Autumn 2022: PPHA 31002 Statistics for Data Analysis I

Dr. Eyal Frank	Dr. Sheng-Hao Lo	Dr. Daniel Moskowitz
eyalfrank@uchicago.edu	shenghaolo@uchicago.edu	dmoskowitz@uchicago.edu

Course Overview:

This is the first quarter of the statistics sequence at the Harris School. This course aims to provide students with a basic understanding of statistical analysis for policy research. We recognize that Harris students come from a variety of academic backgrounds and real-world experiences. This course makes no assumptions about prior knowledge of statistics or data science, apart from basic mathematics and coding skills introduced in Math Camp. Students with more advanced backgrounds in statistics or econometrics may wish to instead consider PPHA 31202, Advanced Statistics for Data Analysis I.

Course Objectives:

- ☑ Become an adept and responsible consumer of quantitative data analysis.
- ☑ Understand how to *do* basic quantitative analysis in your own policy research.

To achieve these goals, we are going to help you develop rigorous statistical thinking and quantitative reasoning skills that will allow you to make appropriate inferences from data. We will develop a powerful toolkit for conducting quantitative analysis, and that toolkit is going to serve as a foundation for even more sophisticated tools that you will build upon in Stats II and your other Harris courses.

Health & Safety Considerations:

We are all members of the same community of learning, and we derive considerable benefits from our interactions with one another inside the classroom. However, in the context of an ongoing pandemic, our interactions with one another are not without risk. To mitigate this risk, students are expected to abide by all UChicago health policies, including self-monitoring for symptoms and self-isolating when required to do so.

Course Format & Materials:

It is imperative that you keep up with the course material; the nature of the quarter system—with only nine weeks of instruction—means even a short lapse in attention to the course can make it extremely difficult to catch up with the material. Note that we do not enforce attendance or reading: *you, as an adult, are responsible for your education*. We are certainly here to help and ensure that you have the resources you need to succeed, but ultimately the onus is on you to figure out how you learn best and then implement those practices. We provide multiple resources and modes of learning, because we know that there is enormous variation in how students learn successfully. Each week students should attend, view, or read the following:

(1) A **twice-weekly lecture** at the Keller Center on Tuesdays and Thursdays. These lectures and the accompanying slides (available on Canvas) will be the main conceptual resource for completing assignments and studying for assessments. These sessions will also include the demonstration of statistical concepts through practice problems, coding concepts through occasional R tutorials, and ample time for questions. *Note: to ensure a high-quality and safe experience compliant with public health rules, you are only permitted to attend the course section to which you are assigned unless you have written permission from your instructor to attend a different section*.

(2) A weekly **recitation with a teaching assistant**. In these recitations, TAs provide instruction that is helpful in providing you with the coding and practical knowledge required to complete homework assignments. Recitations also provide you with another opportunity to ask questions about the material. *Note: as with the session with your instructor,* you *may only attend the*

recitation section to which you are assigned. A recording of the remote TA recitation section will be posted to Canvas each week if you are unable to attend your assigned session.

(3) A **strongly recommended weekly reading assignment**. For most weeks, we will post a recommended reading to Canvas, as many students learn better with reading material that accompanies the lecture. While these reading assignments are not per se required, they are strongly recommended, especially for any student who is not exceling in the course. In addition to the strongly recommended readings posted to Canvas, we also provide supplemental reading suggestions from two textbooks: David Salsburg, *The Lady Tasting Tea* and Charles Wheelan, *Naked Statistics: Stripping the Dread from Data*. These resources can be very useful for any topics for which you think you need additional practice. We will provide a suggested reading list by topic for the supplemental textbooks on Canvas.

The course is divided into eight units:

(1) Introduction
(2) Probability Theory
(3) Random Variables
(4) Summary Statistics & Exploratory Data Analysis
(5) Regression for Prediction & Description
(6) Hypothesis Testing
(7) Simulations
(8) Experiments

Course Assignments & Assessment:

The composition of your overall grade is as follows:

Four Homework Assignments	56%
Weekly Mini-Assignments	5%
Two Mid-Quarter Quizzes	4%
Final Exam	35%

(Tentative) Assignment/Assessment Schedule:

Homework #1	10/10
Quiz #1	10/28
Homework #2	10/24
Homework #3	11/7
Quiz #2	11/18
Homework #4	12/2
Final Exam	TBD
Mini-Assignments	Weekly

*The above dates indicate the date of an exam/quiz or the due date of a homework assignment.

Homework Assignments:

There are four homework assignments due throughout the quarter. The first homework assignment is worth 14% of your overall grade and will be graded based on completion (i.e., whether the student made a good-faith effort and provided answers for all questions). This first assignment will allow you to become familiar with our expectations for how to complete, format, and submit an assignment. For the remaining three assignments, we will expect you to follow all of the formatting requirements when submitting assignments; you will be penalized on your grade if you fail to do so. Of these three remaining assignments, the

assignment with your highest score will be worth 17% of your overall grade, the assignment with your lowest score will be worth 11% of your overall grade, and the assignment with your median score will be worth 14% of your overall grade. This weighting scheme is designed to minimize the consequences of a single bad assignment and ensure that your best work is rewarded.

All assignments should be submitted through Gradescope on Canvas in PDF format by 11:59pm on the stated due date. These assignments should be formatted in a clean and professional way; graders are not expected to sift through a disorganized writeup to find your answers. *Note: the answers that you upload to Canvas will be the answers that are graded*. You are responsible for ensuring that you have uploaded the most up-to-date and most complete file.

We will be using Gradescope to manage assignments and grading this quarter. You can find the Gradescope shortcut on the left side of your Canvas menu. You must submit a PDF version of your assignment at that link. We will not consider submissions if they are not uploaded to Gradescope. You are required to submit a single PDF, which contains your write-up to the homework assignment as well as your R code appended. There are multiple ways to include your R code in your PDF:

1) Run the line of code, `knitr::stitch('myscript.r')`, which will save a PDF in your working directory called `myscript.pdf`. Then, append that PDF of your code to the PDF of your write-up.
2) If you encounter difficulties using method (1), you can take screenshots of your entire .R script and copy and paste these screenshots into your homework PDF with the rest of your assignment. Be sure the screenshots are legible if you use this option.
3) If you are so inclined, use RMarkdown to knit your code and the rest of your assignment together into a single PDF. Please note that we cannot provide technical support if you encounter problems while knitting your PDF. If you are unable to troubleshoot these issues, you should use one of the above options to format your write-up and code into a single PDF submission.

A late assignment will be penalized by a 20% grade reduction if it is submitted within 24 hours of the due date/time, and an assignment submitted 24-48 hours after the due date/time will be penalized by a 40% grade reduction. Assignments will not be accepted more than 48 hours after the due date/time. *For the final assignment (HW4), late submissions will not be accepted.* We plan to release the HW4 solutions soon after the due date/time to allow students to prepare for the final exam. Late submissions for this assignment will receive a score of 0.

The collaboration policy is clearly stated on each assignment. Your work should only reflect your knowledge and your effort. That means, for instance, that your write-up to an assignment should always and entirely be written in your own words. If your write-up contains the same language as someone else's write-up in the class (including one of your group members), that is considered a violation of the collaboration policy. On occasion, there might be parts of an assignment in which collaboration is *not* permitted. The assignment instructions will clearly state when this is the case. You must work on your own for these parts of the assignment. If you are ever unsure about what is permissible, proper attribution, and other academic (dis)honesty issues, please ask the teaching staff!

Weekly Mini-Assignments:

Each week we will provide a very short assignment that demonstrates concepts from the course in R. These are generally self-guided coding tutorials. Mini-assignments are typically due by 6:59am CT on Tuesdays and *graded based on completion*.

Quizzes:

We will also administer two quizzes about one-third and two-thirds of the way through the quarter. The quizzes are short exams and provide you with a sense of the degree to which you are grasping the course material. Each quiz is only worth on average 2% of your overall grade (your lowest quiz score is worth 1% of your overall grade and your highest quiz score is worth 3% of your overall grade); they are designed to be diagnostic rather than punitive if you are struggling to understand the material. *Note*: the quizzes will be

available for "check out" from Canvas on the scheduled day. You will have a designated time limit to complete the quiz after you check it out. Make sure that you set aside a period of time to take the quiz on each of the two quiz days. If you happen to forget to take the quiz, you will receive a score of zero.

Final Exam:

The final exam will be a comprehensive exam, potentially including questions related to any material covered throughout the quarter. The exam will be administered in-person, so you should ensure that you will be in Chicago during the exam period (the date of the final exam is set by the Harris registrar sometime during the fall quarter). We will provide more details on the final exam as it approaches.

Time Management & Time Expectations:

One of the most challenging adjustments for many students in their first quarter of their graduate studies is time management. It is extremely important that you provide yourself ample time to complete all assignments and to prepare for all assessments.

Mini-assignments are designed to be completed generally in less than 20 minutes. As a reminder, mini-assignments are graded based on completion, so a student who makes a good-faith effort to correctly complete the assignment and submits it on time should receive full credit.

Homework assignments are worth the majority of your grade in the course, and, as a result, they require a substantial amount of time. For many students, the homework assignments may take in excess of 10 hours to complete. With that time commitment in mind, it is extremely important that you start homework assignments sufficiently far in advance of the due date. Starting early ensures that you have ample time to complete the assignment and ample opportunity to seek help from your peers, TAs, instructors, and other course resources should you need assistance.

Other Course Policies, Information, and Resources

Emergencies or Assignment Conflicts, including Religious Exemptions:

In general, due dates and assessment dates are not subject to change out of fairness to your fellow students. That said, we understand there are sometimes circumstances that may require exceptions. Please do *not* email your instructors regarding these requests. Instead, please direct requests to the Dean of Students & Director of Student Affairs, Kate Biddle (kbiddle@uchicago.edu). Dean Biddle will make a determination and if necessary your instructor will work with you regarding an appropriate accommodation.

Communication & Weekly Newsletter:

Communication from instructors to students will happen primarily through the posting of materials on Canvas, including postings to Announcements and the Piazza discussion board. Each Sunday throughout the quarter, you will receive a weekly newsletter in the form of a Canvas Announcement. The newsletter will convey important information about all of the relevant items for the coming week. You should use this newsletter as a reference for upcoming deadlines and planning for the week. Please note that *you are responsible for reading all Canvas Announcements related to the course*. To ensure receipt, you may wish to confirm that you have email notifications enabled for Canvas Announcements.

As there are many students in this sequence, emailing your instructor directly is an ineffective way to have either a logistical or a pedagogical issue resolved. Therefore, we suggest and request that communication from students take the following forms:

> *Questions regarding scheduling and other course logistics* should be directed to the head administrative TA: Sushant Banjara: sbanjara@uchicago.edu.
>
> *Questions regarding course material* may be posted to the course Piazza page, a forum that is monitored by the teaching assistants and instructors. Please note that, while we strive to respond

expeditiously to student questions posted on Piazza, you should *not* expect always to receive prompt replies, especially if your question is posted on the weekend or after normal business hours. So, please do not expect to receive a quick response at 9:00pm on a Friday evening.

Office Hours:

Each week, your instructors dedicate time for additional instruction through their office hours. You are encouraged to attend your instructor's office hours (rather than another instructor's office hours). However, if your schedule precludes you from attending your instructor's office hours, you may attend the office hours of another instructor. To attend instructor office hours, a student or small group of students (we encourage you to attend office hours in a small group!) can reserve an available time slot to meet with the instructor. More information, including time slots for scheduled office hours, is available on the course Canvas site.

In addition, the TAs will hold regularly scheduled drop-in office hours that you are encouraged to attend. These office hours are useful for asking questions about the course material, homework assignments, and troubleshooting R code. Additional information about TA office hours is also available on Canvas.

Re-Grading Policy:

A re-grade request must be submitted via email within 7 days from when the assignment or assessment is returned to you. To request a re-grade, you must complete a re-grade request form (available on Canvas), and email it to your instructor and the head TA. In all cases, *the entire assignment will be re-graded*, not just the question or specific part that pertains to your grievance. As a result, the re-grade can (and often does) result in a lower overall grade on the assignment or exam.

Statistical Software:

This course will require you to follow lectures and complete assignments using the statistical software R (and its companion software RStudio). From the outset of the course, we will be using R, so make sure it is fully working on your computer before the quarter begins. Fortunately, there is no charge for use of R, RStudio, or any of the R libraries that we will use in the course.

Coding Lab for Public Policy:

Coding Lab for Public Policy provides students with introductory coding instruction in R. Enrollment in this non-credit course is encouraged but not required. A Coding Self-Assessment was included with the Harris Orientation Modules to help students determine whether they would benefit from Coding Lab. Students are not required to submit the self-assessment for review; it is a tool to help students determine how much assistance they need to be successful in the Core. If you have questions about Coding Lab or the Coding Self-Assessment, reach out to your Academic Advisor.

Core Tutoring Program:

In addition to the Coding Lab, Harris offers 10 hours of free tutoring for students enrolled in core classes. Students can get assistance with understanding statistical concepts, reinforcing computational rules, and with coding in R. Harris Student Affairs will share information about how to access tutoring in Week 3 of the quarter. Any questions should be directed to HarrisStudentAffairs@uchicago.edu.

Ethical Academic Conduct:

All University of Chicago students are expected to uphold the highest standards of academic integrity and honesty. Among other things, this means that students shall not represent another's work as their own, use impermissible materials during exams, or otherwise gain an unfair academic advantage. All students suspected of academic dishonesty will be reported to the Harris Dean of Students for investigation and adjudication. The disciplinary process can result in sanctions up to and including suspension or expulsion

from the University. At the instructors' discretion, the student may receive a failing grade for the course regardless of their performance on other elements of the course.

Please note that the collaboration policy is explicitly stated on every homework assignment. Not abiding by this policy is considered an ethical violation.

For additional resources on academic integrity at the Harris School and the University of Chicago, see:

- University Student Manual
- Harris Student Policies
- University of Chicago University Policies and Regulations on Academic Honesty & Plagiarism

Recording & Deletion and Copyright & Course Content Policies:

The Recording and Deletion Policies for the current academic year can be found in the Student Manual under Petitions, Audio & Video Recording on Campus:

- Do not record, share, or disseminate any course sessions, videos, transcripts, audio, or chats.
- Do not share links for the course to those not currently enrolled.
- Any Zoom cloud recordings will be deleted 90 days after the completion of the recording.

All course materials (including, but not limited to, class lectures and discussions, handouts, examinations, study guides and web materials) and the intellectual content of the course itself are protected by United States Federal Copyright Law. Students are permitted to make notes solely for their own private educational use. Students and all other persons are expressly forbidden from recording lectures or discussions and from distributing or selling lectures notes and all other course materials without the prior written permission of the instructors. Because the instructors own the copyright to the classroom presentations and all course materials, any notes taken during those presentations and subsequently sold or distributed to others would constitute an unauthorized derivative work and expose the person or persons involved to individual copyright infringement actions by the instructors.