

# POS 7113: Quantitative Research Methodology II

Tulane University  
Center for Inter-American Policy & Research  
Richardson Building,  
Second Floor, Room M  
New Orleans, LA 70112

Last updated: February 26, 2018.

Download last version [here](#).

Not intended as a definitive version

**Professor:** Hector Bahamonde

e: [hbahamonde@tulane.edu](mailto:hbahamonde@tulane.edu)

w: [www.HectorBahamonde.com](http://www.HectorBahamonde.com)

**Class meetings:** DAY TIME.

**Location:** TBA.

**Office Hours:** Make an appointment [here](#).

## *Overview and Objectives*

*What is the influence of income on the probability of voting? Is there any systematic relationship between democracy and trade? Is it true that correlation does not imply causation?* To consider these kind of statements in a rigorous way we will learn a technique called *ordinary least regression*, the workhorse of applied quantitative methodology in social sciences. At the end of the course you will become an educated consumer and producer of quantitative analysis.

This [graduate-level course](#) offers an introduction to linear models in political science. This course will hopefully prepare you for the things you will encounter when you (attempt to) publish quantitative work with linear models. This course will provide you with a systematic approach to assessing, fixing and presenting your linear model results. Though we focus almost exclusively on linear models, some nonlinear models will also be introduced.

During the semester, we will focus on a number of assumptions, applications, violations to those assumptions, and their corresponding solutions. It is the intention of this course to be both theoretically relevant and practically oriented at the same time. As you will quickly realize, you will “learn by doing.” Besides learning quantitative methodology, you will also learn a number of computer skills, mainly **R** and **L<sup>A</sup>T<sub>E</sub>X**. Therefore, if you are not already with those tools, try to be patient with your self.

I hope this course catches your attention, in the expectation that you continue taking more methods and/or computing courses. Most of all, I hope you see how applied statistics are both an art and a science. **Welcome!**

## *Prerequisites*

There are no formal prerequisites for this course. Only good predisposition to learn and spend quality-time doing the readings and problem sets. That said, basic math, calculus, probability and linear algebra is assumed at the level covered in any “Math Camp,” or equivalently at the level of *Essential Mathematics for Political and Social Research* (Gill, 2006).

## Software

It is not enough to just *read* about statistics. The best way to learn applied statistics is by *doing*. For this reason, in this course you will sit in the driving seat and actually estimate (a lot of) statistical models using **R**. **R** is free and compatible with essentially all kinds of machines. One of **R**'s main virtues from the grad-student point of view is that the base package and all of the add-ons (called “packages” in **R**) are free. You can download the base package of **R** from the *Comprehensive R Archive Network (CRAN)* [website](#). You can find a virtually endless set of resources for **R** on the Internet, including this [Getting Started With R](#) page. If you are completely new to **R**, you should complete this online short course, [Try R](#). Finally, I suggest a **R** interphase called RStudio. [RStudio](#) makes **R** easier to use. It includes a code editor, debugging and visualization tools.

Quantitative methodology is not only about estimating the “best” model (if such a thing even exists) but also about presenting your results in an effective way. For this reason, all problem sets and final presentations should be carried out in  $\text{\LaTeX}$ .  $\text{\LaTeX}$  is a high-quality typesetting system; it includes features designed for the production of technical and scientific documentation.  $\text{\LaTeX}$  is the *de facto* standard for the communication and publication of scientific documents. Download the last version [here](#).

You are free to use other softwares such as STATA, MathLab, SAS, etc. However, I might provide assistance only for the former. That said, all in-class examples are going to be taught in **R**, which is by far the most common statistical tool nowadays.

## Course Texts

- Fox, John. (2015) *Applied Regression Analysis and Generalized Linear Models*. 3rd ed. Thousand Oaks, CA: Sage Publications, Inc.
- Fox, John and Sanford Weisberg. (2011) *An R Companion to Applied Regression*. 2nd ed. Thousand Oaks, CA: Sage Publications, Inc.

## Course Learning Objectives

Upon successful completion of this course, you will be able to:

- Acquire an understanding of the main tools related to linear regression.
- Use the linear method up to the point where you feel comfortable doing analysis in your own research.
- Consume *critically* methodological-oriented literature.

## Classroom Etiquette

- Please, do not eat during class. Beverages are fine.
- No computers, phones, or any other electronic devices may be used in lecture for any reason—no exceptions. Any such devices on your person must be off (e.g., not merely on silent) and put completely away. Those who do not respect this requirement will be asked to leave class.
- Attendance is mandatory (and part of your participation grade). If you missed a class, please get the notes from another student. I do not offer make up sessions for students who are absent.
- Please, follow the *Email Etiquette* I have [posted](#) in my website.

## Requirements and Evaluations

### 1. Problem Sets: 15 %.

I expect you to keep up with the readings over the course of the semester. I employ an interactive lecture style, and you will need to have done the readings in order to participate. There will be a number of pop quizzes during the semester, particularly during the first part of the semester. Quizzes will be short (3-5 minutes), completed at any point of the class, and designed to make sure everyone is keeping up with the readings and lecture. There will be no make-up quizzes. If you are absent (or late) from class that day, you will get a 0 on that quiz.

When reading the class materials, you should locate the main argument, strengths, weaknesses, and other issues that are of concern. As you read through the material, think about the following questions: *What is the cause and what is the effect? What makes the theory 'move,' is it individuals? institutions? (ir)rationality? Does/do the author/s have a strong research design/methodology to support the paper's argument?*

On average, students are expected to put in approximately 10-12 hours of work per week for a four-credit class, as per U.S. Department of Education guidelines. Since you will be spending 2.5 hours in the classroom, this means you should be working about 7.5-9.5 hours per week for this course **outside** of the classroom. If you find that you are spending more than 12 hours per week on the class, please see me to discuss strategies to read more efficiently.

### 2. Student Project, Date: 50 %.

An empirical paper that applies methods learned in this class to a research question of their choice. The paper should be 10-15 pages in length and focus on the research question, data, empirical strategy, results, and conclusions. Literature reviews, background, lengthy motivations, etc. should be omitted or may be included as an appendix.

You should be able to convince me that your empirical strategy is solid. For that, you should provide graphs, tables, robustness checks, among other tools.

The paper should be very clear about the kind of model used, and it should use proper mathematical notation. The paper should be written in  $\text{\LaTeX}$ . Documents in *Word* will not be accepted.

You also need to submit a copy of your **R** code. I should be able to replicate your exact results. That is, I should be able to hit "run" in *RStudio* and obtain the exact same output. Failure to replicate your results will discount points out of grade.

Students are free to choose any topic they want. As long as it has a strong data-analytical component, it's fine by me. I strongly advice you to get the data in advance. As you will quickly realize, the real world does not work as when you do homeworks, e.g. the data you will originally find, will be messy, with lots of inconsistencies, variables will not be labeled, among other programs. My advice: *start early*.

### 3. Problem Sets, Date: 30%.

This is a methodological course, developing skills in understanding and applying statistical methods. You can only learn statistics by doing statistics and therefore the homework for this course is extensive, including weekly homework assignments. The assignments consist of analytical problems, computer simulations, and data analysis.

In this section you will learn how to recode a variable, merge datasets, summarize variables, plot a distribution, among other more theoretical issues. And while we will not focus on mathematical proofs, you will spend time thinking about and working on the mechanics of linear regression.

4. **Final Exam:** 20 %.

There will be a single, take-home final exam in the course that will require that you apply the knowledge gained in the course to particular methodological problem. The format of the exam will be discussed by the end of the semester. However, you can expect the test to be an exercise in guided replication and primarily involves data analysis and interpretation. Exam questions will be drawn both from the readings and lectures. The final exam is set by the registrar. Hence, both place and time are TBA. There will not be exceptions. We will also schedule an in-class review session for **date**.

### *Grading*

This course will be grade according to the following scale: A:  $\geq 93$ , A-: 90-92, B+: 87-89, B: 83-86, B-: 80-82, C+: 77-79, C: 73-76, C-: 70-72, D+: 67-69, D: 63-66, D-: 60-62, and F:  $\leq 59$ .

### *Disputing Grades*

I am happy to go over any exam or paper with you. Request for re-grading, though, must be done in writing. Please refer to my [re-grading policy](#).

### *Academic Integrity*

In accordance with Tulane University policy on Academic Integrity, you are expected to fully comply with the school's [policies](#).

### *Students with Disabilities*

Students with disabilities who require accommodation should check with the [Goldman Center for Student Accessibility](#).

### *Absence from Exams*

There will be no make-up exams unless you have a *documented medical* emergency. If at all possible, I need to be notified before the exam of your inability to take it. Absence from an exam because of travel plans will not be excused. Make travel plans accordingly.

### *Office Hours*

I have an open-doors policy, feel free to stop by my office at any time. However, you might want to minimize the risks that I am not there, or can't meet you that day. I advise you then to [schedule time with me](#) using my automatic scheduler. I think fixed office hours do not work because ... well, they are *fixed*. I prefer flexibility. Hence, you can see me any day/time that's available during the week. Do not send me a reminder as I will receive an alert: If the time spot is available, I am happy to see you there.

## *Schedule*

Each entry represents a single topic. Readings are designated either as suggested (★) or supplemental (-). This should serve as a nice reference to which you can return if the intricacies of a particular topic have faded from your memory.

### 1. Preliminary Material

- ★ Fox (2015), Chapters 1 & 2.
- ★ Fox and Weisberg (2011), Chapters 1 & 2.

### 2. OLS II: Effective Presentation

- ☐ Factors and contrasts; quasi-variances and graphical displays
- ☐ Interactions and effect displays
- ☐ Standardization and relative importance
- ★ Armstrong II (2013)
- ★ Berry, Golder and Milton (2012)
- ★ Silber, Rosenbaum and Ross (1995)
- Brambor, Clark and Golder (2006)
- Braumoeller (2004)
- Firth and Menzies (2004)
- Kam and Franzese (2007)

### 3. In-class Lab I

- ☐ Factors and contrasts.
- ☐ Interactions.
- ☐ Relative Importance.

### 4. Linearity: Diagnostics, Transformations and Polynomials

- ☐ Diagnosing linearity through residual plots.
- ☐ Fixing non-linearity with data transformations and polynomials.
- ☐ Linearity and ordinal variables.
- ★ Fox (2015) Chapters 4 & 12 (Sections 12.3-12.5)
- ★ Fox and Weisberg (2011) Chapter 3
- ★ Jacoby (1999)
- Box and Tidwell (1962)
- Breiman and Friedman (1985a,b), Pregibon and Vardi (1985), Buja and Kass (1985), Fowlkes and Kettering (1985)

### 5. Re-sampling Techniques and Regression

- ☐ Bootstrapping and Jackknifing.
- ☐ Cross-validation.

- ★ Fox (2015) Chapter 21
- Stone (1974)
- Efron and Tibshirani (1993)
- Davison and Hinkley (1997)
- Ronchetti, Field and Blanchard (1997)

## 6. Model Selection

- ☐ Theoretical issues in model searching and post-data model construction.
- ☐ Model selection criteria and multi-model inference.
- ☐ Subset selection models.

- ★ Fox (2015) Chapter 22
- ★ Leamer (1983)
- ★ Burnham and Anderson (2004)
- ★ Leamer and Leonard (1983)
- ★ Box (1976), Box and Hunter (1962)
- Freedman (1991b,a), Berk (1991), Blalock (1991), Mason (1991)
- Miller (2002), Breiman (1992), Breiman and Spector (1992)

## 7. Non-Linearity, Smoothing and Splines

- ☐ Nonparametric Smoothing - Lowess.
- ☐ Inference for regression smoothers.
- ☐ Regression Splines.
- ☐ Generalized Additive Models.

- ★ Fox (2015) Chapters 17 & 18
- ★ James et al. (2013) Chapter 7
- Keele (2008) Chapters 2-6

## 8. Flexible Models: Tree-based Regression, Multivariate Adaptive Regression Splines

- ☐ Fundamentals of flexible models.
- ☐ Automatic variable selection.
- ☐ Inference and effects in statistical learning models.
- ☐ When (and when not) to use these kinds of models.

- ★ Montgomery and Olivella (forthcoming)
- ★ James et al. (2013) Chapter 7
- ★ Berk (2016) Section 3.14

## 9. In-Class Lab II

- ☐ Non-linearity transformations.
- ☐ Polynomials.

- ☐ Smoothers and splines.
- ☐ Trees and other flexible methods.

#### 10. Regression Discontinuity Designs

- ★ Cattaneo, Idrobo and Titiunik (2017).
- ★ Calonico, Cattaneo and Titiunik (2015) .
- ★ Keele (2015).
- ★ Sekhon and Titiunik (2016).

#### 11. Finite Mixture Models

- ★ Imai and Tingley (Forthcoming).
- ★ Grun and Leisch (2008).
- ★ Grun and Leisch (2007).

#### 12. Missing Data and Multiple Imputation

- ☐ Whats the problem with missing data?
- ☐ When can we fix it?
- ☐ How do we impute the data and use those imputations?
- ★ Fox (2015) Chapter 20
- ★ van Buuren and Groothuis-Oudshoorn (2011)
- ★ Honaker and King (2010)
- ★ Cranmer and Gill (2013)
- ★ Akande, Li and Reiter (forthcoming)
- ★ Xia and Yang (2016)
- ★ Resseguier, Giorgi and Paoletti (2011)
- Schafer (1997)
- Rubin (1987)

#### 13. In-class Lab IV

- ☐ RDD.
- ☐ Mixture Models.
- ☐ Missing Data and Multiple Imputation.

#### 14. Outliers and Influential Data: Diagnostics

- ☐ Outliers, leverage and influential data.
- ☐ Hat values, standardized residuals, Cook's D.
- ☐ M-estimation (and extension) and iterative re-weighted least squares.
- ☐ Diagnostics for outliers revisited.
- ★ Fox (2015) Chapter 11.
- ★ Fox and Weisberg (2011) Chapter 6 (pp 101-201).

- ★ Andersen (2008).
- ★ Fox (2015) Chapter 19.
- Cantoni and Ronchetti (2001).
- Rousseeuw and Leroy (1987).
- Jasso (1985, 1996), Kahn and Udry (1986).

15. Non-constant error variance and collinearity: Diagnostics and Fixes

- ☐ Residual plots.
- ☐ ML transformations of Y.
- ☐ Weighted least squares.
- ☐ Heteroskedastic linear regression.
- ☐ Robust standard errors.
- ★ Fox (2015) Chapters 12 & 13.
- ★ Fox and Weisberg (2011) Chapters 3 & 6.
- ★ Long and Ervin (2000).
- ★ King and Roberts (2015).
- Harvey (1976).
- Cribari-Neto (2004), Cribari-Neto, Souza and Vasconcellos (2007), Cribari-Neto and da Silva (2011).

16. Critiques of the Linear Regression Model

- ☐ How important are the assumptions behind OLS Regression?
- ☐ How should we appropriately use regression models?
- ☐ The importance of sampling to inference.
- ★ Berk (2004).