

APPLIED SOCIAL RESEARCH — 2

DATA 755

Thursdays, 6:30 - 8:15 pm on ZOOM

Charles F. Turner

Student E-Mail: Soc755.CFTurner@GMail.com

Website: www.PopEpi.org

Office Hours

Office Hours via ZOOM: 1pm to 3pm Thursdays, appointment required

See syllabus to make other arrangements

The aim of this course is to provide students with a hands-on learning experience conducting multivariate statistical analyses of large datasets. The course will emphasize the conduct and interpretation of data analyses for substantively important questions. While I do not intend to turn students into software gurus, all students will be required to master the technical details required to use one of the most common statistical software packages employed in social, econometric, demographic, biostatistical, and other quantitative fields of study, Stata. This software has a *much* gentler learning curve than R, and it provides a GUI and built-in documentation. By the conclusion of the course, students will be sufficiently fluent in the major tools of multivariate statistics to conduct independent analytic work in Stata or whatever language they choose.

Students will find this course easier if they have already obtained a good foundation in statistical theory. Soc 710 and 712 are prerequisites for registering for this course.

There are no required texts for this course since the entire set of Stata manuals are bundled as PDFs with the program. However, students may find the following texts to be useful: ***A Gentle Introduction to Stata*** (by Alan Acock, Stata press.), ***Statistics with Stata*** (by Lee Hamilton, Duxbury Press), *Applied Survey Data Analysis, Second Edition* (by Steven Heeringa et al., Chapman & Hall/CRC). These texts are listed in increasing order of technical difficulty; they are available at Amazon.com and www.stata.com. If available, I recommend that you purchase used books; editions for Stata versions 8 or later are acceptable. Four additional primers on statistical inference, multivariate linear, and logistic regression are suggested as reference material for our course work (Schroeder, *Understanding Regression Analysis*, Lewis-Beck, *Applied Regression*, Menard, *Applied Logistic Regression Analysis*, and Treiman, *Quantitative Data Analysis*).

PDF files and/or xeroxed readings may also be made available to the class to illustrate analytic techniques and strategies for data interpretation.

DATASETS

Class exercises and student projects will use a large composite dataset (N = 2,061,980) extracted from the 1997 thru 2018 National Health Interview Survey. NHIS is the USA's largest in-person household health survey. In the words of CDC, "NHIS data are used to: Monitor progress towards national health objectives; Evaluate health policies and programs; and Track changes in health behaviors and health care use."

GRADING. Two exercises will be assigned during the course to test students' mastery of the material taught in class. These exercises will be designed to be completed in 2-3 hours and students will be given one week to turn them in. Performance on these assignments will count for 40 percent of the final grade. All students will also be required to submit one larger data analysis project at the end of the term (see below). This project will count for 60 percent of the final grade.

PROJECT. All students are required to submit a data analysis project that uses the techniques taught in this class to answer a substantively important research question. The analysis must use the class NHIS dataset or another public dataset approved by the instructor. Before beginning work on this project, all students must meet with the instructor to discuss and receive approval for their proposed research topic. These meetings should be completed on or before October 31st.

COMPUTER ACCOUNTS. I have purchased and distributed 6-month licenses for STATA IC edition to each member of the class. If you have not received your license, please let me (Turner) know ASAP.

PLAGIARISM. The use of other people's work or ideas without appropriate acknowledgment is a serious breach of the standards of academic scholarship. Students who engage in such behavior may be given a grade of F, and they may be subject to other disciplinary action. The university's academic integrity policy is published in full at:

web.cuny.edu/academics/info-central/policies/academic-integrity.pdf

CONTACTING ME. The best way to contact me is to send E-Mail to Soc755.CFTurner@GMail.com. Voicemail may **only** be left on 917-243-4806. If you wish to have a conversation with me, please email or voicemail me with your telephone number and times you are available to talk.

COURSE TOPICS

Sept. 1 Overview of course. Introduction to hardware, software, and datasets.
Discussion of substantive interests of class members.

Sept. 8 Exploring large datasets: First steps, formulating strategies, efficient data processing, and good programming hygiene.

ANALYSIS OF METRIC DATA

Sept. 15 Basic Concepts and the Case of Two Metric Variables

Sept. 22 Multiple Linear Regression

Oct. 6 **NOTE: Sept. 29 – Classes follow Monday Schedule**

Oct. 13 Non-Linear Regression

ANALYSIS OF CATEGORICAL DATA

Oct. 20 Understanding the Logic of Tables

Oct. 27 Logistic Regression

Nov. 3

Nov, 10

SPECIAL TOPICS

Nov. 17 Special Topics and Work on Projects

Nov. 24

STUDENT PRESENTATIONS

Dec. 1 Students present their projects

Dec. 8