# SOSC 5340: Econometric Approaches to Social Science Research

Spring 2021

**Lecture Time: Tuesday 9:00 - 11:50AM**

*This version prepared on Jan 29, 2021*


| | Instructor | Teaching Assistant |
|---|---|---|
| | **ZHANG, Han** | **Yin, Yabin** |
| Office | Room 2379, Academic Building, Lift 15 | Room 3001, Academic Building, Lift 4 |
| Email | zhangh@ust.hk | yyinak@connect.ust.hk |
| Office Hour | Tuesday 2:00PM to 3:00PM | TBD |


## Online Teaching

We will use Zoom for online teaching until further notice from the university. Zoom link can be found on Canvas. Teaching can be difficult without face-to-face interactions. Therefore, asking questions promptly is even more important. Please do not hesitate to stop me and ask questions if you have any question regarding the course material and teaching.

## Prerequisite

This course is the second of the two-semester graduate-level applied social science statistics sequence (after SOSC 5090).

- Students are expected to be familiar with the materials covered in SOSC 5090 or its equivalent.
- Students should also be able to use at least one statistical programming language. We will use R in lecture and tutorials. All course contents and exercises can be implemented in other languages such as Stata.

## Goal

Upon finishing the course, students should be able to:

1. Understand the core math concepts of statistical estimation and inference
2. Be familiar with applied regression modeling
3. Be able to articulate the challenges in establishing causal arguments in social sciences research, and understand some of the statistical methods used to address these challenges.
4. Build a solid, reproducible research pipeline to go from raw data to the final paper.
5. Gain hands-on experience of writing and presenting a research paper that aims to eventually be publishable in an academic journal.

## Course Outline (Tentative)

| Week | Topic | Item Dues |
|---|---|---|
| 1 (Feb 2) | Review of Statistical Inference | |
| 2 (Feb 9) | Linear regression | |
| 3 (Feb 16) | Logistic/probit | |
| 4 (Feb 23) | Generalized linear models | Assignment 1 |
| 5 (Mar 2) | Machine learning basics | Paper presentations |
| 6 (Mar 9) | Causal inference: counterfactual framework | Assignment 2 |
| 7 (Mar 16) | Causal inference in experiments | Preliminary draft of final paper |
| 8 (Mar 23) | Causal inference in observational studies | Paper presentations |
| 9 (Mar 30) | Diff-in-diff and fixed effect model | Assignment 3 |
| 10 (Apr 13) | Instrumental variables | Paper presentations |
| 11 (Apr 20) | Regression discontinuity | |
| 12 (Apr 27) | Fuzzy regression discontinuity | Assignment 4 |
| 13 (May 4) | Wrap up | Presentations of final paper |

## Grading Policy

Your score will be accessed based on the following five components (no mid-term and final exams):

| | |
|---|---|
| Attendance | 5% |
| Assignments | 35% |
| Presentation of an academic article (15 min) | 10% |
| Presentation of your final paper (20 min) | 15% |
| Write-up of your final paper | 35% |

1. Homework assignment: short coding homework to make sure that you know how to run models we covered in the lectures. Our TA will hold tutorial sections to teach you how to run these models before assignments.
    - All assignments are due at the beginning of the class. You can turn it in online through Canvas.
2. Presentation of an article: You will select one article from a list of articles provided by the instructor and present it to the class. Each student has to choose a different article. This exercise prepares you a chance to practice giving a presentation at an academic conference.
3. Final paper:
    - Preliminary draft of the final paper: double spaced, 4-6 pages. Describe the background, hypotheses, data, and methods you plan to use.
    - The final paper is due two weeks after the final class, on **May 18, 11:59 PM**. Double spaced; at least 20 pages including everything, i.e., title, abstract, texts, tables, figures, and references.
4. Presentation of your final paper: Your presentation will be on **May 4**, the last day of the class.

### Late submission policy

Late delivery of due items will be marked down 75% if received within 1 day of the due date, and 50% if received within 3 days of the due date; you will receive zero credit if the due item is not delivered within 3 days of the due date. Contact the instructor if there are rare unforeseen circumstances.

### Re-grade policy

If you want to dispute a grade, please submit your argument in writing along with your assignment. We will evaluate the merit of your argument as well as perform a full reassessment of your entire assignment. This means that your grade may end up lower than it was originally.

## Course Materials

There is no required textbook. Most of my lecture contents are drawn from the three books below:

- Angrist, J. D. & Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. It's available for free download at here.
- Aronow, P. M. & Miller, B. T. (2019). *Foundations of Agnostic Statistics*. New York: Cambridge University Press. A "modern" interpretation of statistics without relying on regressions.
- Hansen, Bruce (2020). *Econometrics*. Free at the author's website `https://www.ssc.wisc.edu/~bhansen/econometrics/`

Other recommended textbooks:

- Applied regression modeling:
  - Fox, J. (2015). *Applied Regression Analysis and Generalized Linear Models*. Los Angeles: SAGE Publications, Inc, third edition. A classical (and traditional) treatment of applied regression modeling.
  - Gelman, A. & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge ; New York: Cambridge University Press, 1 edition edition. More on hierarchical models.
  - Wooldridge, J. M. (2015). *Introductory Econometrics: A Modern Approach*. Boston, MA: Cengage Learning, 6 edition edition. Classical econometric textbook on regressions.
- Econometrics:
  - Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press. More advanced topics.
- Causal Inference:
  - Imbens, G. W. & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. New York: Cambridge University Press, 1 edition edition. A detailed treatment causal inference topics in social sciences.
- Machine Learning:
  - James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An Introduction to Statistical Learning: With Applications in R*. New York: Springer, second edition. Introductory level; with R code examples. Free for downloading at the author's website.
  - Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., & Tibshirani, R. (2009). *The Elements of Statistical Learning*, volume 2. Springer. Intermediate level; recommended for people not from a CS background but from a statistics background. Free for download at the author's website.

## Programming Resources

We will use R lectures and tutorials. Using Stata or other languages (e.g., Python or Julia) is acceptable. R is freely available for download and runs on Macintosh, Windows, and Linux computers. Students are strongly encouraged to use Rstudio, another freely available software package that has numerous features to make data analysis easier.

Additional resources for learning R:

- Download R: `https://www.r-project.org/`
- Download Rstudio: `https://www.rstudio.com/`

- R syntax:
  - R for beginners. Good overview.
  - Software Carpentry course on R
  - Princeton's Computing for Data Analysis in the Social Sciences workshops
- Using R for regression and causal inference:
  - Princeton R Tutorial. Quick start on using R for regression modeling.
  - UCLA IDRE. Detailed resources for regression modeling with R.
- Visualization:
  - R graphics with ggplot2 workshop notes at Harvard
  - Healy, K. (2018). *Data Visualization: A Practical Introduction.* Princeton, NJ: Princeton University Press, 1 edition edition
  - Andrew Gelman: Lets Practice What We Preach: Turning Tables into Graphs
- Reproducibility:
  - Gentzkow and Shapiro, "Code and Data for the Social Sciences: A Practitioners Guide". Detailed advices on organizing your codes and data to enhance reproducibility.
- If you already know Stata:
  - A Quick Introduction to R (for Stata Users)