Statistics for Sociological Research I, Spring 2014, TTH 2-3:15, ASY 1101

Instructor:	Meredith A. Kleykamp, ASY4125, Phone:405-6409, kleykamp@umd.edu
TA:	Daniel Standridge, dstandri@umd.edu
Office Hours:	Kleykamp T 11-12 TH 1-2, and by appointment Standridge, Wed. TBA, and by appointment
Textbook:	Introduction to the Practice of Statistics, 7th ed. by David S. Moore, George P. McCabe, & Bruce A. Craig, Additional readings noted on class schedule
Other possibly useful resources:	Common Errors in Statistics by Phillip Good & James Hardin 2006. The Tao of Statistics: A Path to Understanding (With No Math) by Dana K. Keller. A Gentle Introduction to Stata [3rd edition] by Alan Acock There are several useful Stata texts for those who anticipate doing extensive quantitative work on the following site: Stata books. I find Cameron and Trivedi and Long and Freese are indespensible resources in my own work. Academic Technology Services at UCLA provides guidance for many data management and analysis problems you may encounter.
Other requirements:	Calculator with $\sqrt{x}$, exp, and ln functions that does not store formulas Regular access to computer with Stata software (I'm a PC user and not much help with a Mac. You can use the BSOS computer labs when open, or purchase your own copy of Stata. Please ask me about which flavor of Stata would be best for you to purchase given your data needs. You can use another program, but we cannot offer assistance with how to use it.) Regular access to ELMS/Canvas and e-mail
Prerequisites/corequisites:*	college algebra course undergraduate statistics course or the SOCY winterterm statistic course Coreq: Stata computing course (1-credit with Alan Neustadtl.) Without taking this you will be responsible for learning Stata outside of any commands/syntax I provide.

*You should review the math refresher on the course Canvas site and evaluate your preparedness for the basic math and algebra in this course. If you are not comfortable with these concepts you will find the math in this course challenging. You should review college algebra skills and be prepared to use those skills in this course. A college algebra course is a pre-requisite for the course. If it's been a long time since you took it-you should expect to spend ample time refreshing yourself on those skills.

Course Overview:

This course introduces basic descriptive and inferential statistics, and various techniques of data analysis and their application in sociology. This course introduces you to both statistics and data analysis (they go hand in hand), but perhaps not in equal proportion. The two basic objectives of statistical methods are data reduction and statistical inference. Data reduction involves the summarization of complex, large and multivariate datasets with an eye to the identification of patterns or associations. Statistical inference quantifies uncertainty about those summaries and associations, uncertainties that arise because data are subject to random variation. More practically in sociology, inferential statistics allow us to draw conclusions from data from random samples drawn from a population of interest. Along the way, we will discuss challenges to the simplistic assumptions of many models and what is to be done in the face of a generally messy data world. We will do all of this using statistical software to facilitate the visual display of patterns in our data, and the testing of hypotheses derived form sociological theory.

The course is divided into a few sections. It may not follow the order of your previous statistical courses, and this is known and intentional. First we discuss basic tools for data reduction, description, and analysis. Next we review some basic probability theory, enough to motivate and situate the connections between randomness, sampling, and inference about a whole from a part (i.e. inference). We then spend the rest of the course on estimation and elaboration of the basic regression model.

The main objectives of the course then are to:

- provide a foundation in data analysis and statistical inference
- to prepare students for more advanced graduate courses in social statistics
- to develop students' capacity to read and critically evaluate published research in the social sciences (i.e. you should be able to read articles in AJS and ASR)
- to begin to develop students' ability to carry out their own research using quantitative data and statistical models

Homework

Homework will be due on most Thursdays, with due dates shown on the class schedule. There are 10 homeworks worth 10 points each. Late or missing homework is not accepted so that we may provide solutions to homeworks in a timely manner. The course is fast paced, and cumulative. We need to post solutions quickly so that you all can review your work, correct errors of understanding and move ahead.

We will rely on using statistical software for at least some portion of your homework assignments. For any homework involving the use of software, you must submit a do-file that contains only the commands required to run your analysis and produce your results, and the log file of this do-file that shows what you did to generate your results along with the relevant results. "Relevant" is the operative word here; please only include the commands and output required to replicate your work, along with comments to explain what you are doing in those commands in the log file (i.e. if you make 15 coding errors and fix something on the 16th try, edit out the bad 15 lines of code for your do and log files.) Homeworks will be graded not only on the correctness of the answer, but the clarity of your presentation. This may seem like overkill, but part of developing into successful data analysts is cultivating a habit of making your work reproducible (what do files allow), documenting what you actually did (log files), and comprehensible (parsimonious and effective presentation of results in a write-up).

I will instruct with Stata, and prefer that you use it too. You may use whatever program you prefer (SAS or R), but we cannot help with other software programs if you choose to use them. SAS and STATA are available on the shared network drive (i:/) at most computers in the Art-Sociology Building, and at several campus computer centers, such as the OACS center in 225-227-229-231 LeFrak Hall. For anyone considering purchasing Stata for personal use, please consult with me before buying it, so I can advise you on which version best suits your long-term needs (i.e don't just immediately buy the cheapest or 2nd cheapest version (it may not meet your long-term needs.)

Lectures

Lectures typically cover topics in the text and more. I will post lecture notes on Canvas and will do my best to have them up before class whenever possible, but I make no guarantees. This is done as a favor to you , and is not an obligation for me. Please let me know if you find errors, and accept my apologies for their presence in advance.

Examinations

There will be a one-hour fifteen minute midterm during a class period (Note the date on the schedule now-it's the Thursday before Spring Break. Please do not make any travel plans that take you away from campus this day). You will also be required to conduct your own empirical analysis using the statistical tools for this course and submit a paper on this analysis. For those who already have an idea for a second year paper using available data, you may find the final paper option helpful in jumpstarting your work. For others, I will will offer some suggestions for available public data that you might find useful, in particular, the American Time Use dataset. The written final paper will due at the final exam time scheduled for this class by the university, in this case 5/15 by 12:30pm (earlier submissions would be welcomed.) You may bring a single sheet of notes and calculator to the midterm. All work for the final paper is to be conducted independently. You will need to identify a dataset and topic early, as the homeworks will ask that you begin to use the data as part of your weekly work. The idea is that you will be able to get frequent feedback on basic errors along the way. But this requires you to do some quick decision making about what data to use.

The University of Maryland, College Park has a nationally recognized Code of Academic Integrity, administered by the Student Honor Council. As a student you are responsible for upholding these standards for this course. It is very important for you to be aware of the consequences of cheating, fabrication, facilitation, and plagiarism. For more information on the Code of Academic Integrity or the Student Honor Council, please visit http://www.shc.umd.edu. I shudder to think of this being a problem with graduate students, but you are an apprentice in a profession driven largely by internal policing of ethical standards.

Grading

40% of your grade is based on your homework: 4% each assignment for 10 assignments
30% of your grade is based on the Midterm Exam
30% of your grade is based on the Final Exam or Final Paper

Grades will be assigned as follows:
90.0 - 100 = A or A-
80.0 - 89.9 = B+ or B or B-
70.0 - 79.9 = C+ or C or C-
60.0 - 69.9 = D+ or D or D-
0 - 59.9 = F

If you disagree with a grade you received (on a homework or exam), please come to my office hours.

Students come to this class with a range of statistical backgrounds and aptitudes. This means that some students might work harder and improve more than others, yet still receive a lower grade. To recognize improvement, I reserve the option to slightly overweight the final and underweight the midterm for students who show improvement in their test scores.

Incompletes

University policy allows for incompletes in unusual cases where students have completed most of the class expectations, but cannot finish the class due to unforeseen circumstances. I do not give out incompletes except in these unusual circumstances. In general, avoid incompletes if at all possible; students who take incompletes seldom do as well as if they had completed the course on time.

Accommodations:

See me in the first week of class if you have a documented disability to make arrangements for accomodation of that disability. Please also let me know if you have any life circumstances that may affect your semester (i.e. a chronic illness, going through a difficult divorce, having a baby, etc...)
Similarly, see me if any class-related activity, including exams, conflict with religious observances of a generally recognized nature that you are under obligation to participate in.

Communication with Instructor and TA:

Office hours are times when we are guaranteed to be available (barring an emergency); we can make appointments to meet if you are unable to come to regularly scheduled office hours. Math can be difficult to explain over email so it's usually best to try and solve issues in person whenever possible. Quick points of clarification are obviously different from fundamental questions about material, and much easier to answer efficiently over email.

Something to consider/unsolicited advice

For those of you who expect to do quantitative research in your career, you might use this course as an opportunity to learn how to either use the equation editor in MS Word, or, better yet, learn to use LaTeX, a typesetting language that lets you deal much more easily with equations and formulas than does MS Word. It also let's you make prettier documents. There is certainly a STEEP learning curve, but if you think you may do more technical work, it may be worth learning the ropes now, using homework assignments as the testbed for writing "real " papers with LaTeX. I'm happy to hold an optional session on getting started with LaTeXfor anyone who might be interested.

Now is the time to establish good habits in using statistical software, doing data analysis and writing up the results. Establishing a workflow early is good practice, and helps you avoid pitfalls later on. A good resource on this is "The Workflow of Data Analysis Using Stata" by J. Scott Long. It may seem like overkill right now, but it's a good idea to anticipate some of the issues you'll face as you begin working on your own research. This piece is also totally worth reading: http://www.kieranhealy.org/files/misc/workflow-apps.pdf. Just don't be too distracted by the beautiful aesthetic quality of the document, or the emphasis on R (you can use Stata and LaTeXeasily) lest you forget the important message of the article: find and use tools that help you get work done, not create more of it.

SOCY601 Schedule of Course Meetings

Exam dates are inflexible, but the schedule of lectures and homeworks may be adjusted as I see fit. You should have read and reviewed the material scheduled on a given day before the class meeting. Homework is due at the beginning of class.

Schedule:

1.28 (T) ..... Introduction
1.30 (TH) ..... Graphical Summaries (IPS1.1), Wainer, MEET IN LeFrak computer Lab 5

2.04 (T) ..... Numerical Summaries & the Normal distribution (IPS 1.2-1.3)
2.06 (TH) ..... American TIme Use Data-Guest Liana Sayer

2.11 (T) ..... Scatterplots and correlations (IPS 2.1-2.2), HW1
2.13 (TH) ..... Bivariate regression and transformations (IPS 2.3-2.4)

2.18 (T) ..... 2-way or contingency tables (IPS 2.5), Kastellec and Leoni, IP 2.6 on your own
2.20 (TH) ..... Model Thinking (IPS 3), HW2

2.25 (T) ..... Randomness and probability (IPS 4.1-4.4)
2.27 (TH) ..... Sampling distribution: means (IPS 5.1), HW3

3.04 (T) ..... Sampling distribution: counts and proportions (IPS 5.2)
3.06 (TH) ..... Estimating with confidence: inference about a mean (IPS 6.1), HW4

3.11 (T) ..... Hypothesis testing about a mean (IPS 6.2)
3.13 (TH) ..... MIDTERM in class

3.18 (T) ..... SPRING BREAK
3.20 (TH) ..... SPRING BREAK

3.25 (T) ..... Power and Stargazing (IPS 6.3-6.4), Leahey
3.27 (TH) ..... Inference in practice: 1 sample (IPS 7.1), HW5

4.01 (T) ..... Inference in practice: 2 samples (IPS 7.2)
4.03 (TH) ..... Inference with proportions, 1 & 2 sample (IPS 8.1, 8.2), HW6

4.08 (T) ..... ANOVA or *How you can tell who in the room is a psychologist* (IPS 12&13)
4.10 (TH) ..... Measures of fit for contingency tables: χ^2 (IPS 9), HW7

4.15 (T) ..... Inference for regression (IPS 10)
4.17 (TH) ..... Multiple regression (IPS 11), HW8

4.22 (T) ..... Challenges in using regression: why you need theory (supplement, Freedman experiment)
4.24 (TH) ..... Dummy variables (supplement), HW9

4.29 (T) ..... Interactions, HW10
5.01 (TH) ..... Practical worked example, in-class

5.08 (T) ..... Logistic Regression: overview and introduction (IPS14), Morgan and Teachman
5.10 (TH) ..... Some tricks of the trade: Preparing professional-quality tables, figures, and papers.

5.13 (T) ..... No Class, Kleykamp at NSF

5.19 (Monday) ..... Final paper, due by 12:30pm