GUSH

Guide to Harvard Statistics

last edited: Oct. 2024

Introduction

Hello person-who-might-be-interested-in-statistics!

The Group for Undergraduates in Statistics at Harvard (GUSH) has created this short (uh…it's gotten a lot longer) booklet as a way to introduce students to the wonderful world of statistics at Harvard. We understand that students who might be interested in statistics courses come from all kinds of math/statistics backgrounds and with all kinds of goals and concentration interests, and we hope that this guide will help you understand what statistics looks like at Harvard and how you can explore and get started with different courses, research opportunities, and concentration options.

While we'll try to be as comprehensive as possible, this guide is by no means meant to serve as a replacement for other advising support like talking to concentration advisors, professors, and fellow students. That being said, feel free to reach out to us with any questions you might have – we're here to help make your experience with statistics at Harvard as enjoyable as possible!

With lots of love,
Harvard GUSH
harvardgush@gmail.com
gushclub.org

What is statistics? Why should I study statistics?

Ask different people, and you'll probably get a wide range of definitions. At Harvard, the statistics department will expose you to various main ideas: probability, where you'll learn to reason about and "control" random events, numbers, and objects; statistical inference, where you'll apply your knowledge about probability to draw inferences from data; and a variety of other areas that use some combination of the two (e.g., machine learning, biostatistics, data science). Harvard's own website notes that "the Department of Statistics helps students acquire the conceptual, computational, and mathematical tools for quantifying uncertainty and making sense of complex data arising from many applications."

Especially as the influence of data and its analysis grows, statistics has become more important than ever. Learning statistics will give you better insight into how "statistical analyses" come to be: through the concentration, you can learn what a p-value is, what confidence intervals in political polls capture, how to compute how much you'll win from a gambling game, what linear regression is really "doing," and much more.

From our own experience (and this can easily be confirmed by most people at Harvard), the things you'll learn in the statistics department are super useful – and often necessary – for CS classes (e.g., CS 121, 124, 181, etc.), for econ classes (e.g., Ec 1011, 1123, 1126), for internships/interviews/jobs (e.g., in finance, SWE, biotech, data science), and for so many other things that we simply don't have room to list here.

Besides this, we've found – and we hope you will too – that statistics is a beautiful way of using math to learn about the world and other subjects. You'll also find that learning other subjects makes you a better statistician.

Courses in Statistics and Related Fields

Mathematics

Math forms the backbone of statistics, and a solid foundation in linear algebra and calculus is extremely helpful and almost necessary for most statistics courses at the level of Stat 110 and beyond.

Math Ma/Mb: Introduction to Functions and Calculus

Math Ma/Mb are designed for students who aren't ready for Math 1a but want/need to learn calculus. The two course sequence – which roughly covers precalculus and calculus – first teaches the mathematical fundamentals for calculus and then covers the content of Math 1a. After taking Math Ma/Mb, students are ready to take Math 1b.

Math 1a: Introduction to Calculus

Math 1a is the introductory calculus class at Harvard and is the first half of a one year sequence in single-variable calculus. In 1a, differential calculus is the focus. The class begins from elementary mathematical modeling, highlights the central role of rates of change to motivate the concept of derivatives, and culminates in the fundamental theorem of calculus and applications (including integrals and integral rules).

Given that differential calculus is a prerequisite for Stat 110, if you have never seen this material (or are very rusty), you should take this class. In particular, if you didn't take AP Calculus BC, then you probably will have to take this class. If you took it but are rusty, then refer to your Math Placement test to see if you still need to take it. In any case, before taking

Stat 110, it would be helpful to at least review derivative rules and the fundamental theorem of calculus, as well as integral rules.

Math 1b: Integration, Series and Differential Equations

The second half of the first-year single-variable calculus splits time between integral calculus, series, and differential equations. The first two thirds of the course contain material that is relevant for Stat 110, such as Taylor series, convergence / divergence tests for series, and integrals. Even so, a lot of the material on integrals relevant for STAT 110 is covered in Math 1a.

While a proactive student could reasonably go from Math 1a to STAT 110 by self-studying the aforementioned relevant material, it can still be smart to take Math 1b before taking Stat 110 if you have little to no familiarity with series and integration. Any questions regarding mathematical placement should be referred to the Math department. It is additionally highly recommended that you refer to your Math Placement test results; these will give you a gauge as to where you should begin.

Math 21a: Multivariable Calculus

Math 21a is an applied introduction to multivariable calculus. You'll spend a lot of time extending the knowledge of single-variable calculus you've built to multiple dimensions. For instance, we might be interested in taking a *partial derivative* of a function of multiple variables, which captures how the value of a function changes if we change only one input infinitesimally and hold the other inputs constant. You'll also learn to take integrals over more complex surfaces.

Math 21b: Linear Algebra and Differential Equations

Math 21b is an applied introduction to linear algebra. As opposed to 22a/25a/55a, Math 21b focuses primarily on the computational aspects of linear algebra. You'll learn how to work with linear transformations of multidimensional space, manipulate matrices meaningfully, and how to connect matrix operations to differential equations.

Math 22: Vector Calculus and Linear Algebra

If you're interested in dipping your toes into theoretical math but aren't too keen on having most of your free time consumed by actually doing math, then you might want to consider Math 22a (linear algebra) and Math 22b (multivariable calculus). The class assumes no prior knowledge in proof-writing and starts from the basics, providing a solid foundation for anyone who is interested in learning how to write proofs. Most students have not previously taken linear algebra or multivariable calculus, although students who did have prior experience with these topics in a non-proof context did not find the Math 22 material repetitive.

Math 25: Theoretical Linear Algebra and Real Analysis

A fast-paced and rigorous introduction to linear algebra in the fall (Math 25a) and to real analysis in the spring (Math 25b). Math 25 problem sets will eat up a decent chunk of your free time, but you'll leave the class with a solid and theoretical understanding of linear algebra and real analysis, and you'll feel confident writing proofs. The only truly necessary prerequisite is single-variable calculus at the level of AP Calculus BC, and the course does not expect students to have any prior experience with proofs. Some students come in having taken courses at the level of 21a/b and having done some competition math, but many students come in having done only one or none.

Math 55: Studies in Algebra/Group Theory/Analysis

Math 55a covers linear algebra, group theory, and some basic representation theory whereas Math 55b covers topology as well as some real and complex analysis. Both semesters are fast-paced and very rigorous, so be prepared to spend a lot of time on problem sets. The midterm is typically given before the add-drop deadline, which makes it a good diagnostic for whether this class is right for you. There are no formal prerequisites, but experience in writing proofs would be helpful. While all topics in this class are foundational to anyone wanting to pursue further study in pure mathematics, statistics majors will typically only encounter a handful of the topics in their coursework.

Computer Science

CS 50: Introduction to Computer Science

Ah yes, the classic Harvard CS 50 experience. Taught by the incredibly theatrical David Malan, CS 50 has one of the highest course enrollments at Harvard. This course is geared towards beginner programmers and covers many languages (C, Python, SQL, HTML, CSS, JavaScript) and introductory topics (data structures, algorithms, memory, databases, web development) in one semester. Many would say CS 50 is too fast-paced for a complete beginner who has never touched CS before, but if you dedicate time and effort to working on the labs then it's a fruitful experience. If you have taken AP Computer Science in the past then the topics might be too basic, but be sure to check the websites of past CS 50 iterations to evaluate whether the course content could be beneficial for you.

CS 181: Machine Learning

You've probably heard the term "machine learning" countless times, but what actually is machine learning? CS 181 gives a fundamental (albeit perhaps sometimes unglamorous) introduction to what machine learning is. You'll learn about the foundations and mathematics of machine learning, starting from basic clustering and regression methods, which then give rise to more flashy ideas like neural networks. While you might not learn all of the state-of-the-art ML ideas (e.g., CS 181 usually doesn't cover transformers in depth), the course gives you enough background to pick these up. Being familiar with manipulating expectations and probabilities is a must, as is working with matrices and vectors. Likelihood ratios figure fairly largely throughout the course, so having taken or concurrently taking Stat 111 can be helpful (and vice versa), though it certainly is not necessary.

CS/Stat 184: Introduction to Reinforcement Learning

Are you curious about how computers beat humans at Go, chess, and Starcraft? Have you wondered what gives a language model chatbot its "personality"? Reinforcement learning (RL) lies at the heart of these achievements. This subfield of machine learning investigates how an agent can learn to make decisions in an interactive environment. STAT 184 aims to give you the theoretical and programming skills to understand and develop RL algorithms. The course begins with simple environments where you can solve for the optimal strategy exactly. Then it explores more and more general cases that require more complex estimation and optimization techniques. The homeworks are a mix of mathematical problem solving and programming; There is a final project (no exam).

Statistics

Stat 110: Introduction to Probability

Stat 110 is what most people would consider to be the flagship course of the Harvard statistics department, and for good reason— the material you'll learn in this class is fundamental not just for any future work you'll do in statistics, both theoretical and applied, but also in related quantitative subjects like computer science, economics, engineering, and mathematics. It is also by far the largest course in the Stats department, enrolling over 500 students every fall.

For many students, this may be their first exposure to probability theory— while statistics courses like AP Statistics and Stat 100 focus more on practical applications of statistics for the purpose of data analysis, Stat 110 provides you with the theoretical probability toolbox that you'll need to be able to understand and quantify randomness and uncertainty in the world.

The majority of the content in Stat 110 revolves around understanding various probability distributions that are commonly used in statistics and data science, and the weekly problem sets and exams are designed to challenge you and help you to develop the problem-solving skills and mathematical maturity that will help you succeed in more advanced statistics courses.

While prior experience with advanced math and probability can be helpful, the only required background knowledge is single-variable calculus at the Math 1b level and familiarity with matrices.

Stat 111: Introduction to Statistical Inference

After having taken Stat 110, you'll hopefully be familiar with modeling various random phenomena using random variables, which you can then analyze using a variety of tools from the course. Stat 111 connects these tools with data and the methods/ideas that most people associate with "statistics": hypothesis testing, confidence intervals, estimation, and linear/logistic regression.

As opposed to Stat 110, most of the problem sets involve some coding in R (you can also use Python, but there isn't as much support). It's not necessary at all though to be familiar with R; if you're familiar at all with coding, you'll pick up R skill through the class. GUSH also typically runs a short R workshop before the class begins.

The main prerequisite for Stat 111, unsurprisingly, is Stat 110. The course assumes deep familiarity with the content from Stat 110 (in particular, details about the named distributions from Stat 110 and tools like Adam's law and Eve's law).

Stat 117: Data Analysis in Modern Biostatistics

As the longest-standing and most "advanced" undergraduate biostatistics course offered, Stat 117 benefits from both previous iterations of the course and Giovanni's expertise in the field. Stat 117 runs as a 30-person seminar style course where lecture notes and discussion prompts are posted beforehand. Attendance is required but participation is informal during the discussion-heavy classes—beware the 9am!

Content-wise, Stat 117 covers a wide range of biostatistical tools through the lens of biomarker cancer research. All projects and biweekly assignments center around two gene expression datasets; there are no exams. Giovanni provides all of the limited biomedical background needed and many additional resources and papers to understand the deeper statistical underpinnings of tools used in the course.

Familiarity with Stat 110, 111, and 139 is helpful for understanding and applying (in R) the tools in Stat 117.

Stat 139: Introduction to Linear Models

Stat 139 is the final course in the introductory statistics sequence of 110/111/139, and is probably the most applied out of the three. Like the course title suggests, the majority of the course is devoted to understanding the theory and applications of linear regression and related modeling techniques (e.g. LASSO and Ridge regularization, variable selection, etc.) and learning how to implement them in R. If you're interested in doing projects and research in applied statistics, Stat 139 will provide you with a solid foundation in common modeling methods that you'll use in fields including but not limited to biostatistics, sociology, and economics.

Stat/CS 109a is another undergraduate-level course which covers a similar set of topics. However, Stat 139 is more theoretical, has a larger emphasis on linear regression and its variations, and assignments are completed in R instead of Python.

Stat 210: Probability I

Stat 110's older sibling. Stat 210 is the first graduate statistics course most students take, and it's usually also taught by Joe, so you can expect a fairly similar lecture style. It's run somewhat like an upper-level undergraduate course: problem sets are still weekly, the class is large but is run like a well-oiled machine, and there is a midterm and a final exam.

Notably, in contrast to analogous graduate probability courses at most other universities, Stat 210 places greater emphasis on reasoning by representation (i.e., learning the relationship between different common distributions) instead of on measure theory, though measure theory still comes up from time to time.

Make sure you're solid on your Stat 110 content: familiarity with your Stat 110 distributions and stories, Adam's/Eve's law, LOTUS, etc. is assumed. .

Stat 211: Statistical Inference I

Stat 111's older sibling. While Stat 111 introduces you to the principles of statistical inference (likelihoods, estimators, hypothesis testing), Stat 211 dives into these topics on a much deeper and much more technical level. You'll prove a lot of the cool theorems taken for granted in Stat 111 (e.g., consistency and asymptotic normality of the MLE) and spend a *lot* more time on hypothesis testing (e.g., how do you define and find the *best* or most *powerful* test for a certain problem? what if we want to run *lots* of tests at the same time?). Expect lots of Taylor expansions and tricky homework/exam questions.

Stat 212: Probability II

Stat 212 is supposed to be a continuation of Stat 210, though it feels much more like a standard math course than one of Joe's statistics courses. Content is covered quickly in lecture, but problem sets are typically assigned biweekly. In Stat 212, you'll work closely with martingales and stochastic processes and potentially cover select advanced topics chosen by the instructor.

Math at the level of real analysis is essentially necessary, and knowing some measure theory from Math 114 can be super helpful.

Stat 213: Statistical Inference II

Stat 213 is a pretty natural continuation of Stat 211. You'll learn *when* the theorems in Stat 211 break down (e.g., when your likelihood function isn't "nice") and powerful tools to use in these situations (e.g., quadratic mean differentiability). You'll also be introduced to a powerful new perspective for asymptotic analysis: while any "decent" estimator/test should become perfect as the sample size grows to infinity, if we make the problem "harder" over time, certain estimators/tests seem to be "better."

Stat 244: Linear and Generalized Linear Models

At a high level, Stat 244 combines the content of Stat 139 and Stat 149 (fittingly, 44 is also the average of 39 and 49). The first half of Stat 244 covers linear models, and the second half covers generalized linear models. While these topics find common applications in industry, the focus of Stat 244 is primarily theoretical. The linear models aspect of Stat 244 also differs from Stat 139 in this regard; while Stat 139 includes coding in R, the biweekly problem sets in Stat 244 do not involve model implementation.

The course listing recommends Stat 210/211 as prior background, though concepts from these classes are not really used in the class. Stat 110, Stat 111, and familiarity with linear algebra is usually sufficient preparation for the course.

The Concentration

As data grows in size, complexity, and importance, interest in statistics and related yields has continued to grow. In 2008, Harvard's Department of Statistics had just 20 total concentrators; since 2018, we've consistently had over 200 total concentrators.

The concentration itself is built upon the foundational courses Stat 110 (Introduction to Probability) and Stat 111 (Introduction to Statistical Inference), which all concentrators are required to take for a letter grade by junior year (they're also often official or unofficial prerequisites for electives). Stat 139, which teaches concentrators to implement and interpret many frequently used statistical techniques, is also required for all concentrators.[1]

The exact requirements of the concentration (linked here) will differ slightly depending on which *concentration track* you pick. Ultimately, these tracks don't matter much: you can list them on your resumé, but it won't appear on your diploma.

[1] It's also possible to replace Stat 110 and/or Stat 111 with their graduate course equivalents, Stat 210/211, though almost all students will take the undergraduate version(s) before the graduate version(s). Similarly, it's possible to replace Stat 139 with its graduate equivalent, Stat 244.

Planning Your Statistics Courses

There's no single "correct" or "best" way, timeline, or pace to take statistics courses at Harvard. Some people will come to Harvard with tons of math experience and already determined to do statistics – which is great if that's the case – but so many (more than you'd expect!) eventual statistics concentrators come to Harvard not yet ready to dive right into statistics courses, whether it be because of their math background or their academic interests at the time. This is 200% fine, and it is 300% possible to finish the statistics concentration, even if you find later in college that you want to concentrate in statistics. For the sake of hopefully providing some clarity, we've listed some considerations for timing your courses, but these are in no way authoritative, necessary, required, or even suggestions:

A great deal of timing statistics courses centers directly or indirectly around Stat 110. Most concentrators take Stat 110 in the fall of their sophomore year. Though there's no need to rush through the course, it can be nice to knock out Stat 110 once you're ready because it'll open up a wide range of courses that list it (officially or unofficially) as a prerequisite: e.g., Stat 111/171/210, CS 124/181, Ec 1011a/1126.

Freshman Year:

Freshman year can be a great time (a) to explore various courses related to (and unrelated to!) statistics and (b) to take math courses to satisfy the math requirement and prepare for statistics courses. There's no need to take a statistics course in your freshman year if it doesn't make sense for you, but some future concentrators will take:

- Stat 100. Stat 100 is an application-driven introduction to statistics. While the course is not a required course for the concentration, it does (a) count toward the total number of statistics courses required for concentrators, (b) give good practice with R that can be helpful for Stat 111, and (c) help introduce *what* statistics is. It is, however, very different from Stat 110 and Stat 111 in terms of style and content – it should not be thought of as a prerequisite for those courses.

- Stat 110. Some students with strong math backgrounds will take Stat 110 their freshman year. If you're one of these students, it'll give you a nice head start and give you early exposure to statistics, but we want to reiterate that most students do *not* take Stat 110 until sophomore year or later.
- Math courses. Stat 110 will require some familiarity with calculus, so it can be helpful for students to take Math 1a and 1b in their freshman year if they are rusty or have not taken calculus in high school. A proactive student can take Stat 110 with just Math 1a (or Ma/Mb) under their belt: note that the calculus portion of 110 requires knowledge of integration and some knowledge of series and Taylor expansions.

  Students placed into Math 21+ will often take one of Math 21ab, 22ab, 25ab, 55ab: these satisfy the concentration's two course math requirements of linear algebra and multivariate calculus/real analysis. Additionally, while Stat 110 doesn't require formal knowledge of proofs, having experience with proof-based reasoning can be helpful for Stat 110.

  It is also possible to satisfy the math requirements with other courses. Some concentrators, for instance, will take Math 112 instead of 22b for a formal introduction to real analysis.

Sophomore Year:

Most concentrators – though, again, we want to emphasize this isn't true for all concentrators! – will take Stat 110 and Stat 111 in the fall and spring of their sophomore year, respectively.

Junior and Senior Year:

The statistics concentration requires that both Stat 110 and 111 are taken by the end of your junior year. Apart from those courses, students' journeys through the statistics concentration varies a ton. A few considerations:

- At this point, students still have the linear models (Stat 139) requirement: there isn't as much uniformity as to when students satisfy this requirement.
- If you are considering pursuing the concurrent master's program, you will need to have satisfied the three course requirements for the program by the end of the fall semester of your junior year. Because of this, many of these students will take Stat 110 and 111 in their sophomore year and Stat 210 in the fall of their junior year, though this is certainly not the only way that students satisfy the concurrent master's application requirements.

Sample Course Plans

Student starting in Ma

	Freshman	Sophomore	Junior	Senior
Fall	Math Ma CS 50	Math 1b Stat 110	Math 21b	Stat 139 Stat 108
Spring	Math Mb Stat 100	Math 21a CS 51	Stat 111	Stat 149 Stat 171

Student starting in 1a

	Freshman	Sophomore	Junior	Senior
Fall	Math 1a	Math 21a Stat 110	Stat 139	CS 109a Stat 99r
Spring	Math 1b Stat 100	Math 21b Stat 111	CS 32 Stat 171	Stat 141 Stat 99r

Student starting in 21a, on the data science track

	Freshman	Sophomore	Junior	Senior
Fall	Math 21a CS 50	Stat 110 CS 109a	Stat 139 CS 61	
Spring	Math 21b	Stat 111 CS 181	Stat 171 CS 124	CS 51

Student starting in 22a, completing the concurrent master's program

	Freshman	Sophomore	Junior	Senior
Fall	Math 22a CS 50	Stat 110 CS 109a	Stat 210 Stat 139	Stat 211 Stat 244
Spring	Math 22b	Stat 111 CS 181	Stat 220	Stat 288

Student starting in 25a, accelerating through statistics courses

	Freshman	Sophomore	Junior	Senior
Fall	Math 25a CS 50 Stat 110	Stat 210	Stat 211 Stat 244	Stat 99r
Spring	Math 25b Stat 111	Stat 212 CS 181	Stat 213	Stat 99r

Concurrent Master's (A.B./A.M.)

A handful of students each year (e.g., sixteen in the Class of 2024) are enrolled in the concurrent master's program, which allows students to earn their bachelor's degree (A.B.) and their master's degree (A.M.) in four years; that is, they'll graduate with their undergraduate class but with an additional master's degree in statistics.

To enter the concurrent master's program, students are required to formally apply to the program at the end of their junior fall (rules are slightly different for off-cycle students) through GSAS. In order to be accepted, students must have completed at least three statistics courses, which must include: (1) Stat 110 or 210 (with an A- or higher), (2) Stat 111 or 211 (with an A- or higher), and (3) at least one 200-level course (with a B or higher) if not Stat 210 or 211.

For the application, you'll need to fill out a plan of study, meet with Dr. Glickman (the director of the concurrent master's program in statistics) and get two letters of recommendation. The two letters of recommendation sound much scarier than they actually are: usually, the statistics department just wants them to vouch for your ability to finish the concurrent master's program.

Statistics is not the only field in which you can pursue the concurrent master's program. A number of SEAS fields (e.g., computer science, applied math, applied physics) have A.B./S.M. concurrent master's programs, and there are also a handful of other fields in GSAS (e.g., pure math) that offer the A.B./A.M. concurrent master's.

Research

Getting involved in research can be a super valuable and rewarding way to augment the rest of your statistics journey at Harvard. Broadly speaking, research is split into applied research and theoretical research (though this split isn't as clean as we might make it seem). Below, we outline what applied and theoretical research look like, but we want to emphasize that research – by its very nature – is highly variable, so your experience may not perfectly match our descriptions.

Applied Research

Applied research often consists of writing code to clean, analyze, and visualize complex datasets. Often, you'll work with a PI (often a faculty member), postdocs, and graduate students to accomplish data-related tasks, such as producing tables and figures, implementing estimators and algorithms, and performing exploratory analyses.

The process of applied research is often very iterative: you'll write code to accomplish some task, get feedback on what changes to your result might be desirable, and then figure out how to adjust your code to accomplish these changes.

Usually, for applied research, you'll join a project as a member of a team: the amount of control you have on the overall direction of a project can vary by lab. Sometimes, you might be in charge of a subproject; sometimes, you might jump around from different parts of the project, providing coding support as necessary. That is, applied research can take the form of being a research assistant (RA) for a professor, but it can also consist of more independent work that is supervised by the professor.

Theoretical Research

Theoretical research is more geared toward proving certain statements rigorously. This might involve proving lemmas as intermediate steps toward proving a desired theorem/result. Often, you'll work closely with a faculty member who can guide you through intermediate lemmas and provide advice for proving certain results.

Theoretical research can seem daunting at first, but in some ways, it's not too different from a problem set for a theory class. Usually, you'll have small things that you try to prove between each meeting with an advisor; research meetings can then be used to think through tricky parts of a proof with your advisor and review the general direction of a research project.

As opposed to a problem set, however, theoretical research projects usually have much more lofty goals, so keeping the big picture in mind is important while working on smaller intermediate proofs. Additionally, while the claims that you're asked to prove on a problem set are (hopefully) known to be true, it's not always immediately clear that a certain lemma or theorem in a research project is true: because of this, simulation, intuition, and heuristics play a large role in research. This part of research – trying to figure out what things are actually true and what aren't – can be immensely frustrating but can also lead to new research ideas and insights.

Theoretical research often has a higher barrier to entry than applied research. Many advisors will often require that you have taken certain courses to work with them, since their research might involve concepts only covered in upper-level or graduate courses.

Finding An Advisor

Like the rest of this section, there's no one way to find an advisor/lab. We've included some considerations:

- The first step in finding an advisor/lab is often to figure out what you're interested in. For applied research, this might look like finding an application area that you're interested in (e.g., genomics, behavioral economics, or voting behavior). For theoretical research, this might look like finding a subfield that you're interested in (e.g., free probability, high-dimensional inference, conformal inference, etc.). You can definitely also shape your interests around advisors/labs you would want to work with, but it's worth considering whether you're really interested in the relevant topics – it'll make research a lot more engaging!
- There's a few ways to find an advisor/lab. You might want to work with a professor you've taken a class with, you might browse through faculty members/labs on a department website, and some departments (e.g., the Department of Economics or the School of Public Health) have programs for matching students to a project during different parts of the year.
- If there's no formal process for matching students with a project/advisor/lab, you'll often have to reach out on your own! You'll often want to know some things about the project/advisor/lab, both (a) to make sure you're actually interested and (b) to demonstrate that you know what you're signing up for. Then, you can reach out, asking if there are any spots for undergraduate students (sometimes there aren't!).

Funding and Getting Course Credit

Research can be great on its own, but it's even better when you can get paid or get course credit for your work! There's a few different opportunities for different parts of the year:

Faculty/Department Funding

A few faculty/labs are able to pay research assistants directly for the work that they do, though this is more common in departments outside of statistics and happens more often for applied research (e.g., many economics professors pay their RAs on an hourly basis). Often, even these groups will still ask undergraduate RAs to apply for funding (i.e., the opportunities listed below).

Stat 91r

Stat 91r is one way to receive concentration credit for term-time research. It must be taken as SAT/UNS and can count as one of the related courses for the concentration. To "take" Stat 91r, you'll need to already have a faculty advisor, fill out the Stat 91r proposal form, receive permission from Dr. Rader, and enroll in the course. There isn't actually any formal course content; rather, at the end of the semester, your faculty advisor will give you a SAT or UNS grade.

Notably, you may *not* enroll in Stat 91r and be paid for your research work. i.e., you cannot be enrolled in Stat 91r and receive a stipend from HCRP.

HCRP/Herchel Smith

HCRP is a research stipend offered by Harvard for term-time (up to $1000) or summer research (up to $6000) supervised by a Harvard faculty member. The HCRP application consists of a research plan (4-5 double-spaced pages) that outlines your project's objectives, importance, and methods and a letter of recommendation written by the faculty member who will supervise your research.

Herchel Smith is a research stipend offered by Harvard for summer research. As opposed to HCRP, Herschel Smith allows for a non-Harvard affiliated faculty advisor and provides a slightly higher stipend. The application is similar to that of HCRP but also includes a short essay about how receiving the stipend will help you achieve your academic and professional goals.

Research Village (PRISE/SPUDS/BLISS/KRANIUM/etc.)

The Harvard Summer Undergraduate Research Village (HSURV for short) is a collection of different residential research programs hosted by Harvard that run during the summer. Each program is focused on a different discipline – SPUDS is focused on data science, BLISS is focused on social sciences, etc. – and run slightly differently, but most programs consist of research work with an advisor/mentor/lab, occasional professional development/faculty presentations, and a final presentation at the conclusion of the program.

Many of the programs will ask that students arrange advisors/labs/projects themselves before submitting their applications (the specific project is usually not binding), though a few programs have students apply to specific projects.

Students receive room and board for the summer and a moderate stipend (around $3000). HSURV puts a great emphasis on the community aspect of the program: Harvard pays for various activities throughout the summer that are separate from research (e.g., trips to amusement parks, attractions, festivals), there are (required) opening and closing dinners, and students all live in the same house for the duration of the summer (e.g., most years, all students live in Winthrop).

Acknowledgements

This guide itself is massively inspired by the SPS Guide to Physics and the Harvard Chemistry Club's Guide to Chemistry, and we can only hope to provide a portion of the guidance that those manuals have given students over the years.

This guide wouldn't have been possible without the invaluable help from a great number of individuals who volunteered their time to give helpful feedback, write blurbs for classes, and provide suggestions.

GUSH takes responsibility for any mistakes made in this guide and caution that much of the text in this guide is opinion and is based on our own experiences during our times at Harvard; i.e., you might have a completely different experience than us!

Go forth and study statistics!

GUSH