RPOS 517: Quantitative Methods

Professor: Matthew C. Ingram
Office: Milne Hall 314-A
Dept: 518-442-3248
Offc: 518-442-3940
email: mingram@albany.edu

Location: DR-014
Time: M 1:40-5:20
Office hours: TTh 9:00-10:30
or by appointment
Contact Center: HU-016

Course Description (Brief). This is the core, required course in quantitative research methods for doctoral students in the Department of Political Science. As such, it is intended to provide you with a basic expertise in statistics for social science so that you can be a literate participant and contributor in the discipline. If you decide you are interested in quantitative methods, this course should provide you with the fundamentals to then decide what other methods might interest you and to capably enter and participate in those courses. If you decide you are not interested in quantitative methods and this may be your last course in this area, then this course should provide you with the basic skills to communicate with and evaluate other colleagues who employ these methods in their work.

Course Description (Extended).
Let's be honest – few students look forward to studying research methods.
Yes, this course involves some math.
Yes, it also involves some computer programming.
No, this is not too hard.
Research methods are about careful, critical thinking. If we want to draw good conclusions about what political systems are like, what causes what, how things operate, etc., we should reflect on the way we reach our conclusions. That is, we should be self-conscious – not insecure, but rather self-aware – in the way we draw conclusions. In short, research methods are about *thinking about the way we think*. Some of the material in this course is conceptual, overlapping with fields in philosophy of science and epistemology – how to identify valuable knowledge and know how to generate valuable knowledge of our own. Casual, loose, unsystematic thought is likely to give us a flawed picture of the very things we seek to understand clearly. So, its important to have careful methods for observing political phenomena, analyzing patterns in our observations, and reaching conclusions about causation, even prediction.

In this spirit of self-conscious thinking, this course offers an introduction to the tools of political analysis, emphasizing basic statistical methods, but also integrating non-mathematical tools. Careful thinking involves lots of things beyond statistics, such as defining concepts clearly (Week 3). For instance, what do we mean by "democracy"? Is the concept just about voting for representatives? Or does it mean more than that, perhaps real fairness and equality? Also, much work in political science is non-quantitative (i.e., interpretive or qualitative), and there is a rich and evolving tradition in "mixed methods", employing both quantitative and qualitative techniques. Towards the middle of the course (Week 7), we spend a session on using quantitative tools for case selection in order to systematically integrate the case(s) chosen for qualitative research within the larger sample of observations studied quantitatively.

Several factors have led to a dramatic increase in the use of statistics to study political phenomena. First, increasing availability of large data sets allows scholars to test hypotheses that previously could not be tested. Second, advances in computing have made quantitative methods relatively accessible. The **Stata** software and **R** and **LaTeX** environments in this course are examples of this, and we will spend some time familiarizing ourselves with these

tools, that is, "tooling up". Third, quantitative methods have proven to be powerful and flexible tools for social scientists. With many observations we inevitably need a system to summarize and organize the data for us, and thus we turn to statistics. This course uses a hands-on, applied approach to these statistics, including learning how to use **Stata** and **R** statistical software. After the course, you should feel comfortable reading much of the political science research published in academic books and journals, confident in your ability to critique this work, and well on your way to designing and producing work of your own. That is, this course prepares you to contribute evidence of your own to the growing field of knowledge in our discipline and other social sciences.

It is important to recognize that the statistical methods highlighted in this course are not the only way to go about studying the political world. The availability of large sets of data and the accessibility of computing software make for a powerful historical juncture, and in this juncture statistics seems like a reasonable way to study politics. Some might even say that you cannot say anything meaningful about the world without a large enough number of observations and statistical methods.
I do not share that view.
Indeed, I am fairly promiscuous in my own research methods, borrowing from different traditions depending on the question that interests me. At the end of the course, we'll briefly discuss other, complementary methods, and ways of systematically integrating quantitative and qualitative methods in your research. Hopefully, the range of methods you see, including statistics, will inspire you to continue studying methods. Again, this course is an introduction, and my hope is that you'll take more courses like it in the future. Perhaps you'll become a methodologist! At a minimum, I hope you continue to think about the way you think. Unlike other areas of your life, in methods be self-conscious and promiscuous!

Pre-requisites. For political science graduate students (MA and PhD), the main requirement for this course is RPOS 516. Otherwise, there are no pre-requisites. That said, any background in research design, research methods, math, statistics, or programming will be helpful. If it has been some time since you last took a mathematics or statistics course, please see the recommended reading under the first section, "Introduction", in the course outline below.

Readings.
The main text for this course is:

MB Bailey, Michael. 2016. *Real Stats: Using Econometrics for Political Science and Public Policy*. Oxford: Oxford University Press.

The companion website is:

- Real Stats Companion Website

Copies of this text are available at Mary Jane's bookstore. You can also find used and new versions at various online booksellers. The text will be supplemented periodically with journal articles and other short pieces. Please see the course outline below for a detailed list of the reading assignments. Articles and other short pieces are available via the library's electronic databases, but I will also place them on Blackboard. If unavailable online, I will provide these materials in class. We meet once per week, and you should have all reading done prior to each class meeting.

Requirements and Grading.
The requirements for this class and approximate weight of these requirements are:

1 Attendance and Participation (see below)
2 Weekly Assignments: 25%

3 Two take-home mid-term exams: 35%
4 Replication project 25%
6 Review of colleague's research plan 15%
7 Tutorial (extra credit 2%)

1. Attendance and Participation.
Active engagement with the material is critical to your success in class, especially in graduate school. For this reason, you should be present in class and come to each class session prepared to actively participate in the discussions, exercises, and other in-class activities. Some of this class requires brief lectures, but the class is oriented primarily towards applied, hands-on engagement with the material and active discussion. The small class size should facilitate closer interaction than in larger classes. If you have two unexplained absences from class, you will receive a failing grade for the semester. For absences to be explained, you must have a reasonable explanation related to extraordinary circumstances (e.g., family or medical emergency) and documentation of these circumstances.

Separately, classroom conduct can also affect how your participation is evaluated. You are expected to promote a classroom environment that makes it easy for your colleagues to engage with the material. In this regard, please keep distractions to a minimum. With regards to technology in the classroom, please turn your phones and other electronic devices off during class. Texting and other messaging during class is unacceptable. Laptop use is allowed and encouraged for taking notes and for working on other activities relevant to class, but sending emails, messaging, checking social media, or watching videos online is unacceptable. If you are texting, on social media sites, watching videos, or otherwise using technology inappropriately in the classroom, you will be asked to leave for the day. If you are asked to leave, please do so promptly and without disrupting the class further; if you would like to discuss why you were asked to leave, please contact me by email and we can discuss what happened at a later time. If this happens a second time, you will receive a failing grade for the class, just as if you had two unexplained absences from class. If you have extraordinary circumstances that justify keeping your phone on during class (e.g., family or medical urgency), please let me know ahead of time, before class starts that day.

2. Weekly Assignments.
You will receive short assignments **before** each seminar meeting that you have to complete either before or during class, as indicated in each individual assignment. These assignments are identified in the class schedule below under the heading "DO" for each week, and the assignments will either be sent via e-mail or available for download on the course website. You must submit the at-home portion of any assignment via **Blackboard by noon (12:00PM) on Sunday** prior to the class in which the assignment is due. The in-class portion of the assignments will be collected at the end of our class meetings. For assignments involving work in Stata, you must submit the related ".do" file, following the template provided. Group work is encouraged on in-class assignments; if you work in a group, you must still submit an individual assignment and identify who you worked with on that assignment.

3. Mid-Term Exams.
In weeks 5 and 10, you will receive exams to take home and complete on your own. These assignments must be submitted as PDF files via Blackboard on the indicated due dates (usually end of the week).

4. Replication Project.
Replicating (i.e., reproducing) other scholars' work is a key element of the scientific process. To engage with quantitative social scientific studies, you will replicate (reproduce) a study of your choice or from a list of suggested articles using the methods you are leaning in our course. This assignment will also give you some insight on how to conduct your own data analysis. By week 8, you will identify a scholarly article from a political science journal that uses quantitative methods (from bivariate associations to multivariate linear regression) and

for which replication data are publicly available. You will then complete the following steps:

- Retrieve the article, its replication data, and identify the main results
- Write an outline of your replication plan
- conduct the replication of the main analysis in the article and write a replication memo, summarizing the core results from the article and comparing those with the core results from your own analysis.

5. Research Plan
To facilitate your use of the methods learned in this course, you will compose a research plan that will help you work towards a publishable paper (e.g., in POS 695-696). This research plan is also similar to the type of document you might submit to pre-register a study at an academic journal. For details on pre-registration, see *Political Analysis*, Volume 21, Issue 1, and the 2014 call for proposals from Comparative Political Studies. Your document needs to contain a clear statement of your research question, a summary of the theoretical and empirical motivations for your question, a clear statement of working hypotheses and the empirical implications of these hypotheses, a clear explanation of the data and methods you will use in your research, and a summary of preliminary or anticipated answers to your question. Given the nature of this course, the greatest emphasis will be placed on the methods (i.e., the techniques you will use to analyze your data). Considering this emphasis, you will need to show data and demonstrate your methods. If you already have the data you will be analyzing, then this is a good opportunity to take a first look at those data and what results might look like. If you do not have any data, then you need to simulate data (i.e., generate fake or "dummy" data) given what you know about the features of your data (e.g., level of measurement, range, central tendency, dispersion). In either case, you then need to take examine at least a sample of these data and show what these data look like (descriptive statistics) and at least report the specific explanatory techniques and format of the results on which you would base your future conclusions (inferential statistics). You will submit the document (15-20 double-spaced pages) and the data to me by May 1, the penultimate day of class. I will then send the document to a randomly assigned colleague for review (see below).

6. Review of colleague's research plan.
Reviewing others' work (e.g., article manuscripts, grant proposals) will be an important part of your work as a scholar or professional. To prepare for this and practice your applied use of the methods covered in this course, you will provide constructive feedback to a colleague's research plan (and will receive such feedback from one of your colleagues). In the final week of class, you will produce a 2-page review of a randomly assigned research plan from someone else in class. I will provide additional guidelines on this assignment at that time. Your colleague will receive your review of their work after the last day of class.

7. Tutorial.
You will be asked to give a brief (e.g., 10 minutes) tutorial to teach the rest of us a computing skill you have learned on your own and have found useful in your work. Please see separate instructions and sign-up sheet for tutorials.

Grading.
Final grades will be assigned according to the following guidelines.

Grade	Score	Grade	Score
A	93-100	C+	77-79.99
A-	90-92.99	C	73-76.99
B+	87-89.99	C-	70-72.99
B	83-86.99	D	60-69.99
B-	80-82.99	E	0-59.99

Other Policies.

Email.
I expect you to check your email. You are responsible for material sent by email.

Late Work and Missed Exams.
All work delivered in class must be turned in within the first 10 minutes of class on the day it is due, or by 5pm if there is no class on the due date or no other established deadline. Without a legitimate (e.g., medical or family emergency) and documented explanation, late work will be penalized one letter grade (10%) for each day it is late, and it is considered late if turned in beyond the time limits above (i.e., after the first 10 minutes of class, or after 5pm on days there is no class; this includes weekends and holidays). I stop deducting points after 5 days (50%), so even if you are more than 5 days late on an assignment, it is better to turn something in rather than to have a zero for that assignment. No late work will be accepted after the last day of class. No late exams will be given.

Academic Integrity. All students must familiarize themselves with the Standards of Academic Integrity on the Universitys website and pledge to observe its tenets in all written and oral work, including oral presentations, quizzes and exams, and drafts and final versions of essays. The full standards and examples of dishonest behavior are available here.
The most common violation of academic integrity is plagiarism or cheating.
My advice is simple: dont do it.
Dont even think about doing it.
Plagiarism is the use of someone else's words or ideas without giving the original author credit by citing him or her. If you use someone else's language directly, you must use quotation marks. If you rely on another person's ideas in creating your argument or shaping your research, you must provide a citation that explicitly acknowledges the source of those ideas. If you have any questions about plagiarism, please contact me before you submit the assignment for grading. Plagiarism or cheating will result in a failing grade for the assignment and the submission of your name to the Office of Conflict Resolution at the very minimum. Ignorance of this policy will not provide a defense to the application of this policy.

Americans with Disabilities Act (ADA). Qualified students with disabilities needing appropriate academic adjustments should contact me as soon as possible to ensure your needs are met in a timely manner.

Miscellaneous. If you feel you need any help or simply want clarification on any of the material, please do not hesitate to raise your question in class or approach me outside of class. I hold regular office hours twice per week on main campus (see page 1). If you cannot arrange to come talk with me during these hours, please email me or contact the Department of Political Science administrative offices so that we can set up an appointment.

Changes to Syllabus. This syllabus may be modified throughout the semester. I will announce any changes. Current version of syllabus can always be found on Blackboard.

Additional Resources:

- Books:
  - Statistics
    - OpenIntro Statistics: https://www.openintro.org/stat/textbook.php
  - Econometrics
    - Gujarati, Damodar, and Dawn Porter. *Basic Econometrics*. 5th ed. New York: McGraw-Hill.

    - Angrist, Joshua D., and Jrn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.
  - Math for Social Sciences
    - Moore, Will H., and David A. Siegel. 2013. *A Mathematics Courts for Political and Social Research*. Princeton: Princeton University Press.
    - Gill, Jeff. 2006. *Essential Mathematics for Political and Social Research*. Cambridge, UK: Cambridge University Press.
- Videos:
  - OpenIntro: https://www.openintro.org/stat/videos.php
  - Johannes Karreth: http://www.jkarreth.net/rpos517-s16.html
  - David Siegel's companion videos for Moore and Siegel (2013; see above): here
- Data:
  - Harvard Dataverse
  - ICPSR
  - Journal of Peace Research, Replication Datasets
  - Individual author sites:
    - Michael Bailey, Georgetown University
    - Rafael La Porta, Dartmouth College
    - Eric Neumayer, University of Exeter
    - Paul Poast, University of Chicago

Course Outline

Week 1 (Jan 23)

1 Introduction

Materials: None

Objectives:

- introductions (instructor, TA, students)
- review syllabus and clarify course content, structure, and expectations
- identify materials and computing requirements
- address any early questions
- begin addressing computing resources (will do more of this over next 2 weeks)

Computing Resources

Key Questions:

- What are elements of a good work flow or production process?
- What are some commonly used and valuable tools I should integrate into my work?
- Why should I learn a programming language?
- What does "free and open-source software" (FOSS) mean?

Stata

Stata is a commercial software package for statistical analysis. For the sake of coordination and complementarity, in 2016 Rockefeller College decided to emphasize instruction in Stata in all its core, graduate methods courses (PAD504, POS517, PAD705, PAD725).

All computers in university labs on campus including the computer labs in Husted Hall, Draper Hall, and Dewey Library have Stata installed. Please note, however, that you cannot make modifications to Stata on these computers because you would need administrator privileges to do so. Thus, making any modifications to administrative settings (e.g., installing new packages or updates; registering Stata in order have access to it via Python, IPython, and Jupyter notebooks; see below) will not be possible for you on these university computers. Still, you should be able to access and use Stata without any cost if you only work on university computers.

Some of you may want to work at home, on the road, or elsewhere at off-campus locations, or may want to modify Stata as administrator. To do this, you have two options.

- purchase Stata
  - I do not personally promote commercial, proprietary software for statistical analysis, especially in an educational context (if I have not said this already, I will be repeating this message throughout semester)
  - If you would like to purchase Stata, please check student pricing options here. If you decide to make the purchase, I would recommend Stata/IC. This version has a much larger capacity for observations than "Small Stata", and there is no increase from this version to the next (Stata/SE). Also, the version installed on

campus computers is Stata/IC, so you are likely to have fewer compatibility issues if you purchase the same version.
  - The benefits of purchasing your own copy of Stata are: (1) you can then make modifications to basic settings as the administrator, and (2) you have access to Stata on your own computer(s) where you are (home, traveling, etc.).
- remotely access XStata
  - Access version of Stata from off-campus locations using VPN
  - More on this later as I get updates on equipment updates on main campus
  - Keep in mind that this is still a university resource, so you will still not be able to make adjustments as administrator

For additional resources in using Stata, see the following:

- UCLA Institute for Digital Research and Education (IDRE): here
- Stata video tutorial (also see other resources under "Training"): here

Why Stata?
Stata is a commercial, proprietary software package for statistical analysis. There are many other statistical packages out there including other commercial ones (e.g., SAS, SPSS, MATLAB) and non-commercial options (e.g., R, Python). For the sake of coordination and complementarity, in 2016 Rockefeller College decided to emphasize instruction in Stata in all its core methods courses (in both PADP and POS). Thus, Stata is the core statistical software used in this course. However, we will complement Stata with **R** (see below).

Alternatives: learning about FLOSS, R, Python, Jupyter

- FLOSS
  - At GNU philosophy, browse web site but also focus on:
    - What is "free software"?
    - What is difference between 'free' and 'open-source'??
    - What is "copyleft"?
    - Why we should care about about Free Software and Education?
- Python
  - For overview, see here
  - To download, I recommend Anaconda (see notes on Jupyter below)
  - To install packages:
    - Windows: install from command line using 'pip install', e.g. 'pip install ipystata'
    - Mac:
    - for details, see:
      - installing python packages
    - for some introductory videos on Python, search online or see my POS619 course site here
  - For integrated environment, use Jupyter Notebooks (see below)
- R
  - For overview and download options, see here
  - For new users, I recommend RStudio; can download for free here

  - To install packages:
    * use 'install.packages()'
    * and also need to open package using 'library()'
    * for some introductory videos on Python, search online or see my POS619 course site here
  - For integrated environment, use Jupyter Notebooks (see below)
- For addition resources using R, see the following:
  - R Project (links to manuals, FAQs, etc.): here
  - R Journal
  - Quick R
  -
- **Using Jupyter**

  - Steps:
    * Download and install Anaconda (includes Python, Spyder (IDE for Python), IPython, and Jupyter
    * At command line, install nbconvert and pandoc:
      · 'pip install nbconvert'
      · 'pip install pandoc'; if this does not work, download pandoc manually here
    * for details, see:
      · installing python packages
      · setting up ipystata
  - For integrated environment, use Jupyter Notebooks
- **Using Stata with Python**

  - To install packages:
    * Windows: install from command line using 'pip install', e.g. 'pip install ipystata'
    * Mac:
    * for details, see:
      · installing python packages
      · setting up ipystata
  - For integrated environment, use Jupyter Notebooks
- **Using R with Python**

  - Download Jupyter Notebooks (as part of Anaconda) and use R Essentials (includes 80 common R packages)
    * Windows: install from command line using 'conda install -c r r-essentials'
    * Mac:
    * for details, see:
      · using R with conda
  - For integrated environment, use Jupyter Notebooks

LaTeX and Text Editors
There is a quick comparison of different TeX editors here.

Operating Systems: Windows, OS, or Linux? Distributions of every tool discussed thus far are available for Windows, OS, and Linux. Most individuals in this class will be working in either Windows or OS environments. In some cases (e.g., Stata Automation mode of ipystata in Python), a distribution is available for Windows but not for OS or Linux. In such cases, you might consider a Windows emulator. If you find yourself in this situation, I recommend WINE.

Week 2 (Jan 30)

2 Big Picture: Two Revolutions

Readings:

- Bailey, ch. 1, ch. 16
- Shmueli (2010)

Summary:
Two revolutions that have recently marked social science research, as well as research in other disciplines, are: (1) causal identification, and (2) data access, reproducibility, and transparency. In this (e.g., DA-RT). In this section, we will focus primarily on the causal identification revolution and only briefly introduce the second revolution, which we will address in greater depth at a later point in the course. Thus, here we focus on understanding what we mean by causation or causality, how an emphasis on causation is separate from other types of propositions in scientific research, and how social scientists approach causation.

Terms (from Bailey, p.23, and others):

constant	experimental study	observational study
control group	generalizability	prediction
correlation	independent variable	randomization
dependent variable	intercept	scatterplot
endogenous	internal validity	slope coefficient
error term	model	treatment
exogenous	modeling	treatment group

Week 3 (Feb 6)

3 Concepts and Concept Formation

Readings:

- Sartori 1970
- Gerring 1999
- Gerring and Barresi 2003
- Collier et al. 2006

Summary:
This week we take a close look at concept formation. For the variables that we use in statistics to make any sense, they must be reasonable translations or operationalizations of our concepts. Thus, we must understand some basic principles of concept formation before we can assess data, measurement, and variables.

Week 4 (Feb 13)

4 Measurement, Data, and Good Data Practices

Readings:

- Stevens
- Bailey, ch. 2

Summary:

This section assesses the transition from concept to variable, providing tools for a better understanding of good measurement practices and good data handling practices.
Here, we revisit the second revolution from Week 2, namely, the reproducibility revolution.

Reproducibility

Readings:

- Bailey, ch. 2
- Stodden et al. (2016)
- ASA 2016 Statement and Comments

Key Questions:

- What is reproducible research?
- Why should we care about reproducibility?
- How can we best attain reproducibility?
- What are elements of a good work flow or production process?

Summary:

Notes: Reproducible Research

- Work Flow
  - Identify common or even repetitive aspects of this process
  - Develop a system that makes this process more efficient and *works for you*
  - In this course, we will use the following:
    - Statistical software: Stata
    - Markdown and LaTeX
- Notebooks
  - Statistical software: Stata
  -

TeXEditors

There is a quick comparison of different TeX editors here.

Week 5 (Feb 20)

Mid-Term Exam #1 handed out at end of class; due back by end of week

5 Bivariate Regression

Readings:

- Bailey, ch. 3

Summary:

In this section, we return in greater detail to the core model introduced in Ch. 2:

$$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

Specifically, we take a closer look at the following topics:

- interpretation of the models coefficients (beta0 and beta1)
- randomness of these coefficients
- bias and precision of these estimates
- outliers
- goodness of fit of the overall model

DO:

- at home: Assignment #2 (Ch. 3, Exercises 1-3)
- in class: Ch. 3, Exercises 4 and 5

Additional resources

- Optimization (e.g., minimizing sum squared residuals, SSR; see Bailey, p.49):
  - Bailey, ch. 14
  - Moore and Siegel, ch. 8, ch. 16
  - Videos
    - David Siegel's videos to complement Moore and Siegel Chapter 8 and Chapter 16
- Probability Distributions
  - Bailey, Appendix F–I
  - Moore and Siegel, ch. 10, ch. 11
  - Videos
    - David Siegel's videos to complement Moore and Siegel Chapter 10 and Chapter 11

Week 6 (Feb 27)

6 Hypothesis Testing

Readings:

- Bailey, ch. 4
- Wasserstein and Lazar (web link here: ASA 2016 Statement)

Suggested readings:

- Wasserstein and Lazar 2016 (ASA 2016 Statement), and full set of supplementary materials
- Gelman and Carlin 2014
- Gelman and Loken 2014
- Gelman 2004 (quick read)

Summary:
In this section, we examine statistical significance, power calculations, and various core elements of statistical inference.

TBA

DO:

- at home: Assignment #3 (Ch. 4, Exercises 1, 3, 4, and 6)
- in class: Ch. 4, Exercises 2 and 5

Week 7 (Mar 6)

7 Multivariate Linear Regression

Readings:

- Bailey, ch. 5
- Kastellec and Leoni
- Beck

Summary:
TBA

DO:

- at home:
- in class:

KEY IDEAS & SKILLS:

- multivariate OLS as method to address endogeneity
- omitted variable bias
- variance of β
  - bivariate OLS vs. multivariate OLS (check formulas below)
  - assumption in both: ϵ are (1) uncorrelated and (2) homoscedastic
- factors that affect $var(\beta_j)$:
  - model fit (see goodness of fit below, e.g., variance of regression, $sigma^2$)
  - N
  - variation in X_j (as variation increases, var(β) decreases)
  - multicollinearity (noise decreases precision, and as precision decreases risk of Type II error (false negative) increases)
- assessing multicollinearity (plots, VIF, auxiliary regressions)
- problems associated with multicollinearity (not bias; precision)
- assessing goodness of fit (add Adjusted R^2 to previous measures: variance of regression (σ^2), s.e. of regression (σ; Root MSE in Stata), plot, and R^2)
- model specification
  - inclusion and exclusion of variables
    - inclusion of irrelevant variables does not affect bias, but reduces precision
  - model fishing
  - replicability
- replication

FORMULAS:

Concept	Formula
var(β), bivariate OLS	$var(\beta) = \sigma^2/(N * var(X_j))$
var(β), multivariate OLS	$var(\beta_j) = \sigma^2/(N * var(X_j)(1 - R_j^2))$
VIF	$VIF = 1/(1 - R_j^2)$

NO CLASS MARCH 13: SPRING BREAK, MARCH 13-17

Week 8 (March 20)

8 Diagnostics and Evaluation

Readings:

- Bailey, Ch. 14, esp. 14.4-14.6; review sections 2.2, 3.6, 3.7, 3.8, 4.4, 4.5, 5.4, 5.5
- Kennedy
- Kastellec and Leoni
- Seawright and Gerring

Suggested(optional):

- Model Evaluation
  - Comments on and reply from Kennedy: Journal of Economic Surveys, Sep. 2002, Vol. 16., No. 4)
- Case Selection
  - Seawright, ch. 4 and 5
  - Center for Qualitative and Multi-Method Research
    - Symposium: Case Selection, Case Studies, and Causal Inference (Newsletter of Qualitative and Multi-Method Research, Fall 2008, Vol. 6, No. 2)
  - Gerring
- Transparency and Reproducibility
  - Data Access and Research Transparency (DA-RT)
  - Berkeley Initiative for Transparency in the Social Sciences (BITSS)

Summary:
TBA
Notes:
- cover case selection opportunities here
- cover p-hacking and model fishing again here (review ASA statement and supplements)
- cover 10 best practices here (e.g., Giles; Kennedy 2002; responses to Kennedy)

Week 9 (March 27)

9 Dummies

Readings:

- Bailey, ch. 6

Summary:

Week 10 (Apr 3)

10 Deeper Dive: Interactions

Readings:

- Review Bailey, ch. 6
- Braumoeller 2004
- Brambor, Clark, and Golder 2006

Summary:

NO CLASS APRIL 10: PASSOVER, APRIL 10-12 (M-W)

Week 11 (Apr 17)

11 Transformations and Polynomial Models

Readings:

- Bailey, ch. 7

Summary:

At this stage in semester, we have a few choices about what topics to consider over remaining weeks. Students will play a role in deciding by voting on preferred topics.

Among potential topics we could consider are the following:

- Advanced OLS (Bailey, Ch. 14)
- Panel Data and Fixed Effects (Bailey, Ch. 8)
- DID and Generalized Synthetic Controls (see Xu 2017)
- Dichotomous Dependent Variables (Bailey, Ch. 12)
- Data cleaning, organization, transformation
- Graphing
- Computing in R and RStudio
- Stata, R, and Jupyter
- Remote computing
- Other?

Week 12 (Apr 24)

12 Advanced OLS

Readings:

- (Bailey, 2016, ch. 14)

Summary:

Week 13 (May 1)

13 Dichotomous Dependent Variables

Readings:

- (Bailey, 2016, ch. 12)

Summary:

Week 14 (May 8)

Last day of class

14 Reflections and Extensions

- observational data and causation
- challenges in applied quantitative analysis
- extensions and preview of advanced methods and suggested courses (based on student interest)
- this last section can include grouped data (e.g., repeated measures in panel data, or nested data), interdependent data structures (temporal, relational, geographic dependence), logistic models, count models, etc.
- discuss fixed, random, WB effects here (assign mats from 619, but also Ellhorst 2003 to discuss fixed effect and random effect versions of "variable intercept" models, vs "variable coefficient" models

Readings:

- TBA

Summary:
TBA

DO:

- at home:
- in class:

Personal Webpage — ResearchGate — SSRN — YouTube — Google

References

Bailey, M. A. (2016). *Real Stats: Using Econometrics for Political Science and Public Policy*. Oxford University Press.

Beck, N. (2010). Making regression and related output more helpful to users. *The Political Methodologist 18*, 4–9.

Gerring, J. Case selection for casestudy analysis: Qualitative and quantitative techniques. In H. E. Brady, D. Collier, and J. M. Box-Steffensmeier (Eds.), *The Oxford Handbook of Political Methodology*. Oxford, UK: Oxford University Press.

Kastellec, J. P. and E. L. Leoni (2007). Using graphs instead of tables in political science. *Perspectives on Politics 5*, 755–771.

Kennedy, P. (2002). Sinning in the basement: What are the rules? the ten commandments of applied econometrics. *Journal of Economic Surveys 16*, 569–589.

Moore, W. H. and D. A. Siegel (2013). *A Mathematics Course for Political and Social Research*. Princeton: Princeton University Press.

Seawright, J. (2016). *Multi-Method Social Science: Combining Qualitative and Quantitative Tools*. Cambridge: Cambridge University Press.

Seawright, J. and J. Gerring (2008). Case selection techniques in case study research. *Political Research Quarterly 61*(2), 294–308.

Shmueli, G. (2010, 08). To explain or to predict? *Statist. Sci. 25*(3), 289–310.

Stodden, V., M. McNutt, D. H. Bailey, E. Deelman, Y. Gil, B. Hanson, M. A. Heroux, J. P. Ioannidis, and M. Taufer (2016). Enhancing reproducibility for computational methods. *Science 354*(6317), 1240–1241.

Wasserstein, R. L. and N. A. Lazar (2016). The asa's statement on p-values: Context, process, and purpose. *The American Statistician 70*(2), 129–133.