S-052: Intermediate and Advanced Statistical Methods for Applied Educational Research
Harvard Graduate School of Education, Spring 2021

Instructor: Andrew Ho (he/him)
Andrew_Ho@gse.harvard.edu

Faculty Assistant: Wendy Angus (she/her)
Wendy_Angus@gse.harvard.edu

Links to all synchronous and asynchronous content are available on Canvas:
https://canvas.harvard.edu/courses/85822

Synchronous meeting times (on Zoom: https://canvas.harvard.edu/courses/85822/external_tools/53244**)**

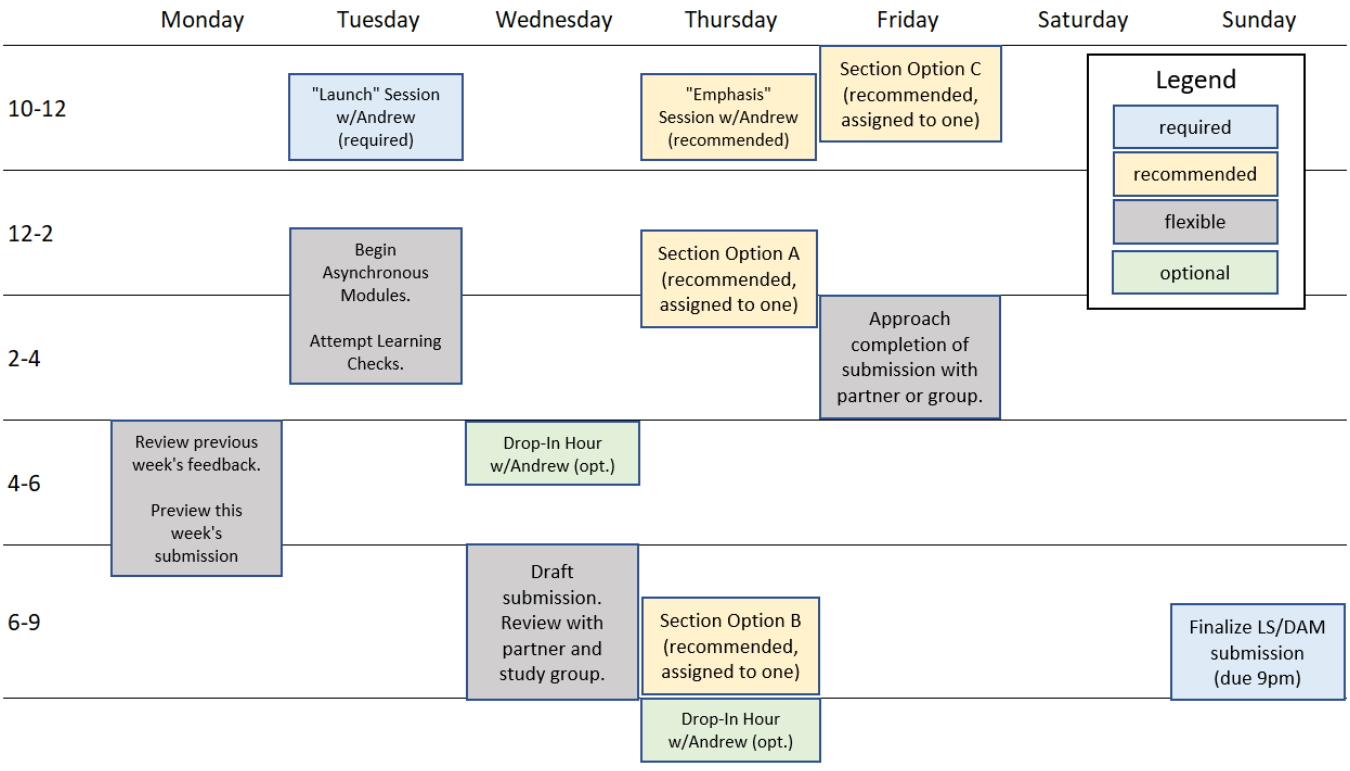
Weekly required synchronous "Launch" meetings (assigned to either Squad A or Squad B):
Squad A: Mondays 9pm-10:15pm ET
Squad B: Tuesdays: 10:30am-11:45am ET

Weekly recommended synchronous "Emphasis" meetings
Squad A: Wednesdays 9pm-10:15pm ET
Squad B: Thursdays: 10:30am-11:45am ET

Andrew's optional drop-in hours ("Please Drop By My" Office Hours)
Wednesdays, 4pm-5pm ET
Thursdays, 9pm-10pm ET

Weekly recommended 75-minute synchronous sections with Teaching Fellows (assigned)
Thursdays 1:30-2:45pm ET, Thursdays, 7:30-8:45pm ET, or Fridays 9-10:15am ET

An illustrative week for a Squad B student in S-052 (all times Eastern). Squad A sessions meet Monday/Wednesday at 9pm instead.



Course overview

Welcome to *S-052: Intermediate and Advanced Statistical Methods for Applied Educational Research*. This course is an integrated continuation of the fall course, *S-040*, and is part of the *HGSE* school-wide network of courses in quantitative methods. The *S-040* and *S-052* courses form the cornerstone of a sequence of courses in applied statistical methods for consumers and producers of rigorous educational, social, and psychological research.

The course is designed to develop and extend the data-analytic skills that you began to acquire in earlier courses and to help you learn to communicate your findings clearly to audiences of other empirical researchers, scholars, policy-makers, practitioners, students, and parents. We have designed *S-052* to contribute to the diverse data-analytic toolkit that you will need in order to perform sensible and useful analyses of complex educational, psychological, and social data.

Core topics such as multiple regression analysis, introduced in *S-040*, continue to be the foundation of *S-052*. However, we extend your use of these techniques to cover a wider variety of conditions encountered in the world of real data-analysis, including multilevel models and selected multivariate methods. A listing of major course topics is provided later in this document.

True to its name, *S-052* is an *applied* (not a *theoretical*) course in which you will learn by observing and engaging in the authentic activities of real applied data analysis. We will model the use of new statistical techniques in class, and then you will apply these new techniques to real problems using real data in "data-analytic memos" and a significant final take-home assignment that is affectionately known as, "the celebration of learning."

In all these assignments, we ask you to interpret the outcomes of your data-analyses in words, and to communicate these interpretations clearly and concisely in writing. And you will acquire the basic programming skills necessary for hands-on data analysis in *Stata* or *R*.

Library reserves and other text resources (on Canvas)

Many students learn well from textbooks. Unfortunately few textbooks cover the range of methods we teach in this course. Instead we provide a range of resources you can find on Canvas, here. Folks who wish to invest in a paper copy may find the first volume of Sophia Rabe-Hesketh and Anders Skrondal's textbook useful, *Multilevel and longitudinal modeling using Stata, Volume I*, 3rd edition (RH&S). We have some of the chapters already available on Canvas reserve. You can also purchase the text at the Stata Bookstore for around $65, cheaper if you have a discount code. Amazon and the Coop tend to be more expensive. It is a slightly advanced but highly applied text that offers excellent support for about 2/3 of the course material, including linear, logistic, and multilevel regression. You can see the coverage here. We will also be using a free online chapter from their second volume on logistic regression, here.

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata, Volumes I and II* (3rd ed.). College Station, TX: Stata Press. [Only Volume I is recommended for this course.]

A more general textbook on applied regression and multivariate methods is Klienbaum et al. (2013), which you can purchase or rent at Amazon and other locations, e.g., here. For survival analysis, we also use an excerpt from my a textbook written by my colleagues, Judy Singer and John Willett. It is linked in our course reserves on Canvas, here. For more links, see our overview of course content in this next section.

Overview of course content (This is subject to adjustment, particularly toward the course's end)

1. Fitting Sensible Taxonomies of Multiple Regression Models (RH&S, Beck&Beck):
(a) *Deciding Which Regression Models to Fit*. Addressing research questions by fitting taxonomies of multiple regression models and determining a sensible "final" model. Reviewing the specification of regression models in which a single substantive construct is represented by a system of "dummy" predictors. Reviewing the notion of a statistical interaction between predictors. This introductory section provides an opportunity to review your prior learning about multiple regression analysis.

(b) *Testing Complex Hypotheses About Regression Parameters*. Comparing nested multiple regression models. Using the *General Linear Hypothesis Tess* (ΔR^2 *test*) to conduct formal tests of the joint effect of several predictors simultaneously on an outcome.

(c) *Detecting Influential Observations and Assessing their Effects on Model Fit and Parameter Estimation.* Introducing the notion of influence statistics. Conducting sensitivity analyses.

(d) *Checking the Assumptions on the Residuals.* Understanding the importance of the assumptions on the residuals in a multiple regression analysis. Graphical methods for assessing distributional assumptions.

(e) *Interpreting and Reporting Findings*. Using fitted plots to display and interpret the size and direction of detected effects for prototypical individuals in the population, especially in the presence of statistical interactions.

(f) *Dealing Empirically with Non-Linear Outcome/Predictor Relationships*. Using logarithmic and other transformations to linearize the outcome/predictor relationship.

(g) *Causal Inference and Counterfactual Reasoning:* Introducing the fundamental problem of causal inference and distinguishing between causal vs. predictive interpretation of regression coefficients.

2. Basic Logistic ("Binomial Logit") Regression Analysis (Pampel, 2000; RH&S):
(a) *Modeling the Relationship Between a Dichotomous Outcome and Predictors using a Linear Probability Model*. The problematic impact of specifying a linear regression model when the outcome is dichotomous – appropriate interpretation of fitted values and problems in the residual distribution.

(b) *Modeling the Relationship Between a Dichotomous Outcome and Predictors with Logistic Regression ("Logit") Analysis*. Using a non-linear logistic (or "logit") function to represent the hypothesized relationship between a dichotomous outcome and predictors. Goodness-of-fit statistics for logistic regression analysis.

(c) *Fitting Taxonomies of Nested Logistic Regression Models*. Addressing research questions about the prediction of dichotomous outcomes by fitting and comparing nested logistic regression models using a *General Linear Hypothesis* (χ^2) *Test.*

(d) *Interpreting Fitted Logistic Regression Models*. Using fitted *odds, odds-ratios,* and fitted trend lines plotted for prototypical individuals in the population to demonstrate the size and direction of an effect detected via logistic regression analysis.

3. Survival Analysis (Singer & Willett, 2003, Chapter 10, Course Reserves):
(a) *Discrete-Time Survival Analysis*. Using logit analysis to examine the occurrence and timing of events in a person's life. Introducing the concepts of hazard and survivor probability, and the discrete-time hazard model.

(b) *Adding Predictors to the Discrete-Time Hazard Model.* Using prototypical fitted hazard and survivor functions, and predicted median lifetimes, to interpret findings.

4. Multilevel Models (RH&S Ch. 2, 3):
(a) *Introducing the "Fixed Effects" Model*. A useful and robust approach using dummy variables for groups to account for multilevel structures.

(b) *Introducing the "Random Effects" Multilevel Model*. Contrasting a "fixed effects" model with a multilevel "random effects" regression model to account for the grouping of individuals within higher-level "units."

(c) *Interpreting variance components and intraclass correlations.* Fitting the multilevel model using random-effects regression analysis. Partitioning residual variance into its within-group and between-group components, estimating and interpreting the intraclass correlation.

(d) *Using the Multilevel Regression Model to Analyze Longitudinal Data*. Extending the random intercepts model to a random slopes model for analyzing individual change over time.

5. Multivariate Methods—Reliability, Principal Components Analysis (Dunteman)
(a) *Reliability and Classical Test Theory*. Traditional strategies for forming data-composites -- standardization of indicators, creating a weighted linear composite. Measurement error and internal-consistency reliability (Cronbach's α).

(b) *Using Principal Components Analysis (PCA) To Form An "Ideal" Data-Composite*. Introducing principal components analysis as an alternative to classical item-analysis in data-compositing. Creating a composite that accounts for maximum variance.

(c) *Using PCA To Evaluate the Multivariate Structure of Data.* Using the eigenvalues and a scree plot to estimate how many "dimensions of information" underlie a given set of indicators. Interpreting the underlying dimensions numerically, graphically and substantively.

6. Cluster Analysis (time permitting)
(a) *Exploratory Cluster Analysis of Variables.* Automated exploratory analyses to extend PCA so that interesting "coherent" clusters of indicators can be detected within a larger defined group of variables.

(b) *Exploratory Cluster Analysis of Individuals*. Automated exploratory analyses that group individuals together into discrete clusters so that "profiles" and "types" of behavior are revealed. Using the *Pseudo-F Statistic* to determine the number of clusters, and the *Tree Diagram* to display the clustering detected.

(c) *Using the Cluster Analysis of Individuals to Detect Multivariate Outliers*. Which of these people are most like the others, and which of these people don't belong?

7. Factor Analysis and Structural Equation Models (time permitting):
(a) *Basic Factor Model*. Using path diagrams to describe multivariate hypotheses about relationships among multiple indicators, constructs and measurement errors in a first-order factor model. Representing the path model as a system of inter-linked statistical models and understanding the role played by the model parameters.

(b) *Structural Equation Model (SEM) and Confirmatory Factor Analysis (CFA).* Fitting the hypothesized model to data and assessing its fit. Testing and interpreting critical parameters in the model.

Course activities and participation

Most of our time—both synchronously and asynchronously—will be spent learning how to do data analysis. Sometimes understanding is enhanced by knowing more about the mathematical underpinnings, but we also offer straightforward conceptual explanations that do not sacrifice intellectual rigor.

We will devote time to illustrating how to present results in words, tables and figures. Good data analysis is craft knowledge. It involves more than using software to generate reams of output. Thoughtful analysis can be difficult and messy, raising delicate problems of model specification and parameter interpretation. We'll confront such issues directly, offering concrete advice for sound decision making.

Class participation is an important part of learning, even in a relatively large lecture course like S-052. If you have a question, it's likely that others do as well. I encourage active participation with live, open-edit documents as well as raised hands. If students make efforts in and out of class to engage actively with course content to the benefit of themselves and their fellow students, I may factor this into grades that fall near grading cutpoints. For in-class participation, please do not be offended if I defer your contribution to another time, if I feel addressing the question may take us too far astray.

Meeting times and the attendance policy

Class will start promptly at the official start time. Please sign on at or before the appointed time. I take attendance digitally through in-class online "learning checks." We use attendance data formatively, to ensure that students are staying on pace and to direct additional support.

Online class recordings

All class meetings are recorded on zoom and available on Canvas. "Launch" sessions are required; "Emphasis" sessions are expected and recommended but optional. We appreciate your commitment to synchronous attendance as we endeavor to build a coherent and lasting community. In the event of unexpected conflicts, simply let us know. All class sessions are recorded.

Professional behavior in a digital age

All of us are likely to face distractions during synchronous meeting times. Sometimes these distractions are in our control, and other times they are not. We encourage you to stay focused and engaged during synchronous sessions, even as we understand when other events or priorities arise unexpectedly. I encourage you to read this review article by our HGSE colleague Susan Dynarski about the risks of distraction to learning and do your best to monitor your engagement accordingly.

Statistical computing with Stata 16 or R

Statistical computing is an essential part of S-052. To support your learning, the quantitative methods sequence at HGSE supports both Stata 16 and R. We will provide all programming language for both Stata and R, so no independent programming tasks will be necessary excepting students who opt to take on course projects. On balance, we recommend Stata, and our slides are often Stata-based. We assume that everyone has used a statistical program to fit models. We do not assume that you have used Stata or R before. We recommend recent versions of Stata, but if you have Stata 12 and up, you should be fine. For students who agree to use the free statistical program, R, we will provide limited support, but we will allow you to submit assignments for which you have used R to fit models and estimate statistics. We recommend Stata for all but the most ambitious or already-R-fluent students, as you will essentially be "on your own" for programming support.

Again, we provide all code necessary for assignments, so we will not cover programming in lecture or section except where it directly relates to course content. Students will run provided code themselves and

interpret output. Teaching fellows can field separate questions about coding for independent projects during their office hours.

You can download Stata and R here: https://its.gse.harvard.edu/research-software-support

Ongoing Assessments: Learning Checks, Learning Submissions, and Data Analytic Memos

We use three types of assessments to help you and us monitor and support your learning.

Learning Checks. In synchronous class sessions and asynchronous modules, we will use "learning checks," 1-3 structured questions that you will discuss with your fellow students and then submit responses. Correct answers will be provided immediately after your response to support your learning. We will not provide feedback on learning checks but use your responses simply to track attendance and progress. The goal is to to give you and us a sense of your learning and engagement.

Learning Submissions. At the end of the first two weeks of class, and every other week thereafter, students will submit responses to a brief series of questions on Canvas. These must be submitted by 9pm on Sunday. Example walkthroughs and solutions will be provided immediately after your submission to support your learning via immediate feedback. Like learning checks, we will not provide feedback on learning checks, but we encourage you to give them careful consideration and due effort to help us monitor and support your learning.

Data Analytic Memos. The best way to understand statistical analysis is to actually conduct statistical analysis. To help you develop your skills, we distribute and provide feedback for 6 assignments. In honor of my predecessor, John Willett, we call these "Data Analytic Memos" (DAMs) which he referred to lovingly as "those DAM things." These will be due every other Sunday at 9pm. To maximize your learning, we encourage you to open each DAM early (a good habit is to open it the day after you submit the previous DAM) and reference it often as you engage with synchronous and asynchronous material. We encourage you to have it complete or near-complete before the weekend on which it is due.

Design and Purpose of S-052 Assessments

Assessment Type	Length	Completion	Collaboration	Feedback	Purpose
Learning Checks	Very brief	In synchronous sessions	Individual submission after consultation in breakout rooms	Correct answers provided immediately.	Formative Attendance.
Learning Submissions	Moderate	Due Sundays at 9pm ET in Week 1, 2, 4…	Individual submission after optional consultation with colleagues	Correct answers provided immediately.	Formative. Engagement.
Data Analytic Memos	Extended	Due Sundays at 9pm ET in Week 3, 5, 7…	Partnered submission	Detailed, approximately 1 week later.	Formative and summative.

Learning Submission Due Dates (individual, Sundays 9pm ET): 1/31, 2/7, 2/21, 3/7, 3/28, 4/11, and 4/25.
DAM Due Dates (partnered, Sundays 9pm ET): 2/14, 2/28, 3/14, 4/4, 4/18, and 5/2.

Culminating Assessments: Final "Celebration of Learning" and Optional Course Projects

The final "celebration of learning" is a partly collaborative and partly individual affair that will be posted during the examination period beginning Wednesday, May 5. It is roughly the length of a long DAM. It consists of two documents. The first document, the "evidentiary materials," contains an extended amount

of Stata/R code and output. The second document is a list of questions similar to those in assignments. The first document may be discussed in groups of any size. The second document, once opened, must be completed individually, and no subsequent discussion of the evidentiary materials nor the content of the questions is allowed until the celebration is complete.

Most students begin the first day of the celebration by reviewing the first document individually for a short time, then transitioning to group discussion to ensure that all members understand the evidentiary materials. Students generally transition to the individual component of the celebration after they are comfortable with the evidentiary material.

The questions sum to the magnitude of a large DAM, without coding. To provide flexibility, the assessment window is open from Monday, May 5 through Wednesday, May 12. Like assignments, we encourage you to plan in advance to ensure that final celebrations will be submitted on time. Please contact me early if you have any conflicts with this date.

For students who wish to apply the skills they develop in this class to their own data, we encourage the submission of a course project. The course project will usually be 5-10 written pages, including a statement of the research questions, a methods section describing the analysis, and a writeup of the results and conclusions. Students who opt to submit a course project must still complete the celebration of learning, but they may answer fewer (around 60%) of the questions. Students who wish to develop their project over the course of the semester may also offset some DAM parts with progress reports on their project. I will provide more details about this option early in the semester.

Slack, collaboration, and study groups

We will be piloting Slack for collaboration and communication in this course. Please access the S-052 Slack channel on the Canvas site here: https://canvas.harvard.edu/courses/85822/external_tools/73360 . Standard Slack norms are here:

- Be good citizens of Slack: answer questions, assume best intent, try not to sidetrack conversations.
- Understand **public and private channels**, and try to communicate with the proper audience.
- Use **threads** to organize smaller group discussions around specific topics.
- **Format messages** so that they're easier to read.
- **Use emoji reactions** liberally to acknowledge messages while keeping channels from becoming inundated with reaction messages.
- Use the search function to try to find an answer, reducing duplicate questions.
- Manage **notifications** to reduce information overload and maintain focus.
- Remember that there are situations when it's best to move the conversation off of Slack messaging. Students can meet face-to-face using Zoom or **Slack calls.**

To mimic statistical work in the real world and to provide a chance for you to use statistical language actively, I mandate completion of DAMs in partnerships throughout the course.

We mandate collaboration for at least three reasons. First, learning statistics is like learning a language. To learn it, one must "speak" it actively and in a genuine context with other individuals. Second, collaborative statistical analysis is the norm and individual work is the exception in the world of statistical practice. Third, my experience has been that, on average, students who work in partnerships and groups both perform better and enjoy themselves more than students who work individually. Statistical collaboration is a case where the whole is greater than the sum of its parts.

Beyond partnerships, larger study groups can be helpful to you as you prepare to do the assignments, both in terms of how to approach the work and in terms of how to think about important concepts. **However, students must submit work as pairs, not group work. Papers should be written in your own words—your text should reflect your own understanding of the material.**

A couple of rules will help to avoid misunderstandings and violations. First, never send electronic documents with your responses to members outside of the partnership. Second, never compose collaborative documents with members outside of the partnership; beyond the partnership, do not cowrite answers to be shared.

Each group will undoubtedly develop its own structure; nevertheless, here are a few suggestions:

- Groups with six or more members become less useful and may be harder to organize because finding common meeting times becomes increasingly problematic.
- Plan at least one session of 1½ to 2 hours (early enough so that there is sufficient time if an additional session is necessary). After 2 hours of statistics, everyone's eyes will be glazing over.
- Schedule the meetings so that you have sufficient time afterwards to write in pairs or individually. When we read your assignments, we focus on what you say and how you say it. The assignments require not only statistical techniques but skills in analyzing and reporting the material.
- Use the groups to ask questions, try out interpretations, and so on. Often one person can explain something that makes you see something in a new way. Different people have different insights and strengths – some are good programmers, some ask good questions, others value contextual analysis—and you can learn from listening to what others in a group have to offer.
- Be sensitive to the distinction between collaboration to plan for and interpret the assignment and collaboration to write up the assignment. The former is encouraged; the latter is forbidden beyond, when applicable, your partner. If the distinction begins to feel murky, refocus your group's work on lecture content and course materials.

Grading

You will be evaluated based on your engagement with learning checks and learning submissions (15%), the learning you demonstrate on DAMs (50%), and the final celebration and (if applicable) your final project (35%).

We use arithmetic computations to arrive at a first approximation of your course grade. We then conduct a number of checks to ensure that no individual assignment or score takes on undue weight. We consider your growth as well as your average performance. And we look at your whole portfolio of work when assigning course grades. For more details, see our handout entitled, *How We Evaluate Assignments*.

Students may choose to take the course on a satisfactory/unsatisfactory basis on the condition that can find another partner who can take it on this basis. Satisfactory performance requires course attendance, an average of B or better, and completion of all assignments and the final celebration.

Avoiding plagiarism

Please read the School's policy on plagiarism in the HGSE Student Handbook, which includes the statement, "Students who submit work either not their own or without clear attribution to the original source, for whatever reason, face sanctions up to and including dismissal and expulsion." Attention to this policy is particularly important in a course like S-052, in which collaboration with other students is encouraged. If you work closely with other students during the planning of your analyses—a process that I encourage and fully support—recognize the other students' contributions explicitly in your written account (a footnote is fine for this purpose). This helps avoid the natural questions that arise when

similarities are detected at grading. **If you have any questions about what constitutes appropriate collaboration, or how to define what constitutes your own work, please see me or a Teaching Fellow.**

Accommodating Students with Disabilities

I and the Harvard Graduate School of Education strive to make all learning experiences accessible by providing reasonable accommodations for students with disabilities. If you anticipate or experience academic barriers based on your disability (including mental health, cognitive, learning, sensory, physical, chronic or temporary medical conditions) please contact KellyAnn Robinson, Ph.D., Associate Director of Student Support Services in the Office of Student Affairs, as soon as possible, to explore what arrangements need to be made to assure access to course work and classroom learning experience. Please do not hestitate to contact me if there are additional accommodations you would find useful.

Our use of electronic data on Google and Canvas

I am always trying to improve my teaching and your learning. As part of my effort to improve my teaching and your learning, I will use data from online resources to provide feedback to myself and, of course, to you.

You should be aware that all these resources record data from your interactions in their server logs. Sometimes this will seem obvious to you, such as when you submit an assignment, answer an assessment question, or ask a question. Other times it will seem less obvious, such as when you log in and download a handout. It is important for you to understand that, while all these data exist, I will always make it clear when and how I will use this data for grading.

You should also be aware that, like all educational data collected in the natural course of an educational process, these data may support future research endeavors, provided that your identity is masked, or exemptions required by federal law apply.

For more details, see Harvard's Canvas Privacy Policy, linked here.

(Boilerplate) Required Information for Students Included in a Zoom Classroom Recording

- Instructors may use Zoom to record class sessions.
- If an instructor uses Zoom to record a class session, Zoom provides audio and visual indicators to inform you when the recording starts, stops, is in progress, and is paused or unpaused.
- You, as a student, may not yourself record a class session. (More generally, you should not record a class session using any other technology.)
- You have the option to appear in an audio-only mode, such that your webcam is disabled (turned off) during the class.
- You have the option to access Zoom class sessions using a pseudonym.
- In order to facilitate class participation, you are expected to communicate any pseudonym to your instructor in advance of the class.
- Links to class session recordings, if available, will be posted in the Zoom meetings section of the Canvas course webpage. More generally, any class recording must only be posted inside of the Canvas site for the course.
- Links to Zoom class session recordings will be removed and videos deleted at the end of the academic term.
- You may not disclose the link to a class session recording or copies of recordings to anyone, for any reason. It is available to your class only.

Frequently Asked Questions: Is S-052 right for me?

Can I attend synchronous sessions as a guest or an auditor?
Unfortunately not. Synchronous attendance at lectures and sections is restricted to enrolled students. Because I have explicit learning goals related to creating community among enrolled students, I do not allow formal auditing or guest attendance. Students interested in asynchronous resources may email my assistant to receive guest access to the course website. I allow this access in certain cases, such as for alumni and for interested students with conflicts who cannot otherwise enroll.

I am not a HGSE student. When should I file a cross-registration petition?
Cross-registration opens on January 19 at http://my.harvard.edu. I ask that you file your petition by Friday, January 22, at 12PM. When you cross-register, you may email me to let me know how you have satisfied the prerequisites, or I will reach out to you directly. There is no enrollment limit in the course.

I intend to cross-register. Do I need to demonstrate my prerequisites now?
Not before January 24. You should decide whether this course is best for you, including whether you meet prerequisites, below. Then you should file a cross-registration petition by Friday, January 24, at 12PM. I will then contact you to determine your eligibility. If you meet the requirements, I will approve you. If we decide that another course is better for you, I will not.

What are the prerequisites for S-052?
Successful completion of S-030 or S-040 (A- or A) or the equivalent. We expect you to have successfully, a) fit a regression model, b) with an interaction term, c) to real data, d) with a computer program, and e) interpreted the statistical significance of the coefficient for the interaction term as well as, f) written out the meaning of the coefficient for the interaction term in writing, g) in the context of a research question.

Is S-030 a better choice for me?
Maybe! S-030, offered in the same semester as S-052 by Dr. Hadas Eidelman, is designed precisely for students with limited exposure to multiple regression. The depth of coverage of multiple regression in S-030 is considerable. In contrast, coverage of multiple regression in S-052 is limited to a brief review. Students with limited past exposure who wish to develop real comfort and expertise with multiple regression will be better off in S-030 than in S-052. S-052 offers a much broader introduction to more advanced statistical methods but cannot compensate for a student's limited exposure to multiple regression on its own. Note: S-040 students should enroll in S-052, not S-030.

I just took an introductory statistics course (e.g., in SPH, BIO201 or ID201) this fall. Is that sufficient?
An A- or an A in a recent, rigorous introductory statistics course like BIO201 or ID201 will suffice as a prerequisite for S-052 if you have estimated and interpreted regression models with interaction terms as described above. However, S-030 remains complementary, and we strongly suggest that you consider it. Again, S-030 is a deep dive into applied multiple regression, and it builds analytical and interpretive skills that a typical introductory statistics class does not.

I earned a B+ in an introductory statistics course like BIO201 or ID201. Should I take S-052?
No. S-052 moves quickly, demands deep conceptual understanding, and covers several advanced topics. We recommend that students at this level take advantage of S-030 as an option to truly master foundational regression analysis first.

I don't meet the prerequisites, but I really want to learn survival analysis/multilevel modeling/principal components analysis. Should I take the course?
We do not recommend this. Our goal is not superficial acquisition of methods but deep conceptual mastery. For this, the foundations are necessary, and we recommend S-030 as a rigorous alternative.