Harris School of Public Policy
University of Chicago
PPHA 31002

Autumn 2024: Statistics for Data Analysis I
Syllabus, Subject to Change

Dr. Eyal Frank
eyalfrank@uchicago.edu

Dr. Sheng-Hao Lo
shenghaolo@uchicago.edu

Dr. Daniel Moskowitz
dmoskowitz@uchicago.edu

Last Updated: October 1, 2024

Background and Goals

PPHA 31002 Statistics for Data Analysis I is the first quarter of the core statistics sequence at the Harris School. PPHA 31002 partially fulfills the statistics course requirements for the MPP and MSCAPP degrees, and this course serves as a prerequisite for PPHA 31102 Statistics for Data Analysis II. In lieu of PPHA 31002 & 31102, students may instead fulfill the statistics course requirements with PPHA 31202 & 31302, Advanced Statistics for Data Analysis I & II.

This course aims to provide students with a basic understanding of statistical analysis for policy research. The course focuses primarily on descriptive and, to a lesser extent, predictive inference. Descriptive inference tools allow us to understand the state of the world and diagnose policy problems. Measuring and describing the world in which we live is one of the most important tools for a policy analyst to determine the most pressing problems in society that warrant attention and scarce resources. The two primary objectives for the course and the broader statistics sequence are:

☑ 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. Upon completion of the core statistics sequence, students should be able to competently:

☑ Summarize, describe, and visualize data while avoiding common errors and pitfalls.

☑ Assess the validity of descriptive, predictive, and causal claims.

☑ Understand and quantify uncertainty.

☑ Analyze data in R and conduct careful quantitative analyses.

☑ Understand how to use evidence to inform policy decisions.

Prerequisites

There are no formal prerequisites for this course. 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 instead consider enrolling in PPHA 31202, Advanced Statistics for Data Analysis I.

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 that even a short lapse in attention to the course can make it extremely difficult to catch up with the material. *You, as an adult, are ultimately 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, we urge students to attend, view, or read:

(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 time for questions. *Note: as per the Harris School's attendance policy, you are required to attend lecture for core courses. However, you are only permitted to attend the course section to which you are assigned unless you have written permission (e.g., email) from your instructor.*

(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 lectures, 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 units in the course, 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 course unit for the supplemental textbooks on Canvas.

Note that 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/RStudio, so make sure both are fully installed and working on your computer before the quarter begins. Fortunately, R, RStudio, and all R libraries that we will use in the course are available free of charge.

Course Topics

The course is divided into seven units:

(1) Introduction

(2) Probability Theory & Random Variables

(3) Summary Statistics & Exploratory Data Analysis

(4) Regression for Prediction & Description

(5) Hypothesis Testing

(6) Simulations

(7) Experiments

The readings and lecture slides for each unit are posted on the Modules page of the course Canvas site.

Course Assignments & Assessment:

The composition of your overall grade is calculated based on the percentages indicated in Table 1. Details on how the grade is calculated for each of these components is discussed further below. As a core course in the Harris School, we must follow the strict grading curve indicated in Table 2. In other words, after your numerical score is calculated at the end of quarter, the students with the top 12.5% of overall scores will receive a grade of A, the 25% of students with the next highest scores will receive an A-, and so on. Because of the nature of a strict curve, your numerical scores may not translate into a letter grade in the way to which you might be accustomed. To help provide a sense of where you are situated in the grade distribution, we will provide you with basic summary statistics, including quantiles, for the midterm exam and each homework (excluding homework #0). *To ensure your academic success in the course, the most important thing that you must do is complete and submit on time all assignments and assessments.* The course schedule (*subject to change*) with the (due) date of each homework, diagnostic, and exam is provided in Table 3.

Table 1 – Grade Composition

Component	Share
Attendance	5%
Homework #0	2%
Homework #1-4	45%
Diagnostics #1-3	3%
Midterm Exam	5% or 10%
Final Exam	35% or 40%

Table 2 – Grade Distribution

Grade	Share of Students
A	12.5%
A-	25%
B+	25%
B	25%
B- or below	12.5%

Table 3 – Course Schedule

Item	Date
Homework #0	10/7
Homework #1	10/14
Diagnostic #1	10/15-10/16
Homework #2	10/28
Diagnostic #2	10/29-10/30
Midterm Exam	10/31
Homework #3	11/18
Diagnostic #3	11/19-11/20
Homework #4	12/6
Final Exam	12/11 9am-11am

Attendance

As per the Harris School's policy, students are *required* to attend lecture sessions for core courses. Class attendance is worth 5% of your overall grade. There is no penalty to your grade for missing up to two class sessions to allow for illness, emergencies, and other similar circumstances. Students do *not* need to communicate with the teaching staff regarding absences. For each additional absence beyond the two non-penalty absences, students will experience a one percentage point deduction to their attendance grade component. Students who are chronically absent will also be reported to the Dean of Students Office and may be denied credit for the class. Students should come to class prepared to engage with the material. We will begin tallying absences as part of your grade in week 3 of the quarter, as students' course schedules are sometimes in flux during the first two weeks.

Homework Assignments:

The first homework assignment (homework #0) for the course is a "practice" assignment, which is graded based on completion (i.e., students who make a good-faith effort and provide answers for all questions will receive full credit). This practice assignment should allow students to become familiar with our expectations for how to complete, format, and submit an assignment. This assignment is not lengthy and is designed for students to be able to complete it relatively quickly.

There are four subsequent homework assignments (homework #1-4) due throughout the quarter. After the practice assignment, we expect you to abide by all formatting and submission requirements; we will deduct points for failing to do so. Your lowest homework score will be dropped, and the remaining three assignments will each be worth 15% of your overall grade. This grading scheme ensures that a single bad assignment (or a failure to submit an assignment) has no impact on a student's final grade.

We will be using Gradescope for all homework submissions this quarter. You can find the Gradescope shortcut on the left sidebar of the Canvas course site. You must follow *all* submission instructions, which are provided on Canvas with each assignment. Homework assignments are due at 11:45pm on the due date. We will provide a 14-minute grace period in which assignments submitted by 11:59pm on the due date will not be marked as late. Late submissions after the grace period but within the subsequent 24-hour period will be penalized by a 20% grade reduction. Assignments submitted after this 24-hour late period will not be accepted. *For the final assignment (HW4), late submissions (beyond the 14-minute grace period) will not be accepted.* We plan to release the HW4 solutions soon after the deadline to allow students to begin preparing immediately for the final exam. Any late submission for this assignment will receive a score of zero.

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 (including one of your group members), that is considered a violation of the collaboration policy. If you are ever unsure about what is permissible, proper attribution, and other academic (dis)honesty issues, please ask the teaching staff!

Diagnostics:

To help you assess your learning throughout the quarter, we will provide three required diagnostic exercises. The diagnostics are individual exercises designed to provide you with a sense of your grasp of the course material. They are graded based on completion, meaning that students who submit their answers through Canvas by the deadline will receive full credit for the diagnostic. While collaborating with peers on assignments provides many learning benefits, one drawback is that students may occasionally rely too much on group members and, thus, have a weaker understanding of the course material than they realize. The diagnostics are a tool intended to gauge your understanding of the material, so that you have a better idea of what you know and what you don't know. In our experience as instructors, scores on these individual exercises are more strongly correlated with students' scores on the final exam than homework scores are.

Each of these three exercises will be made available immediately after the HW1, HW2, and HW3 deadlines. Note that these are *closed-book, closed-note individual* exercises to mimic an exam setting. Because these exercises are graded based on completion (i.e., you will receive 100% credit if you complete it), it is in your interest to follow these rules so that you receive an accurate assessment of your learning. You will have a 48-hour period starting on the scheduled day to complete and submit your diagnostic through Canvas. Make sure that you set aside a period of time to complete the exercise when it becomes available prior to the submission deadline. If you fail to complete a diagnostic submission by the deadline, you will receive a score of zero.

Midterm and Final Exams:

The midterm exam will take place in class on October 31. Students are expected to be in attendance for the exam. The exam will include material that we have covered through that point in the quarter. We will provide more details on the midterm exam, including the format, closer to the date of the exam.

The final exam will take place on December 11, 2024 at 9:00am in the Keller Center. It will be a comprehensive exam, potentially including questions related to any material covered throughout the quarter. The final exam is an important and required component of your assessment. As such, students must take the exam in order to receive a letter grade in the class; in other words, students will not be exempted from the

exam. All students should ensure that they will be present in-person during the university's exam week, as the final exam cannot be taken remotely. We will provide more details on the final exam as it approaches.

The midterm exam is worth 5% or 10% of your overall grade, and the final exam is worth 35% or 40% of your overall grade. If you receive a higher score on the midterm than the final exam, your midterm exam weight will be 10% and your final exam weight will be 35%. On the other hand, if you receive a higher score on the final than the midterm exam, your midterm exam will be worth 5% and your final exam will be worth 40% of your overall grade. This weighting scheme is designed to reward students' best work in the course.

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 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 assessment.

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. Homework assignments are worth nearly half of your overall grade in the course, and, as a result, they require a substantial amount of time and effort. For many students, the homework assignments will 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 rare circumstances that may require exceptions. Please do not email your instructors about these personal situations. Instead, you should email your academic advisor regarding circumstances that may warrant an exception. Your advisor will then communicate with the Dean of Students and/or your instructor to determine the appropriate accommodation if the situation merits one. Importantly, any such request must be communicated sufficiently *in advance of the due date* to be considered.

Electronics Policy:

Because there is substantial evidence that student use of laptops and phones in the classroom negatively affects learning outcomes (for evidence, see here, here, here, and here), student use of these devices will generally not be permitted during class unless otherwise indicated. There are exceptions for hand-written note-taking on tablets laid flat on the desk as well as disability accommodations that allow these electronics.

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 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* should 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 say, 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.

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. An optional coding self-assessment was included with the Harris Orientation Modules to help students determine whether they would benefit from Coding Lab. 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 8 hours of free tutoring for support with core courses and coding in Stata, R, and Python. Tutoring will be available starting week 2 of the quarter. More information about the program is available on the Harris Student Handbook Canvas site. Any questions should be directed to your academic advisor or harrisdeanofstudents@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 academic misconduct and is subject to the sanctions described above.

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

Use of Generative Artificial Intelligence:

Generative AI tools, such as ChatGPT, are incredibly powerful and potentially transformative. Students are permitted to use these tools for assignments in the course as they would any other internet resource. With that in mind, students should remember that:

(1) Generative AI tools, like other internet resources, are sometimes inaccurate. Large language models can generate remarkably coherent and correct-sounding text that is profoundly inaccurate. There is some evidence that these models are especially prone to producing inaccurate responses to prompts in the quantitative domain. As a result, compared to traditional internet resources (the sources of which are known), these tools may confuse and mislead rather than provide guidance towards the solution. Students are always *fully responsible for the accuracy of their written work* for the course.

(2) Just as it is not acceptable to copy and paste text into an assignment from a traditional internet resource, it is not acceptable to copy and paste text from generative AI tools. As the collaboration policy states, "*Your* write-up and code should only reflect *your* understanding of the material. As such, these should be written in *your own words*." These principles and this policy are applicable to the generative AI context as well. To comply with the university policies on academic honesty, you must always provide proper documentation and citation of the generative AI tools that you used. When you are not sure about appropriate usage, please ask for clarification.

(3) Over-reliance on these tools may be detrimental to student learning. Of note, students will not have access to these tools (or other resources) for the midterm and final exam, which together constitute nearly 50% of the overall grade. Students who depend on these tools to complete assignments may score lower than they otherwise would have on the final exam.

Recording & Deletion and Copyright & Course Content Policies:

As per the Harris School's policies for core courses, we will record the twice-weekly lecture and make that recording available to students enrolled in the course through the course Canvas site. The recording from only a single lecture section will be posted, meaning that it may not be the recording of your section or even of your instructor. The lecture recording will only include the audio and the screen capture of the computer displaying slides. We, unfortunately, do not have the ability to record video that legibly captures writing on the white board. In addition, while we always hope to provide a recording with clear audio, we make no warranty about the quality of the audio or the auto-generated captions from that audio. In other words, the lecture recordings are not a substitute for regular lecture attendance. Please keep in mind the following course, school, and university policies regarding the recording of lectures:

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

The full Recording and Deletion Policies can be found in the Student Manual under Petitions, Audio & Video Recording on Campus.

All course materials (including, but not limited to, class lectures and discussions, handouts, assignments, assessments, 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.