RPAD/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
web: mattingram.net

Location: HS-004
Time: Mon 5:45-9:25
Office hours: TTh 12:30-2:00
or by appointment
Contact Center: HU-016
(main campus)

TA: César Rentería
email: crenteria@albany.edu
office hours: Fri, 4-7pm, HS-004

Acknowledgments: This syllabus is the result of a wide range of sources, including my own training and research, courses and materials generated by others, and multiple discussions with colleagues and students. For their help and materials shared along the way, I thank Michael Bailey, Richard Ball, Alok Bohara, Shawn Bushway, Wendy Hansen, Shannon Jenkins, Johannes Karreth, Kevin McIntyre, Norm Medeiros, Philip Pollock, César Rentería, Doug Roscoe, Bill Stanley, as well as multiple materials and ideas from ICPSR, Project TIER, Open Stats Lab (OSL), SFI, and various methods courses and institutes.

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 covers basic data management, measurement, hypothesis testing, bivariate least-squares regression, multivariate least-squares regression, and some extensions. The course is designed 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 (e.g., time-series analysis, panel data, maximum likelihood estimation, network analysis, spatial analysis, hierarchical models, causal identification, etc.). 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.
I am well aware that many students approach this course with some trepidation.
Yes, this course involves some math.
Yes, it also involves some computer programming.
No, this is not too hard.
Yes, you can do it.
You might even find that it is fun and rewarding in ways you did not expect.

In brief, this class is about careful, critical thinking. If we want to draw sound conclusions about why certain policies are adopted and not others, why some policies succeed while others fail, what political systems are like, what causes what, how things operate, etc., then we should reflect on the way we reach our conclusions. That is, we should be self-conscious – not insecure, but rather self-aware and mindful – in the way we draw conclusions. In short, research methods is about thinking about the way you 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, it's important to have careful methods for observing phenomena, analyzing patterns in our observations, and reaching conclusions about the range of variation in the world, causation, and 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, R, and \LaTeX{} software in this course are examples of this, and we will spend some time familiarizing ourselves with these tools. That is, part of this course explicitly involves “tooling up” with computing resources commonly used by researchers. Third, quantitative methods have proven to be powerful and flexible tools for social scientists. With large data sets and 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. After the course, you should feel comfortable reading much of the social science research published in academic books and journals in your field, 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 political science, public administration, sociology, and many others disciplines.

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 in any given project. At various points, but especially in the section on mixed-methods, we’ll 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 pre-
requisite for this course is RPOS 516, in which you should have covered research design principles and drafted a research plan and a funding proposal. Otherwise, there are no pre-requisites. In public administration, RPAD 704 is very similar to RPOS 516. Public administration students likely have taken 505, which will help you in the first few weeks of the course. 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:


The companion website is:
- Real Stats Companion Website

Copies of this text are available at the campus 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 that is scheduled for that day done prior to start of class.

I also provide additional recommended readings each week. These are not required, but I offer them as another resource you can turn to if you find you would like to read something else beyond the assigned reading.

Requirements and Grading.
The requirements for this class and their approximate weight are:
1. Attendance and Participation (10%; but see below)
2. Weekly Assignments: 35%
3. Two take-home mid-term exams: 20% (10% each)
4. Replication project 15%
5. Review of colleague’s replication 5%
6. Preregistration exercise 10%
7. Tutorial 5%

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 relatively small class size should facilitate closer interaction than in larger classes. Since each session is nearly four hours and we move quickly from week to week, if you have more than two unexplained absences from class, you will receive a failing grade for the semester. For absences to be explained, you must have two things: (i) a reasonable explanation related to extraordinary circumstances (e.g., family or medical emergency) and (ii) documentation of these circumstances (e.g., note from doctor).

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 I will be happy to discuss what happened at a later time. If this happens a second time, you will receive a zero (0) for the participation component of class, and the highest grade you can earn for a semester will be a “B”. 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 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. There are two different kinds of “DO” assignments: “at home” and “in class”. 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; however, if you work in a group, you must still submit an individual assignment (.do file) 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 using the template provided. These assignments must be submitted as a single file (PDF) via Blackboard on the indicated due dates (usually end of the week). Dates may vary slightly depending on progress of activities in class.

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 learning in this 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 an academic 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:

(1) Retrieve the article and its data
(2) Identify the article’s main results
(3) Write an outline of your plan to reproduce the main results
(4) Conduct the replication of the main analysis in the article and write a report of your replication; this report should first summarize the data, methods, and core results from the original article, identify the data and methods used in the replication, identify the results you obtained, and then compare the original results with the core results from your own analysis.

Note that a “replication” can mean many things, including a step-by-step reproduction of the original findings using the original data and original approach, an approximation of the original results using the same approach but new or different data, and an approximation of
the original results using the same data but a different approach. We’ll discuss these options during semester (see King 1991).

5. Review of colleague’s replication.
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 on their replication (and you will receive such feedback from one of your colleagues). Towards the end of the semester, you will produce a 2-page review of a randomly assigned replication 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.

6. Preregistration exercise
To facilitate your use of the methods learned in this course and your understanding of issues related to research transparency, you will preregister a research plan of your own. Ideally, the research plan you preregister will be one of sincere interest to you (e.g., thesis or dissertation) in order to help you advance towards completion of your research project (e.g., in POS 695-696). You should feel free to use any research plan you developed in RPOS 516 or RPAD 704 as part of this preregistration exercise. A template will be provided to guide you through the exercise.

This preregistration document 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.

7. Tutorial.
You will be asked to give a brief (e.g., 10 minutes) tutorial to teach the rest of us a new computing skill or new tool you have learned on your own and have found useful in your work. The tutorials are design to help us learn from each other. I learn something new every year.

The main requirement of the tutorials is that you use the software we are using in class, primarily Stata and LaTeX. If you move beyond these tools, you may do so only if you are using a free, open-source tool, e.g., R, Python, Julia. All tutorials will take place after Week 7, and I prefer to have no more than 3 tutorials on any given day. Dates are available on a first-come basis, so sign up early or you will have less flexibility as to when you give your tutorial in class. Please see separate instructions and sign-up sheet for tutorials.

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

<table>
<thead>
<tr>
<th>Grade</th>
<th>Score</th>
<th>Grade</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>93-100</td>
<td>C+</td>
<td>77-79.99</td>
</tr>
<tr>
<td>A-</td>
<td>90-92.99</td>
<td>C</td>
<td>73-76.99</td>
</tr>
<tr>
<td>B+</td>
<td>87-89.99</td>
<td>C-</td>
<td>70-72.99</td>
</tr>
<tr>
<td>B</td>
<td>83-86.99</td>
<td>D</td>
<td>60-69.99</td>
</tr>
<tr>
<td>B-</td>
<td>80-82.99</td>
<td>E</td>
<td>0-59.99</td>
</tr>
</tbody>
</table>

Other Policies.

Email.
I expect you to check your email. You are responsible for material sent by email.
Late Work and Missed Exams.
All at-home 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. All other work must be turned in according to deadlines established in class and course schedule below. All late work is penalized according to rules above.

Academic Integrity. All students must familiarize themselves with the Standards of Academic Integrity on the University’s 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: don't do it. Don't 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). Reasonable accommodations will be provided for students with documented physical, sensory, systemic, medical, cognitive, learning and mental health (psychiatric) disabilities. If you believe you have a disability requiring accommodation in this class, please notify the Disability Resource Center (518 - 442 - 5490; drc@albany.edu). Upon verification and after the registration process is complete, the DRC will provide you with a letter that informs the course instructor (me) that you are a student with a disability registered with the DRC and list the recommended reasonable accommodations.

The statement above also appears on our University website as part of our Statement of Reasonable Accommodation Policy In Response to the Americans with Disabilities Act that can be found at the following link: [here].

Counseling and Psychological Services (CAPS). If you are going through a difficult time (e.g., stress, anxiety, depression, relationship issues, traumatic experience), there are resources on campus to help you. The office on campus that houses these resources is Counseling and Psychological Services (CAPS). Services are free during the academic year, and confidential. Phone: 518-442-5800. Email: consultation@albany.edu. Web: http://www.albany.edu/counseling_center/

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. I retain the right to modify the syllabus and give notice in class of any modifications in a timely fashion. You are responsible for being aware of such notices and for identifying any changes. The current version of syllabus can always be found on Blackboard. If in doubt, check date in footer on first page of syllabus.

**Additional Resources:**

There are numerous outside resources from any number of disciplines or research fields that you could use to help you along in this course. Here are some that I have often turned to myself.

- **Open Resources:**
  - Textbooks:
    * [OpenIntro Statistics](#)
  - Videos:
    * [OpenIntro Statistics](#)
  - Software:
    * R
    * Python
    * Julia
  - Data:
    * [Open Data NY](#)
    * [NYC Open Data](#)
    * [Open Government](#) (U.S. Federal Government)
  - Exercises and Tutorials:
    * [Open Stats Lab (OSL)](#)
  - Collaboration and Development Platforms
    * [Open Science Framework](#)
    * [JupyterHub](#)
- **Books:**
  - Econometrics
    * Fox, John, Regression Diagnostics (SAGE; "little green book" series)
    * Fox, John, Applied Regression Analysis & Generalized Linear Models (SAGE, 3rd ed)
  - Math for Social Sciences
• Videos:
  – 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

• Centers, Institutes, Projects
  – Project TIER, Haverford College, PA
  – Center for Open Science, Charlottesville, Virginia
  – Berkeley Initiative for Transparency in the Social Sciences, BITSS

• Fun Challenges and Competitions
  – Center for Open Science, Preregistration Challenge
  – Predict U.S. Supreme Court decisions at FantasySCOTUS
  – Predict next Supreme Court justice at FantasyJustice
# Course Outline

## Week 1 (Jan 29)

### 1 Introduction

**Reading:**
- Ingram, “Notes on Computing” ([here](#))
- Ingram, “Introduction to Stata” ([slides here](#))
- Ingram, “Introduction to LaTeX” ([slides here](#))

**Objectives:**
- introductions (instructor, TA, students)
- highlight methods and self-esteem
- review syllabus and clarify course content, structure, and expectations; major goal today is clarify what to expect over next 14-15 weeks
- identify materials and computing requirements
- address any early questions
- Assignment 1: data management exercise is homework for next week (use Stata)

**DO:**
- at home: review readings above and introductory videos and other materials on computing resources
- in class: begin addressing computing resources (will do more of this over next 2 weeks)
  - Introduction to **Stata** (see [workshop materials](#))
  - Introduction to **LaTeX** (see [workshop materials](#))
Week 2 (Feb 5)

NO CLASS MONDAY, MARCH 5: PROF. INGRAM OUT OF TOWN
Need to identify alternate meeting date this week: Wed or Fri after 12:30

2 Assignment 01: Data Management Exercises Using Stata

Readings:
• Assignment 01 materials

Summary:
In this section, we will cover some basic data management skills in Stata. The assignment is graded on a “credit/no credit” basis, so as long as you show effort to complete the assignment and turn in your .do file documenting that effort you will receive credit.

DO:
• at home:
  – review introductory materials on Stata from last week and online
  – gather all materials for Assignment .01 (Blackboard)
  – complete Assignment .01 on your own or in groups
  – submit your own individual .do file on Blackboard by Sunday at noon prior to next class
• in class:
  – go over Assignment .01
  – clarify all Stata syntax

KEY IDEAS & SKILLS:
• familiarization with Stata
• core Stata data management commands
  – clear
  – cd
  – use
  – import
  – browse
  – summarize
  – describe
  – destring
  – tostring
  – keep
  – drop
  – if
  – generate
  – replace
  – collapse
3 Concepts and Concept Formation

Readings:
- Sartori 1970
- Collier and Mahon (1993)
- Collier and Levitsky (1996)
- Munck and Verkuilen (2002)

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. Ideally, our research should rely on good concepts, we string these concepts together to build good propositions/arguments/theories, our concepts lead to good measures, and our measures and theories lead to good, testable hypotheses that guide our quantitative analysis.

DO:

KEY IDEAS & SKILLS:
- concept structure and concept trees
- min-max approach
  - Sartori’s ladder of abstraction (generality)
  - classical subtypes
- diminished sub-types approach
- precising
- shifting overarching concept
- family resemblance
- radial concepts

Additional resources
Week 3 (Feb 12)

4 Measurement, Data, and Good Data Practices

Readings:
- Stevens
- Bailey, ch. 1 and 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.

DO:
- at home: Assignment #1A (Ch. 2, Exercise 1)
- in class: go over Assignment #1A plus Ch. 2, Exercise 2
- at home, after class: Assignment #1B (Ch. 3, Exercises 3 and 4)

KEY IDEAS:

Additional resources

5 Workflow, Research Integrity and Transparency, and Reproducibility

Materials:
- “Research Fails”
  - Herndon and Pollin (2013)
  - Bhattacharjee (2013)
  - Aschwanden and Koerth-Baker (2016)
  - Retraction Watch (skim entries)
- Bailey, ch. 2
- Stodden et al. (2016)
- Video: Transparency and Reproducibility, Bill Jacoby and Arthur Lupia (ICPSR 2017); esp after 11:20, and esp. after 23:40

Key Questions:
- What is reproducible research?
- Why should we care about reproducibility?
- How can we best attain reproducibility?
- What are elements of a good workflow 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

\LaTeX Editors
There is a quick comparison of different \TeX editors [here](#).

6 Computing 2

Readings:
• Review Ingram, “Introduction to \LaTeX”
• Review workshop materials on Workflow in Stata \LaTeX and MarkDoc [here](#)
• Review “Introduction to R”
• Review workshop materials on Workflow in R using \LaTeX and Sweave [here](#)

Summary:
In this section, we will cover additional computing resources, including R and general principles of workflow and research transparency.

DO:
• at home: review materials under readings above
• in class:
  – integrate Stata and \LaTeX using MarkDoc
  – integrate R and \LaTeX using Sweave in RStudio

KEY IDEAS & SKILLS:
• familiarization with \LaTeX
• begin familiarization with R and RStudio
• workflow principles
• research ethics, integrity, and transparency
Week 4 (Feb 19)

7  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 (\( \beta_0 \) and \( \beta_1 \))
- randomness of these coefficients
- bias and precision of these estimates
- outliers
- goodness of fit of the overall model

DO:
- at home: Assignment #2A (Ch. 3, Exercise 1)
- in class: go over Assignment #2A plus Ch. 3, Exercises 2 and 3
- at home, after class: Assignment #2B (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 5 (Feb 26)
Mid-Term Exam #1 handed out at end of class; due back by end of week

8 Hypothesis Testing

Readings:
- Bailey, ch. 4
- Wasserstein and Lazar (web link here: ASA 2016 Statement)
- Daniel J. Benjamin and Johnson (web link here: Nature Human Behavior)
- Resnick (web link here: Vox.com)

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)
- McShane, Gelman, et al. (2017)

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 6 (Mar 5)

9 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 $\beta$
  - bivariate OLS vs. multivariate OLS (check formulas below)
  - assumption in both: $\epsilon$ are (1) uncorrelated and (2) homoscedastic
- factors that affect $\text{var}(\beta_j)$:
  - model fit (see goodness of fit below, e.g., variance of regression, $\sigma^2$)
  - $N$
  - variation in $X_j$ (as variation increases, $\text{var}(\beta)$ 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 ($\sigma^2$), s.e. of regression ($\sigma$; 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:
<table>
<thead>
<tr>
<th>Concept</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{var}(\beta)), bivariate OLS</td>
<td>(\text{var}(\beta) = \sigma^2/(N \times \text{var}(X_j)))</td>
</tr>
<tr>
<td>(\text{var}(\beta)), multivariate OLS</td>
<td>(\text{var}(\beta_j) = \sigma^2/(N \times \text{var}(X_j)(1 - R^2_j)))</td>
</tr>
<tr>
<td>VIF</td>
<td>(VIF = 1/(1 - R^2_j))</td>
</tr>
</tbody>
</table>

NO CLASS MARCH 12: SPRING BREAK, MARCH 12 – 16
Week 7 (Mar 19)

**MIDTERM POINT**

10  OLS Assumptions, Diagnostics, and Model Evaluation

Readings and Materials:
- Fox
- Lewis-Beck
- 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
- ASA 2016 Statement and Comments
- Kastellec and Leoni

Suggested (optional):
- Model Evaluation
- 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 8 (March 26)

11 Case Selection and Mixed-Methods Designs

Readings:
- Seawright and Gerring

Suggested (optional):
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

Summary:
This week we revisit model diagnostics, using the tools from last week as potential quantitative tools for case selection to inform mixed-methods research designs.

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 (April 2)

12 Dummies

Readings:
- Bailey, ch. 6

Summary:
13  Deeper Dive: Interactions

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

Recommended readings:
- Esarey and Sumner; web link here
- Franzese and Kam

Summary:
Week 11 (Apr 16)

14 Transformations and Polynomial Models

Readings:
- Bailey (2016, ch. 7)

Recommended:
- Lewis-Beck and Lewis-Beck 2016, 86-93

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 23)

15  Advanced OLS

Readings:
• (Bailey, 2016, ch. 14)

Summary:

Week 13 (Apr 30)

16  Dichotomous Dependent Variables

Readings:
• (Bailey, 2016, ch. 12)

Summary:

Week 14 (May 7)

**Last day of class**

17  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


