PA 542: Advanced Data Analysis II

Spring 2020

Michael D. Siciliano

office: 2116 AEH

ph: 312-413-5177

email: sicilian@uic.edu

Class: Monday 3:30pm – 6:15pm, 2202 AEH

Office Hours: I am always willing to meet. I am in my office most days and am happy to meet via skype in the evenings. Email me to set up a time that is convenient for you.

I. LEARNING AND COURSE OBJECTIVES:

This course builds on a foundation in statistics, linear regression, and data analysis. The course covers a series of topics and methods critical to the analysis of social science data including count regression models, regression discontinuity designs, instrumental variables, panel data methods, multilevel models, longitudinal multilevel models, and text analysis. The emphasis of this course is on conceptual understanding and the appropriate application of methods given one's sampling strategy, data, and theory. The course will entail a series of problem sets to help the student become familiar with the methods and get their hands dirty working with data. My goal is that students will not only become proficient in the methods covered in the course but also gain the ability and confidence necessary to go out and learn new methods (i) as needed for their own research and, (ii) as the field of quantitative methods progresses.

A major component of the class will be a replication study where the student recreates the analysis and results of a recently published journal article. It is my hope that you will take your replication paper and extend those results through the application of new methodology or the addition of new variables - this is a great way to produce a quality conference paper and/or journal article. For interested students, I hold optional "labs" a few times during the semester. These informal labs will provide an opportunity to clear up any issues from prior lectures, practice implementing data analysis techniques, and work on the replication study.

II. SOFTWARE REQUIREMENTS

I do not require you to adopt a particular statistical program – if you are proficient in a given software you are free to continue using it for certain aspects of the course. I have worked with SAS and Stata before and have taught courses using SPSS and R. However, I conduct most of my work in R and the predecessor of this course, PA 541, was taught using R. While I allow you to work with software of your choosing, the examples in class and modeling approaches I discuss will be implemented in R. Some techniques we will cover (e.g., text analysis) may not be possible in other statistical packages, so you will need to learn a little R at some point. I will also provide tips and tools for better workflow procedures, data handling, and graphics in R. Thus, I would encourage all students, even those who consider themselves proficient in another statistical package to learn and use R. We can discuss software and the limitations of certain packages on the first day of class. Below are instructions for installing R and RStudio. RStudio is an integrated development environment. RStudio includes a console, syntax-highlighting editor, as well as tools for plotting, history, debugging and workspace management. I highly recommend using RStudio.

Installing R

1. Install R First
    a. Go to https://www.r-project.org/ and click 'download R'
    b. Choose the 0-Cloud mirror from the list of CRAN mirrors.
    c. Choose the download for you system (Mac or Windows)
    d. Click 'install R for the first time' and then download and choose the default options.
2. Install RStudio
    a. Go to https://www.rstudio.com/products/rstudio/download/ and choose the Free RStudio Desktop version to download and appropriate system (Mac or Windows)
    b. Once downloaded, run the installation and open RStudio.
    c. You should be up and running!

III. READINGS

There is no required textbook for this class. Each week I will assign two sets of readings: one that covers the method and another that provides an example of the application of the method. The application readings are generally drawn from recent articles in top journals in policy, administration, and political science. We can discuss different textbooks that you may want to have on your shelf as reference tools.

IV. REQUIREMENTS AND GRADES

1. Replication Paper

During the course of the semester you will be working on a replication paper. This will require you to find a recently published (within the last 5 years) article in your field of study and to recreate the findings as presented in the journal. You will turn this replication in as your final paper. Note: more and more journals, especially in economics and political science, require authors to provide replication materials (i.e., the data and the analysis script needed to reproduce the results). You will need to choose an article that does not provide the script. Replication papers are due during finals week. Papers should be between 5,000 and 7,500 words including tables, citations, etc… I will discuss this assignment in more detail in class. The paper you replicate should use one of the methods we discuss in this course or a similar/more advanced method.

2. Problem Sets

Most weeks you will have a problem set to work on. The problem set will focus on applying the methods to real data and interpreting and using the output. I encourage you to work with your classmates on these assignments but to ultimately complete and write-up the results individually.

3. New methods presentations

As noted, a major goal of this course is for you to gain confidence in your ability to learn and implement new methods. Not only because of the range of methods that exist outside of this course, but because quantitative methodology is constantly expanding. Each student, working in teams of two, will be required to present on an advanced methodological topic not covered in this course. You will choose your method in consultation with me and will be required to present the method to class. You will also prepare for the class a high level summary of the method (this is a 3-5 page document that discusses the method and how to implement it in R; this could include an example analysis and interpretation of the results). Some options include spatial analysis, path models/SEM, qualitative comparative analysis, survival analysis, regression trees/splines, meta-analysis, machine learning, etc….Presentations will be given the last week of class.

4. Midterm

There will be an in-class midterm during week 9. The midterm will cover material from all preceding weeks. The midterm will focus on the student's ability to interpret model results. It will be a closed book exam.

Your grade will be calculated as follows:

Component	Percentage of overall grade	Due date
Replication Paper	30%	May 5th
Problem Sets	25%	To be scheduled
New Methods Presentation	10%	April 28th
Midterm	35%	March 10th

V. COURSE POLICIES AND OTHER INFORMATION:

Academic Integrity: As an academic community, UIC is committed to providing an environment in which research, learning, and scholarship can flourish and in which all endeavors are guided by academic and professional integrity. All members of the campus community—students, staff, faculty, administrators—share the responsibility of insuring that these standards are upheld so that such an environment exists. Instances of academic misconduct by students shall be handled pursuant to the Student Disciplinary Policy. The Student Disciplinary Policy is available online at https://dos.uic.edu/wp-content/uploads/sites/262/2018/10/DOS-Student-Disciplinary-Policy-2018-2019-FINAL.pdf.

Special Needs: UIC and the Department of Public Administration are committed to maintaining a barrier-free environment so individuals with disabilities can fully access programs, services and all activities on campus. The Office of Disability Services works to ensure the accessibility of UIC programs, classes, and services to students with disabilities. Services are available for students who have documented disabilities, including vision or hearing impairments and emotional or physical disabilities. Students with disability/access needs or questions may contact the Office of Disability Services at (312) 413-2183 (voice) or (312) 413-0123 (TTY only). Please feel free to contact me if you need any special accommodations.

Diversity and Inclusion: It is my goal that people from diverse backgrounds and perspectives be included in and served by this course, the Department of Public Administration, and the University of Illinois at Chicago. I believe that the diversity that the students bring to this class is a resource that can be used to improve student learning and perspectives. Course materials and activities are designed to be respectful of all types of diversity: gender identity, sexuality, disability, age, socioeconomic status, ethnicity, race, nationality, religion, values, and culture. It is my intent that this class be an environment where all students feel safe to express their perspectives and

opinions. Please let me know if something said or done by me, the TA, guest lecturers, or other students, is troubling or causes discomfort or offense. If this occurs, please feel free to discuss the situation with me privately, to raise your concern in class, or to let me know about the issue through a trusted sources (e.g., another student or faculty member or your academic advisor).

Campus Advocacy Network: Under the Title IX law you have the right to an education that is free from any form of gender-based violence and discrimination. Crimes of sexual assault, domestic violence, sexual harassment, and stalking are against the law and can be prevented. For more information or for confidential victim-services and advocacy contact UIC's Campus Advocacy Network at 312-413-1025 or visit http://can.uic.edu/. To make a report to UIC's Title IX office, contact Rebecca Gordon, EdD at TitleIX@uic.edu or (312) 996-5657.

VI. WEEKLY SCHEDULE (readings subject to change)

Week 1 - January 14

Topics: Course intro and review. Linear models, qualitative predictors, interaction effects, and assumption testing for regression models.

Methods Readings: None. But please review your notes and readings from a previous course in statistics.

Application Readings:

King, Gary (2006) Publication, Publication, *PS: Political Science and Politics*, Vol. XXXIX, No. 1, 119-125 [This provides some insight on how you can extend your replication paper into a published work]

Optional R Readings: See folder on Blackboard.

Week 2 - January 21

Topics: Data Wrangling in R; Review of generalized linear models and logistic regression.

Methods Readings:

Moore, David S. and McCabe, George P., "Introduction to the Practice of Statistics", Chapter 16 – Logistic Regression. P. 1-16

dplyr tutorial: https://cran.r-project.org/web/packages/dplyr/vignettes/dplyr.html

Application Readings:

Krebs, T. B., & Pelissero, J. P. (2010). What Influences City Council Adoption and Support for Reinventing Government? Environmental or Institutional Factors? Public Administration Review, 70(2), 258-267.

Week 3 – January 28

Topics: Count Regression models (Poisson, negative binomial, zero inflated)

Methods Readings:

Atkins, David C. and Gallop, Robert J. (2007) Rethinking how family researchers model infrequent outcomes: A tutorial on count regression and zero-inflated models.

Coxe, S., West, S. G., & Aiken, L. S. (2009). The Analysis of Count Data: A Gentle Introduction to Poisson Regression and its Alternatives. Journal of Personality Assessment, 91(2), 121-136.

Application Readings:

Amirkhanyan, Anna A., Kenneth J. Meier, Jr Laurence J. O'Toole, Mueen A. Dakhwe, and Shawn Janzen. 2017. Management and Performance in US Nursing Homes. *Journal of Public Administration Research and Theory* 28 (1):33-49.

Week 4 - February 4

Topics: Ordinal and multinomial logistic regression.

Methods Readings:

Kleinbaum et al. (2014) Applied Regression Analysis and Other Multivariate Methods, 5th ed. Chapter 23.

Application Readings:

Weber, Rachel, Stephanie Farmer, and Mary Donoghue. "Predicting School Closures in an Era of Austerity: The Case of Chicago." Urban Affairs Review (2018): 1078087418802359.

Week 5 - February 11

Topics: Techniques for causal modeling and evaluation: Instrumental variables and Regression discontinuity designs.

Methods Readings:

Angrist, J. D., & Krueger, A. B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. The Journal of Economic Perspectives, 15(4), 69-85.

Bennett, D. A. (2010). An Introduction to Instrumental Variables Analysis: Part 1. Neuroepidemiology, 35(3), 237-240.

Application Readings:

Angrist, J. D., & Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. The Quarterly Journal of Economics, 114(2), 533-575.

**Optional Methods Readings on Regression Discontinuity:

Trochim, William M.K., The Regression Discontinuity Design: An Introduction

Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel MJ Vermeersch. Impact Evaluation in Practice. The World Bank, 2016. Chapter 6: Regression Discontinuity

**Optional Application Readings on Regression Discontinuity:

Gormley Jr., W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The Effects of Universal Pre-K on Cognitive Development. Developmental Psychology, 41(6), 872-884.

Week 6 - February 18

Topics: Multi-Level Models (Hierarchical Linear Models) part 1

Methods Readings:

Hox, Joop. (2010). "Multilevel Analysis: Techniques and Applications", Second Edition. New York: Routledge. Chapter 2.

Hofmann, D. A. (1997). An Overview of the Logic and Rationale of Hierarchical Linear Models. Journal of Management, 23(6), 723-744.

Application Readings:

Kelleher, Christine A., and Susan Webb Yackee. 2009. A Political Consequence of Contracting: Organized Interests and State Agency Decision Making. *Journal of Public Administration Research and Theory* 19 (3):579-602.

Week 7 – February 25

Topics: Multi-Level Models (Hierarchical Linear Models) part 2

Methods Readings:

Hayes, A. F. (2006). A Primer on Multilevel Modeling. Human Communication Research, 32(4), 385-410.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. . Psychological Methods, 12(2), 121-138.

Application Readings:

Kubrin, C. E., & Stewart, E. A. (2006). Predicting who reoffends: the neglected role of neighborhood context in recidivism studies. Criminology, 44(1), 165-197.

Week 8 – March 3

Topics: Catch up on any remaining topics. Review for midterm.

There are no assigned readings for this week. We will spend time working on data analyses and interpretation of the methods we have discussed so far. This would be a good time to work on your replication papers.

Week 9 - March 10

Topics: **Midterm

Week 10 – March 17

Topics: Panel data; difference in difference models; fixed effects methods

Methods Readings:

Dougherty, Christopher. (2011) "Introduction to Econometrics", 4th ed. Oxford University Press: New York. Chapter 14 - Introduction to Panel Methods.

Angrist, Joshua D., and Jörn-Steffen Pischke. Mastering'metrics: The path from cause to effect. Princeton University Press, 2014. Chapter 5: Difference-in-Differences.

Application Readings:

Song, Hummy, Anita L. Tucker, Karen L. Murrell, and David R. Vinson. (forthcoming) Closing the Productivity Gap: Improving Worker Productivity Through Public Relative Performance Feedback and Validation of Best Practices. *Management Science*

Week 11 - March 24

******SPRING BREAK******

Week 12 – March 31

Topics: Multi-Level Models for repeated measures part 1

Readings:

Singer, Judith D. and Willett, John B. (2003). "Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence." Oxford University Press: New York. Chapters 1-4

Application Readings:

None this week

Week 13 – April 7

Topics: Multi-Level Models for repeated measures part 2

Methods Readings: TBA

Application Readings: TBA

Week 14 - April 14

Topics: Text as Data; regular expressions; string manipulation

Methods Readings: TBA

Theory and Application Readings:

Hollibaugh, G. E. (2018). The Use of Text as Data Methods in Public Administration: A Review and an Application to Agency Priorities. Journal of Public Administration Research and Theory, 29(3), 474-490. doi:10.1093/jopart/muy045

Week 15 - April 21

Topics: Repertory grid methodology

Methods Readings:

Fransella, Bell, & Bannister. (2003). A Manual For the Repertory Grid Methodology, 2nd Ed. Chapters 1-3.

Look over documents and code on: http://docu.openrepgrid.org/index.html

Application Readings:

Kearns, K. P., Bell, D., Deem, B., & McShane, L. (2012). How Nonprofit Leaders Evaluate Funding Sources: An Exploratory Study of Nonprofit Leaders. Nonprofit and Voluntary Sector Quarterly.

Krackhardt, D., & Kilduff, M. (1990). Friendship Patterns and Culture: The Control of Organizational Diversity. American Anthropologist, 92(1), 142-154.

Week 16 - April 28

****Due**: New Methods Presentations

Finals Week – May 5

****Due**: Replication Paper