Quantitative Analysis and Empirical Methods II (API-202 C)

Harvard Kennedy School

Course Syllabus – Spring 2022

Draft Version: Updated 1/12/2021

Faculty

Benjamin Schneer

E-mail: benjamin_schneer@hks.harvard.edu

OHs: Thursdays 1:00 PM – 3:00 PM EST

Teaching Team

Teaching Fellow:	**Sophie Hill**	sophie_hill@g.harvard.edu
	OHs: TBD	
Course Assistants:	**Kevin Dong**	kevindong@hks.harvard.edu
	OHs: TBD	
	Catherine Darin	cdarin@hks.harvard.edu
	OH: TBD	
Faculty Assistant:	**Jahaida Jesurum**	jahaida_jesurum@hks.harvard.edu

Class Meetings

Classes: Tuesday and Thursday 10:30 AM – 11:45 AM EST

Review Sessions: Friday 10:30 AM – 11:45 AM EST

Description

This course, which is a continuation of Quantitative Analysis and Empirical Methods (API-201), equips you with the tools to conduct and understand empirical analysis of public policy problems that you might encounter in a professional environment. The emphasis is on empirical applications that will give you experience gathering and working with data along with developing empirical models for causal inference and for prediction.

The course is designed with two objectives in mind. The first is to provide you with the ability to assess empirical analysis done by others so that you can make intelligent decisions about how to use that analysis in the design of public policy. The second is to provide you with the skills necessary to perform empirical analysis on your own, or to participate on a team involved in such an empirical analysis. Part of this goal involves developing facility with statistical software. To that end, you will learn to conduct statistical analyses in **R** and to translate your findings into policy recommendations.

The course is broadly divided into three parts: **(1) Data and Models**, **(2) Causal Inference**, and **(3) Prediction and Machine Learning**.

Prerequisites

This course is designed for MPP students, and is a continuation of the content from API-201. All students who register for this course must have taken, or exempted from, API-201.

Course Sections

Students are encouraged to determine which section of API-202 they will be in as early as possible. The deadline to switch between Z and non-Z sections is February 18. If you switch over, we will send your homework grades to the section that you have switched over to. A list of Frequently Asked Questions (FAQs) related to the Z-section can be found here.

R Statistical Software

Completing problem sets will require the use of R. In order to begin the semester with a basic familiarity with R, a series of tutorials will be shared in the weeks prior to the beginning of the semester, along with an assignment that must be submitted before the start of class. This assignment is graded for completion.

We will be using RStudio in this course, which provides a usable front end to work with the R language. Use of RStudio is free and downloadable at www.rstudio.com. R training sessions will be held at the start of the semester, and support from the teaching team for R-related issues will be available throughout the semester.

Resources: Letter to Students | R Textbook | Cheat Sheet

Grading

Your final grade will be based on the following criteria:

- R Winter Assignment (Completion) 5%
- Class Participation/Engagement 10%
- Problem Sets 30%
- End of Unit Assessments 30%
- Final Assessment 25%

Pre-Semester Exercise

Before the start of the semester, you will complete an exercise designed to give you experience working with the **R** statistical package. This exercise will be graded for completion only.

Problem Sets (30%)

One problem set will be assigned for each unit (for a total of three problem sets) and will be **due on Thursdays at 9:00 AM**. Problem sets will be posted on the course website, as will suggested answers. They will be graded on a three-point scale:

- 3 points = check-plus
- 2 points = check
- 1 point = check-minus

Problem sets not received by the deadline will be considered late. There will be no credit for late assignments.. Your two highest assessment problem set grades will each comprise 13% of your total grade, and your lowest will comprise 4% of your total grade

Under the Harvard Kennedy School Academic Code, the problem sets for this course are "Type II" assignments unless indicated otherwise. **You are encouraged to work in a study group, but must submit your own solutions**. Examples of assignments that are not in accordance with the HKS academic code include reprints of substantially identical assignments, printouts of substantially identical Excel tables or R script files, and copies of solutions from previous years. Violations of the Academic Code are a serious violation of academic and professional standards and can lead to a failing grade in the course, failure to graduate, and even expulsion from the University. The Kennedy

School Academic Code is available at:
https://www.hks.harvard.edu/educational-programs/academic-calendars-policies/student-handbook/general-regulations-and-1

Instructions for submitting problem sets:

- Turn them in electronically via the Canvas course page.
- Submit them by **9:00 AM** on the day they are due (Thursday). Assignments submitted after class begins will be considered late.
- Indicate on the cover page the names of the classmates you worked with.

Class participation and engagement (10%)

Your willingness to contribute to the overall learning of your classmates is an important component of this class. To encourage this, your participation will be used in determining your grade in the course. Note that this assessment includes both the quantity and quality of your contributions.

Engagement in class: A strong engagement in class means attending class regularly and punctually, engaging actively in all in-class activities, and in general contributing to a positive learning atmosphere in the classroom. My aim is to make our class a highly collaborative environment, using a combination of discussions, group work, and interactive voting. I hope to encourage you to ask questions about the topic at hand, provide explanations to your peers, and draw from your personal and professional experiences to inform your thinking.

You are expected to attend and engage in all live classes. If you cannot attend a live class, please send your teaching team an email before class (except for medical emergency reasons).

Engagement out of class: Students are encouraged to pose questions about the material, R programming, or other relevant issues on the Canvass discussion board. You are also encouraged to help answer questions posed by your classmates where applicable. Teaching staff will also monitor the site to help answer questions and share ideas.

Quizzes (30%)

There will be three quizzes/assessments during the semester, one during each unit in the course. Your two highest quiz/assessment grades will each comprise 13% of your grade, and your lowest grade will comprise 4% of your grade. These will be completed on Canvas rather than in class.

Final Assessment (25%)

You will have one final take-home exam, which you will have 24 hours to complete. It will be cumulative, and include questions on the concepts of the course in addition to work using R.

Regrade policy

Requests for reconsideration of grades on assessments are not encouraged, and will be accepted only in writing, with a clear statement of what has been incorrectly graded, and within two business days of receiving your graded exam. Please submit your full exam so grading on all questions can be reconsidered.

All course activities, including class meetings, homework assignments, and exams, are subject to the HKS Academic Code (available here) and the HKS Code of Conduct (available here).

Letter Grades

The numerical grades for each exam and for each component of the course (problem sets, final exercise, and class participation and engagement) will be standardized (i.e. curved) and then an overall score for the course will be calculated for each student. This overall score will be translated into a final course letter grade using the Dean's Recommended Grade Distribution.

Other Items

Recording Classes

Classes will be video-recorded, and recordings will be available to provide you with the option of reviewing the class so you can clarify or deepen your understanding of a particular concept. As a member of our learning community and to stimulate risk-taking and vigorous debate in class, you are expected not to make any recordings available outside of our learning community.

Readings

The course material is self-contained and there is no required textbook for the course. There will be some required readings and some optional readings. The optional readings are available for students desiring additional reference material.

W: Wooldridge, J.M. Introductory Econometrics, Thompson South-Western, 2006 (Available here)

AP: Angrist, J. D. and Pischke, J. Mostly Harmless Econometrics, Princeton University Press, 2009. (Available here)

MM: Angrist, J. D. and Pischke, J-S., Mastering 'Metrics, Princeton University Press, 2015.

JWHT: James, G., Witten, D. Hastie, T., and Tibshirani, R. An Introduction to Statistical Learning with Applications in R, Springer, 2013. (Available here and here)

SW: Stock, J. and Watson, M. Introduction to Econometrics, Alison-Wesley, 2007 (Available here)

Any edition of each is fine. Note that, due to differences in editions, the reading selections listed in the course schedule may differ from the edition you use. You should be able to find a corresponding reading, but please get in touch if you have trouble finding the appropriate reading.

Course Schedule

This schedule is subject to (modest) change. Please see the Canvas course website for updates. Note that all readings not in bold are optional.

	Class	Date	Topic	Assignment
Data	1	**January 25**	Answering Questions with Data: Description and Prediction	
	2	**January 27**	Answering Questions with Data: Causality	
	3	**February 1**	Data Collection I: Sampling, Surveys and Observational Data	
	4	**February 3**	Data Collection II: RCTs, Internal and External Validity	
	5	**February 8**	Data Collection III: Measurement and Missing Data	
	6	**February 10**	Introduction to Regression: Bivariate and Multivariate OLS	Quiz #1
	7	**February 15**	Introduction to Regression: R-Squared and Standard Errors	
	8	**February 17**	Introduction to Regression: Transforming Explanatory Variables	PS #1
	9	**February 22**	Causal Inference I: Assumptions and Biases; Potential Outcomes	
Causal Inference	10	**February 24**	Causal Inference II: Estimating Causal Effects with Experimental Data	
	11	**March 1**	Regression and Causality I: Basics	
	12	**March 3**	Regression and Causality II: Fixed Effects	Quiz #2
	13	**March 8**	RDD	
	14	**March 10**	IV	PS #2
			Spring Break	
	15	**March 22**	Matching	
	16	**March 24**	Differences-in-Differences (DID) and Synthetic Controls	
Prediction & ML	17	**March 29**	Case: Jamaica Path	
	18	**March 31**	Prediction I: Introduction to Prediction, Regression versus Classification	
	19	**April 5**	Prediction II: Lasso and Ridge Regression	
	20	**April 7**	Prediction III: Trees and Random Forest	Quiz #3
	21	**April 12**	Case: OSHA	
	22	**April 14**	Wrap Up and Conclusion	PS #3
		TBD	**Final Assessment**	

Detailed Course Plan

Class 1: Answering Questions with Data: Description and Prediction

- Predictive: JWHT: Ch. 1 and 2
- Descriptive: Shrider, Emily, Melissa Kollar, Frances Chen, and Jessica Semega. "Income and Poverty in the United States: 2020." U.S. Census Bureau, September 2021.

Class 2: Answering Questions with Data: Causality

- MM: Introduction
- AP: Ch. 1
- 2021 Nobel Prize article by Jörn-Steffen Pishcke

Class 3: Data Collection I: Sampling, Surveys and Observational Data

- Mercer, Andrew, Arnold Lau, and Courtney Kennedy. "For Weighting Online Opt-In Samples, What Matters Most? - Pew Research Center Methods, January 26, 2018. (p. 1 only)
- Shrider, Emily, Melissa Kollar, Frances Chen, and Jessica Semega. "Income and Poverty in the United States: 2020." U.S. Census Bureau, September 2021. (repeat from Lecture 1)
- SIPP User Guide: Ch. 1, 2.1, and 7

Class 4: Data Collection II: RCTs, Internal and External Validity

- MM: Ch. 1
- AP: Ch. 2
- Glennerster, Rachel, and Kudzai Takavarasha. Running Randomized Evaluations. Princeton University Press, 2013: Ch. 4.1, 4.2, and 6.1

Class 5: Data Collection III: Measurement and Missing Data

- TBD

Class 6: Introduction to Regression: Bivariate and Multivariate OLS

- W: Ch. 1.3, 2.1, 2.2, 2.5, 2.6 3.1-3.3, 4.1-4.5

Class 7: Introduction to Regression: R-Squared and Standard Errors

- W: Ch. 7.1-7.4

Class 8: Introduction to Regression: Transforming Explanatory Variables

- W: Ch. 17.1 6.2, 7.5,

Class 9: Causal Inference I: Assumptions and Biases; Potential Outcomes

Class 10: Causal Inference II: Estimating Causal Effects with Experimental Data

- Mosteller, M. (1995). The Tennessee study of class size in the early school grades, The Future of Children, Critical Issues for Children and Youth, Vol 5. Num. 2, Summer/Fall.

Class 11: Regression and Causality I: Basics

Class 12: Regression and Causality II: Fixed Effects

- AP: Ch. 5.1, SW: 10.3

Class 13: RDD

- AP: Ch. 6

Class 14: IV

- W: Ch. 15.1 - 15.3; AP: Ch. 4.1, 4.4

Class 15: Matching

- TBD

Class 16: Differences-in-Differences (DID) and Synthetic Controls

- Card, David, and Alan B. Krueger. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *The American Economic Review* 84, no. 4 (1994): 772.
- Card, David. "The Impact of the Mariel Boatlift on the Miami Labor Market." *ILR Review* 43, no. 2 (January 1, 1990): 245–57.
- MM: Ch. 5
- AP: Ch. 5.2
- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association*105, no. 490 (June 2010): 493–505.
- Abadie, Alberto. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." *Journal of Economic Literature* 59, no. 2 (June 1, 2021): 391–425.

Class 17: Case: Jamaica Path

- Jamaica Path Case

Class 18: Prediction I: Introduction to Prediction, Regression versus Classification

- TBD

Class 19: Prediction II: Lasso and Ridge Regression

- Mullainathan, Sendhil, and Jann Spiess. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31, no. 2 (2017): 87–106.
- JWHT: 2.1-2.2, 4.1, 4.4, 4.5
- JWHT: 5.1, 8.1

Class 20: Prediction III: Decision Trees and Random Forest

- JWHT: Ch. 8
- Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. "Human Decisions and Machine Predictions." *The Quarterly Journal of Economics* 133, no. 1 (2018): 237–93.

Class 21: Case: OSHA

- OSHA Case

Class 22: Wrap Up and Conclusion