UAlbany / Spring 2016 / Tuesdays 1:15pm–4:55pm / Husted 006

RPOS 517: Quantitative Research Methods

Dr. Johannes Karreth

Office: Milne 300G (Center for Policy Research)
Office hours: Tuesdays 5:00pm–7:00pm and Wednesdays 11am–12pm; and other times by appointment via email.
Email: jkarreth@albany.edu
Course website: http://www.jkarreth.net/rpos517.html (Contains links to all materials mentioned in this syllabus)

Teaching assistant: Nakissa Jahanbani
TA office: Milne Graduate Student Lounge (1st floor).
TA office hours: Thursdays 1pm–2pm; and other times by appointment via email.
TA email: njahanbani@albany.edu

Course description and objectives

Welcome to RPOS/RPAD 517! Assuming little to no participant familiarity with quantitative methods–and even some trepidation toward statistics–this course provides an introduction to basic statistical methods for the social sciences. Upon successful completion, course participants will be better consumers of statistical methods, have acquired competence and understanding of introductory statistics and be able to apply these methods to social science problems of interest to them. The course also prepares participants for taking more advanced and specialized quantitative methods courses. Special emphasis is placed on applied data analysis. Participants will work with a variety of hands-on exercises to prepare them to conduct their own research projects.

Although the purpose of this class is to insure participants acquire a fair degree of mathematical/statistical knowledge, no prior statistical knowledge is assumed. While the course has a core of set goals, we will slow down or speed up according to participant needs. Variation in enthusiasm and ability will be respected. At the other extreme, if participants have had appropriate statistical background, we will move more quickly.

Upon successful completion of this course, participants will have acquired an understanding of:

- how quantitative methods can contribute to study social and political phenomena, help make inferences about relationships, and contribute to evaluating theories
- differences between experimental and observational data and their implications for interpreting quantitative analyses
- how to describe quantitative data
- how to make inferences and test hypotheses using quantitative data
- how to identify, assess, and interpret relationships among variables
- the logic and assumptions of linear regression modeling
- diagnostics of linear regression models
- common problems in fitting linear regression models to empirical data
- limitations of contemporary quantitative approaches to social science

Participants will acquire practical skills in:

- using software for data management, analysis, and creating presentable summaries of findings
- documenting a reproducible workflow from beginning to end
- building on a core set of skills to learn new tools and commands in other, subsequent courses

Prerequisites

The prerequisite for POS Ph.D. students is RPOS 516 (*Introduction to Political Inquiry*). Exceptions may be granted on an individual basis and for graduate students in programs other than POS.

Communication

We use Blackboard exclusively for communication in this course. If I have any **announcements**, I will make them via Blackboard. Blackboard forwards announcements to your `albany.edu` email address. If you have any **questions** for me related to the course (about assignments, being stuck with an exercise, or any other problems), please use the appropriate **discussion board** on our Blackboard site rather than emailing me. This is not because I do not want to respond to your emails; rather, there is a high chance that you are not the only participant with a question. Using Blackboard for asking & answering questions makes any exchanges available to the whole seminar and much more efficient.

Materials

The readings for this course are split into technical (mostly textbook) readings [T], lighter background readings [B], and applications [A] of the techniques as you will encounter them in your work as social scientists. [T] readings are required (unless otherwise noted). [B] readings are strongly recommended. You should read at least one [A] article for each week and you need to read those that are highlighted orange (i.e. required). My goal is to keep the financial cost of this course as low as possible for you. But you may find it worthwhile to purchase some of the recommended resources; in that case, consult with me first to see what might make the most sense for you.

Books

You **need** to have access to three books and one website for this course:

OI Diez, D. M., Barr, C. D., and Çetinkaya Rundel, M. (2015). *OpenIntro Statistics, Third Edition*. OpenIntro.org. This is an open-source, but high-quality textbook on introductory statistics. You can download a PDF copy for free at `www.openintro.org` or purchase a paperback for $10 at `amazon.com`. The companion website at `www.openintro.org` provides plenty of useful materials to work through the examples used in the book.

AR Fox, J. (2008). *Applied Regression Analysis and Generalized Linear Models*. Sage Publications, Thousand Oaks. This is a more technical textbook on regression analysis that will serve you far beyond this course. While we will not work through all 22 chapters of the book, you will be able to study them independently after you successfully complete our course. The companion website at `tinyurl.com/arabook` provides useful appendices, data sets, and exercises. *Note: A third edition of this book was just published. You will be able to follow the seminar using either the second or third edition.*

NS Wheelan, C. (2013). *Naked Statistics: Stripping the Dread from the Data*. W. W. Norton, New York, NY. This light background reading addresses many "so what" questions in a broader context. It will be useful to contextualize the methods we learn in our course.

QR Kabacoff, R. Quick-R. Available at `statmethods.net`. This website offers well-explained computer code to complete most, if not all, of the data analysis tasks we work on in this course.

The required books will help you successfully complete this course. The following **recommended** books will be useful if you find the required books to be too narrow and want to learn more about the math behind the concepts we discuss and the programming language we use.

MC *Math background reading:* Moore, W. H. and Siegel, D. A. (2013). *A Mathematics Course for Political and Social Research*. Princeton University Press, Princeton, NJ. This book is excellent at explaining the mathematical concepts used in the techniques we explore. It features useful "why should I care?" sections for each concept it introduces.

RC *Software guide:* Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression, Second Edition*. Sage, Thousand Oaks. This companion to *Applied Regression Analysis* introduces the R software we use in this course. If you would like a lengthier treatment of the methods explained on statmethods.net (and more), I recommend this book. You do not need to purchase the book to be successful in this course, but you might find it and its companion website at tinyurl.com/carbook useful.

DS *Basic introduction to Statistics and R:* Field, A., Miles, J., and Field, Z. (2012). *Discovering Statistics Using R*. SAGE Publications, Thousand Oaks. If you learn better with a lightly written textbook with many applied examples, tests, and definitions, you will like this book. It covers much of the content of our course. While it lacks the mathematical depth of the main texts for our course, it can be quite helpful as a step-by-step background reading that combines concepts and execution in R.

EA *Econometrics in general:* Greene, W. H. (2012). *Econometric Analysis, Seventh Edition*. Prentice Hall, Upper Saddle River, NJ. This book is a good investment if you realize you will be using quantitative methods throughout your career; it will be useful as a backup when you keep learning new methods.

There are dozens of other quantitative methods books out there, though, and I'd like you to be aware of some that you will likely find helpful if you continue using quantitative methods in your career. Come see me during office hours to chat about other books you may want to look into—I have a good number of books in my shelf that I can recommend based on your specific applications and interest.

Articles

Articles are available through the UAlbany library ⟨http://library.albany.edu⟩. The library website offers helpful tutorials on how to retrieve scholarly articles here: ⟨http://library.albany.edu/usered/find/index.html⟩. Any book chapters/excerpts not available through the library are posted on Blackboard ⟨https://blackboard.albany.edu⟩.

Videos

I will introduce some of the topics in this course through video recordings. You are required to watch these videos before the day to which they are assigned. Take good notes and write down any questions you have before you come to class meetings. Videos usually elaborate on ideas that are either not in the reading for a particular day or not addressed in sufficient detail in the readings. For each day, videos are linked on the course website at ⟨jkarreth.net/rpos517.html⟩.

Computing

We will primarily rely on R, a highly versatile open-source software package that you can download for free. Please refer to the respective tutorial from Day 1 for instructions on how to set up an R on your personal computer. Resources on using Stata will also be made available in the course for self-study for participants interested in broadening their software skills. **To complete this course, you need to bring your laptop computer to every seminar meeting.**

Tutorials. For each day, I provide detailed tutorials that you can access through the course website. Read these tutorials before class and work through them on your own. In the assignments for each class, I will ask you to complete tasks that will use the skills you encounter in these tutorials.

Reproducible computing. All work you do as a social scientist, particularly any data analysis you use to reach conclusions, needs to be reproducible. To this end, our course puts special emphasis on techniques and tools that help you make your research reproducible. Scripts and data analysis notebooks are some of these tools. If you'd like to have a closer look at resources and techniques for reproducible research, I recommend the following print books:

- Stodden, V., Leisch, F., and Peng, R. D. (2014). *Implementing Reproducible Research*. Chapman and Hall/CRC, Boca Raton, FL

- Gandrud, C. (2013). *Reproducible Research with R and RStudio*. Chapman and Hall/CRC, Boca Raton, FL [If you know you will mainly use R in your work.]
- Long, J. S. (2009). *The Workflow of Data Analysis Using Stata*. Stata Press, College Station, TX [If you know you will mainly use Stata in your work.]

Requirements and assignments

As in every graduate seminar, you need to be prepared for each course meeting. In the schedule below, five tasks are listed for each day. The first three, READ, WATCH, and DO, need to be completed before the session. TERMS TO KNOW are for your self-assessment purposes: you should be familiar with these terms after the session. **Use these terms to guide your reading, especially of the AR book.** In addition, SOFTWARE SKILLS lists the data analysis steps you should be able to conduct on your own after the session.

1. Weekly assignments (20%)

You will receive short assignments **before** each seminar meeting that you have to complete before or during class, as indicated in the individual assignment. Specific assignments are noted under DO for each week and available for download on the course website. You submit the at-home part of these assignments via **Blackboard by 8am on Monday mornings** before class. The in-class portion of assignments will be collected at the end of our class meetings. For assignments involving work in R, you have to submit these assignments as RMarkdown data analysis notebooks, following the template `rpos517_assignment_template.Rmd` on the course website. Group work is encouraged on in-class assignments, but you need to indicate on the assignment with whom you worked on them (if applicable).

2. Two take-home midterm exams (30%)

In weeks 5 and 10, you will receive exams that I ask you to complete at home on your own. These assignments must be submitted as PDF files via Blackboard on the indicated due dates.

3. Replication project (25%)

Replicating (or more precisely, 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 our course. This assignment will also give you some insight on how to conduct your own data analysis. By week 4, you will identify a scholarly article from a political science journal that uses quantitative methods (from bivariate associations to multiple linear regression) and for which replication data is publicly available. You will then complete the following steps:

1. Retrieve the article, its replication data, and identify the main results
2. Write an outline of your replication plan
3. Conduct the replication analysis of the main analysis in the article and write a replication memo, summarizing your findings

4. Research plan (15%)

To facilitate your use of the methods learned in this course, you will compose a research plan that will help you write a publishable paper. This research plan is also similar to the type of document you would submit to pre-register a study at a journal. See issue 1, volume 21 of *Political Analysis* and the call for proposals for a forthcoming issue of *Comparative Political Studies* for details on pre-registrations. Your document needs to contain a summary of your research question, preliminary answer(s), research design, and a data analysis plan. In addition, you need to collect/compile and submit the data for this project. You will submit the document (8–10 single-spaced pages) and the data to me on May 6. I will then send the document to a randomly assigned colleague for review.

5. Review of a colleague's research plan (10%)

Reviewing others' work (e.g., journal articles, grants, or commissioned studies) will be an important part of your work as a scholar or professional. To prepare you for this and to practice your use of statistical methods, you will provide constructive feedback to a colleague's research plan. Between May 7 and May 13, you will produce a 2-page review of a randomly assigned colleague's research plan. I will provide guidelines for this review. Your colleague will receive your review after May 13.

Grading

Weekly assignments	20 pts
Two midterm exams	30 pts
Replication project	25 pts
Research plan	15 pts
Review	10 pts
Total	100 pts

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

Class policies

Attendance. Success in this course will heavily depend on students' engagement with the course materials. All students are expected to attend all classes and to complete all required readings and assignments when due.

Disability accommodations. Reasonable accommodations will be provided for students with documented physical, sensory, systemic, cognitive, learning and psychiatric disabilities. If you believe you have a disability requiring accommodation in this class, please notify the Director of the Disability Resource Center (Business Administration Building Room 120, 442-5490). That office will provide the course instructor with verification of your disability, and will recommend appropriate accommodations. If you wish to discuss academic accommodations for this course, please also inform me as soon as possible. While I will make every effort to accommodate valid requests, students should not expect that, if they do poorly on an exam or other assignment, to claim, at that time, the need of an accommodation. This statement is to preclude that problem, and allow people with a need for accommodations to be treated fairly and appropriately.

Religious observance. I will also make every effort to accommodate religious observances and obligations. You are asked to bring any possible conflicts to my attention as soon as possible. Students should not expect that, if they do poorly on an exam or other assignment, to claim, at that time, the need of an accommodation. This statement is to preclude that problem, and allow people with a need for accommodations to be treated fairly and appropriately.

Academic honesty. All students of the University at Albany are responsible for knowing and adhering to the academic integrity policy of our institution. You must read and familiarize yourself with this policy at ⟨http://www.albany.edu/graduatebulletin/requirements_degree.htm#standards_integrity⟩ before the course begins. Violations of this policy may include: cheating, plagiarism, aid of academic dishonesty, fabrication, lying, bribery, and threatening behavior. The policy also prohibits submitting material that was previously or is currently being submitted in another course. **If you are involved in plagiarism, the penalty will be failure in the course and you will be reported to judicial affairs. In this one regard there are no second chances. If you are not sure if something violates standards—ask. If you are not sure whether you can submit work that is related to something you have done or are doing in another**

course—ask. If you are not sure whether to cite or not to cite—cite. Every student must complete the following tutorial: ⟨http://library.albany.edu/usered/plagiarism/index.html⟩.

Schedule

This syllabus is a general outline for the class and may be modified as the course progresses. Please check your email daily for notifications about any changes or class cancellations. The latest version of the syllabus is always posted at `http://www.jkarreth.net/rpos517.html`; please consult that link for up-to-date information on assignments and dates. All videos, assignments, and other materials mentioned in the syllabus are posted on that site as well.

Day 1 (January 26)

Research design. Questions and models. Experimental vs observational data.
Using your computer as a scientific workstation.

READ:

- T: Chapter 1 in Morgan, S. L. and Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, New York. [Blackboard]
- T: OI sections 1.1, 1.2, 1.3, 1.4, 1.5
- B: NS ch. 1

WATCH:

- Video: Welcome!
- Video: Questions & models
- Video: Computing

DO:

- **At home:** Make sure your computer has Microsoft Word or Libre Office installed.
- **At home:** Download and install R
- **At home:** Download and install RStudio
- **At home:** Download and install LaTeX following the instructions under "Installation" at `jkarreth.net/latex.html`.
- **At home:** Read pp. 1–14 of the tutorial `Intro to R`
- **Bring** your laptop to class.
- **Bring** a hard copy of your research design assignment from RPOS 516 (if you took that seminar).
- **In class:** Work through pp. 15–18 of the tutorial `Intro to R`

TERMS TO KNOW:

- Causality
- Observational data
- Experimental data
- Treatment effect
- Console
- Object-based language
- Comment
- Script
- Dataset
- RMarkdown
- Package/library in R

SOFTWARE SKILLS:

- Install R and RStudio
- Install and load packages in R
- Open, edit, execute, and save an R script file
- Open and compile a template for an RMarkdown data analysis notebook in R

Day 2 (February 2)

Working with data. Descriptive & univariate statistics.

READ:

T: OI sections 1.6, 1.7, 1.8
B: NS chs. 2 & 3
A: Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, 103(2684):677–680
A: RC ch. 2 [Blackboard]

WATCH:

- Videos: Working with Data & Descriptive and Univariate Statistics

DO:

- **At home:** Read the tutorial `Intro to Data`
- **At home & in class:** Complete the tasks listed in `Assignments Day 2`

TERMS TO KNOW:

- Data set/matrix/frame
- Vector
- Summary statistic
- Observation/case
- Variable (continuous, categorical, ordinal)
- Association
- Independence
- Population
- Sample
- Bias
- Explanatory variable
- Response variable
- Confounding variable
- Experimental design
- Scatterplot
- Mean vs. median
- Histogram
- Density
- Skew
- Modality
- Variance
- Standard deviation
- Quartiles
- Column & row proportion
- Bar plot
- Box plot
- Simulation

SOFTWARE SKILLS:

- Open a dataset in Excel/Open Office/Google Drive
- Import a dataset into R
- Summarize variables and datasets
- Create a well-designed, editable document with descriptive statistics and graphs

Day 3 (February 9)

Probability & distributions.

READ:

T: OI sections 2.1, 2.5, 3.1, 3.2 (alternative advanced reading: MC chs. 9–11)
B: NS ch. 5

WATCH:

- Videos: Probability & Distributions

DO:

- **At home:** Read the tutorial `Probability & Distributions`
- **At home & in class:** Complete the tasks listed in `Assignments Day 3`

TERMS TO KNOW:

- Probability of an outcome
- Law of large numbers
- Expected value
- Normal distribution
- Parameters of a distribution
- Z-Score
- Percentiles
- Normal probability table
- 68-95-99.7 rule

SOFTWARE SKILLS:

- Plot a distribution with histograms, density plots, and normal probability ("QQ") plots
- Sample data from a distribution
- Document and organize code

Day 4 (February 16)

Inference & hypothesis testing.

READ:

T: OI sections 4.1 through 4.4 as needed after watching the videos
T: OI sections 4.5.4, 4.5.5
B: NS chs. 8 & 9
A: Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, 2(3):196–217
A: Goodman, S. (2008). A Dirty Dozen: Twelve P-Value Misconceptions. *Seminars in Hematology*, 45(3):135–140
A: Gill, J. (1999). The Insignificance of Null Hypothesis Significance Testing. *Political Research Quarterly*, 52(3):647–674

WATCH:

- Videos: Inference & hypothesis testing

DO:

- **At home:** Read the tutorial `Inference & hypothesis testing`
- **At home & in class:** Complete the tasks listed in `Assignments Day 4`
- **By 11:59pm on Saturday (February 13):** Via the link provided on Blackboard, sign up for an article you have chosen to replicate.
- **By 8pm on Wednesday (February 17):** Submit the structured summary of your replication article on Blackboard.
- **By 8pm on Saturday (February 20):** Submit your data exploration document on Blackboard.

TERMS TO KNOW:

- Population mean
- Sample mean
- Point estimate
- Sampling variation
- Population parameter
- Sampling distribution
- Standard error of an estimate
- Confidence interval
- Central limit theorem
- Margin of error
- Null hypothesis
- Alternative hypothesis
- Type 1 error
- Type 2 error
- p-value
- One-sided and two-sided hypothesis tests
- Unbiased estimate
- Statistical versus practical/substantive significance

SOFTWARE SKILLS:

- Sample from a distribution
- Calculate and plot confidence around a mean

Day 5 (February 23)

Associations between variables.

READ:

T: OI sections 5.1 through 5.5 and 6.1 through 6.4

A: Bennett, S. E. and Resnick, D. (1990). The Implications of Nonvoting for Democracy in the United States. *American Journal of Political Science*, 34(3):771–802

A: Gibson, J. L. (2002). Truth, Justice, and Reconciliation: Judging the Fairness of Amnesty in South Africa. *American Journal of Political Science*, 46(3):540–556

A: Licklider, R. (1995). The Consequences of Negotiated Settlements in Civil Wars, 1945-1993. *The American Political Science Review*, 89(3):pp. 681–690

A: Mutz, D. C. and Reeves, B. (2005). The New Videomalaise: Effects of Televised Incivility on Political Trust. *American Political Science Review*, 99(1):1–15

DO:

- **At home:** Read the tutorial `Associations between variables`
- **At home & in class:** Complete the tasks listed in `Assignments Day 5`
- **At home:** Submit your replication plan on Blackboard before noon
- **Exam:** Pick up midterm exam in class on Tuesday and submit to Blackboard by 8am on Monday (February 29)

TERMS TO KNOW:

- Difference of means test
- Central Limit Theorem for normal data
- Degrees of freedom
- t-distribution
- Two sample (Student's) t-test
- ANOVA and F-test
- Hypothesis test for a proportion
- chi-square distribution
- chi-square test statistic

SOFTWARE SKILLS:

- Conduct and present a difference-of-means test
- Conduct and present a test for differences in two proportions

Day 6 (March 1)

Bivariate regression.

READ:

T: AR ch. 5.1 & 6.1 (Plan some time on getting used to the notation and different style of the book)
T: OI ch. 7 (recommended as background)
B: NS ch. 11
A: Coppedge, M. (1998). The Dynamic Diversity of Latin American Party Systems. *Party Politics*, 4(4):547–568 (Figure 1)
A: Hurwitz, J. and Peffley, M. (1997). Public Perceptions of Race and Crime: The Role of Racial Stereotypes. *American Journal of Political Science*, 41(2):375–401
A: Segal, J. A. and Cover, A. D. (1989). Ideological Values and the Votes of U.S. Supreme Court Justices. *The American Political Science Review*, 83(2):557–565

WATCH:

- Video: Intro to Regression

DO:

- **At home:** Read the tutorial `Intro to Regression`
- **At home & in class:** Complete the tasks listed in `Assignments Day 6`

TERMS TO KNOW:

- Explanatory variable
- Response variable
- Residuals
- Correlation coefficient
- Intercept
- Slope coefficient
- Least-squares fit
- $y_i = \alpha + \beta X_i + \varepsilon_i$
- $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
- $\hat{y}$
- Least-squares coefficient
- Regression sum of squares
- Total sum of squares
- Residual sum of squares
- Linearity (A1)
- Constant error variance (A2)
- Normality of the errors (A3)
- Independence of observations (A4)
- No correlation between X and ε (A5)
- Variation in X (A6)
- Linear estimator (P1)
- Unbiased estimator (P2)
- Efficient estimator (P3)
- Gauss-Markov theorem
- Normal distribution of α and β
- Standard error of β

SOFTWARE SKILLS:

- Create a scatterplot of two variables with a line of best fit
- Calculate the correlation coefficient of two variables
- Estimate a linear regression model with one predictor
- Create a residual plot
- Summarize and present regression results in a well-designed document

Day 7 (March 8)

Review.
Data management.

READ:

T: Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59(10):1–23

DO:

- **At home:** Read the tutorial `Data management`
- **At home & in class:** Complete the tasks listed in `Assignments Day 7`

TERMS TO KNOW:

- Data set/matrix/frame
- Filtering/subsetting
- Transforming
- Aggregating
- Sorting
- Summarizing

SOFTWARE SKILLS:

- Import datasets from different sources into R
- Clean a dataset for data analysis
- Merge two datasets with a common identifier
- Collapse a dataset

Day 8 (March 22)

Multiple regression.

READ:

T: AR ch. 5.2 & 6.2

A: Franklin, M. N. (1999). Electoral Engineering and Cross-National Turnout Differences: What Role for Compulsory Voting? *British Journal of Political Science*, 29(1):205–216

A: Stasavage, D. (2005). Democracy and Education Spending in Africa. *American Journal of Political Science*, 49(2):343–358

A: Knack, S. (1995). Does "Motor Voter" Work? Evidence from State-Level Data. *Journal of Politics*, 57(3):796–811

A: Poe, S. C. and Tate, C. N. (1994). Repression of Human Rights to Personal Integrity in the 1980s: A Global Analysis. *American Political Science Review*, 88(4):853–872

DO:

- **At home:** Read the tutorial `Multiple Regression`
- **At home & in class:** Complete the tasks listed in `Assignments Day 8`
- **At home:** Begin outlining your project proposal.

TERMS TO KNOW:

- Partial coefficient
- "Holding a value constant"
- Standard error of the regression
- Squared multiple correlation R^2
- Standardized regression coefficient
- Omnibus F-test
- Collinearity
- Descriptive and causal interpretation of β

SOFTWARE SKILLS:

- Estimate a regression model with multiple predictors
- Present regression results graphically
- Calculate standardized regression coefficients

Day 9 (March 29)

Dealing with unusual & influential data.

READ:

T: AR ch. 11

A: Choi, S.-W. (2009). The Effect of Outliers on Regression Analysis: Regime Type and Foreign Direct Investment. *Quarterly Journal of Political Science*, 4(2):153–165

A: Li, Q. (2009). Outlier, Measurement, and the Democracy-FDI Controversy. *Quarterly Journal of Political Science*, 4(2):167–181

A: Wolak, J., Lowery, D., and Gray, V. (2001). California Dreaming: Replicating the ESA Model, Unusual Cases, and Comparative State Political Analysis. *State Politics & Policy Quarterly*, 1(3):255–272

A: Harden, J. J. and Desmarais, B. A. (2011). Linear Models with Outliers: Choosing between Conditional-Mean and Conditional-Median Methods. *State Politics & Policy Quarterly*, 11(4):371–389

DO:

- **At home:** Begin conducting your replication analysis.
- **At home:** Read the tutorial `Regression Diagnostics I`
- **At home & in class:** Complete the tasks listed in `Assignments Day 9`

TERMS TO KNOW:

- Outlier
- Leverage
- Influence
- Hat-values
- Studentized residuals
- Cook's D statistic

SOFTWARE SKILLS:

- Diagnose outliers
- Assess outliers, leverage, and influence in one combined plot
- Create added-variable plots

Day 10 (April 5)

Diagnosing and dealing violations of the OLS assumptions, including endogeneity.

READ:

T: AR chs. 12 & 13, but only sections 12.1, 12.2.1, 12.2.3, 12.3.1, 13.1, 13.2.1.
B: NS ch. 12
A: King, G. and Roberts, M. E. (2015). How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About it. *Political Analysis*, 23(2):159–179 Online at `http://j.mp/InK5jU`

DO:

- **At home:** Read the tutorial `Regression Diagnostics II`
- **At home & in class:** Complete the tasks listed in `Assignments Day 10`
- **At home:** Continue working on your replication analysis.
- **Exam:** Pick up midterm exam in class on Tuesday and submit to Blackboard by 8am on Monday (April 11).

TERMS TO KNOW:

- Non-normally distributed errors
- Nonconstant error variance (heteroskedasticity)
- White's robust standard error and its limitations
- Nonlinearity
- Collinearity
- Variance inflation factor
- Limitations of "fixes" to collinearity
- Endogeneity

SOFTWARE SKILLS:

- Conduct numerical and graphical checks for violations of the OLS assumptions
- Create component-plus-residual plots
- Transform variables
- Calculate variance inflation factors
- Calculate "robust" standard errors for a regression model and add them to a table

Day 11 (April 12)

Moderating relationships: interaction terms.

READ:

T: AR ch. 7

A: Brambor, T., Clark, W. R., and Golder, M. (2006). Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis*, 14(1):63–82

A: Hainmueller, J., Mummolo, J., and Xu, Y. (2016). How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice. *Working paper*, available at `ssrn.com`.

A: Braumoeller, B. F. (2004). Hypothesis Testing and Multiplicative Interaction Terms. *International Organization*, 58(4):807–820

A: Bearce, D. H. (2008). Not Complements, But Substitutes: Fixed Exchange Rate Commitments, Central Bank Independence, and External Currency Stability. *International Studies Quarterly*, 52(4):807–824

A: Sides, J. and Citrin, J. (2007). European Opinion About Immigration: The Role of Identities, Interests and Information. *British Journal of Political Science*, 37(3):477–504

A: Golder, S. N. (2006). Pre-Electoral Coalition Formation in Parliamentary Democracies. *British Journal of Political Science*, 36(2):193–212

DO:

- **At home:** Create your replication table and/or graph and write your replication memo.
- **At home:** Read the tutorial `Interaction terms`
- **At home & in class:** Complete the tasks listed in `Assignments Day 11`

TERMS TO KNOW:

- Dummy variable
- Dichotomous factor
- Polytomous factor
- Interaction term
- Constitutive terms
- Principle of marginality
- Centering variables
- Marginal effects

SOFTWARE SKILLS:

- Estimate linear regressions with interaction terms
- Present and interpret interaction terms numerically and graphically
- Create marginal effects plots for interactions

Day 12 (April 19)

Model fit. Model checking. Functional forms. Variable selection. And: how to use your (completed) replication analysis to inform your research.

READ:

T: AR sections 13.2.1, 13.2.2, p. 451, & 22 (pp. 607-610, 618-622, 626-630)

T: Sections 8.1–8.3 in Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York, NY [Blackboard]

B: NS ch. 12

A: Herndon, T., Ash, M., and Pollin, R. (2014). Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Cambridge Journal of Economics*, 38(2):257–279

A: Dafoe, A. (2014). Science Deserves Better: The Imperative to Share Complete Replication Files. *PS: Political Science & Politics*, 47(1):60–66

A: Ward, M. D. and Hoff, P. D. (2007). Persistent Patterns of International Commerce. *Journal of Peace Research*, 44(2):157–175

DO:

- **At home:** Submit your replication memo on Blackboard by noon.
- **At home & in class:** Complete the tasks listed in `Assignments Day 12`

TERMS TO KNOW:

- Adjusted R^2
- Problems with criteria for model selection
- Model validation
- Training and validation samples
- Fake-data simulation

SOFTWARE SKILLS:

- Fit quadratic and polynomial terms in a linear regression
- Describe nonlinear relationships graphically using locally weighted regression

Day 13 (April 26)

Generalized linear models.

READ:

T: AR chs. 14 & 15

A: Fox, J. (2003). Effect Displays in R for Generalised Linear Models. *Journal of Statistical Software*, 8(15):1–27

A: King, G., Tomz, M., and Wittenberg, J. (2000). Making the Most of Statistical Analyses: Improving Interpretation and Presentation. *American Journal of Political Science*, 44(2):347–361

A: Zorn, C. (2005). A Solution to Separation in Binary Response Models. *Political Analysis*, 13(2):157–170

A: Bell, M. S. and Miller, N. L. (2015). Questioning the Effect of Nuclear Weapons on Conflict. *Journal of Conflict Resolution*, Forthcoming

DO:

- **At home:** Finish your project proposal.
- **At home & in class:** Complete the tasks listed in `Assignments Day 13`

TERMS TO KNOW:

- Categorical response variables
- Dichotomous response variables
- Linear probability model
- Probability of the response (π)
- Linear predictor
- Transformations: logit and probit
- Unobserved/latent variable formulation
- $\frac{exp(\mathbf{X}\beta)}{1+exp(\mathbf{X}\beta)}$
- Maximum likelihood estimation
- Deviance
- Log-likelihood
- Separation/separability
- Ordered logit and probit
- Multinomial logit

SOFTWARE SKILLS:

- Estimate generalized linear models using maximum likelihood
- Present estimates using predicted outcomes (probabilities)
- Diagnose problems with generalized linear models

Day 14 (May 3)

Returning to where we started: observational data and causality.
Issues and problems in applied quantitative research.
Preview of advanced methods and pointers to further resources (based on student interests & project plans):

- Approaches to grouped data (repeated measures; nested data)
- Measuring latent concepts

READ:

A: King, G. (2006). Publication, Publication. *PS: Political Science and Politics*, 39(1):119–125
A: Schrodt, P. A. (2014). Seven Deadly Sins of Contemporary Quantitative Political Analysis. *Journal of Peace Research*, 51(2):287–300
A: Kastellec, J. P. and Leoni, E. L. (2007). Using Graphs Instead of Tables in Political Science. *Perspectives on Politics*, 5(4):755–771
A: Gigerenzer, G. (2004). Mindless Statistics. *The Journal of Socio-Economics*, 33(5):587–606
A: Gerber, A. S., Green, D. P., and Nickerson, D. (2001). Testing for Publication Bias in Political Science. *Political Analysis*, 9(4):385–392
A: Gelman, A. (2014). Experimental Reasoning in Social Science. In Teele, D. L., editor, *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences*, pages 185–195. Yale University Press, New Haven, CT
A: Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., Glennerster, R., Green, D. P., Humphreys, M., Imbens, G., Laitin, D., Madon, T., Nelson, L., Nosek, B. A., Petersen, M., Sedlmayr, R., Simmons, J. P., Simonsohn, U., and Van der Laan, M. (2014). Promoting Transparency in Social Science Research. *Science*, 343(6166):30–31
A: Levitt, S. D. and List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, 53(1):1–18
A: Keele, L. and Stevenson, R. (2014). The Perils of the All Cause Model. *Unpublished manuscript, Society for Political Methodology*

DO:

- Submit your project proposal **without your name on it** on Blackboard by 5pm on Friday, May 6.
- Look for a proposal to review in your email inbox on Saturday, May 7.
- Review the proposal I send you, read through the `guidelines for writing reviews`, and submit the review on Blackboard by 5pm on May 13.

TERMS TO KNOW:

- Publication bias
- File-drawer problem
- Garbage-can model
- Explanation and prediction
- Reanalysis
- "Cult of Stars and P-Values"
- Statistical and experimental controls
- Randomized trial
- Pre-registration
- Fixed effects
- Latent variables

SOFTWARE SKILLS:

- Transporting your new R skills to other software packages
- Handouts for working through our whole seminar in Stata

Acknowledgments

This course is influenced by and benefitting from syllabi and materials from (in no particular order) Sally Friedman, Dave Armstrong, Sara Mitchell, Jennifer Wolak, Ryan Bakker, Ann Karreth, Tim McDaniel, Andrew Gelman, Chris Zorn, and Catherine de Vries.

Disclaimer

Last updated: April 12, 2016. This syllabus is a general outline for the class and may be modified as the course progresses. Please check your email daily for notifications about any changes or class cancellations, and subscribe to the discussion boards on Blackboard for notifications about messages relevant to the class. The latest version of the syllabus is always posted at `http://www.jkarreth.net/rpos517.html`; please consult that link for up-to-date information on assignments and dates.