POLI 784: Regression Models

Spring 2020

Ted Enamorado Simon Hoellerbauer

Department of Political Science
University of North Carolina at Chapel Hill

This course is the second course in applied statistical methods for social scientists. Students will learn a variety of basic *cross-section* regression models including linear regression model, structural equation and instrumental variables models, discrete choice models, and models for missing data. Unlike traditional courses on applied regression modeling, we will emphasize the connections between these methods and causal inference, which is a primary goal of social science research.

1 Contact Information

Ted Enamorado (Instructor)
Office: Hamilton Hall 304
Email: tede@email.unc.edu

Simon Hoellerbauer (Teaching Assistant)
Office: Hamilton Hall 451
Email: hoellers@unc.edu

2 Logistics

- Lectures: Mondays and Wednesdays (Hamilton Hall 351) 03:35–04:50 pm
- Lab sessions: Fridays (Dey Hall 401) 10:10–11:00 am (first session: January 10th)
- Ted's Office Hours: Mondays and Wednesdays, 9:30–11:00 am (or just drop by or make an appointment)
- Simon's Office Hours (Hamilton 150): Wednesdays (5:30–7:00 pm) and Fridays (1:30–3:00 pm)

3 Questions and Announcements

In addition to lab sessions and office hours, please use the **Piazza** Discussion Board at `https://piazza.com/` when asking questions about lectures, problem sets, and other course materials. This allows all students to benefit from the discussion and to help each other understand the

materials. Both students and instructors are encouraged to participate in discussions and answer any questions that are posted.

To join the POLI 784 Piazza site, click on "Search Your Classes" from the Piazza homepage. After specifying UNC Chapel Hill as your school, search for "`POLI784`" You will then be prompted to enter your `unc.edu` email address to confirm your registration. All class announcements will be posted on Piazza and Sakai will still be used for hosting all class materials.

Some useful tips for Piazza include:

- Piazza has apps available for the iOS and Android platforms. The apps are free downloads and provide complete access to all of Piazza's message board features.
- To insert LaTeX-formatted text in a post, place a double dollar sign ($$) on both ends of the relevant text, or click the fx button in the Details toolbar above your post.
- To add formatted **R** code to a post, click the "pre" button in the Details toolbar above your post. A grey text box will open up where you can paste code from **R**.
- You can classify a post using pre-selected tags, or you can generate your own by prepending a hash (#) to your chosen label. Posts can then be sorted by these tags using the search bar in the left-hand column.

4 Course Requirements

Your final grade is based on lab attendance, the problem sets, and two exams:

- **Labs** (10%): There will be a weekly lab session where you will work through a guided practical exercise designed to help with understanding of the materials covered that week and the problem sets. You are required to attend ALL lab sessions.
- **Problem sets** (50%): There will be four problem sets throughout the semester. Each problem set will equally contribute to the final grade and contain both analytical and data analysis questions. The following instructions will apply to all problem sets:
  - *Group Collaboration.* Each student will be assigned to a group. Groups are required to work together and produce a single set of solutions to each problem set. To facilitate individual learning, all group members should contribute to all problem set questions rather than dividing them up. You must work through the whole problem set together with your group, rather than working separately on different portions of the problem sets. There is to be no collaboration between groups, aside from public posts on Piazza. Groups will be reassigned for each problem set.
  - *Submission.* Submit your answers and code as a PDF file (**R Marcdown** is highly recommended) to Sakai and bring its printed copy to class. Please ensure your code adheres to the tidyverse's **R** Style Guide rules (`https://style.tidyverse.org/`). Late submission will not be accepted unless you obtain a prior approval from the instructor.
- **Exams** (40%): Two closed-book exams. Dates: March 4 and May 4 (both from 3:35 to 4:50 pm, location Hamilton 351), covering the first and second half of the course materials, respectively. Each exam is equally weighted (20% each).

Grading Scale

Score	Grade	Score	Grade	Score	Grade	Score	Grade
$> 90\%$	H	$\geq 75\%$	P	$\geq 60\%$	L	$< 60\%$	F

5 Statistical Software

In this course, we support a statistical computing environment, called **R**. **R** is available for any platform and without charge at http://www.r-project.org/. We choose **R** for its flexibility and power. However, students may use other statistical software such as STATA, Python, Julia, etc., for the problem sets and the final project, but at their own risk; that is, we will not be able to answer your software-related questions. Of course, there will be no penalty for using different statistical software. What matters is the analysis you present rather than the software you use.

6 Textbooks

There is no single textbook for this course. However, you may find the relevant parts of the following textbooks useful. You do NOT need to buy any of them.

1. Probability and Statistics

   David A. Freedman. *Statistical Models: Theory and Practice*. Cambridge University Press, Cambridge, 2005.

   Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, New York, 2005.

2. Econometrics

   Jeffrey M. Wooldridge. *Introductory Econometrics: A Modern Approach*. South-Western Cengage Learning, Mason, OH, 5th edition, 2012.

   Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, MA, 2nd edition, 2010.

   Fumio Hayashi. *Econometrics*. Princeton University Press, Princeton, 2000.

   Bruce Hansen. *Econometrics*. Manuscript, 2019. URL `https://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf`.

3. Causal Inference

   Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015.

   Stephen L. Morgan and Christopher Winship. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, New York, 2007.

   Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, 2009.

4. Statistical learning

James Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning : with Applications in R*. Springer New York Inc., New York, NY, 2013. URL `http://faculty.marshall.usc.edu/gareth-james/ISL/`.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, 2001. URL `https://web.stanford.edu/~hastie/ElemStatLearn/`.

7 Other Class Policies

- **Attendance:** you will not be graded directly on your attendance to the class lecture. However, I strongly suggest students expecting to receive an H in this course attend all lectures, and your in-class work grade will reflect missed sessions. Lab attendance *is* mandatory.
- **Technology in the classroom:** You will frequently make use of computers in this course, during some lecture periods and during software training. Please be respectful to your instructor and your peers by using your computers only for class-related purposes. Put your phone away before class starts and do not bring it out.
- **Students with disabilities:** Students with disabilities needing academic accommodation should 1) contact the office of Learning Disabilities at UNC (`http://www.unc.edu/depts/lds/index.html`); 2) bring a letter to the instructor indicating the need for accommodation and what type. This should be done during the first week of class.
- **Religious observances:** Some students may wish to take part in religious observances that occur during this semester. If you have a religious observance that conflicts with your participation in the course, please meet with the instructor before the end of the second week of the semester to discuss appropriate accommodations.
- **Academic honesty:** As noted above, while collaboration is expected on assignments, plagiarism will not be tolerated. This includes (but is not limited to) using someone else's code to conduct the analysis required for the assignments, copying responses from others, and using others' work without attribution when writing your final paper. More generally, please review UNC's policies regarding academic honesty, which you can learn about at `https://studentconduct.unc.edu/`. If you have any question regarding this issue, please feel free to ask me.

8 Course Outline

We will cover some of the following topics as time permits!

1. Statistical Modeling with Cross-section Data
   (a) Simple Regression
   (b) Multiple Regression
   (c) Regression Diagnostics
   (d) Model Selection and Validation

(e) Instrumental Variables

(f) Extension for the Linear Regression: Ridge and LASSO

(g) Missing Data and Multiple Imputation

2. Causal Inference (Potential Outcomes)

   (a) Causality and Regression

   (b) Permutation Tests

   (c) Mediation Analysis

   (d) Regression Discontinuity Designs (Sharp, Fuzzy, Kink)

3. Maximum Likelihood

   (a) Likelihood Theory

   (b) Discrete Choice Models

   (c) Resampling Methods

4. Introduction Statistical Modeling with Longitudinal Data

   (a) Fixed Effects, First Differences, and Difference-in-Differences

   (b) Random Effects

5. Introduction to Statistical Learning with Unlabeled Data

   (a) Clustering Analysis: k-means (and its variants)

   (b) The Expectation–Maximization (EM) Algorithm

   (c) Clustering Analysis (revisited): Mixture Models

   (d) Text Classification via Mixture Models