

Columbia University

Sociology W3020: Social Statistics

Summer 2016, Session II

Instructor: Adam Obeng

Email: adam.obeng@columbia.edu

Office hours: Wednesdays 0900–1000 in 612 Knox Hall (register 24 hours in advance via Google Calendar)

Course time: Tuesdays and Thursdays, 0900–1210 Websites: Courseworks, Piazza

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.

This is a course in statistics applied to the social sciences. The aim is for you to have an intuitive understanding of the basic descriptive and inferential techniques most commonly used in the social sciences. The course is practical, and will not cover the statistical theory in much detail. 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 open source statistical programming language R. 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.

Course Objectives:

In this course on statistics, you 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 R
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:

Evaluation

1. Participation: 15%
 - i.e. Piazza polls and feedback
2. Five assignments (R/A&F exercises): 35%, 7% per assignment
 - due every week on Tuesday
3. Midterm: 20%
 - due 2016-07-21
4. Final: 30%
 - due 2016-08-11

Participation

The participation grade will be composed of traditional in-class participation as well as formative online in-class polls. You don't have to get these in-class exercises right, but you do have to respond. Although I would like to encourage you to speak in class, I appreciate that not everyone is an extrovert, so if you've got something to say you'll still be able to participate without the pressure of face-to-face interaction.

Assignments

The problem sets are designed to help you grasp statistical concepts and perform commonly-used analyses. They also help you formulate certain hypotheses and test them with survey data using appropriate statistical techniques in R. 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 we have very limited time in the summer session, and 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.

Problem sets will be due at the beginning of class every Tuesday except for the day of the first class. All assignments must be submitted by their respective deadlines, usually via Google Forms.

Of course, I understand that you're juggling a lot of responsibilities, so you will be allowed one 48-hour no-questions-asked extension for the assignments. Any other late, missing or otherwise non-compliant assignments will lose one letter grade per day they remain so. If there is a genuine reason that prevented your assignment from being submitted, we will work with your Dean to decide what to do.

That said, if you can anticipate that you will need more time for assignments for any reason, please get in touch as soon as possible, and I'll do my best to accommodate.

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.

Textbook

The class textbook is Agresti, Alan and Barbara Finlay. 2009. *Statistical Methods for the Social Sciences*, 4th Edition. Upper Saddle River, NJ: Prentice Hall. (ISBN 9780130272959)

The textbook is available on reserve at the Lehman Social Sciences Library and also for purchase at Book Culture on West 112th St.

Course Policies:

Attendance

Doing the reading in advanced and coming to class is essential. The reason for this is pedagogical: the textbook goes into more detail than we will in class and it also presents material with more run-up and in a different style. The trick to intuitively understanding abstract concepts is to work out what sort of analogy you can apply to them. For example, if you're trying to understand derivatives (not required for this class!), you might picture the tangent to a slope, or the relationship between an object's speed and its position, or perhaps even the arithmetical relationship between functional forms. Therefore, reading the textbook will both prepare you for the material so you're not encountering it for the first time, and will present you with a different explanation which might make more intuitive sense to you.

Office hours

I am happy to meet with you individually to answer questions related to all aspects of the course, and to 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.

In order to make the best use of our time, you must sign up for office hours via Google Calendar at least 24 hours in advance, and at that time also let me know what you'd like to talk about.

Technology

We will do a substantial amount of data analysis this semester with the aid of a software package called R. You may have used statistical software packages like SPSS or State before, but R is different in a number of respects, and worth learning for that reason. R is installed on machines in Columbia's computer labs so normally this class is conducted in a computer lab. Unfortunately, we don't have access to a lab this time. Therefore, it is essential that you bring a personal laptop with to every class, and I will help you set up R and RStudio. If this will be a problem, please let me know as soon as possible and I will endeavour to come up with a solution.

Schedule

Class 1: Introducing social statistics

Tuesday 2016-06-05

Class 2: Descriptive and inferential statistics

Thursday 2016-06-07

- Readings: A&F Chapters 1, 2, 3

Class 3: Probability and estimation

Tuesday 2016-06-12

- Readings: A&F Chapters 4, 5
- Assignment 1 due

Class 4: Hypothesis testing and statistical inference

Thursday 2016-06-14

- Readings: A&F Chapter 6

Class 5: Comparisons of two groups

Tuesday 2016-06-19

- Readings: A&F Chapter 7
- Assignment 2 due

Class 6: Associations among categorical variables

Thursday 2016-06-21

- Readings: A&F Chapter 8
- Midterm due

Class 7: Linear regression and correlation

Tuesday 2016-06-26

- Readings: A&F Chapter 9
- Assignment 3 due

Class 8: Statistical controls

Thursday 2016-06-28 - Readings: A&F Chapter 10

Class 9: Multiple regression

Tuesday 2016-07-02

- Readings: A&F Chapter 11
- Assignment 4 due

Class 10: Dummy variables and interactions

Thursday 2016-07-04

- Readings: A&F Chapter 12

Class 11: Collinearity and misspecification

Tuesday 2016-07-09

- Readings: A&F Chapter 14
- Assignment 5 due

Class 12: Logistic regression and wrap-up

Thursday 2016-07-11

- Readings: A&F Chapter 15
- Final due

Material used in producing this course is derived (with permission) from the version taught by Van Tran in Fall 2013.