ED 287: Data Analysis in Education
University of California, Irvine
Spring, 2013

Instructor: Thurston Domina
Email: tdomina@uci.edu
Office: EDUC 2062

Mondays 1:00-3:50, EDUC 2024
Website: https://eee.uci.edu/11s/12480
Office Hours: Thursdays, 2-4

This is an applied course in quantitative research methods, designed for students who have a background in statistical reasoning but who have limited hands-on experience with large-scale data analysis. My primary goal for this course is to give students the analytical tools necessary to undertake independent secondary data analysis projects.

The course will have a lecture component as well as a workshop component. Topics for in-class lectures will include:

- existing sources for educational data,
- the Stata software package,
- data management and organization,
- applying basic statistical concepts to educational research, and
- brief introductions to advanced techniques in quantitative analysis.

The course’s workshop component will allow students to work in small groups to develop a research project and apply concepts discussed in the lectures to their collective work. At the beginning of the quarter, I will introduce students to several sources for secondary data. Each student will choose a data source and then work with their colleagues who have chosen the same data source to formulate a research question that can be answered using these data and conduct a series of analyses to answer their research question. Students will present their works in progress to the class, so that we can work together to address data limitations, talk through the problems that arise during the course of quantitative research projects, and identify new avenues for potential analyses.
Course software and texts:
In designing the course, I assume that you have some familiarity with basic statistical concepts, and that you have access to an introductory statistics textbook. If you don’t have a statistics textbook on your shelf, I recommend Moore and McCabe, *Introduction to the Practice of Statistics* (Fifth Edition).

I encourage students to purchase a Stata/IC license, which is a relatively affordable and very powerful package. (Stata offers a very good deal on the software for UCI students and faculty, [http://licenses.nacs.uci.edu/detail2.cfm?ProductID=57](http://licenses.nacs.uci.edu/detail2.cfm?ProductID=57).) The Stata software license comes with a short manual entitled *Getting Started With Stata*.

In addition, I will assign readings from a variety of sources. The following are particularly worth owning:


Assignments and grading:
Weekly homework assignments will insure that students are making progress both individually and collectively and help make the connections between the course lectures and these projects clear. Keeping up with these assignments will be essential to your success in the course. *Homework will make up 30% of your final grade. All assignments are due at 9AM before the class session in which they are listed.*

In addition, each student will make several informal workshop presentations to the class. In these presentations, students will outline their research questions, provide an introduction to their data, and walk the class through their analyses. At the beginning of each class session, I will invite a handful of students to present based on their homework. I see these presentations as opportunities for the class to talk through the challenges and questions that arise in the course of the research, and to brainstorm solutions to these. *Your presentations, together with the contributions that you provide to other students during the course of these workshop presentations will make up 15% of your final grade.*

Each student will work with a small group of peers to prepare a final working paper, in which they summarize their research project and present their findings. This truncated research paper need not include a comprehensive literature review. However, it should include: (1) a clear research question; (2) enough background information to help the reader understand why this is an important question; (3) a comprehensive data and methods section; (4) a summary of the
project’s findings, including appropriate descriptive, bivariate, and multivariate analyses; and (5) a discussion section spelling out the paper’s implications and limitations. The working paper will make up 40% of your final grade. It will be due at 12PM on Wednesday, June 12.

On the last day of the class, each student will participate in a formal presentation of the group work they’ve done over the quarter. These group presentations, which will resemble a conference presentation in style and substance, will make up 15% of your final grade.

Academic honesty:

I will strictly enforce the academic honesty principles laid out in the UCI Principles of Community (http://www.editor.uci.edu/catalogue/appx/appx.2.htm):

Students have responsibility for:

1. Refraining from cheating and plagiarism.
2. Refusing to aid or abet any form of academic dishonesty.
3. Notifying professors and/or appropriate administrative officials about observed incidents of academic misconduct. The anonymity of a student reporting an incident of academic dishonesty will be protected.

Students who witness plagiarism, cheating, or other forms of academic dishonesty should contact me. I will report all instances of academic dishonesty to the appropriate Associate Dean, and will not give credit for plagiarized work.
Seminar schedule:

April 1, 2013: Introduction

  a. Review syllabus
  b. Introduction to quantitative research
  c. Existing data sources
     i. Head Start Impact Study (HSIS)  
     ii. Early Childhood Longitudinal Study – Kindergarten (ECLS-K)  
         (http://nces.ed.gov/ecls/dataproducts.asp)
     iii. Education Longitudinal Study (ELS)  
          (http://nces.ed.gov/surveys/els2002/avail_data.asp)
     iv. California Motivation Project (CAMP Math)  
         (http://www.gse.uci.edu/research/california_motivation_project.php)
     vi. Texas Higher Education Opportunity Project (THEOP) administrative data  
         (http://opr.princeton.edu/archive/theop/)

April 8, 2013: Starting out

  HOMEWORK: Individual 2-page memo: Review the datasets described above. Identify one dataset that interests you, and describe a topic that you would like to investigate using that dataset. Describe the proposed dataset’s universe, sampling scheme, key variables, and strengths and weaknesses for addressing your topic of interest.

  READING: Murnane & Willett, Ch. 1-4.

  Class discussion:

  Moving from research subject to research question

  a. Designing an analysis (outcome/ treatment /controls; dependent variable/independent variables)
  b. Variable selection
  c. Making causal claims

April 15, 2013: Data and data management
**HOMEWORK:** Individual 2-page memo: Propose a research question that you would like to answer using your dataset. Is the question that you are proposing to answer causal or descriptive? Identify at least one published paper that answers a related research question. To what extent does the design in the paper that you have identified enable causal claims?

**READING (required):**

University of Wisconsin SSCC “Stata for Researchers”
http://www.ssc.wisc.edu/sscc/pubs/sfr-intro.htm

**READING (recommended):**

Acock, Ch. 1-4, 12.

**Class discussion:**

a. Introduction to Stata  
b. Basics of data management  
c. Unit of analysis and file shape  
d. Data headaches: Sampling, study design, weighting, etc.

**April 22, 2013: Measurement and more data management**

**HOMEWORK:**

**Group:** Work together to propose research question and research design. In addition, work together to build file, label variables, set missing to missing, compute summary scores. Write a short memo describing research question and design. What unique contributions can you make? Include descriptive statistics for key variables.

**Individual:** Identify a paper that attempts to measure one of the constructs that are central to your group project. In 1 page, discuss the key challenges in measuring this construct. How do the authors of the paper that you identify address those challenges? What are the strengths and weaknesses of their approach?

**READING:**

“Validity and reliability” http://www.soc.iastate.edu/sapp/soc302relval.html

**READING (reference):**

Acock, Ch. 5-11.

Class discussion:

a. More on Stata and good work habits
b. Measurement, measurement error, validity, etc.
c. Variable types
d. Distribution

April 29, 2013: Graphing and bivariate analysis

HOMEWORK:

Individual: Identify an article that addresses a question related to your group project’s research question using a multivariate analysis technique. Reproduce one or more key tables; summarize the logic underlying these tables as well as their findings and important conceptual limitations.

Group: Get to know your data and the nature of the relationships among your variables. Create the key variables for your analyses and analyze their measurement properties. Run a bunch of exploratory analyses and write up what you’ve learned. How do these preliminary findings change the way you think about the final analyses?

READING:

Rodriguez, German. “Stata Tutorial” (esp. section on Stata graphics) http://data.princeton.edu/stata/

READING (reference):


Class discussion:

a. Visualizing relationships
b. Statistics and hypothesis testing
c. Correlations and bivariate plots: Relationships between continuous variables
d. Chi-square: Relationship between categorical variables
e. T-tests and other tests for means: Relationship between continuous and categorical variables
f. Scatterplots, regression lines
May 6, 2013: Regression

**HOMEWORK:**

**Individual:** Identify an article that addresses a question related to your group project’s research question using an advanced multivariate analysis technique (such as an interaction, nonlinear relationship, etc.) Reproduce one or more key tables, summarize the logic underlying them and their key findings, and describe their important conceptual limitations.

**Group:** Keep at the preliminary analyses. Run a bunch of different regression models to understand key relationships in your data, and experiment with different methods to improve model fit and theoretical relevance. Write up a brief memo describing these preliminary analyses and their implications, considering their relevance for the final paper.

**Reading (recommended):**

Murnane & Willett, Ch. 5-12.

**Class discussion:**

a. Conceptual introduction  
b. Building a regression model  
c. Reading a regression output  
d. Logistic regression and other solutions for categorical data  
e. Regression diagnostics  
f. Interaction effects  
g. Modeling nonlinearity  
h. Quantile regression and distributional analysis

May 13, 2013: Estimating causal effects

**HOMEWORK:**

**Individual:** Identify an article that addresses a question related to your group project’s research question using an experimental or quasi-experimental research design. Reproduce one or more key tables, summarize the logic underlying them and their key findings, and describe their important conceptual limitations.

**Group:** Review preliminary analyses from last few weeks. Decide which of these analyses should go into your paper. Continue to finalize analyses. In addition, start to think about how you might extend upon and improve these analyses. What are future directions that you might consider?
READING:

Miller, Ch. 1-4; Murnane & Willett, Ch. 13.

Class discussion:

a. Randomized experiments
b. Natural experiments
c. Regression discontinuity
d. Instrumental-variables estimation
e. Panel data; time-series
f. Propensity score matching and weighting

May 20, 2013: Presenting analyses and interpreting data

HOMEWORK: Start drafting paper and preparing presentation.

READING:


Class discussion:

a. Table construction
b. Visual representations
c. Writing about quantitative data analysis

May 27, 2013: No class this week, because of Memorial Day. But each group should plan to meet with me at least once during the course of this week as you work on your paper and presentation.

June 3, 2013: Group presentations

June 12, 2013: Final paper due.