Overview of API 201

API-201 introduces a range of quantitative tools that are commonly used to inform public policy issues. It provides an introduction to descriptive statistics, probability theory, statistical inference, and decision analysis.

Our goal is that 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 in the U.S. 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.

4. Use the decision analysis framework as one tool to make personal and professional decisions, and to think about policy problems.

The course content is divided into four broad units: (1) Descriptive Statistics, (2) Probability, (3) Statistical Inference, and (4) Regression. Section Z also provides you with the opportunity to become familiar with statistical programming to analyze and present quantitative data.

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

**How does the Z Section differ from the A-B-C Sections?** Section Z follows a similar syllabus to Sections A-B-C and covers similar material but assumes a greater comfort level with mathematics and a stronger interest in learning to use statistical software. It will proceed faster, allowing more time for applications, in-depth discussions, and a number of advanced topics. The exams and problem sets will be more difficult than those of the other sections. Section Z problem sets will use Stata (or R) for data analysis, as opposed to Excel. 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 or A- will be required in either Sections A, B, C, or Z of API-201. 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 an exemption exam during orientation week.

Students may switch from the Z section to the A-B-C sections (but not into the Z section). Students who switch by October 8 at 12:00pm are expected to take the midterm exam in the A, B, or C sections. Students who switch between October 8 at 12:00pm and October 18 at 12:00pm are expected to take the midterm in the Z section (on October 11), and this midterm grade will count as the midterm grade in the A-B-C sections. No switches will be allowed after October 18 at 12:00pm. The following table summarizes the three scenarios:

<table>
<thead>
<tr>
<th>If you switch to A-B-C:</th>
<th>You will take:</th>
<th>Note:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before 10/8 at 12pm</td>
<td>A-B-C midterm</td>
<td></td>
</tr>
<tr>
<td>Between 10/8 at 12pm &amp; 10/18 at 12pm</td>
<td>Z midterm</td>
<td>Z midterm will count as midterm grade</td>
</tr>
<tr>
<td>After 10/18 at 12pm</td>
<td>Z midterm</td>
<td>You must remain in Z for the semester</td>
</tr>
</tbody>
</table>

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 the A-B-C sections, you will be assigned to your original A-B-C section.
Logistics & Requirements

Prerequisites. There are no prerequisites for this course. Some of the problem sets will be time consuming and be quite challenging, particularly if you have never worked with statistics or statistical software before. Give yourself ample time and consider working in a group.

Class Attendance. We expect that you will arrive on time to class and do your best to attend every class. If you need to miss a class due to an emergency, it is your responsibility to obtain missed notes and course announcements from another student; if you miss class, do not email the instructors, as we do not take attendance. All lecture notes will be posted on the course website after lecture.

Review Sections. Friday review sections will be designed to review statistical concepts, cover programming techniques, and clarify any material. A regular review section will be split in half: The first half (11:45am-12:20pm) is designated for statistical concepts, and the second half (12:25-1:00pm) is for covering programming. Students are welcome to attend one or both parts. Attendance is optional but strongly encouraged.

Laptop and Cell Phone Policy. Responsible use of laptops is permitted in class. Please be considerate and silence and put away phones during class.

Suggested Textbooks. 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, along with API 201 A-B-C, there is one recommended (not required) textbook:


  This text can be rented, bought used, or bought new, all online. A newer edition (9th Edition) has been recently released, but the 8th Edition as well as older editions are acceptable. There is no need for Z Section students to purchase any of the online supplementary subscriptions.

  In addition, we recommend the purchase of:


  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.
**Statistical Software Packages.** Section Z will use Stata (and R, as an option) to conduct statistical analyses throughout the course. Problem sets will contain exercises designed to get you to practice the basics of Stata (or R). Stata will be officially supported by the teaching fellow and course assistants, although both the teaching fellow and head course instructor can answer questions about the use of R.

Other languages—for example, Python or MATLAB—are widely used for research-intensive applications, such as programming your own estimators. However, we will use Stata (with R as optional) in this course because it is currently the default language for statistical analysis, is more accessible for those without a formal programming background, allows you to quickly estimate and compare regression models, and is used in many courses at HKS. Stata will be used in API 202, including API 202-Z.

All HKS lab computers have Stata installed. Students may also visit the Stata website directly to purchase the software on their laptop; there is a discount for using an HKS email to make the purchase. Details can be found here: [https://www.stata.com/order/new/edu/gradplans/student-pricing](https://www.stata.com/order/new/edu/gradplans/student-pricing). R is open source and can be downloaded for free at [https://www.r-project.org](https://www.r-project.org). R Studio, a user-friendly interface to R, can be downloaded at [https://www.rstudio.com](https://www.rstudio.com). Harvard’s Institute for Quantitative Social Sciences holds regular workshops to introduce Stata and R. More details can be found here: [https://dss.iq.harvard.edu/workshop-registration](https://dss.iq.harvard.edu/workshop-registration).

*Students concerned about or unable to invest the time and effort needed use a statistical software language (such as Stata or R) should opt into the A-B-C Sections instead, which use Excel.*

**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](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 10 problem sets, which are due at the beginning of class on the date they are due in hard copy. 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), ✓ (good faith attempt for all parts), or ✓ − (substantial parts not attempted or misguided). There is no “Problem Set 0,” as in the A-B-C sections.

All students must comply with the HKS Academic Code. You are encouraged to work on problem sets in groups. You must write the names of your group members on your problem set. In addition, you submit your own hand- or type-written 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 photocopies or excerpts that have been “cut and pasted” from other group members’ assignments, printouts of substantially identical Stata or R tables or graphs, Stata or 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 and subject to potential bias. Class participation will therefore only help, rather than hurt, students, and will be used primarily to bump up students 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+.

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.

The final exercise is due in class on Tuesday, December 4. More details and helpful links are available on the course website. Students are encouraged to start thinking early about their topic, group composition, and potential data.

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

Midterms will be given in class on Thursday, October 11, and Thursday, November 15. These dates are different than the 201 A-B-C midterm dates. In addition, due to circumstances outside of our control, please note that the first midterm on October 11 is scheduled for the day prior to the final assignment for MLD-220M (negotiations module). 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 **Tuesday, December 18, from 9am to noon**. The final exam cannot be rescheduled. We have no discretion to change 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 are closed book/note. A formula sheet will be provided. Calculators may be used, but statistical functions on them may not. Calculators that allow text storage are not permitted. Smartphones are not permitted. Students who arrive late will not be given additional time.

**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, and you might end up having more points deducted than if you hadn’t 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 6 (Lecture 1):** Introduction and Class Overview

- A separate Stata and R Setup Session led by the Teaching Fellow will be held in the afternoon, 4-6pm in room W332. Recommended for students relatively new to either program (or both). Details will be posted on Canvas.

**September 11 (Lecture 2):** Descriptive Statistics

- Moore, McCabe, Craig 8th Ed. (MMC) 1.3, 2.3, 2.7

**September 13 (Lecture 3):** Probability I

- MMC 4.1, 4.2, 4.5
- Problem Set #1 Due

**September 18 (Lecture 4):** Probability II

- MMC 4.5
September 20 (Lecture 5): Bayes’ Rule
- Article on monograms
- Problem Set #2 Due

September 25 (Lecture 6): Application: Public Pensions in Mexico
- Case study on Public Pensions in Mexico

September 27 (Lecture 7): Discrete Distributions
- MMC 4.3 (through page 255), 4.4, 5.2
- Problem Set #3 Due

October 2 (Lecture 8): Continuous Distributions
- MMC 1.3, 4.3 (from page 255 onward)

October 4 (Lecture 9): Introduction to Decision Analysis
- Case study
- Problem Set #4 Due

October 9 (Lecture 10): Sampling
- Readings TBA

October 11 Midterm
- No Friday section this week. Midterm review session will be held earlier in week.

October 16 (Lecture 11): Introduction to Inference and the Central Limit Theorem
- MMC 3.3, 5.1

October 18 (Lecture 12): Inference with Means I
- MMC 6.1, 6.2, 7.1
- Problem Set #5 Due

October 23 (Lecture 13): Inference for Means II
- MMC 7.2, 8.2, 6.3
October 25 (Lecture 14):  Inference for Proportions I
- MMC 8.1
- Problem Set #6 Due

October 30 (Lecture 15):  Inference for Proportions II
- MMC 8.2, 6.4

November 1 (Lecture 16):  Paired Data
- MMC 7.1, 8.2 (review)
- Problem Set #7 Due

November 6 (Lecture 17):  One-ANOVA
- MMC 12.1, 12.2

November 8 (Lecture 18):  Application: Oregon Health Care Experiment
- Oregon Health Care experiment case study
- Problem Set #8 Due

November 13 (Lecture 19):  Chi-Square Tests
- MMC 9.1, 9.2, 9.3

November 15 (Lecture 19):  Midterm
- No Friday section this week. Midterm review session will be held earlier in week.

November 20 (Lecture 20):  Regression Analysis I
- MMC 2.3, 2.4, 10.1 (p. 563–580)
- Problem Set #9 Due

November 21-23:  THANKSGIVING HOLIDAY RECESS – NO CLASS

November 27 (Lecture 21):  Regression Analysis II
- MMC 11.1, 11.2
- Final Exercise Executive Summaries Due
November 29 (Lecture 22):  Regression Analysis III

- MMC 6.3
- Problem Set #10 Due

December 4:  Final Exercises Due/In-Class Final Presentations

December 6:  Final Exam Review

December 18:  Final Exam (9am - noon)