

EPPS 6316
Applied Regression
SPRING 2019

Instructor: Vito D’Orazio

Time and Location: Tuesday 7:00-9:45pm, GR 3.402B

Contact: dorazio@utdallas.edu 972-883-6212

Office Hours: Wednesday, 9:00-11:00am, GR 3.108D

Teaching Assistant: Austin Kingsolver

Contact: akingsolver@utdallas.edu

Office Hours: Tuesday 4:00-6:00, GR 3.314

Course Goals and Overview

This course provides a survey of the regression models estimated using Ordinary Least Squares (OLS), with an emphasis on using regression models to test hypotheses about social phenomena. This application-focused course presents examples drawn from political science, public policy, economics, and sociology, introduces the basic concepts and interpretation of regression models, and basic methods of inference. We will do a lot of analysis with statistical software, both in class and for homework assignments.

There are several learning objectives for this course. The first is to understand what OLS is, when to use it, and what are the assumptions that underly it. The second is to be able to identify when one of those assumptions is violated and to correct for the violation, as is appropriate. The third learning objective is to interpret the estimates of OLS in both a mathematical and a substantive sense. Finally, there is also the objective of learning to work with real data using either R or Stata. These learning objectives will be assessed through class discussion, homework assignments, quizzes, and the final paper.

The prerequisite for this course is EPPS 6313 or its equivalent. If you have not taken the prerequisite, please contact Professor D’Orazio.

Course Requirements

You are expected to read all assigned readings in detail and to be prepared to discuss the readings in class. Since this is a methods course, much of what is read contains math. Do not skip over the math—take the time to understand what the author is saying, even if you do not fully grasp all the equations and symbols.

Final Grade Composition

Attendance and participation	10%
Homework assignments	30%
Final paper	35%
Quizzes	25%

Attendance and participation (10%)

Show up on time, be attentive, and participate to get full credit. Although this is a methods class, we will have discussion and I expect all students to participate.

Homework assignments (30%)

Homework assignments will be made available on eLearning. For each assignment, you are to submit two files: (1) the report and (2) the replication code. All reports must be in pdf format. The replication code must be a plain text file (.do, .R, or .txt) meaning that it can be opened and read with a plain text editor. If the replication code does not replicate your results, there will be a one point deduction on the assignment.

You are to submit each assignment through eLearning prior to class on the day the assignment is due. Also, bring a hardcopy of the pdf report to class on the day the assignment is due.

Final project (35%)

For the final project, you are to write a conference-quality research paper with an emphasis on data exploration and analysis. The paper is to include the following sections:

- Abstract
- Introduction
- Literature review and theory, leading to explicitly stated hypotheses
- Research design, with a discussion of the method used, the unit of observation, variables, scope, and data sources
- Empirical results, with an emphasis on displaying the results and assessing the hypothesis
- Conclusion

Since this is a methods course, most of your effort should be spent on the research design and empirical results portion of the paper. The literature review and theory must be present, but it may be minimal.

You may conduct (1) an original analysis, or (2) a replication and extension of an existing research paper. If you choose the replication and extension, the Harvard Dataverse is an excellent resource for finding replication code and data (<https://dataverse.harvard.edu/>).

Your extension should be something new: it could be the testing of a new variable that makes theoretical sense and for which you've derived testable hypotheses; running diagnostics and various sensitivity analyses; assessing the hypotheses using alternate measures of key concepts to see if the results still hold; etc.

The final project is broken out into three separate assignments. Please submit each of the following through eLearning.

1. Proposal (due March 12)

- The proposal should be two pages at most. Describe the paper you want to write: the hypotheses, the research design, the data, and the models.
- If conducting a replication, also submit a pdf copy of the original research article.
- Before writing the proposal, you should talk about your topic and strategy with me. I may ask you to revise your proposal if it is not appropriate.

2. Data (due April 16)

- Your data should be fully prepared and ready for analysis.
- The data paper describes your data, including the unit of observation, the variables, any coding decisions, the data sources, etc. Also, include exploratory or descriptive visualizations.

3. Final paper (due 11:59pm on May 6)

- The final paper should include 3 files: a pdf report, the data, and the replication code for reproducing all graphics and results in your paper.

Quizzes (25%)

There will be three in-class quizzes worth 5, 10, and 10 points, respectively. These quizzes will be closed book and will not require a computer. The first is a rather straight-forward vocabulary quiz on the third day of class. The second and third quizzes are in the style of the MQE (Parts A, B and D). You will be asked conceptual questions (e.g., what is an outlier, how to detect it, and options for handling it), applied questions (e.g., here are some regression results, use them to assess these hypotheses), and computational questions (here is some data, use it to assess a hypothesis).

Grade Scale

A	≥ 93%	B	= 83-86%	C	73-76%
A-	= 90-92%	B-	= 80-82%	F	≤ 72%
B+	= 87-89%	C+	= 77-79%		

See <http://catalog.utdallas.edu/2016/graduate/policies> for additional information about university grading policies.

UT Dallas Syllabus Policies and Procedures

The information contained in the following link constitutes the University's policies and procedures segment of the course syllabus: <http://go.utdallas.edu/syllabus-policies>.

Required Texts

1. Treiman, D. J. (2009). *Quantitative data analysis: Doing social research to test ideas*. John Wiley & Sons. ISBN: 978-0470380031

Optional Texts

These texts are optional. Note that they may be available electronically through the UTD library. I have found each of these texts useful in different ways. Chatterjee and Hadi is very accessible with many simple examples. Fox is more statistical, and feels more like a math text, but still understandable and much more complete. Angrist and Pischke is great for understanding causal inference.

1. Chatterjee, S. and Hadi, A. S. (2015). *Regression analysis by example*. John Wiley & Sons, fifth edition
2. Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton
3. Fox, J. (2016). *Applied Regression Analysis and Generalized Linear Models*. Sage, 3rd edition

Course Outline

Day 1: Introductions, Syllabus, and Levels of Measurement January 15

- Stevens, S. (1946). On the theory of scales of measurement. *Science*, 103(2684):677–680

Day 2: Experimental and Observational Data January 22

- Angrist and Pischke (AP), Chapters 1 and 2
- Chatterjee and Hadi (CH), Chapter 1

Day 3: Tables January 29

- Quiz 1 on notation and vocabulary
- Treiman, Chapters 1, 2, 3

- Day 4: Data and Visualizations** **February 5**
- Homework 1 is due
 - Treiman, Chapter 4
- Day 5: Simple Regression** **February 12**
- Treiman Chapter 5
 - CH Chapter 2: Simple Linear Regression
- Day 6: Multiple Regression I** **February 19**
- Homework 2 is due
 - Treiman Chapter 6
 - CH Chapter 3: Multiple Linear Regression
- Day 7: Multiple Regression II** **February 26**
- Homework 3 is due
 - Treiman Chapter 7
 - CH Chapter 4: Regression Diagnostics: Detection of Model Violations
- Day 8: Quiz, Visualizations and Presenting Regression** **March 5**
- Quiz 2 on basics of OLS
- Day 9: Dummy Variables and Interaction Terms** **March 12**
- Proposals for final project due
 - CH Chapter 5: Qualitative Variables as Predictors
 - Brambor, T., Clark, W. R., and Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis*, 14(1):63–82
 - Berry, W. D., Golder, M., and Milton, D. (2012). Improving tests of theories positing interaction. *The Journal of Politics*, 74(3):653–671
- Day 10: Multicollinearity** **March 26**
- CH Chapter 9: Analysis of Collinear Data

Day 11: Multiple Imputation and Regression Diagnostics **April 2**

- Homework 4 is due
- Treiman chapter 8, 10

Day 12: Heteroscedasticity and Autocorrelation **April 9**

- Homework 5 is due
- CH Chapter 6: Transformation of Variables
- CH Chapter 8: The Problem of Correlated Errors

Day 13: Logistic Regression **April 16**

- Data paper due
- Treiman chapter 13
- CH Chapter 12: Logistic Regression

Day 14: Random and Fixed Effects **April 23**

- Homework 6 is due
- Treiman chapter 15

Day 15: Quiz **April 30**

- Quiz 3. The final quiz will be given in the style of the MQE. You will have the entire class to complete the test, but will likely not need that much time.

***Final papers are due by 11:59pm on Monday, May 6.**