Advanced Quantitative Methods I (API-209)

Harvard Kennedy School

Course Syllabus – Fall 2023

Aug 1 version

Key people:

Faculty:	**Dan Levy:** dan_levy@harvard.edu
Teaching Fellow:	**Dom Valentino:** dvalentino@g.harvard.edu
Course Assistants:	**Callan Corcoran:** callancorcoran@hks.harvard.edu
	Shreya Dubey: shreyadubey@hks.harvard.edu
	Angela Perez: angelaperez@hks.harvard.edu
Faculty Assistant:	**Victoria Barnum:** victoria_barnum@hks.harvard.edu

Course Description:

The goal of this course is to prepare you to analyze public policy issues using statistics. Key topics in the course are in the areas of probability theory, estimation, hypothesis testing, regression analysis, causal inference and prediction. While many students taking this class will have already taken courses in statistical inference and regression analysis, this course will place a much stronger emphasis than typical courses on conceptually understanding the underlying methods. Since the course is targeted to first-year students in the MPA-ID program, we will not shy away from using mathematical tools, but the emphasis of the course will be on the conceptual understanding and application of the tools rather than on the math or the mechanics behind the tools. So for example, when studying hypothesis testing, we will place a heavier emphasis on what the test is doing, when to use it and how to interpret its results, than on mathematical proofs underpinning the test or on mechanical repetitions of the calculations involved in conducting the test.

The ultimate goal is that by the end of this course you will be able to:

1. Conceptually understand the statistical methods studied in the course and be able to apply them to a wide range of public policy issues.
2. Interpret the results of statistical analyses and think critically about the potential issues that arise when trying to draw conclusions from such results.
3. Conduct statistical analyses using a statistical package called R.

Class Meetings:

Classes:	Tuesdays and Thursdays, 1:30 p.m. – 2:45 p.m. in L140
Review sessions:	Fridays, 9:00 a.m. - 10:15 a.m. in R306 and 10:30 a.m. - 11:45 a.m. in R306

Office Hours:

- Fridays, 2:00 PM – 5:00 PM, in person in R136 or virtually through Zoom. Please sign up in advance.
- Office hours for TF and CAs will be posted on the Canvas course website.

Prerequisites:

The main prerequisite for this course is an understanding of calculus at the level of a typical undergraduate calculus course. However, as far as mathematics goes, the most important prerequisite is a certain level of "mathematical sophistication," i.e., comfort in dealing with mathematical constructs and arguments. Experience with computer programming is helpful.

Target Students:

This course is required for first-year students in the MPA/ID program. Students not in the MPA/ID program will be admitted only with permission of the instructor. Familiarity with the concepts indicated in the background section of the schedule (last column of table below) will be assumed. Students looking for a more mathematical course should consider EC2110 and EC2115 (offered by the Economics Department). Students looking for a less mathematical course should consider API-201 and API-205 (both offered here at HKS). You might also want to look at a draft list of other stats courses you could take at Harvard (put together by students in the PhD in Health Policy program). If you have a basic background in probability and economics and are looking for a course that broadens your analytical toolkit, you might consider taking API-318, an elective I am also teaching this Fall.

Note: In general, I strive to welcome to the class non-MPA/ID students for whom API-209 is the best option. But unfortunately, I can only accept a limited number of students. Hence, unless you have compelling reasons to take this particular course, I suggest you consider taking another course.

Texts and Other Course Materials:

Key Recommended Textbook:

- *Introductory Econometrics: A Modern Approach*, by Jeffrey Wooldridge, South-Western, Fourth Edition, 2009. [W]. This textbook will be used mainly for regression analysis, and may be a useful book for API-210. This book is available for purchase at the Harvard Coop and other bookstores.

Background Textbooks:

- *Mathematical Statistics with Applications*, by Wackerly, Mendenhall and Scheaffer, Duxbury, Seventh Edition, 2008. [WMS]. This textbook is mainly for the first half of the course. Recommended if you do not already have a good mathematical statistics textbook.
- *Probability and Statistics*, by DeGroot and Schervish, Addison Wesley: This textbook is similar to WMS both in style and mathematical sophistication, but is organized slightly differently.
- *Mind on Statistics, by Utts and Heckard, Fifth Edition, 2015, Cengage.* This book is at a much lower mathematical level than WMS but provides more examples of the use of statistics in everyday life.

R: We will be using R as the statistical software for the course. Here are some resources that might be helpful: Summer Letter explaining use of R | Textbook | Cheat Sheet.

Handouts:

Handouts will be available for almost every class. The main goal of the handouts is to facilitate the process of taking notes so that you can fully engage in class. They are not meant to substitute for class attendance or for studying the assigned reading material. Handouts will contain blank spaces for you to fill in during class, usually in response to questions.

Course website:

Course materials will be posted on the Canvas course website for API-209. The link is here: https://canvas.harvard.edu/courses/125666.

Grading:

The main goal in this course is to help you learn concepts, develop skills, and ultimately change the way you think about the world. To achieve this, I expect you to exhibit the highest professional and ethical standards for every activity you undertake in the course.

The grading policy described below seeks to be consistent with these high expectations of you. I recognize that given the world that we live in, it might sometimes not be possible for you to do some of the work expected of you, and I have tried to build some flexibility into the grading policy to allow for this. If you feel you are in trouble, communicate with me or any member of our teaching team as early as you can. The single most important thing you can do is be honest and transparent.

The class grade will be based on the following criteria:

10% - Problem sets
15% - Class participation and engagement
15% - Final Exercise
25% - Midterm exam
35% - Final exam

Problem Sets (10%)

Problem sets will be assigned almost every week. They will give you hands-on experience with the analytic techniques introduced in class. You should plan to spend approximately 8-10 hours on each problem set. 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 problem sets submitted late. The lowest problem set grade will be dropped when calculating the average grade for the problem sets.

We abide by the Harvard Kennedy School Academic Code (available here) for all aspects of the course. In terms of problem sets, unless explicitly written otherwise, the norms are the following: You are free (and encouraged) to discuss problem sets with your classmates. However, you must hand in your own unique written work and code in all cases. Any copy/paste of another's work is plagiarism. In other words, you can work with your classmate(s), sitting side-by-side and going through the problem set question-by-question, but you must each type your own answers and your own code. Your answers may be similar but they must not be identical, or even identical'ish. 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. I take this issue very seriously. If you have questions about the degree of collaboration allowed or about any other aspect of the Academic Code, please come to see me.

Instructions for submitting problem sets:

- Turn them in electronically via the Canvas course page.
- Submit them by 10:00 am on the day they are due.
- Indicate on the cover page the names of the classmates you worked with.

Class participation and engagement (15%)

This component of the grade is meant to reward the things you do to contribute to your own learning and to the learning of your classmates. Below are some of the key vehicles for doing so in this course. For all these, both quantity and quality will count.

- *Engagement in class*: I strongly believe that student participation can substantially enrich the learning experience for both the students and the instructor. In this spirit, you are encouraged to ask questions and to share with the class any relevant insights you may have from your work experience or from previous exposure to the topic at hand. Effective class participation requires that you come prepared for class, and that you think about how your comment or question may enrich the learning experience of your classmates. I also ask that the questions and comments be brief and related to the topic being discussed. I will sometimes need to defer questions for a future class or office hours. A strong engagement in class also means attending class regularly and punctually, engaging actively in both classwide and small group activities, and in general contributing to a positive learning atmosphere in the classroom.
- *Engagement outside of class*: A key goal of the course is to help you use statistical tools in the real world. Hence, I encourage you to make connections between the tools we study in class and anything relevant that you encounter outside of the classroom using the Slack workspace for our course. We will also make Twitter (#api209) an option for course participation. Examples of such postings include references to news articles, a report that you worked on before coming to the Kennedy School, reflections on connections with things you have studied in your other core courses, or perhaps some event or controversy in which statistics is playing a key role. You can contribute both by starting a discussion and by moving it forward. Finally, you are encouraged to challenge one another's thinking and to debate conflicting perspectives.

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

Final Exercise (15%)

The final exercise will require applying some of the statistical tools learned in class using a real data set. More details will be provided later in the course.

Midterm Exam (25%):

Please note that the midterm exam will be held on **Friday**, Oct 6, from 8:00 AM to 10:00 AM. It will be a two-stage exam (see details below).

Final Exam (35%):

Please note that the final exam will be held on **Monday**, Dec 11, from 9:00 AM to 12:00 PM. It will be a two-stage exam (see details below).

All students are expected to take the exams on the assigned days. We will adhere to the Registrar's policy regarding rescheduling of exams (i.e., only to be done in case of documented health-related or personal emergencies).

Two-stage exams: The midterm and final exams will be two-stage exams. During Stage 1, you will be asked to complete the exam individually. After Stage 1, the exams will be collected and you will be given a second exam that will contain a subset of the questions from the original exam. During Stage 2, you will be asked to work with a group, reach consensus answers, and submit one copy of the exam for the whole group. If your group grade is higher than your individual grade for that subset of questions, your grade for those questions will be 90% of your Stage 1 score plus 10% of your Stage 2 score. If your Stage 2 grade is lower

than your Stage 1 grade, we will not incorporate Stage 2 into your score and your exam grade will simply be your Stage 1 grade.
The main reason we conduct a second stage of the exam is to allow you to learn more during the exam. Traditional exams tend to be summative rather than formative, and two-stage exams represent an opportunity to redress this imbalance. The process of discussing your answers with your teammates is a significant learning opportunity and supports the kind of collaborative learning that we encourage.

Regrade Policy
Requests for reconsideration of grades on exams are not encouraged, and will be accepted only in writing, with a clear statement of what has been incorrectly graded, and within one week of receiving your graded exam. When a request regrade is submitted, the entire exam will be regraded.

All course activities, including class meetings, problem sets, and exams are subject to the HKS Academic Code and Code of Conduct.

Letter Grades
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 for two purposes. First, to provide you with the option of reviewing the class so you can clarify or deepen your understanding of a particular concept. Second, to help me improve my teaching. The recordings will be kept in a protected page that is accessible to you only via the course site. As a member of our learning community, and to support honest exchange, risk- taking and vigorous debate in class, you are expected to never make any recordings available outside of our learning community. If you are uncomfortable with classes being recorded, please come and speak with me.

Use of Data: Data will be collected in various forms in this course. Some forms of data collection will be obvious to you (such as when responding to a question on a survey) but others might not be (such as someone from our teaching team recording class participation or the Canvas course website system recording activity while you are logged in). Whatever the form of data collection, I pledge to use the data to help improve my teaching and ultimately your learning. This includes using your responses to online quizzes to tailor a class better to the backgrounds and learning needs of students in the class, conducting research about the effectiveness of a particular teaching approach, etc. I also pledge to keep your data confidential so that it can only be used for the purposes of improving teaching and learning or to help you and other students connect with future professional opportunities.

Teachly: I strive to teach in a classroom in which everyone feels welcome to participate in class. To help me get to know you better, keep track of classroom participation patterns and create an inclusive classroom environment, I will be using a tool developed by faculty and students at the Harvard Kennedy School called Teachly. For more details, please see the list of Frequently Asked Questions (FAQs) about Teachly.

Use of Generative AI in this course: Please read these guidelines for using generative AI in this course.

Tentative Schedule:

Tentative schedule is below. Adjustments may need to be made and will be communicated as soon as they are available. W refers to readings from Wooldridge. For a refresher on the concepts indicated in the background section of the schedule (last column of table below), see Deb Hughes Hallett's excellent handouts (click on "Math Camp" link here: https://canvas.harvard.edu/courses/120357).

READINGS

Orr, Larry L. 1999. Social Experiments: Evaluating Public Programs with Experimental Methods (Thousand Oaks, CA: Sage Publications), pp. 103-115.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002." Statistical Conclusion Validity and Internal Validity," and "Construct Validity and External Validity." Chapters 1, 2, and 3 in *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Boston: Houghton Mifflin). pp. 33-47, 53-55, 83-96.

Stokey and Zeckhauser, A Primer for Policy Analysis, *Chapter 12*

The Economist, "Signifying Nothing?" January 31, 2004

The Economist, "Vaccine Economics: A Bigger Dose", August 8, 2020.

CASES

HKS Case 2011.0: Providing Pensions for the Poor: Targeting Cash Transfers for the Elderly in Mexico

HBS Case 9-68-019: Improving Worker Safety in the Era of Machine Learning, Case A.

HKS Case 2166.0: Evaluating the Impact of Solar Lamps in Uganda.

Date	Class	General Topic	Specific Topics (Tentative List)	Problem Set Due	Readings	Background Material
29-Aug	1	COURSE OVERVIEW	• Syllabus Overview •Key Concepts in this Course		W [Appendix C.1]	Basic concepts in descriptive statistics: Mean, Variance, Std Deviation, Mode, Covariance, Correlation, Conditional Mean [W: Appendix B.3]
31-Aug	2	PROBABILITY: Probability, Conditional Probability, Bayes' Rule	•Key Probability Concepts •Bayes Rule •Application: Should women in their 40s have regular mammograms		NYT article about mammograms	Key concepts in probability: Axioms of Probability [W: App B.1]; •Random Variables and Probability Distributions (pdf, cdf) • Expected Value and Variance of Random Variables [WMS: 3.1-3.3, 4.1-4.3; W: B1, B3], Conditional Expectation
5-Sep	3	PROBABILITY: Decision Analysis	•The Value of Perfect Information •Allowing for Risk Aversion •General Framework	PS 1	Stokey and Zeckhauser (Ch 12), The Economist (2020)	
7-Sep	4	PROBABILITY: Mexico's Pension System	• Targeting performance •Assessing the Options •Other Issues		Case Study about Mexico's Pension System	
12-Sep	5	ESTIMATION: Sampling Distribution	•Estimators and Sampling Fluctuations •Sampling Distribution	PS2	W [Appendix: C.2, C.3]	Normal Distribution [ALZ: 3.7.1]; Binomial Distribution [ALZ: 3.6.2]
14-Sep	6	ESTIMATION: Estimation	•Point Estimators •Confidence Intervals		W [Appendix: C.2, C.3, C.5]	
19-Sep	7	HYPOTHESIS TESTING: Conceptual Framework	•Intuition Behind Hypothesis Testing •Key Ideas behind Hypothesis Testing •Framework for Conducting hypothesis testing	PS3	W [Appendix: C.6]	
21-Sep	8	HYPOTHESIS TESTING: Hypothesis tests involving the normal and t distributions	•Hypothesis Tests Involving the Normal distribution •Hypothesis tests involving the t distribution•Hypothesis tests in the context of regression analysis		W [Appendix: C.6]	Main distributions used for hypothesis testing: Normal, t, Chi-Square, and F [W: B.5]
26-Sep	9	HYPOTHESIS TESTING: Chi-Square Tests	•Chi square tests	PS4	W[Appendix: B.5]	Main distributions used for hypothesis testing: Normal, t, Chi-Square, and F [W: B.5]
28-Sep	10	SAMPLING: Statistical Power	•Factors that affect the power of the design •Minimum detectable differences		Orr (1999)	
3-Oct	11	SAMPLING: Data Collection and Sampling	•Challenges with Data Collection •Types of Sampling	PS5		
5-Oct	12	HYPOTHESIS TESTING: Practical significance and Midterm Review	•Practical Significance •Type I and Type II Errors •Practice Questions on Hypothesis Testing		The Economist (2004); W [C.6]	
6-Oct		**MIDTERM EXAM (8:00 AM - 10:00 AM)**				
10-Oct	13	REGRESSION: Introduction and Causal Inferenc	• The Counterfactual •Average Treatment Effects	PS6		
12-Oct	14	REGRESSION: Bivariate Regression I	• Population Regression Function •Sample Regression Function •Ordinary Least Squares		W [1.3, 2.1, 2.2, 2.5, 2.6]	

17-Oct	15	REGRESSION: Bivariate Regression II and Randomized Controlled Trials (RCTs)	•Introduction to Randomized Controlled Trials •Bivariate Regression with a Dummy Variable			
19-Oct	16	REGRESSION: Multiple Regression - Estimation I	• The Model with Two Explanatory Variables •The Model with k Explanatory Variables •The Expected Value of the OLS Estimators	PS7	W [3.1-3.3]	
24-Oct	17	REGRESSION: Omitted Variable Bias	•Omitted Variable Bias - Theory •Omitted Variable Bias - Examples	PS8	W [3.3]	
26-Oct	18	REGRESSION: Multiple Regression III - Goodness of Fit and Standard Errors	•Goodness of fit •Standard Errors of OLS Estimators		W [2.3, 3.4, 3.5, 6.3]	
31-Oct	19	REGRESSION: Multiple Regression Analysis - Inference	• Sampling Distribution of OLS Estimators • Testing hypotheses about a single population parameter: The t test • Testing multiple linear restrictions: The F	PS9	W [4.1-4.5]	
2-Nov	20	REGRESSION: Non-Linear Functional Forms	• Logs		W [7.1-7.4]; W [6.2, A.4]	
7-Nov	21	PREDICTION: The Basics of Prediction for Continuous Variables	•Predicting a Continuous Variable •The Problem of Overfitting •The Basics of Machine Learning	PS10		
9-Nov	22	PREDICTION: The Basics of Prediction for Binary Variables	•Predicting a binary (dummy) variable •OSHA Problem Set			
14-Nov	23	PREDICTION: Prediction III - The OSHA Case	•Introduction to OSHA Case •Assessing Performance of Predictive Algorithms •Prediction and Public Policy	PS11	Case Study about OSHA	
16-Nov	24	OVERALL: Assessing the Effects of an Intervention - The Case of IDinsight	•Evaluation design issues •Interpreting Results: The Big Picture •Other Issues	PS12	Case Study about ID Insight	
21-Nov	25	OVERALL: Aggregating Evidence	•Validity •Aggregating evidence	PS13	Shadish et al. (2004)	
23-Nov		**NO CLASS - Thanksgiving Holiday**				
28-Nov	26	OVERALL: Looking Back and Looking Ahead	•Airport Concepts •Roadmap of course •Five Meta Lessons	PS14		
30-Nov	27	FINAL CLASS: Final Exercise Presentations		Final Exercise		
11-Dec	-	**FINAL EXAM (9:00 AM - 12:00 PM)**				
W refers to readings from Wooldridge.						