**Statistics for Sociologists**

**SOC6302H**

Meetings: Tuesdays 10a-12p (class) and 12p-1p (lab session) in Room 36 (Bloor/Spadina basement)

Instructor: Markus Schafer, PhD

Teaching Assistant: TBA

Phone: 416-946-5900

E-mail: markus.schafer@utoronto.ca; TA TBA

Office: Rm. 374

Office Hours (Schafer): Thursday, 2p-3:30p; and by appointment

Office Hours (TA TBA): by appointment only

**Description**

This course provides a survey of statistics and data analysis for MA sociology students. The course begins with data exploration and description, focusing on how to concisely summarize large and complex datasets. We will focus on the distributional properties of single variables and on bivariate descriptive associations. We then shift to the principles of probability, sampling distributions, and statistical inference, leading to an overview of hypothesis testing in the context of comparing means and proportions. The latter part of the course gives particular emphasis to multiple linear regression as an orienting framework for quantitative sociological research. We close with an introduction to logistic regression.

**Goals of the course**

There are three overarching goals of this course.

1. First, SOC6302 should ***lay a foundation for later courses*** in statistical methods in the U of T’s sociology curriculum and beyond. In many ways, this course serves as a “refresher” for many of the concepts covered in a standard undergraduate social science statistics course.
2. Second, the course intends to ***integrate statistics with data analysis***—preparing you to understand the basic concepts and then to apply statistical reasoning in your own research. To that end, a major goal of the course is to build familiarity with statistical computing in Stata. We will put this idea into practice by managing datasets, constructing syntax files, and executing statistical commands. Assignments requiring students to conduct original analyses will further reinforce these skills.
3. Finally, SOC6302 should help ***draw connections between statistical methods and the practice of sociological research***. Throughout the semester, we will examine how published empirical work applies and communicates statistical reasoning. We will also consider how to clearly and effectively express quantitative ideas in our own written work. Taken together, these efforts should help make us better producers and consumers of social science research.

**Reading**

Main text:

Moore, David S., George P. McCabe, and Bruce A. Craig. 2012. *Introduction to the Practice of Statistics, 7th Edition*. New York: W.H. Freeman and Company. [available on reserve in Mathematical Sciences Library and in Victoria University EJ Pratt Library]

Optional texts:

Acock, Alan C. 2018. *A Gentle Introduction to Stata* (now in its sixth edition). College Station, TX: Stata Press. [various editions available on reserve in Mathematical Sciences Library and in Industrial Relations & Human Resources Library; also available at OISE]

Long, J. Scott. 2009. *The Workflow of Data Analysis Using Stata*. Stata Press: College Station, TX: Stata Press.

Longest, Kyle C. 2012, 2015. *Using Stata for Quantitative Analysis* (1st or 2nd editions). [available as a digital holding in U of T Libraries]

Selected weeks will have additional supplemental and optional readings (available on Quercus).

**Labs and Computing**

This course includes a lab component in which we will work through analyses using the Stata statistical program. Data and computer code for the labs can be found on the course’s Quercus page.

Stata is a powerful and user-friendly program ideal for data management, statistical analysis, and graphic production. The program is available in the department’s computer lab. If you would like to use Stata on your personal computer, the company offers a “GradPlan” which provides software and manuals at a reduced price for enrolled students (see <http://www.stata.com/order/new/edu/gradplans>). The standard version of Stata (Stata/IC 14) is available for as little as $125 USD (annual license), or $295 USD for a perpetual license.

Our focus on computing issues will be somewhat cursory, but numerous resources for learning and mastering Stata are easily available on the web or through published books (see the optional course reading texts). Lab sessions will draw heavily from the optional text, *Using Stata for Quantitative Analysis* (Longest, 2015).

Though all labs will be conducted in Stata, students may feel free to conduct their assignments in alternative programs (e.g., R, SAS, SPSS) so long as they include reproducible syntax scripts.

**Evaluation components**

*I. Problem sets and assignments (collectively worth 90% of course mark)*

Students will turn in four assignments. The assignments are a mix of (a) problems found in the Moore et al. textbook, (b) an assessment of how statistics are used in the sociological literature, and (c) the analysis of data chosen by students.

A hard copy of each assignment must be handed in during class on the day it is due. A 2% late penalty will be added for each workday that the assignment is late. Students may work together on assignments, but work must be unique (e.g., written interpretation must be written independently and not copied, research questions for student-selected data should not be identical). For homework using the computer, all syntax must be included as an appendix.

Assignment 1 – Analyzing distributions, central tendency, and dispersion

Distribution date: TBA

Due date: 1 October

Relative weight towards final mark: 20%

Assignment 2 – Bivariate associations

 Distribution date: TBA

 Due date: 15 October

 Relative weight towards final mark: 20%

Assignment 3 – Inferential techniques

 Distribution date: TBA

 Due date: 19 November

 Relative weight towards final mark: 20%

Assignment 4 – From data description to hypothesis testing with multiple regression analysis

 Distribution date: TBA

 Due date: 13 December

 Relative weight towards final mark: 30%

*II. Presentation of work in progress (worth 10% of course mark)*

The semester will conclude with presentations of work in progress on **3 December**. The goal of these presentations is to briefly communicate the research question of your fourth assignment and to describe what you’ve found thus far. To accommodate all students in the allotted time frame, presentations are required to be no longer than 10 minutes. Students should use presentation software to summarize their findings and ensure that their talk is organized and succinct. Presentations should have an informative title. Finally, be sure to communicate the big picture: why is your question important, timely, or sociologically relevant?

**LATE WORK AND MISSED DEADLINES**

Homework assignments are collected during class on their due date andthe test should be taken as scheduled. The only exception for either deadline is a legitimate, documented reason beyond your control (e.g., illness, family emergency). In cases where there is no legitimate reason for being late, a 2% penalty will be added for each workday that a homework assignment is overdue.

Please notify me promptly if you must miss a deadline and provide documentation as soon as possible.

**ACADEMIC INTEGRITY**

Students are expected to know and adhere to the University’s principles of academic integrity. Any act of plagiarism or other unethical behavior will be addressed in accordance with University guidelines. Please see the “Code of Behaviour on Academic Matters” (<http://www.governingcouncil.utoronto.ca/policies/behaveac.htm>) for specific information on academic integrity at the U of T.

**COURSE SCHEDULE**

***Class 1—10 September***

Introduction to the course; the role of statistics in sociology; potentials and perils of quantitative data; Introduction to Stata (lab)

Reading:

* Miller, Jane E. 2004. *The Chicago Guide to Writing about Numbers*. Chicago: University of Chicago Press.
	+ Chapters 1 - 2
* Raftery, Adrian E. 2001. “Statistics in Sociology, 1950-2000: A Selective Review.” *Sociological Methodology* 31:1-45.

Optional reading:

* Long, J. Scott. 2009. *The Workflow of Data Analysis Using Stata*. Stata Press: College Station, TX.
* Longest, Kyle C. 2012, 2015. *Using Stata for Quantitative Analysis* (1st or 2nd editions).
	+ Chapters 1-3

***Class 2—17 September***

Describing data and analyzing distributions – level of measurement, central tendency, and variability

Reading:

* Moore et al., *Introduction to the Practice of Statistics*
	+ Chapter 1.1 - 1.2
* Miller, Jane E. 2004. *The Chicago Guide to Writing about Numbers*. Chicago: University of Chicago Press.
	+ Chapters 2, 4, and 9 (up to page 190)

Optional reading:

* Tuftey, Edward R. 2001. *The Visual Display of Quantitative Information (2nd ed.)*. Cheshire, CN: Graphics Press.
	+ Chapters 2 - 4
* Velleman, Paul F. and Leland Wilkinson. 1993. “Nominal, Ordinal, Interval, and Ratio Typologies are Misleading.” *American Statistician* 74:65-72.

***Class 3—24 September***

1. Analyzing distributions, con’t

Reading:

* Moore et al., *Introduction to the Practice of Statistics*
	+ Chapter 1.3
1. Descriptive bivariate associations

Reading:

* Moore et al., *Introduction to the Practice of Statistics*
	+ Chapter 2.1 - 2.2

***Class 4—1 October***

**ASSIGNMENT 1 DUE**

Descriptive bivariate associations, con’t

Reading:

* Moore et al., *Introduction to the Practice of Statistics*
	+ Chapter 2.3 - 2.6
* Miller, Jane E. 2004. *The Chicago Guide to Writing about Numbers*. Chicago: University of Chicago Press.
	+ Chapter 9 (pages 190-199)

***Class 5—8 October***

Overview of sampling, probability theory, and sampling distributions

Reading:

* Moore et al., *Introduction to the Practice of Statistics*
	+ Chapters 3-5

***Class 6—15 October***

**ASSIGNMENT 2 DUE**

Introduction to statistical inference: confidence intervals and hypothesis testing

Reading:

* Moore et al., *Introduction to the Practice of Statistics*
	+ Chapters 6 - 7.1
* Cohen, Jacob. 1994. “The Earth is Round (*p* < .05).” *American Psychologist* 49:997-1003

***Class 7—22 October***

Using inference to make comparisons: means

Reading:

* Moore et al., *Introduction to the Practice of Statistics*
	+ Chapters 7.2 and 12
* Miller, Jane E. 2004. *The Chicago Guide to Writing about Numbers*. Chicago: University of Chicago Press.
	+ Chapter 3

***Class 8—29 October***

Using inference to make comparisons: proportions

Reading:

* Moore et al., *Introduction to the Practice of Statistics*
	+ Chapters 8.2 and 9.1 - 9.2

***Class 9—12 November***

Using inference for simple and multiple regression

Reading:

* Moore et al., *Introduction to the Practice of Statistics*
	+ Chapters 10 and 11

***Class 10—19 November***

**ASSIGNMENT 3 DUE**

Further considerations for multiple regression

Reading:

* Miller, Jane E. 2005. *The Chicago Guide to Writing about Multivariate Analysis*. Chicago: University of Chicago Press.
	+ Chapter 9

Optional reading:

* McCloskey, Deirdre N. and Stephen T. Ziliak. 1996. “The Standard Error of Regressions.” *Journal of Economic Literature* 34:97-114
* Mirowsky, John. 2013. “Analyzing Associations between Mental Health and Social Circumstances.” Pp. 143-165 in Handbook of the Sociology of Mental Health, *2nd Ed*, edited by C.S. Aneshensel, J.C. Phelan, and A. Bierman. New York: Springer.

***Class 11—26 November***

Logistic regression

Optional reading:

* Menard, Scott. 2002. *Applied Logistic Regression Analysis, 2nd Ed*. Thousand Oaks, CA: Sage

***Class 12—3 December***

**Works in Progress presentations**

Reading:

* Miller, Jane E. 2005. *The Chicago Guide to Writing about Multivariate Analysis*. Chicago: University of Chicago Press.
	+ Chapters 14-15

**ASSIGNMENT 4 DUE 13 DECEMBER (*please drop off at my office or mailbox by 5p*).**