COLUMBIA UNIVERSITY

Sociology W3020: Social Statistics
Fall 2013

Instructor: Van C. Tran
Phone: 212-854-4115
E-mail: vantran@columbia.edu
Office: 606 Knox Hall
Course time: Mondays & Wednesdays, 8:40-9:55 a.m.
Office hours: Wednesdays, 4:00-6:00 p.m.
Website: https://courseworks.columbia.edu/portal/site/SOCIW3020_001_2013_3

Teaching Assistant: Mr. Sang Won Han
Email: sh3083@columbia.edu
Office hours: Mondays, 4:00-5:00 p.m.
Discussion section: Thursdays, 9 to 10 a.m. starting on September 12, 2013

Course Description:

This course introduces quantitative methods of sociological research for describing and drawing inferences from statistical data. We will focus on basic understanding of statistical concepts and methods of statistical analysis for information drawn from survey data. The course also includes several assignments on analysis of sample survey data, with the objective of equipping students for further quantitative research or senior thesis projects using quantitative methods.

This course touches on some themes also covered in Methods for Social Research (W3010), including the purposes of social science research and the logic involved in pursuing them, notably conceptualization and measurement. We will give some attention to methods of data collection, particularly surveys, which are most often analyzed with quantitative methods.

This course will begin with fairly simple concepts (i.e. mean, probability, etc.) before getting more involved. We will start with distributions of single variables and relationships between pairs of variables before introducing basic multivariate analysis that involve multiple variables. We will pay attention to two key concepts – **data reduction and statistical inference** – both of which are central to analyses of quantitative data.

The course will focus on providing you with an intuitive understanding of statistical summaries and analytical techniques that are applied to survey data. The course will not cover any statistical theory in details. Instead, the focus will be on how statistical techniques are used by sociologists to draw inferences from samples to populations of interests in our research. The course materials will explore real datasets using the statistical package, *Stata*.

This course is open mainly to undergraduates in sociology. For those who would like to gain a deeper understanding of statistical theory, I would recommend that you explore other relevant courses in the Statistics department at Columbia: http://statistics.columbia.edu/.

Course Objectives:

In this course on statistics, we will:

1. Develop an intuitive understanding of statistical summaries, models, and techniques.
2. Explore key statistical concepts and develop analytical skills, including the use of *Stata*.
3. Become an informed reader of statistical evidence in academic research and in the media.
4. Have hands-on experience doing some statistical analyses using survey data and regressions.

Course Requirements:

1. Ten problem sets (40% of final grade)
2. In-class midterm exam (20% of final grade)
3. Take-home final exam (30% of final grade)
4. Attendance & participation (10% of final grade)

Problem Sets

The problem sets are designed to help you grasp statistical concepts and learn how to calculate statistical summaries. They also help you formulate certain hypotheses and test them with survey data using appropriate statistical techniques via *Stata*. Your comfort and facility with statistical analyses will increase with practice, patience and repetition over time. In this regard, statistics are similar to languages, in that you will have to practice to use it or risk forgetting the nuances. Because the course materials are cumulative in nature, you must try to **not** allow yourself to fall behind, as it will take significant time and effort to get caught up again.

Midterm and Final Exam

The midterm exam will cover materials from the first half of the course. The final exam will be cumulative and comprehensive with emphasis on material covered after the midterm.

Course Readings:

Books

Agresti, Alan and Barbara Finlay. 2009. *Statistical Methods for the Social Sciences*, 4th Edition. Upper Saddle River, NJ: Prentice Hall.

Acock, Alan. 2012. *A Gentle Introduction to Stata*, Revised Third Edition. College Station, TX: Stata Press.

We will be reading a chapter a week from the Agresti and Finlay textbook. We will also draw on the Acock text for materials on how to use *Stata*. These books are available on reserve at the Lehman Social Sciences Library and also for purchase at Book Culture on West 112th St.

Course Policies:

1. Doing the reading and coming to class is essential to your comprehension and participation. Please complete assigned readings before each class. Your active participation in class is strongly encouraged. I will set aside some time during my lectures for discussion, so you can raise any questions that you might have then. If you don't understand something, it is often the case that some of your classmates might have the same questions, but they might be too shy to ask them. By raising questions about the reading and lecture materials, you will help all of us to learn better.

2. I am happy to meet with you individually to answer questions related to all aspects of the course. I would also like to get to know you, to learn more about your interests and to see how I can help you succeed in the course. The mechanism for this is my office hours and I encourage you to sign up for a 15-minute appointment with me. Your teaching fellow could be a great resource, so I would also encourage you to approach him during his weekly office hours. Before the midterm and final exam, we will provide extra office hours, but it is always best to plan at least one week in advance if you anticipate needing any help. We also often try to stay a few extra minutes after each class. If you have any "small" questions, then that will also be an excellent time to approach me. We would like each of you to do well in this course, so please do not hesitate to ask questions and to get feedback on your work.

3. Homework assignments should be submitted in hard copy and are due on **Mondays at 5 p.m.** You may submit them in class on Mondays or you may submit them to your teaching fellow **or** to Professor Tran's mailbox in 501 Knox in the department of sociology. Recognizing that all of us have busy lives and multiple commitments, you will be allowed three 48-hour no-question-asked extensions on your problem sets. Beyond these three times, late assignments will receive a lower grade. **No further extensions will be granted, except in the case of serious illness.** If you are ill, you should go to see your physician and provide a note to support any extension request.

4. Technology in the classroom can be both a blessing and a distraction. If you must use your laptop during class, you should turn off your internet browsers and email clients. **Laptops and other electronic devices should be used strictly for note-taking purposes only. We will rely on the honor code for the reinforcement of this rule**, so please help me and your classmates in our effort to create a classroom environment that is conducive to learning and sharing.

5. You should have a hand calculator that will add, subtract, multiply, divide, exponentiate, and take square roots. This will be helpful in doing some of the problems for the exercises, and also essential for in-class examination.

Software Resources:

We will do a substantial amount of data analysis this semester with the aid of a software package called *Stata*. This is available to you via Columbia Computer Labs. Course handouts/examples, readings, and instruction will teach you how to use Stata gradually throughout the course.

TENTATIVE COURSE SCHEDULE AND READINGS

Week 1: Introducing Statistics
Wednesday, September 4
Introducing the use of statistics in sociological research

Readings: Agresti and Finlay, Chapter 1

Week 2: Sampling and Measurement
Monday, September 9
Wednesday, September 11
Introducing survey sampling, measurement of concepts, and implications for statistical analyses

Readings: Agresti and Finlay, Chapter 2

Week 3: Descriptive Statistics
Monday, September 16
Wednesday, September 18
Introducing summary statistics, their properties, graphical and visualization of statistical data

Readings: Agresti and Finlay, Chapter 3
Problem set #1 due on Monday (9/16)

Week 4: Probability Distributions
Monday, September 23
Wednesday, September 25
Introducing basic ideas about probability, probability distributions, and sampling distributions

Readings: Agresti and Finlay, Chapter 4
Problem set #2 due on Monday (9/23)

Week 5: Statistical Inference
Monday, September 30
Wednesday, October 2
Introducing the idea of the point estimate of a population parameter and methods for inference

Readings: Agresti and Finlay, Chapter 5
Problem set #3 due on Monday (9/30)

Week 6: Hypothesis Testing and Statistical Significance
Monday, October 7
Wednesday, October 9
Introducing methods for hypothesis testing, the interpretation of p-values, and statistical significance

Readings: Agresti and Finlay, Chapter 6
Problem set #4 due on Monday (10/7)

Week 7: Comparisons of Two Groups
Monday, October 14
Wednesday, October 16 ****Midterm Exam****
Introducing the t-tests for differences in means and comparisons of group differences

Readings: Agresti and Finlay, Chapter 7
No problem set due this week

Week 8: Associations among Categorical Variables
Monday, October 21
Wednesday, October 23
Introducing the analysis of contingency tables and the chi-square tests for independence

Readings: Agresti and Finlay, Chapter 8
Problem set #5 due on Monday (10/21)

Week 9. Linear regression and Correlation
Monday, October 28
Wednesday, October 30
Introducing linear correlations, correlation coefficient, and least squares method

Readings: Agresti and Finlay, Chapter 9
Problem set #6 due on Monday (10/28)

Week 10. Statistical Controls
Monday, November 4 (academic holiday)
Wednesday, November 6
Introducing statistical controls, spurious associations, and causality

Readings: Agresti and Finlay, Chapter 10
Problem set #7 due on Tuesday (11/5)

Week 11. Multiple Regression Model
Monday, November 11
Wednesday, November 13
Introducing multiple regressions to study the effects of several predictors on an outcome variable

Readings: Agresti and Finlay, Chapter 11
Problem set #8 due on Monday (11/11)

Week 12. Dummy Variables and Interactions
Monday, November 18
Wednesday, November 20
Introducing dummy variables (i.e. categorical variables) and interaction effects

Readings: Agresti and Finlay, Chapter 12
Problem set #9 due on Monday (11/18)

Week 13. Collinearity and Misspecification
Monday, November 25
Monday, November 27
Introducing model specifications, omitted variable bias, and collinearity

Readings: Agresti and Finlay, Chapter 14
Problem set #10 due on Monday (11/25)

Week 14. Logistic Regression Model
Monday, December 2
Wednesday, December 4
Introducing categorical dependent variables and linear regression with binary outcomes

Readings: Agresti and Finlay, Chapter 15
No assignment due this week

Week 15. Review and wrap-up
Monday, December 9
Overview and general discussion of statistical analyses, quantitative methods, and research

****Take-home final exam due on Monday, December 16****

***

A final note on academic integrity and collaboration at Columbia:

The intellectual venture in which we are all engaged requires of faculty and students alike the highest level of personal and academic integrity. As members of an academic community, each one of us bears the responsibility to participate in scholarly discourse and research in a manner characterized by intellectual honesty and scholarly integrity.

In practical terms, this means that, as students, you must be responsible for the full citations of others' ideas in all of your research papers and projects; you must be scrupulously honest when taking your examinations; you must always submit your own work and not that of another student, scholar, or internet agent.

You are allowed to work together on the problem sets, but you must conduct your own analysis, write your own computer code, prepare your own tables, and write up your results individually. **No collaboration will be permitted on the final exam.** If you have any questions about what level of mutual sharing with and learning from your classmates is appropriate, please come to see me or your teaching fellow during our office hours and we can talk in more details. For further information on academic integrity and the policy at Columbia College, please see: http://www.college.columbia.edu/academics/integrity