

**EPPS 6316**  
**Applied Regression**  
**SPRING 2021**

**Instructor:** Vito D’Orazio

**Time and Location:** Tuesday 7:00-9:45pm

**Modality:** Remote, Synchronous

**Email:** dorazio@utdallas.edu

**Office:** GR 3.108D

**Office Phone:** 972-883-6212

**Office Hours:** Individual or group meetings on MS Teams available by appointment.

## Course Modality and Expectations

**Instructional Mode** The instructional mode is *remote*. This means “synchronous online learning at the day and time of the class. The instructor delivers the instruction from home or the office. Students complete the course at a distance.” See <https://covid.utdallas.edu/students-families-info/spring-2021-registration/> for more information.

**Course Platform** The class will meet at our scheduled time on MS Teams. Students can join and view meeting details by selecting the Calendar tab on Teams. For group projects, students will meet on MS Teams at a mutually agreeable time.

**Expectations** Students who do not choose the asynchronous option are expected to attend class at our regularly scheduled time and to participate as if the class were in-person. Class discussion is a critical part of learning, so all students are expected to complete the weekly assignments and come to class prepared to discuss.

**Asynchronous Learning Guidelines** Students may select the asynchronous option at any time. Students who select the asynchronous option must notify the instructor. For each week that the student is engaged in asynchronous learning, the student is required to complete the asynchronous assignment, described below in the Attendance and Participation section.

For more information on asynchronous learning: <https://covid.utdallas.edu/response/faq/#asynchronous>

## COVID-19 Guidelines and Resources

The information contained in the following link lists the University's COVID-19 resources for students and instructors of record.

Please see <http://go.utdallas.edu/syllabus-policies>.

## Course Description

This course provides a survey of Ordinary Least Squares (OLS) regression, with an emphasis on using data and regression models to test hypotheses about social phenomena. This is an applied course, and we will focus on developing the skills needed to estimate, interpret, and analyze regression models appropriately. We will use examples across the social sciences, including political science, public policy, economics, and sociology.

In the first four weeks of class we will learn about different types of data, levels of measurement, and how to assess hypotheses using tables and data visualizations. The second part of class will introduce OLS regression, and how to use it to assess hypotheses. The third part of class focuses on model diagnostics at both the conceptual and applied level. The fourth introduces logistic regression, and fixed and random effects.

Students will gain extensive experience analyzing data. While students are free to complete their assignments using any tools they want, the primary statistical language of this course is R. Why R? One, it is free. Many other tools for data analysis, such as Stata, are proprietary. Two, because it is widely used and often listed (alongside Python) as a requirement for employment. That said, R isn't the right choice for everybody, and students are free to complete assignments using any tool they want.

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

## Learning Objectives

There are four learning objectives for this course. The first is to learn how to assess social science hypotheses using OLS regression and data analysis more broadly. The second is to understand different types of data, and how to estimate and interpret regression models that are appropriate for the kind of data being analyzed. The third, at both a conceptual and applied level, is to understand model diagnostics, such as outlier detection, and methods for addressing associated issues. Finally, the fourth is to learn to use R, or some other tool of the student's choosing, to analyze real data. These learning objectives will be assessed through class discussion, homework assignments, group project, and the final paper.

## Course Requirements

Students 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, and we will be analyzing data, much of what is read contains symbols and math. When reading, try to 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%
Group presentation	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.

For Weeks 2 - 8, our classes will meet virtually for two hours, from 7-9pm. The additional 45 minutes of class time will be shifted to group meetings at a time to be determined by each group. The selected time should also accommodate our asynchronous students, who are required to attend the group meetings. Each week, one member from each group is to complete and upload the Group Meeting Form (available on eLearning), to their group's channel on Teams.

**For asynchronous students**, or for any week where the asynchronous option is selected, participation will be assessed with the asynchronous assignment (in addition to participation in group meetings). Asynchronous students are expected to submit #1-3 *before* class, and #4 *after* class (class is Tuesday, submit #4 by Friday 5pm).

1. Identify and describe one core concept from the reading
2. Raise one thoughtful question about the concept
3. Discuss how you would teach this concept to your classmates
4. Post-class follow-up
  - Watch the recorded lesson on MS Stream
  - Discuss how the instructor taught the concept to the class, what you would have done differently, and what you would have kept the same

**The asynchronous assignment is to be submitted on eLearning, and is in addition to any other assignments due that day.**

**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.

**Final project (35%)**

For the final project, students will complete a replication and extension of an existing research paper. Since this is a methods course, most of the 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. The paper should include the following:

1. Abstract
2. Introduction
3. Literature review and theory, leading to explicitly stated hypotheses
4. Research design, with a discussion of the method used, the unit of observation, variables, scope, and data sources
5. Empirical results, with an emphasis on displaying the results and assessing the hypotheses
6. Conclusion

You can find replication data for many peer-reviewed articles at Harvard's Dataverse. Here are some places to consider, but note that you are not limited to these:

- <https://dataverse.harvard.edu/dataverse/ajps>
- <https://dataverse.harvard.edu/dataverse/jop>
- <https://dataverse.harvard.edu/dataverse/IOJ>
- <https://dataverse.harvard.edu/dataverse/qje>
- <https://dataverse.harvard.edu/dataverse/JPublicPolicy>

Your extension should be something new, but with the same replication data. For the extension, some options include:

- Keep the same dependent variable and assess different hypotheses using different independent variables
- Run diagnostics and sensitivity analyses on the original models

- Assess hypotheses using alternative measures of key concepts
- Test new hypotheses using a different dependent variable
- Test the original hypotheses on interesting subsets of the data

Here are two example of high-quality replication papers in Political Science:

1. Park, J. and Colaresi, M. (2014). Safe across the border: The continued significance of the democratic peace when controlling for stable borders. *International Studies Quarterly*, 58(1):118–125
2. Choi, S.-W. (2011). Re-evaluating capitalist and democratic peace models. *International Studies Quarterly*, 55(3):759–769

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

#### 1. Proposal and initial replication—due **April 13**

- There are four items to submit:
  - (a) Your pdf report
  - (b) The data
  - (c) The replication code
  - (d) A pdf of the article you’re replicating
- This should contain a replication of the original findings and the proposed extension. Note that you do not need to replicate every finding in the original paper. You only need to replicate the findings that you wish to extend. I may ask you to revise your proposal if it is not appropriate.

#### 2. Final paper—due **11:59pm on May 6**

- Same four items to submit:
  - (a) Your pdf report
  - (b) The data
  - (c) The replication code for reproducing all graphics and results in your paper
  - (d) A pdf of the article you’re replicating

### Group Presentation (25%)

The group presentation will be an *exploratory data analysis* (EDA). The basic idea that underlies EDA is to develop and refine hypotheses and insights using data. This chapter provides a good, simple explanation: <https://r4ds.had.co.nz/exploratory-data-analysis.html>.

Each group is to find an interesting dataset, conduct the EDA, and present the results. This presentation should include OLS regressions, but note that they are exploratory and

intended to generate insights rather than assess any individual hypothesis. The purpose of EDA is to develop and refine hypotheses, not to

I encourage you to be creative with your presentation. Your group may give a live presentation with slides, or you may record it in advance. You can create a Shiny app or create and showcase interactive data visualizations however you see appropriate.

Groups will be assigned by the instructor. Students are expected to meet Weeks 2 through 8, for 45 minutes, at a mutually agreeable time. Feel free to invite the instructor to attend! Note that our regular class meetings will last for two hours, not 2:45, so these one hour meetings are intended to comprise that additional 45 minutes of class time for Weeks 2 through 8.

### Grade Scale

A	$\geq 93\%$	B	= 83-86%	C	73-76%
A-	= 90-92%	B-	= 80-82%	F	$\leq 72\%$
B+	= 87-89%	C+	= 77-79%		

See <https://catalog.utdallas.edu/now/graduate/policies/grades> for additional information about university grading policies.

## Required Texts

*Applied Regression Analysis and Generalized Linear Models* is a solid introductory textbook on regression and data analysis. It is comprehensive, and I think the ratio of words:formulas is perfect for this applied class.

1. Fox, J. (2016). *Applied Regression Analysis and Generalized Linear Models*. Sage, 3rd edition

## Optional Texts

These texts are optional—any reading from them will be made available by me. In addition, 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. Angrist and Pischke is great for understanding causal inference. Treiman is best for understanding actual, practical data analysis.

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

## Resources for R and Regression Analysis

Google. Students should learn how to use Google effectively, and use it regularly.

The Institute for Digital Research and Education at UCLA has some excellent resources and examples for using R and Stata (<https://stats.idre.ucla.edu/>). Examples from Fox (2016) and Chatterjee and Hadi (2015) are both available, but only for Stata.

The Summer Institutes in Computational Social Science is also a great resource (<https://compsocialscience.github.io/summer-institute/>). Many lessons on computational social science topics, include a “boot camp” on learning R.

There are also resources available through the UTD library. You may find these helpful to learn R and the basics of OLS regression modeling:

- <https://www.lynda.com/R-training-tutorials/1570-0.html>
- <https://utdallas.edu/library/news/get-ready-to-meet-lynda/>

For an introduction to R, as well as a basic resource, you can use *An Introduction to R* by Venables and Smith. You can access the PDF here (<https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>).

There is also *R for Data Science*, at <https://r4ds.had.co.nz/index.html>. This is a great resource, but note that it uses a set of R libraries that are collectively known as the Tidyverse ([tidyverse.org](https://tidyverse.org)), along with RStudio.

You can download R here (<https://www.r-project.org/>). And you can download RStudio here (<https://rstudio.com/>).

## Course Outline

**Day 1: Introductions, Syllabus, and Levels of Measurement** **January 19**

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

**Day 2: Experimental and Observational Data** **January 26**

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

**Day 3: Tables** **February 2**

- Treiman, Chapters 1, 2, 3

---

<b>Day 4: Data and Visualizations</b>	<b>February 9</b>
<ul style="list-style-type: none"><li>• Homework 1 is due</li><li>• Fox, Chapters 3 and 4</li><li>• Treiman, Chapter 4</li></ul>	
<b>Day 5: CLASS CANCELED</b>	<b>February 16</b>
<b>Day 6: Simple Regression</b>	<b>February 23</b>
<ul style="list-style-type: none"><li>• Fox, Chapter 2</li><li>• CH Chapter 2: Simple Linear Regression</li></ul>	
<b>Day 7: Multiple Regression I</b>	<b>March 2</b>
<ul style="list-style-type: none"><li>• Homework 2 is due</li><li>• Fox, Chapter 5</li><li>• CH Chapter 3: Multiple Linear Regression</li></ul>	
<b>Day 8: Multiple Regression II</b>	<b>March 9</b>
<ul style="list-style-type: none"><li>• Fox, Chapter 6</li><li>• CH Chapter 4: Regression Diagnostics: Detection of Model Violations</li><li>• Homework 3 is due Friday, March 12 by 7pm</li></ul>	
<b>SPRING BREAK</b>	
<b>Day 9: Group Presentations</b>	<b>March 23</b>
<ul style="list-style-type: none"><li>• Submit group presentations before class</li></ul>	
<b>Day 10: Dummy Variables and Interaction Terms</b>	<b>March 30</b>
<ul style="list-style-type: none"><li>• Fox, Chapter 7</li><li>• CH Chapter 5: Qualitative Variables as Predictors</li><li>• Brambor, T., Clark, W. R., and Golder, M. (2006). Understanding interaction models: Improving empirical analyses. <i>Political Analysis</i>, 14(1):63–82</li><li>• Berry, W. D., Golder, M., and Milton, D. (2012). Improving tests of theories positing interaction. <i>The Journal of Politics</i>, 74(3):653–671</li></ul>	

**Day 11: Multicollinearity****April 6**

- Homework 4 is due
- Fox Chapter 13
- CH Chapter 9: Analysis of Collinear Data
- Ch Chapter 10: Working with Collinear Data

**Day 12: Missing Data, Regression Diagnostics****April 13**

- **Project paper # 1 is due**
- Fox, Chapter 11, 20
- Treiman chapter 8, 10

**Day 13: Heteroscedasticity and Autocorrelation****April 20**

- Fox, Chapter 12
- CH Chapter 6: Transformation of Variables
- CH Chapter 8: The Problem of Correlated Errors

**Day 14: Logistic Regression****April 27**

- Homework 5 is due
- Fox, chapter 14
- Treiman chapter 13
- CH Chapter 12: Logistic Regression

**Day 15: Random and Fixed Effects****May 4**

- Fox, chapter 23
- Treiman chapter 15

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

## Class Participation

Regular class participation is expected regardless of course modality. Students who fail to participate in class regularly are inviting scholastic difficulty. A portion of the grade for this course is directly tied to your participation in this class. It also includes engaging in group or other activities during class that solicit your feedback on homework assignments, readings, or materials covered in the lectures (and/or labs). Class participation is documented by faculty. Successful participation is defined as consistently adhering to University requirements, as presented in this syllabus. Failure to comply with these University requirements is a violation of the Student Code of Conduct: <https://policy.utdallas.edu/utdsp5003>.

## Class Recordings

Students are expected to follow appropriate University policies and maintain the security of passwords used to access recorded lectures. Unless the Office of Student AccessAbility has approved the student to record the instruction, students are expressly prohibited from recording any part of this course. Recordings may not be published, reproduced, or shared with those not in the class, or uploaded to other online environments except to implement an approved Office of Student AccessAbility accommodation. Failure to comply with these University requirements is a violation of the Student Code of Conduct: <https://policy.utdallas.edu/utdsp5003>.

The instructor may record meetings of this course. Any recordings will be available to all students registered for this class as they are intended to supplement the classroom experience. Students are expected to follow appropriate University policies and maintain the security of passwords used to access recorded lectures. Unless the Office of Student AccessAbility has approved the student to record the instruction, students are expressly prohibited from recording any part of this course. Recordings may not be published, reproduced, or shared with those not in the class, or uploaded to other online environments except to implement an approved Office of Student AccessAbility accommodation. If the instructor or a UTD school/department/office plans any other uses for the recordings, consent of the students identifiable in the recordings is required prior to such use unless an exception is allowed by law. Failure to comply with these University requirements is a violation of the Student Code of Conduct: <https://policy.utdallas.edu/utdsp5003>.

## Class Materials

The Instructor may provide class materials that will be made available to all students registered for this class as they are intended to supplement the classroom experience. These materials may be downloaded during the course, however, these materials are for registered students' use only. Classroom materials may not be reproduced or shared with those not in class, or uploaded to other online environments except to implement an approved Office of Student AccessAbility accommodation. Failure to comply with these University requirements is a violation of the Student Code of Conduct: <https://policy.utdallas.edu/utdsp5003>.

## Additional Information and Resources

### Technical Requirements

In addition to a confident level of computer and Internet literacy, certain minimum technical requirements must be met to enable a successful learning experience. Please review the important technical requirements on the Getting Started with eLearning webpage: <https://ets.utdallas.edu/elearning/students/current/getting-started>.

### Course Access and Navigation

This course can be accessed using your UT Dallas NetID account on the eLearning website: [elearning.utdallas.edu](https://ets.utdallas.edu).

Please see the course access and navigation section of the Getting Started with eLearning webpage for more information: <https://ets.utdallas.edu/elearning/students/current/getting-started>.

To become familiar with the eLearning tool, please see the Student eLearning Tutorials webpage: <https://ets.utdallas.edu/elearning/students/current/tutorials>

UT Dallas provides eLearning technical support 24 hours a day, 7 days a week. The eLearning Support Center includes a toll-free telephone number for immediate assistance (1-866-588-3192), email request service, and an online chat service: <https://ets.utdallas.edu/elearning/helpdesk>.

### Communication

This course utilizes online tools for interaction and communication. Some external communication tools such as regular email and a web conferencing tool may also be used during the semester. For more details, please visit the Student eLearning Tutorials webpage for video demonstrations on eLearning tools: <https://ets.utdallas.edu/elearning/students/current/tutorials>.

Student emails and discussion board messages will be answered within 3 working days under normal circumstances.

### Distance Learning Student Resources

Online students have access to resources including the McDermott Library, Academic Advising, The Office of Student AccessAbility, and many others. Please see the eLearning Current Students webpage for more information: <https://ets.utdallas.edu/elearning/students/current>.

### Server Unavailability or Other Technical Difficulties

The University is committed to providing a reliable learning management system to all users. However, in the event of any unexpected server outage or any unusual technical difficulty

which prevents students from completing a time sensitive assessment activity, the instructor will provide an appropriate accommodation based on the situation. Students should immediately report any problems to the instructor and also contact the online eLearning Help Desk. The instructor and the eLearning Help Desk will work with the student to resolve any issues at the earliest possible time: <https://ets.utdallas.edu/elearning/helpdesk>.

### **Comet Creed**

This creed was voted on by the UT Dallas student body in 2014. It is a standard that Comets choose to live by and encourage others to do the same:

“As a Comet, I pledge honesty, integrity, and service in all that I do.”

### **Academic Support Resources**

The information contained in the following link lists the University’s academic support resources for all students.

Please go to Academic Support Resources webpage for these policies: <https://provost.utdallas.edu/syllabus-policies/#academic-support-resources>

### **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.

Please go to UT Dallas Syllabus Policies webpage for these policies: <https://go.utdallas.edu/syllabus-policies>.

**The descriptions and timelines contained in this syllabus are subject to change at the discretion of the Professor.**