# SIS 806: Intro to Quantitative Methods in IR

Joe Young

Fall 2022

E-mail: jyoung@american.edu
Web: www.josephkyoung.com
Office Hours: Tues 11:00 a.m. - 2:00 p.m.
Class Hours: Tues 2:30 p.m. - 5:20 p.m.
Office: 208 SIS
Class Room: SIS 348, Zoom 966.177.2306

---

## Course Description

This course is part of an IR methods sequence that aims to provide you the tools necessary to be social researchers. The primary focus of the course is statistical description and analysis. In addition to examining the basics of data analysis, we primarily focus on multiple regression analysis. We explore the assumptions underlying the statistical model, what happens when the assumptions are violated, and how to deal with these problems. Specifically, we examine what happens to the standard regression model in the context of multicollinearity, nonlinearity, heteroscedasticity, non-additivity, measurement error and specification error. Beyond what is often called OLS regression, we discuss regression techniques related to limited dependent variables and times series/panel data.

While some courses on this topic are primarily devoted to mathematical exposition of these techniques, our course is for the applied researcher. We will occasionally delve into math, but the main goal is to help you develop the skills necessary to evaluate hypotheses in a large N context. Additionally, we will spend considerable time learning and utilizing statistical software. While there are a host of different software packages, we will use R as it is powerful, flexible, free and is increasingly used by quantitative social scientists.

## Course Objectives

I have three main objectives for this course. First, the course should help you gain the skills necessary to produce an original quantitative project. Second, you should gain the skills necessary to read and critique quantitative work that appears in peer-reviewed academic journals. Third, the course should serve as a basis for the acquisition of more advanced quantitative methods as your professional needs require.

## Learning Outcomes

- Present descriptive data in a table or figure
- Test hypotheses using multiple regression analysis
- Understand the strengths and limitations of regression analysis
- Accurately interpret regression coefficients

- Know the basic assumptions of the OLS regression model
- Understand the consequences of violating these assumptions
- Given the data generating process of a dependent variable, know the appropriate statistical model and method
- Produce an original research paper applying the techniques learned in the course
- Read and critique quantitative research
- Learn how to use statistical software (R)

## Course Format

Each week we discuss a series of readings. You are responsible for each reading. These readings will then serve as a basis for using the tools/concepts in statistical analysis. We then apply these tools in a directed lab setting using R. Following these labs, you will have an assignment or problem set that reinforces what we learned.

This course will require a great deal of reading and active participation on your part. You must be prepared each day and engage in the discussions, labs, workshops, and other activities, to be successful.

## Required Materials

### Books

### Required Books

Kennedy and Gujarati have been around for a long time. You can use an older or newer version depending on your budget. I'll be teaching from the 2008 version of Kennedy and 2003 version of Gujarati. The Grolemund and Wickham book is the best guide to starting to use R. It is available for free online https://r4ds.had.co.nz or you can buy from your favorite book seller. Other useful reference guides and R help is on the canvas site as well as online.

1. Grolemund, Garrett and Hadley Wickham (2018). *R for data science*.
2. Kennedy, Peter. 2008. *A Guide to Econometrics 6th ed.* Cambridge: MIT Press. ISBN 0-262-61183-X.
3. Gujarati, Damador. 2003. *Basic Econometrics 4th ed*. New York: McGraw-Hill. ISBN 0-07233542-4.

### Recommended Books

1. Berry, William and Mitchell Sanders. 2000. *Understanding Multivariate Regression*. Boulder: Westview Press. ISBN 0-8133-9971-8.
2. Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models* New York: Cambridge University Press. 0-521-68689-X
3. Landner, Jared. 2014. *R for Everyone* Addison-Wesley. 013454692X
4. Morgan, Stephen L., and Christopher Winship. 2014. *Counterfactuals and Causal Inference*. Cambridge University Press.

5. Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion.* Princeton University Press.

All of the books are available online. You can find them on Amazon.com or Addall.com at a reduced price. Since advanced statistics is not always easy, it is helpful to have multiple treatments of the same topic. I find that some books are more useful for understanding certain topics than others. The recommended books provide either more sophisticated or more simplistic treatment of most of the topics we cover. Buying them and using them in conjunction with the required texts would be helpful but not necessary. If you plan to pursue a dissertation that utilizes quantitative methods, then purchasing some/all of these books may be something to consider.

Many of the articles for the class will be on Canvas. You can also find most of the journal articles on JSTOR (www.jstor.org). I will also place readings on Canvas that are not available on JSTOR regardless, and these are noted in the syllabus (*).

**R**

- Free download of R (select Mac or Windows) https://www.r-project.org
- Free download of Rstudio (do not buy the pro version, not needed) https://rstudio.com/products/rstudio/

## Evaluation

The grading scale for this class is as follows:

| | | | |
|---|---|---|---|
| A | = 94+ | C | = 74-76 |
| A- | = 90-93 | C- | = 70-73 |
| B+ | = 87-89 | D+ | = 67-69 |
| B | = 84-86 | D | = 64-66 |
| B- | = 80-83 | D- | = 60-63 |
| C+ | = 77-79 | F | = 59 & below |

### Participation (20%):

I expect that you will be prepared to discuss all of the readings each week. I will assign a grade for your participation each week based on the following scale:

- A (outstanding) your comments were insightful and contributed to collective understanding of the material across the readings
- A- (strong) your comments were sometimes insightful and sometimes average but were not consistent across the readings
- B+ (good) your comments demonstrated that you understood the material but did not extend the discussion or offer new insights
- B (average) you participated but did not seem to fully grasp the material
- B- (poor) you spoke, but your comments were not germane to the material
- C (unacceptable) you didn't say anything…

### Problem Sets (10%):

Most weeks, I will assign you a problem set that implements concepts discussed and demonstrated in class. These problem sets are meant to reinforce the topic from the previous week and to prepare for the exams and the final paper.

### Midterm Exam (30%):

On October 10th, you will receive a take-home mid-term exam covering the theoretical foundations learned in the first half of the semester. This test will be open-note, and you may use any materials on the Internet. All questions will be essay style that require analysis and synthesis of course material. You may not consult your peers or anyone else. You will also be asked to do simple tasks in R. The exam will be due by 5:00 pm to Canvas on October 14th.

### Research Paper (30%):

Detailed instructions are on Canvas for this assignment. The research paper must be 20-25 pages (including references, tables, graphs, etc). It is due on the last day of formal classes, December 5th. The quality of the paper, however, is more important than the quantity.

The paper should have the format as outlined by Barry Weingast here:

https://web.stanford.edu/group/mcnollgast/cgi-bin/wordpress/wp-content/uploads/2013/10/CALTECH.RUL_..pdf

While this is not the only way to write a great paper in academia/social science, it is a great way.

These papers also makes useful suggestions for writing a graduate paper that is potentially publishable:

(a) Wolfinger, Raymond E. 1993. "Tips for Writing Papers." *PS: Political Science & Politics* Vol. 26, No. 1. 87-88.

(b) Van Cott, Donna Lee. 2005. "A Graduate Student's Guide to Publishing Scholarly Journal Articles," *PS: Political Science & Politics* , 38(4):741-743.

(c) Thunder, David. 2004. "Back to Basics: Twelve Rules for Writing a Publishable Article," *PS: Political Science & Politics* 37(3): 493-495.

(d) King, Gary. 2006. "Publication, Publication." *PS: Political Science & Politics* 39(1): 119-125.

This paper by King is also useful for quantitative researchers:

Gary King. 1995. "Replication, Replication." *PS: Political Science & Politics* 28(3): 444-452.

### Research Presentation (10%):

During the final exam period for class, we will have a poster session where the you will present your research in a conference-like situation. You will field questions from visitors, the instructor and other class members. More detailed instructions are on Canvas.

## Academic Integrity Code

Students should be aware of the contents of the AU Academic Integrity Code. I regard violations of this code seriously and will immediately refer the matter to the Dean, should such violations occur. Please see me if you have any questions about the academic violations described in the code either in general or as they apply to particular requirements of this course.

## Additional Support Services

If you experience difficulty in this course for any reason, please do not hesitate to contact me. In addition to the resources of the school, a wide variety of services are available to help you in your efforts to meet the requirements of the course.

- Academic Support Center (x3360, MGC 243): offers study skills workshops, individual instruction, tutor referrals, and services for students with learning disabilities.
- Writing support is available in the ASC Writing Lab or in the Writing Center, Battelle 228.
- Counseling Center (x3500, MGC 214): offers counseling and consultations regarding personal concerns, self-help information, and connections to off-campus mental health resources.
- Disability Support Services (x3315, MGC 206): offers technical and practical support and assistance with accommodations for students with physical, medical or psychological disabilities.

If you quality for accommodations due to a disability, please notify me in a timely manner with a letter from the ASC or DSS so that we can make arrangements to address your needs.

## Tentative Class Schedule

Read the following **before** Tuesday's class session. *Important*: class readings are subject to change given changing circumstances. Any adjustments to this schedule will be posted in a revised syllabus on Canvas.

## Intro and Review

### Week 01, 08/29 - 09/02: Introduction, Review of Statistical Concepts, How to ask good questions

- Introductions
- Discussion
- Gujarati, Appendix A
- Kennedy, Chapters 1 & 2
- Angrist, Joshua D and Pischke, J{"o}rn-Steffen. 2008. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press. Chapter 1.
- Grolemund, Garrett and Hadley Wickham pgs. 37-40, 77-79, 111-116 (Workflow)
- Zinnes, Dina A. 1980. "Three puzzles in search of a researcher: Presidential address." *International Studies Quarterly* 24(3): 315-342.
- Stats Lab #1–Meet R and R Studio, Coding Basics

## Statistical Relationships

### Week 02, 09/05 - 09/09: Concepts, Building blocks of theory, Operationalization

- Goertz, Gary. *Social science concepts: A user's guide*. Princeton University Press, 2012. Chapter 1 (p.1-24)

- Munck, Gerardo L., and Jay Verkuilen. 2002. "Conceptualizing and measuring democracy: Evaluating alternative indices." *Comparative political studies* 35(1): 5-34.
- Mahoney, James, and Gary Goertz. 2006. "A tale of two cultures: Contrasting quantitative and qualitative research." *Political analysis*: 227-249.
- Peruse Vdem project: https://www.v-dem.net/en/
- **BRING SOME CONCEPTS**–Bring a list of concepts that are of interest to you. For example, you may be interested in the relationship between democracy and terrorism or party ideology and economic growth.
- Grolemund, Garrett and Hadley Wickham skim pgs. 119 - 168, Read 43-73
- Stats Lab #2 - R & Data Manipulation (dplyr)
- *DPLYR Homework*

## Regression Analysis

### Week 03, 09/12 - 09/16: Data Analysis and the Regression Model

- Gujarati, Chapters 2-5, pgs. 164-175 (Regression basics)
- *Berry & Sanders, Chapter 2
- Kennedy, Chapter 4 (Interval Estimation and Hypothesis Testing)
- Grolemund, Garrett and Hadley Wickham pgs. 3-30, 81-109
- **BRING SOME MEASURES**–Bring a list of measures that map onto the concepts of interest to you. For example, you may be interested in populism and pandemic response. Bring measures of populism and gov. responses to pandemics.
- Stats Lab #3 - Descriptive stats and visualization

### Week 04, 09/19 - 09/23: Multiple Regression: Estimation

- Gujarati, pp. 202-23, 229-33 (Multiple Regression, Problem of Estimation, Partial Correlation Coefficients)
- Kennedy, Chapter 3 (Classic Linear Regression Model)
- Machain, Carla Martinez, and T. Clifton Morgan. 2013. "The effect of US troop deployment on host states's foreign policy." *Armed forces & society* 39(1): 102-123.
- Linear Regression in R (Read up to "Predicting Linear Models")

https://www.machinelearningplus.com/machine-learning/complete-introduction-linear-regression-r/

Recommended videos to watch:

- Intro to Regression https://www.youtube.com/watch?v=zPG4NjIkCjc
- Calculating Least Squares https://www.youtube.com/watch?v=JvS2triCgOY
- Stats Lab #4 - Multiple Regression Analysis

# Violations of Regression Assumptions

### Week 05, 09/26 - 09/30: Multiple Regression: Inference and Dummy Variables

- Gujarati, Chapter 8 (Multiple Regression: The Problem of Inference)
- Angrist, Joshua D and Pischke, Jorn-Steffen. pgs. 28-67. (Regression Fundamentals)

Dummy Variables

- Gujarati, pp.297-306 (Dummy Variables)
- Kennedy Chapter 14 (Dummy Variables)
- Davenport, Christian, and David A. Armstrong. "Democracy and the violation of human rights: A statistical analysis from 1976 to 1996." *American Journal of Political Science* 48.3 (2004): 538-554.
- Stats Lab #5
  - Dummy variables/factor variables/categorical variables
  - Regression diagnostics

### Week 06, 10/03 - 10/07: Heteroskedasticity and Multicollinearity

- Gujarati, Chapter 11 (Heteroskedasticity)
- Gujarati, Chapter 10 (Multicollinearity)
- Kennedy, Chapter 8 (Violating Assumption Three: Nonspherical disturbances)
- Kennedy, Chapter 11 (Violating Assumption Five: Heteroskedasticity)
- Downs, George and David Rocke. 1979. "Interpreting Heteroskedasticity.'' *American Journal of Political Science* 23: 816-828.
- Lemieux, Peter H. 1978. "A Note on the Detection of Multicollinearity.'' *American Journal of Political Science* 22 (1): 183-186.
- Stats Lab #6
  - Heteroskedasticity and multicolliearity

### Week 07, 10/10 - 10/14: Specification

- Gujarati, pp. 215-217, 506-524
- Kennedy, Chapter 5 (Specification)
- Achen, Christopher. 2002. "Toward a New Political Methodology: Microfoundations and ART.'' *Annual Review of Political Science* 5: 423-450.
- Clarke, Kevin. 2005. "The Phantom Menace: Omitted Variable Bias in Econometric Research.'' *Conflict Management & Peace Science* 22(4): 341-352.
- Ray, James Lee. 2005. "Constructing Mutlivariate Analyses (of Dangerous Dyads)," *Conflict Management & Peace Science* 22(4): 277-292.

- Hafner-Burton, Emilie M. 2005. "Right or robust? The sensitive nature of repression to globalization." *Journal of Peace Research* 42(6): 679-698.
- Stats Lab #7–Hypothesis Testing

### Week 08, 10/17 - 10/21: Interactions

- Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. ÒUnderstanding Interaction Models: Improving Empirical Analyses. *Political Analysis* 14(1):63-82.
- Friedrich, Robert J. 1982. In Defense of Multiplicative Terms in Multiple Regression Equations. *American Journal of Political Science* 26 (November): 797-833.
- Braumoeller, Bear. 2004. "Hypothesis Testing and Multiplicative Interaction Terms.'' *International Organization* 58(4): 807-820.
- https://cran.r-project.org/web/packages/interplot/vignettes/interplot-vignette.html
- Stats Lab #7
  - Interactions
  - Guided exercise

### Week 09, 10/24 - 10/28: Autocorrelation and Measurement Error

- Gujarati, Chapter 12 (Autocorrelation), pp. 524-528
- *Carmines and Zeller, pp. 1-16, 29-32, 37-51
- Kennedy, Chapter 9 (Violating Assumption Four: Autocorrelation and Measurement Error)
- Simmons, Beth A., and Zachary Elkins. "The globalization of liberalization: Policy diffusion in the international political economy." *American political science review* 98.1 (2004): 171-189.
- Linzer, Drew A., and Jeffrey K. Staton. "A global measure of judicial independence, 1948-2012." *Journal of Law and Courts* 3.2 (2015): 223-256.
- Project/Paper Workshop
  - **Bring research question, outline, and data** for paper

## Panel Data, Limited Dependent Variables, Forecasting, & Missing Data

### Week 10, 10/31 - 11/04: Panel Data

- Gujarati, Chapter 16 (Panel Data Regression Models)
- Kennedy, Chapter 17 (Panel Data)
- Beck, Nathaniel and Jonathan Katz. 1995. "What To Do (and Not To Do) with Time- Series Cross-Section Data.'' *American Political Science Review* 89: 634-647.
- Angrist, Joshua D and Pischke, Jorn-Steffen, Chapter 5 (Parallel Worlds).
- Achen, Christopher H. "Why lagged dependent variables can suppress the explanatory power of other independent variables." Ann Arbor 1001.2000 (2000): 48106-1248.

- Kingstone, Peter, and Joseph Young. 2009. "Partisanship and policy choice: What's left for the left in Latin America?." *Political Research Quarterly* 62(1): 29-41.
- Stats Lab #8–Panel Data
  - Estimating time-series cross-sectional models in R
  - https://rstudio-pubs-static.s3.amazonaws.com/372492_3e05f38dd3f248e89cdedd317d603b9a.html

### Week 11, 11/07 - 11/11: Limited Dependent Variables

- Gujarati, Chapter 15 (Qualitative Response Regression Models)
- Kennedy, Chapters 15-16 (Qualitative and Limited Dependent Variables)
- Long and Freese, Limited Dependent Variables, Chapter 2.
- Thomas, Jakana. "Rewarding bad behavior: How governments respond to terrorism in civil war." *American Journal of Political Science* 58.4 (2014): 804-818.
- Stats Lab #9–Logit Models
  - Read Logistic Regression Diagnostics
  - Estimating logit/probit models in R

### Week 12, 11/14 - 11/18: Missing Data

- Schafer, Joseph L., and Maren K. Olsen. "Multiple imputation for multivariate missing-data problems: A data analyst's perspective.'' *Multivariate behavioral research* 33.4 (1998): 545-571.
- Horton, Nicholas J., and Ken P. Kleinman. "Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models.'' *The American Statistician* 61.1 (2007): 79-90.
- Rubin, Donald B. "Multiple imputation after 18+ years.'' *Journal of the American statistical Association* 91.434 (1996): 473-489.
- King, Gary, et al. "Analyzing incomplete political science data: An alternative algorithm for multiple imputation." *American Political Science Review* 95.1 (2001): 49-69.
- Honaker, James, and Gary King. "What to do about missing values in time-series cross-section data." *American Journal of Political Science* 54.2 (2010): 561-581. (skim)
- Stats Lab #10
  - Multiple Imputation

### Week 13, 11/21 - 11/25: Give Thanks! No class.

### Week 14, 11/28 - 12/02: Forecasting and Cross-Validation

- Ward, Michael, Brian Greenhill, and Kristin Bakke. 2010. "The Perils of Policy by P-Value: Predicting Civil Conflicts.'' *Journal of Peace Research* 47(4): 363-375.
- Hill Jr, Daniel W., and Zachary M. Jones. "An empirical evaluation of explanations for state repression." *American Political Science Review* (2014): 661-687.

- Colaresi, Michael and Zuhaib Mahmood. 2017. "Do the Robot: Lessons from Machine Learning to Improve Conflict Forecasting'' *Journal of Peace Research* 54(2).
- Maliniak, Daniel, Ryan Powers, and Barbara F. Walter. "The gender citation gap in international relations." *International Organization* 67.4 (2013): 889-922.
- Stats Lab #11–Cross-validation, Final Models

**Week 15, 12/05 - 12/09: Experiments/Causal Inference, Final Paper Due**

- Druckman, James N., Donald P. Green, James H. Kuklinski, and Arthur Lupia. "An introduction to core concepts." The Cambridge handbook of experimental political science (2011).
- Getmansky, Anna, and Thomas Zeitzoff. "Terrorism and voting: The effect of rocket threat on voting in Israeli elections." *American Political Science Review* 108.3 (2014): 588-604.
- McDermott, Rose. "The ten commandments of experiments." *PS: Political Science & Politics* 46.3 (2013): 605-610.
- Kearns, Erin M., and Joseph K. Young. ""If Torture Is Wrong, What About 24?" Torture and the Hollywood Effect." *Crime & Delinquency* 64.12 (2018): 1568-1589.

**Week 16, 12/12 - 12/16: Poster Presentations**