected and critical to your grade. At the last class meeting, you and your other team members will confidentially rate one another from zero to five. You cannot rate yourself. I will take the ratings you receive from your other team members and calculate your average rating. That average rating is what I will add to your individual score for the course.

***V. Required Texts and Lectures*:**

*“*Statistical Methods for the Social Sciences” Alan Agresti and Barbara Finlay, 2009, 4th Ed., Prentice Hall, NJ. “A Gentle Introduction to Stata” Alan C. Acock, 2006. Stata Press.

Lectures: Importantly, you should note that my lectures complement and supplement the texts. I will teach this course using notes, materials, and exercises that I have developed over several years. Your work in this course should reflect the mix of my class lectures and materials contained in the texts. As is customary, you should attend class, take good lecture notes, participate, and cross-reference lecture and text materials. Again: what you produce for assessment should reflect all the resources that have been made available to you.

***Schedule of Topics***

**Section I: Introduction to Social Data Analysis**

Week 1: March 27 – April 2

A & F: Chapter 1; A: Chapter 1, 2, and 3.

* Why create social data? What is a social data set?
* How social data interacts with statistics, computers, and software.
* Library visit Libe 306, Friday, April 1st & introduction to social data sets.
* Teams assignments announced on Wednesday.

Week 2: April 3 – April 9

A & F: Chapter 2, sections 1, 2, 3, and 5; A: Chapter 4.

* Sampling and measurement
* ***1st examination in CMC lab on Friday, April 8.***

Week 3: April 10 – April 16

A & F: Chapter 3; A: Chapter 5.

* Descriptive statistics
* ***2nd examination in CMC lab on Friday, April 15.***

Week 4: April 17 – April 23

A & F: Chapter 4; A: Chapter 6, sections 6.1 - 6.3.

* Probability distributions
* ***3rd examination in CMC lab on Friday, April 22 and continue every Friday.***
* Write ups of which data set team proposes to analyze for final team paper. Handed to me hard copy.

**Section II: Inferences from Social Data Explorations**

Week 5: April 24 – April 30

A & F: Chapter 5; A: Chapter 7, sections 7.1 - 7.5.

* Statistical Inference: Estimation

Week 6: May 1 – May 7

A & F: Chapter 6.

Midterm break

* Statistical Inference: Significance Tests

Week 7: May 8 – May 14

A & F: Chapter 7; A: Chapter 7, sections 7.6 - 7.10.

* Comparison of Two Groups
* Handout posted on Moodle of my expectations for the final paper and formatting standards.

Week 8: May 15 – May 21

A & F: Chapter 8.

* Associations between socially constructed categorical variables
* Fifteen minute appointments during my office hours of final team papers

Week 9: May 22 – May 28

A & F: Chapter 12, sections 12.1 – 12.2; A: Chapter 9, sections 9.1 – 9.5.

* Comparisons among groups

Week 10: May 29 – June 4

A & F: Chapter 12, sections 12.4, 12.3, 12.8, 12.9; A: Chapter 9, sections 9.6 – 9.9.

* Comparisons among groups
* Case examples and applications to social settings
* The last examination, Exam 10, is during the time of the final. It is not comprehensive and will be exactly the same format as the previous Friday examinations.
* ***Final team papers are due by 5 PM on Wednesday, June 1. No emailed papers will be accepted. A paper that is handed in after 5 PM will only receive 25% of the possible points.***

*Syllabus Statement for Disability Services: This course recognizes that students with documented disabilities might need special accommodations. They are required to speak with Disability Services for Students. Please call extension 4080 to make appropriate arrangements.*