API-201Z
Quantitative Analysis & Empirical Methods
Fall 2021

Lectures Tues/Thurs 1:30-2:45 Starr Auditorium
Review Sections: Fridays noon-1:15pm TBA

Maya Sen, Harvard Kennedy School
Email: maya_sen@hks.harvard.edu
Office Hours: Fridays 1-3pm ET (zoom or in person)
Office: Taubman 356

TF: Summer Rak, Health Policy
Email: srak@g.harvard.edu
Office Hours: TBA

Overview of API 201

API-201 provides an introduction to descriptive statistics, probability theory, statistical inference, and decision analysis, in addition to qualitative methods. The course content is divided into five broad units: (1) Descriptive Statistics, (2) Probability, (3) Statistical Inference, and (4) Regression (if time allows). Section Z also allows students to become familiar with statistical programming to analyze quantitative data.

By the end of this course you will be able to:

1. Take a data set and a broad descriptive policy question (such as "What has happened to incomes across the world in the last 30 years?"), figure out what statistical analysis would be most appropriate to answer the question, conduct such an analysis, and present the findings in a way that is accessible to policymakers.

2. Identify real world policy situations in which the tools of probability can be used, identify which tools are most relevant, and critically consume policy analysis in which probability is used.

3. Critically consume policy studies/papers/reports in which statistical analysis is used.

Non-MPP graduate students may enroll only with the permission of the instructor. This course is not open to undergraduates.

How does the Z Section differ from the non-Z Sections? Section Z covers similar material to the non-Z Sections but assumes a willingness to learn to use statistical software, which can be time consuming, and a greater comfort level with mathematics (but not with statistics). It will proceed faster, allowing more time for applications and slightly more advanced topics. The exams and problem sets will be more difficult than those of the other sections. The Section Z problem sets and final exercise will use R for data analysis. Lastly, Section Z's final exercise is open ended, allowing student groups to write on a topic of their choice.

Please note that in order to enroll in Section Z of API-202 in the Spring, either an A, A-, or B+ will be required in API-201Z. (Either an A or A- will be required in one of the other API 201 sections.) In the absence of these grades, explicit permission from API 202-Z instructor is required. Students are encouraged to talk with the API-201 course head to determine which section is the best fit. First-year MPP students with prior coursework in statistics are strongly encouraged to place out of 201 entirely by taking the exemption exam.

Students may switch from the Z section to another section (but not from the non-Z sections into the Z section). Students who switch by September 30 are expected to take the midterm exam in one of the other API 201 sections. Students who switch between September 30 and October 14 will take the midterm in the Z section (on October 7), and this midterm grade will count as the midterm grade in the other API 201 sections. No switches will be allowed after October 14. The following table summarizes the three scenarios:

If you switch to non-Z:	**You will take:**	**Note:**
Before September 30 at 12pm ET	A-B-C midterm	
Between September 30 at 12pm ET & October 14 at 12pm ET	Z midterm	Z midterm will count as midterm grade
After October 14 at 12pm ET	Z midterm	You must remain in Z for the semester

Section swaps are managed by the MPP Program Office. To switch sections, please email MPP_Program@hks.harvard.edu by the relevant deadline. When switching from Z to one of the non-Z Sections, you will be assigned to your original API 201 section.

Logistics & Requirements

Prerequisites. There are no prerequisites for this course. Some of the problem sets will be time consuming and be challenging, particularly if you have never worked with statistics or statistical software before. You must be willing to commit to the time necessary to do the work.

COVID Policies. We will strictly follow all Harvard/HKS COVID-19 guidelines. So far, we anticipate the class will meet in person this year, but that may change suddenly. Please

monitor any HKS correspondence. Please respect social distancing/masking directives from the university and stay home if you are feeling unwell. We hope to keep everyone healthy and safe!

Class Attendance. We encourage you to do your best to attend every class. If you need to miss a class for whatever reason, please obtain missed notes and course announcements from another student; if you do miss class, there is no need to email the instructors, as we do not take attendance. All lecture notes will be posted on the course website after lecture.

Review Sessions and & R Drop-in Help Sessions. Friday review sections will be designed to review statistical concepts, cover programming techniques, and clarify any material. Attendance is optional but strongly encouraged.

In addition, there will be an R drop-in help session. Students can come with R specific questions and the TF will also prepare some materials to work through.

Laptop and Cell Phone Policy. Responsible use of laptops is permitted in class. Please be considerate in your use of devices.

Suggested Readings. The study of statistics is rapidly moving to online platforms, many of which provide free and excellent overviews of all of the material we cover in this class. In addition, all course lecture notes will be posted to the course website. However, we have some *recommended* (not required) resources:

- *Introduction to the Practice of Statistics*. Moore, McCabe and Craig. W.H. Freeman (Any Edition) (referred to as "MMC").
- *Open Intro Statistics*, 4th edition. Diez, Barr and Çetinkaya-Rundel. The electronic version of the textbook is available for free at http://www.openintro.org/. Printed copies may be purchased online for about $15.

In addition, we recommend the purchase of:

- *The Cartoon Guide to Statistics*. Gonick and Smith. Harper Perennial. ISBN: 0062731025.

The *Cartoon Guide* provides an intuitive, easy-to-follow overview of every single topic we cover in this course. It is strongly recommended, particularly for those new to statistics.

We will have the following readings for case studies (posted on the course website):

- HKS Case Study "Providing Pensions for the Poor: Targeting Cash Transfers for the Elderly in Mexico."
- Baicker et al. "The Oregon Experiment: Effects of Medicaid on Clinical Outcomes." The New England Journal of Medicine. May 2, 2013.

Statistical Software Packages. Section Z will use R, a free, open-source statistical software language, to conduct statistical analyses throughout the course. Problem sets will contain exercises designed introduce you to the basics of R. We teach R because it is open-source and because it has become the default language of statistical analysis.

R can be downloaded for free at https://www.r-project.org. R Studio, a user-friendly interface to R, can be downloaded at https://www.rstudio.com.

Grading

Your grade in this class will be composed of

- 20% - Problem sets
- 10% - Class participation and engagement – see below
- 15% - Final Exercise
- 15% - Midterm #1
- 15% - Midterm #2
- 25% - Final exam

Final letter grades will be determined on a curve that strictly uses the HKS's Recommended Grade Distribution (available at on KNET at https://knet.hks.harvard.edu/DPSA/Registrar/Exams-Grading/Pages/Credit-Grades-Grading.aspx).

Problem Sets. The best way to learn statistics is to practice. We will have 8 problem sets, which are due at the beginning of class on the date they are due *on Canvas*. There will be no credit for late assignments. The lowest problem set grade will be automatically dropped in the calculation of your overall problem set grade. You should plan to spend approximately 6-8 hours on each problem set. Each problem set will be graded as follows: ✓+ (only a few minor mistakes if any), ✓ (mostly correct, good faith attempt for all parts), or ✓− (substantial parts not attempted or misguided). There is no "Problem Set 0."

All students must comply with the HKS Academic Code. You are encouraged to work on problem sets in groups. You must note the names of your group members on your problem set. In addition, you submit **your own** solutions. **It is not acceptable to work on one electronic document as a group and submit identical, or nearly identical versions.** Examples of submissions that are not in accordance with the HKS academic code include document copies or excerpts that have been "cut and pasted" from other group members' assignments, printouts of substantially identical R tables or graphs, R code that has been cut and pasted, and copies of solutions from previous years.

Class participation and engagement. Student participation greatly enriches the learning experience. However, class participation is difficult to assess objectively, penalizes some students, and is subject to potential bias. In addition, the pandemic has created substantial difficulties for members of our community, including those with caregiving responsibilities.

Class participation will therefore only help, rather than hurt, students, and will be used only to bump up students on the grade distribution who are between two grade cut-offs. For example, a student with strong participation who is otherwise on the bubble between a B+ and an A- will be bumped up to an A-, while a student who lacks strong participation will remain at a B+. Upward adjustments will be made after the curve is calculated, ensuring that no students are adversely affected.

Final Exercise. The final exercise is a group-based project that engages all of the skills acquired in the class by asking you to take the statistical skills learned and apply them to the analysis of some original, interesting, or otherwise professionally relevant data set.

Groups should be 3–4 students total. The components of the final exercise include (1) a short write-up of your analyses and results, and (2) a short power-point presentation of the key points. **All students in the group will be assigned the same final grade. Consider this carefully when selecting your group and dividing up and coordinating on tasks.**

The final exercise is due **December 2, 2021**. More details and helpful links are available on the course website. Students are encouraged to start thinking very early about their topic, group composition, and potential data.

Exams. There will be two midterms and a final exam.

Midterms will be on ***October 7***, and ***November 9.*** These dates are different than the non-Z section midterm dates. Midterms will only be rescheduled for individual students in the case of documented health-related or personal emergencies.

The final exam has been scheduled by the HKS Registrar on ***December 9, from 9am to noon***. The final exam cannot be rescheduled. We have no discretion on the time or date of the final exam. The final exam is cumulative and covers material from the entire semester.

Both midterms and the final exam will be open book/note.

For exam preparation, we have posted a wide variety of practice problems and old exams on Canvas. We strongly encourage you to work through as many practice problems as you can, as this will be far more effective than re-watching lectures or reading the textbook. You will learn the most with practice.

Exam Regrade Policy. Requests for regrades will be accepted only in writing, with a clear statement of what has been mis-graded, and why, and within one week of return of your graded exam. We reserve the right to review all of your answers. This means that you might end up having more points deducted than if you had not requested a regrade.

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

Tentative Schedule

This schedule is subject to change, depending on how far we get in each class and how much discussion the topics generate.

September 2 (Lecture 1): Introduction and Class Overview

- A introduction to R Session will be lead by Teaching Fellow during regular review session hours

September 7 (Lecture 2): Causal vs Descriptive vs Predictive Questions

September 9 (Lecture 3): Sample Statistics

- Moore, McCabe, Craig 8th Ed. (MMC) 1.3, 2.3, 2.7 (MMC is recommended only)
- Problem Set #1 Due on September 10 at 5pm ET

September 14 (Lecture 4): Introduction to Probability

- MMC 4.1, 4.2, 4.5

September 16 (Lecture 5): Conditional probability and Independence

- MMC 4.5
- Problem Set #2 Due at 1:30pm ET

September 21 (Lecture 6): Bayes' Rule

September 23 (Lecture 7): Application: Public Pensions in Mexico

- Case study on Public Pensions in Mexico
- Problem Set #3 Due at 1:30pm ET

September 28 (Lecture 8): Introduction to Random Variables

- MMC 4.3 (through page 255), 4.4, 5.2

September 30 (Lecture 9): Continuous Random Variables and the Normal Distribution

- MMC 1.3, 4.3 (from page 255 onward)
- Problem Set #4 Due at 1:30pm ET

October 5 (Lecture 10): Using Random Variables with Decision Theory

October 7 Midterm

- No Friday section this week. Midterm review session will be held earlier in week.

October 12 (Lecture 11): Sampling

- Readings TBA

October 14 (Lecture 12): Introduction to Inference and the Central Limit Theorem

- MMC 3.3, 5.1
- Problem Set #5 Due at 1:30pm ET

October 19 (Lecture 13): Introduction to Hypothesis Testing

- MMC 6.1, 6.2, 7.1

October 21 (Lecture 14): Comparing Groups

- MMC 7.2, 8.2, 6.3
- Problem Set #6 Due at 1:30pm ET

October 26 (Lecture 15): Proportions and Paired ata

- MMC 8.1

October 28 (Lecture 16): Confidence Intervals and Power

- MMC 7.1, 8.2 (review)
- Problem Set #7 Due at 1:30pm ET

November 2 (Lecture 17): Application: Oregon Health Care Experiment

- Oregon Health Care experiment case study

November 4 (Lecture 18): Chi-Square Tests

- MMC 9.1, 9.2, 9.3

November 9 (Lecture 19): Midterm

- No Friday section this week. Midterm review session will be held earlier in week.

November 11: NO CLASS VETERANS DAY

November 16 (Lecture 20): Intro to Regression Analysis

- MMC 2.3, 2.4, 10.1 (p. 563–580)

November 18: Qualitative Lecture I

- TBA
- Problem Set #8 Due at 1:30pm ET

November 23: Qualitative Lecture II

- TBA

November 25: THANKSGIVING HOLIDAY RECESS – NO CLASS

December 30: Qualitative Case Study

- TBA

December 2: Final Exercises Due/In-Class Final Presentations

TBD: Final Exam Review

December 9: Final Exam (9am-noon)