Harvard University PhD Program in Health Policy
Methods for Policy Research Concentration
2023-2024

TABLE OF CONTENTS

INTRODUCTION

General description of the Methods for Policy Research (MPR) concentration:

Training in the MPR concentration will position students to conduct rigorous, policy-relevant research. The concentration includes intensive methodological training that draws from several quantitative fields, including statistics, econometrics, and epidemiology, and also offers opportunities to learn qualitative methods. Students in the MPR concentration will also develop proficiency in experimental and quasi-experimental research designs. The MPR concentration does not require training in a specific discipline of social science, but students are allowed to use two of their distributional requirements to acquire some disciplinary training (e.g., in economics or sociology), and many students pursue additional coursework in a substantive area of interest or advanced levels of statistics. By virtue of this training, MPR students develop a multi-disciplinary toolkit that borrows from the strengths of different quantitative sciences and supports novel applications to health-related questions. Graduates are able to communicate and collaborate effectively with statisticians, economists, epidemiologists, and clinician-investigators, as they lead their independent work. The concentration's broad methodological training flexibly supports scholarship on a wide range of topics, including common areas of interest in health services research related to health care access, quality, costs, and disparities, but also questions focused on clinical decision-making, behavior (e.g., of patients, physicians, or organizations), social determinants of health, and social programs with implications for health. Graduates of the MPR concentration have used their training to study, for example:

POLICY EVALUATION
- [Impact of the ACA Dependent Coverage Insurance Expansions on Prenatal Care and Birth Outcomes](#)

QUALITY OF CARE AND CLINICAL DECISION-MAKING
- [Soft Consolidation In Medicare ACOs: Potential For Higher Prices Without Mergers Or Acquisitions](#)

DISPARITIES
- [Fair Regression for Health Care Spending](#)
- [Association of Nondiscrimination Policies With Mental Health Among Gender Minority Individuals](#)

RANDOMIZED CONTROLLED TRIAL
- [Impact of a Text-Messaging Program on Adolescent Reproductive Health: A Cluster–Randomized Trial in Ghana](#)

POPULATION HEALTH
- [Association of Prenatal Care Services, Maternal Morbidity, and Perinatal Mortality With the Advanced Maternal Age Cutoff of 35 Years](#)

METHODS
- [Birds of a Feather Flock Together: Comparing Controlled Pre–Post Designs](#)
- [Matching and Regression to the Mean in Difference-in-Differences Analysis](#)
- [Testing the Validity of the Single Interrupted Time Series Design](#)

Courses for students in the methods for policy research concentration:

The course requirements for the Methods for Policy Research concentration were selected to provide students with important skills needed for conducting original health policy research; at the completion of coursework, students should be able to propose feasible study designs to answer health policy questions, as well as identify the strengths and limitations of the various designs in their proposed work and in other published studies. They also will develop strong analytical skills, including the technical expertise required to analyze data as well as interpret results, identify the strengths and limitations of analyses, and the broader implications of results for future health policy.

The following course requirements encompass both methods and research design. Students are required to take at least 9 credits of methods. The 9 credits include required coursework in research design, probability, inference, and regression, and additional coursework in other specific areas from which students can choose (e.g., program evaluation, hierarchical modeling, causal inference, machine learning, analysis of survey data, or qualitative methods). **Incoming students should also be aware that many of the intermediate and advanced courses have prerequisites. In addition, some courses are offered only in alternative academic years, and scheduling conflicts do occur. Due to these issues, incoming students should plan a tentative two-year course program early in the first year. They should also consult with more senior students about their tentative program for additional feedback and suggestions.** In planning their required coursework in probability, inference, and regression, students are encouraged to follow a basic sequence of courses *within a particular school or department* (i.e., Biostatistics, Statistics, Economics, or HKS) as opposed to selecting from multiple schools/departments. This is recommended in order to facilitate a more coherent presentation of the fundamentals. Subsequently, students are encouraged to choose courses based on interests and career objectives from among all schools/departments and to expose themselves to the diversity of languages, emphases, and approaches.

OVERVIEW OF MPR AND PROGRAM COURSE REQUIREMENTS

MPR Requirements (9 credits)
- Probability Theory (1 credit)
- Statistical Inference (1 credit)
- Regression (1 credit)
- Impact Evaluation (1 credit)
- MPR Reading Course (1 credit)
- Methods and Analysis (4 credits); must be from at least 2 of the categories:
  - Statistics and Econometrics
  - Survey Research Methods
  - Causal Inference
  - Decision Sciences
  - Epidemiology
  - Machine Learning
  - Network Analysis
  - Program Evaluation
  - Qualitative Research
  - Other

1 credit = 4 units at FAS/HKS/GSE = 5 units at HSPH

Program and Distribution Requirements (6 credits)
- Health Policy Core Course (2 credits)
- Health Policy Research Seminar (1 credit)
- Distribution requirements (3 credits, spanning at least 2 of the following areas)
  - Decision Sciences
  - Economics
  - Ethics
  - History
  - Health Equity and Social Determinants
  - Management
  - Mental Health Policy
  - Political Analysis

Notes:

1) Students in the MPR track are **strongly** encouraged to take a course in Causal Inference. Students are also encouraged to supplement their curriculum with social sciences classes to gain a well-rounded understanding of the health care system. Finally, students should make an effort to regularly attend research design seminars that are held around the university.

2) MPR students may "double count" Decision Sciences courses towards a PhD distribution requirement and an MPR program requirement, with one exception. Students cannot double count both RDS280/285 and API 222 for both MPR requirements as well as distribution.

3) Certain courses are listed more than once under different MPR requirements, but they CANNOT count toward more than one requirement (e.g., BST 222 can count towards either Probability Theory or Statistical Inference, but not both). The classes that are listed under multiple requirements but that only count toward one requirement are:
   - BST 222 counts for Probability Theory or for Statistical Inference
   - Econ 2110 and Gov 2001 count for Statistical Inference or for Regression
   - Econ 2115 counts for Statistical Inference, Regression, or Methods and Analysis (Causal Inference)
   - BST 232 counts for Regression or for Methods and Analysis (Statistics and Econometrics)
   - Statistics 149 counts for Regression or for Methods and Analysis (Statistics and Econometrics)
   - Statistics 186 and Government 2003 count for Statistical Inference or for Methods and Analysis (Causal Inference)

4) To satisfy all MPR, program, and distribution requirements, MPR students should plan on taking 3-4 credits each semester in their first and second years. Students may also take courses beyond the second year if they still have requirements to fulfill or want to take additional coursework.

5) Students may petition the MPR concentration chairs for credit if they have previously taken one of the courses listed in this packet.

6) Students who have taken a class similar to Statistics 110 should contact Mary Beth Landrum. These students may take a more advanced probability course to satisfy the Probability Theory requirement. Alternatively, they may take an additional statistics course instead of another probability course to satisfy the Probability Theory requirement.

7) Students may petition the MPR concentration chairs if they would like a non-listed course to count towards a requirement. They may petition the DGS if they would like a non-listed course to count towards a distribution requirement.

METHODS COURSE REQUIREMENTS IN DETAIL

Class	Instructor	Semester	Time
Probability Theory (1 credit)			
Statistics 110: Introduction to Probability	Blitzstein	Fall	TuTh 1:30–2:45
Statistics 171[1]: Introduction to Stochastic Processes	Sen	Spring	TBA
Statistics 210[1]: Probability I	Blitzstein	Fall	TuTh 10:30–11:45
BST 222: Basics of Statistical Inference	Duan	Fall	MW 11:30–1
Biostatistics 230 / BST 230: Probability 1	Miller	Fall	TuTh 9:45–11:15
Statistical Inference (1 credit)			
Statistics 111: Introduction to Statistical Inference	Blitzstein	Spring	TuTh 1:30–2:45
Statistics 186: Introduction to Causal Inference	Murphy	Spring	TBA
Statistics 211: Statistical Inference I	Janson	Fall	MW 9–10:15
Economics 2110: Econometrics I	Bruich	Fall	MW 1:30–2:45
Economics 2115: Econometric Methods II	Dobbie	Spring	MW 1:30–2:45
Government 2001: Quantitative Social Science Methods I	Imai	Spring	MW 3–4:15
BST 222: Basics of Statistical Inference	Duan	Fall	MW 11:30–1
Biostatistics 231 / BST 231[1]: Statistical Inference I	Wang	Spring	MW 9:45–11:15
API 209: Advanced Quantitative Methods I: Statistics	Levy	Fall	TuTh 1:30–2:45
Regression (1 credit)			
Statistics 139: Introduction to Linear Models	TBA	Fall	TuTh 1:30–2:45
Statistics 149: Introduction to Generalized Linear Models	Glickman	Spring	TBA
Statistics 244[1]: Linear and Generalized Linear Models	TBA	NA	Not offered 2023-2024
Economics 1123: Introduction to Econometrics	Pettenuzzo (F) Bruich (S)	Fall/Spring	(F) MW 9–10:15 (S) TuTh 3–4:15
Economics 2110: Econometrics I	Bruich	Fall	MW 1:30–2:45
Economics 2115: Econometric Methods II	Dobbie	Spring	MW 1:30–2:45
Economics 2120: Principles of Econometrics	TBA	Fall	MW 12–1:15
Government 2002: Quantitative Social Science Methods II	Blackwell	Fall	TuTh 1:30–2:45
BST 211: Regression & Analysis of Variance in Experimental Research	TBA	NA	Not offered 2023-2024
BST 213: Applied Regression for Clinical Research	Orav	Fall	MW 8–9:30
Biostatistics 232 / BST 232: Methods	Nethery	Fall	MW 8–9:30
Biostatistics 235 / BST 235[1]: Advanced Regression and Statistical Learning	Lu	Fall	MW 9:45–11:15
GHP 525: Econometrics for Health Policy	Cheng	Fall	TuTh 8–9:30
API 210: Advanced Quantitative Methods II: Econometric Methods	Dobbie	Spring	TuTh 10:30–11:45
Impact Evaluation (1 credit)			
GHP 228: Econometric Methods in Impact Evaluation	Cohen	Spring	F 8–11:15
MIT 14.387: Advanced Econometrics	Angrist	Fall	TBA
MPR Reading Course (1 credit)			
Health Policy 3080A/B : Graduate Reading Course: Methods for Policy Research	Landrum McWilliams	Fall/Spring	TBA
Statistics 393: Design of Experimental and Non-Experimental Studies[2]	TBA	Spring	TBA
Methods and Analysis (4 credits); must be from at least 2 of the following categories			
Statistics and Econometrics			
Statistics 109A / COMPSCI 109A: Data Science 2: Intro to Data Science	Protopapas Rader	Fall	MWF 9:45–11
Statistics 109B / COMPSCI 109B: Data Science 2: Advanced Topics in Data Science	Protopapas Rader	Spring	MWF 9:45–11
Statistics 120: Introduction to Bayesian Inference and Applications	TBA	NA	Not offered 2023-2024
Statistics 131: Introduction to Time Series and Prediction	Ke	Spring	TBA
Statistics 149: Introduction to Generalized Linear Models	Glickman	Spring	TBA

---

[1] Advanced or very challenging course; please check prerequisites

[2] This course is non-letter graded and therefore does not meet an MPR course requirement, but it may be particularly well suited for G3s and G4s, who have not taken Stat 293, as a reading course. For information about Statistics 293, see below under Causal Inference.

Statistics 151 / EDU S043: Multilevel and Longitudinal Models	Miratrix	Fall	TuTh 12–1:15
Statistics 160: Introduction to Survey Sampling and Estimation	McConville	Spring	TBA
Statistics 220[1]: Bayesian Data Analysis	Liu	Spring	TBA
Economics 2140[1]: Econometric Methods	TBA	Spring	TuTh 1:30–2:45
Sociology 2211: Analysis of Longitudinal Data: Seminar	TBA	NA	Not offered 2023-2024
BST 210: Applied Regression Analysis	Lake (F) Thaweethai (S)	Fall/Spring	(F) TuTh 11:30–1; (S) TuTh 8–9:30
BST 223: Applied Survival Analysis	Haneuse	Spring	TuTh 9:45–11:15
BST 226: Applied Longitudinal Analysis	Fitzmaurice	Spring	TuTh 2–3:30
BST 228: Applied Bayesian Analysis	TBA	NA	Not offered 2023-2024
BST 232: Methods	Nethery	Fall	MW 8–9:30
BST 245[1]: Analysis of Multivariate and Longitudinal Data	Chen	Fall	TuTh 11:30–1
BST 249[1]: Bayesian Methodology in Biostatistics	Trippa	Spring	TuTh 2–3:30
SBS 263: Multilevel Statistical Methods: Concept and Application	Subramanian	Spring	F 9:45–12:45
EDU S052: Intermediate and Advanced Statistical Methods for Applied Educational Research	Ho	Fall	(F) TuTh 10:30–11:45
Survey Research Methods			
Statistics 160: Introduction to Survey Sampling and Estimation	McConville	Spring	TBA
BST 212: Survey Research Methods in Community Health	Mangione	Spring	W 3:45–5:15
BST 239: Health Survey Samples	TBA	NA	Not offered 2023-2024
Causal Inference (highly recommended)			
Economics 2115: Econometric Methods II	Dobbie	Spring	MW 1:30–2:45
Government 2003 / Statistics 286: Causal Inference with Applications	Imai	Fall	MW 10:30–11:45
Statistics 186: Introduction to Causal Inference	Murphy	Spring	TBA
Statistics 293: Design of Experimental and Non-Experimental Studies[3]	Zubizarreta	Spring	TBA
EPI 201: Introduction to Epidemiology - Methods I	Mittleman	Fall	MW 9:45–11:15 or MW 11:30–1
EPI 289[4]: Epidemologic Methods III: Models for Causal Inference	Dickerman Shi	Spring 1	MW 9:45–11:15
Decision Sciences			
RDS 280: Decision Analysis for Health and Medical Practices	Pandya	Fall 2	TuTh 2–3:30
RDS 282: Economic Evaluation of Health Policy and Program Management	Resch	Spring 2	MW 2–3:30
RDS 284: Decision Theory	Hammitt	Fall	MW 11:30–1
RDS 285: Decision Analysis Methods in Public Health and Medicine	Menzies	Spring 1	MW 2–3:30
API 302 / ECON 1415: Analytic Frameworks for Policy	Zeckhauser	NA	Not offered 2023-2024
Epidemiology			
EPI 202: Epidemiologic Methods II: Elements of Epidemiologic Research	Mittleman	Fall 2	(1) TuTh 9:45–11:15; (2) TuTh 11:30–1
EPI 203[5]: Study Design in Epidemiologic Research	Hernandez-Diaz, Lipsitch	Spring 2	TuTh 3:45–5:15
EPI 204: Analysis of Case-Control, Cohort, and Other Epidemiologic Data	Song	Spring 2	MW 9:45–11:15
EPI 207: Advanced Epidemiologic Methods	Robins	Fall 1	MW 3:45–5:15
Machine Learning			
API 222: Machine Learning and Big Data Analytics	Saghafian	Spring	TuTh 12–1:15
BST 263: Statistical Learning	Coull	Spring	TuTh 11:30–1
MIT 6.7910J/9.520J: Statistical Learning Theory and Applications	Poggio	Fall	TuTh 11–12:30
MIT 14.388: Inference on Causal and Structural Parameters Using ML and AI	TBA	NA	Not offered 2023-2024
Government 2004: Introduction to Machine Learning	TBA	NA	Not offered 2023-2024

[3] This course meets jointly with Statistics 393. Statistics 293 requires writing a final paper and is letter-graded, whereas Statistics 393 does not require a final paper and is graded SAT/UNSAT. Only Statistics 293 will satisfy an MPR course requirement, but Statistics 393 may be particularly well suited for G3s and G4s.

[4] The order of the epidemiology sequence goes: EPI 201 -> EPI 202 -> EPI 289 -> EPI 204 -> EPI 207. The sequence must be taken in order.

[5] EPI 203 can be taken as a standalone course without following the epidemiology sequence. It is more conceptual than analytical.

Network Analysis			
BST 267: Introduction to Social and Biological Networks	TBA	NA	Not offered 2023-2024
Program Evaluation			
API 211 / EDU A164: Program Evaluation and Education Policy	Kane	Spring	TBA
HPM 543: Quantitative Methods for Policy Evaluation	Curto	Spring 2	MW 9:45–11:15
Qualitative Research			
GHP 504: Introduction to Qualitative Research for Global Health	Yousafzai	Spring 1	MW 3:45–5:15
SBS 288: Qualitative Research Methods in Public Health	Goldman	Fall 1	F 9:45–12:45
SBS 521: Qualitative Data Analysis for Public Health	Ramanadhan	Fall 2	TuTh 11:30–1
EDU S504: Introduction to Qualitative Research	Duraisingh	Fall	F 9–11:45
EDU S510F: Qualitative Research Methods in Practice	Duraisingh	Spring	TuTh 12-1:15
EDU S513: Introduction to Qualitative Program and Policy Evaluation	Sacks	Spring	M 9–11:45
Other			
SBS 245: Social and Behavioral Research Methods	Chen	Fall	F 2–5
HBSDOC 4070: Design of Field Research Methods	Ely	Fall	Th 9–12
APMTH 121: Optimization: Models and Methods	Levine	Fall	MWF 9–10:15

FACULTY MEMBERS ASSOCIATED WITH THE MPR CONCENTRATION

Concentration Chairs

Mary Beth Landrum
> Professor of Health Policy (Biostatistics), Harvard Medical School

J. Michael McWilliams
> Warren Alpert Foundation Professor of Health Care Policy and Professor of Medicine, Harvard Medical School

Associated Faculty

Michael Barnett
> Associate Professor of Health Policy and Management, Harvard T.H. Chan School of Public Health

Niteesh Choudhry
> Professor of Medicine, Harvard Medical School

Jessica Cohen
> Bruce A. Beal, Robert L. Beal, and Alexander S. Beal Associate Professor of Global Health, Harvard T.H. Chan School of Public Health

Benjamin Le Cook
> Associate Professor of Psychiatry, Harvard Medical School

Arnold M. Epstein
> John H. Foster Professor of Health Policy and Management, Harvard T.H. Chan School of Public Health
> Professor of Medicine and Health Care Policy, Harvard Medical School

Jose Figueroa
> Assistant Professor of Health Policy and Management, Harvard T.H. Chan School of Public Health, and Assistant Professor of Medicine Harvard Medical School

Laura Garabedian
> Assistant Professor of Population Medicine, Harvard Medical School and Harvard Pilgrim Healthcare Institute

Laura Hatfield
> Associate Professor of Health Care Policy (Biostatistics), Harvard Medical School

David Hemenway
> Professor of Health Policy, Harvard T.H. Chan School of Public Health

John Hsu
> Associate Professor of Medicine, Harvard Medical School

Gary King
> Albert J. Weatherhead III University Professor

Nancy Keating
> Professor of Health Care Policy and of Medicine, Harvard Medical School

Bruce Landon
> Professor of Health Care Policy and of Medicine, Harvard Medical School

Dan Levy
> Senior Lecturer in Public Policy, Harvard Kennedy School

Alex McDowell
> Assistant Professor of Medicine, Harvard Medical School

Barbara J. McNeil
> Ridley Watts Professor of Health Care Policy, Harvard Medical School
> Professor of Radiology, Harvard Medical School

Ateev Mehrotra
> Professor of Health Care Policy and of Medicine, Harvard Medical School

Sharon-Lise Normand
> Professor of Health Care Policy (Biostatistics), Harvard Medical School
> Professor in the Department of Biostatistics, Harvard T.H. Chan School of Public Health

Dennis Ross-Degnan
> Associate Professor of Population Medicine, Harvard Medical School and Harvard Pilgrim Healthcare Institute

Stephen Soumerai
> Professor of Population Medicine, Harvard Medical School and Harvard Pilgrim Healthcare Institute

Theodore Svoronos
> Lecturer in Public Policy, Harvard Kennedy School

Alan Zaslavsky
> Professor of Health Care Policy (Statistics), Harvard Medical School

Jose Zubizarreta
> Professor of Health Care Policy, Harvard Medical School

METHODS COURSE DESCRIPTIONS

Probability Theory (1 credit)

Statistics 110. Introduction to Probability
A comprehensive introduction to probability. Basics: sample spaces and events, conditional probability, and Bayes' Theorem. Univariate distributions: density functions, expectation and variance, Normal, t, Binomial, Negative Binomial, Poisson, Beta, and Gamma distributions. Multivariate distributions: joint and conditional distributions, independence, transformations, and Multivariate Normal. Limit laws: law of large numbers, central limit theorem. Markov chains: transition probabilities, stationary distributions, convergence.
Prerequisite: Mathematics 1b or equivalent or above.

Statistics 171. Introduction to Stochastic Processes
An introductory course in stochastic processes. Topics include Markov chains, branching processes, Poisson processes, birth and death processes, Brownian motion, martingales, introduction to stochastic integrals, and their applications.
Prerequisite: Statistics 110 and Mathematics 21a and 21b, or equivalent.

Statistics 210. Probability I
Random variables, measure theory, reasoning by representation. Families of distributions: Multivariate Normal, conjugate, marginals, mixtures. Conditional distributions and expectation. Convergence, laws of large numbers, central limit theorems, and martingales.
Prerequisite: Statistics 110 or equivalent required.

BST 222. Basics of Statistical Inference
This course will provide a basic, yet thorough introduction to the probability theory and mathematical statistics that underlie many of the commonly used techniques in public health research. Topics to be covered include probability distributions (normal, binomial, Poisson), means, variances and expected values, finite sampling distributions, parameter estimation (method of moments, maximum likelihood), confidence intervals, hypothesis testing (likelihood ratio, Wald and score tests). All theoretical material will be motivated with problems from epidemiology, biostatistics, environmental health and other public health areas. This course is aimed towards second year doctoral students in fields other than Biostatistics. Background in algebra and calculus required.
Prerequisite: BST210 or BST213 or PHS 2000A&B.

Biostatistics 230 /BST 230. Probability I
Axiomatic foundations of probability, independence, conditional probability, joint distributions, transformations, moment generating functions, characteristic functions, moment inequalities, sampling distributions, modes of convergence and their interrelationships, laws of large numbers, central limit theorem, and stochastic processes.
Prerequisite: You must be a Biostatistics student or have taken BST222 to register for this course. If you have taken BST222 and are not a Biostatistics student, please ask the instructor for an instructor override.

Statistical Inference (1 credit)

Statistics 111. Introduction to Statistical Inference
The course is designed for undergraduates as their first introduction to rigorous statistical inference. Understanding the foundations will allow you to see more deeply into individual methods and applications, placing them in context and able to learn new ones (and invent new ones!) much faster having understood broad principles of inference.
Prerequisite: Mathematics 19a and 19b or equivalent and Statistics 110.

Statistics 186. Causal Inference
Substantive questions in empirical scientific and policy research are often causal. Does voter outreach increase turnout? Are job training programs effective? Can a universal health insurance program improve people's health? This class will introduce students to both statistical theory and practice of causal inference. As theoretical frameworks, we will discuss potential outcomes, causal graphs, randomization and model-based inference, sensitivity analysis, and partial identification. We will also cover various methodological tools including randomized experiments, regression discontinuity designs, matching, regression, instrumental variables, difference-

in-differences, and dynamic causal models. The course will draw upon examples from political science, economics, education, public health, and other disciplines.
Prerequisite: (1) Stat 110 AND Stat 111 or (2) Gov 2000 AND Gov 2001.

Statistics 211. Statistical Inference I
Foundations of frequentist and Bayesian inference, and decision theory. Likelihood, sufficiency, and ancillarity. Point estimation, unbiasedness, maximum likelihood, method of moments, minimum-variance. Parametric and non-parametric hypothesis testing, confidence intervals. Selective inference: multiple testing, familywise error rate, false discovery rate. Bayesian inference, conjugate priors, credible intervals. Admissibility, Stein's phenomenon, empirical Bayes. Time permitting: post-selection inference and the bootstrap.
Prerequisite: Statistics 111 and 210a or equivalent.

Economics 2110. Econometrics I
Economics 2110 and 2115 comprise a two-course sequence for first-year graduate students seeking training in econometric methods at a level that prepares them to conduct professional empirical research. Economics 2110 (fall) reviews probability and statistics, then covers the fundamentals of modern econometrics, with a focus on regression methods for causal inference in observational and experimental data. Prerequisites: undergraduate courses in probability and statistics, regression analysis, linear algebra, and multivariate calculus.

Economics 2115. Econometric Methods II
Economics 2110 and 2115 comprise a two-course sequence for first-year graduate students seeking training in econometric methods at a level that prepares them to conduct professional empirical research. Economics 2115 (spring) covers topics (different methods) in current empirical research. Faculty members from across the university will teach modules each covering a different method of causal inference, including but not limited to instrumental variables, panel data methods, and regression discontinuity and kink designs. The course will emphasize a mixture of theory and application, with problem sets focused on the replication or extension of recent papers utilizing these methods.
Prerequisite: Economics 2110 or equivalent.

Government 2001. Quantitative Social Science Methods I
This class introduces students to quantitative methods and how they are applied to political science research. It has two overarching goals. First, we focus on the theory of statistical inference - using facts you know to learn about facts you don't know - so that you can truly understand a wide range of methods we introduce, feel comfortable using them in your research, digest new ones invented after class ends, implement them, apply them to your data, interpret the results, and explain them to others. Second, students learn how to publish novel substantive contributions in a scholarly journal. A substantial portion of those in this class publish a revised version of their class paper as their first scholarly journal article. Please see http://j.mp/G2001 for details.

Government 2003. Causal Inference with Applications
Substantive questions in empirical scientific and policy research are typically causal. This class introduces students to both statistical theory and the practice of causal inference. As theoretical frameworks, we discuss potential outcomes, causal graphs, randomization and model-based inference, sensitivity analysis, and partial identification. We also cover important methodological tools, including randomized experiments, regression discontinuity designs, matching, regression, instrumental variables, difference-in-differences, and dynamic causal models.
Prerequisite: Gov 2001 and 2002 or the permission of the instructor.

BST 222. Basics of Statistical Inference
(listed above under Probability Theory)

Biostatistics 231 / BST 231. Statistical Inference I
Exponential families, sufficiency, ancillarity, completeness, method of moments, maximum likelihood, unbiased estimation, Rao-Blackwell and Lehmann-Scheffe theorems, information inequality, Neyman-Pearson theory, likelihood ratio, score and Wald tests, uniformly and locally most powerful tests, asymptotic relative efficiency.
Prerequisite: Biostatistics 230 or signature of instructor required.

API 209. Advanced Quantitative Methods I: Statistics
The goal of this course is to prepare students to analyze public policy issues using statistics. Topics included fall in the areas of probability theory, sampling, estimation, hypothesis testing, and regression analysis. While many students taking this class will have already taken courses in statistics and regression analysis, this course will probably place a much stronger emphasis than typical courses on conceptually understanding the statistical methods. Since the course is targeted to first-year students in the MPA/ID program, we will not shy away from using the

mathematical tools needed to develop the conceptual understanding. But the emphasis of the course will be on the conceptual understanding and application of the tools rather than on the math or the mechanics behind the tools.
Note: This course is open to non-MPA/ID students only by permission of the instructor. May not be taken for credit with API-201.
Prerequisite: Multivariate calculus or linear algebra.

Regression (1 credit)

Statistics 139. Linear Models
An in-depth introduction to statistical methods with linear models and related methods. Topics include group comparisons (t-based methods, non-parametric methods, bootstrapping, analysis of variance), linear regression models and their extensions (ordinary least squares, ridge, LASSO, weighted least squares, multi-level models), model checking and refinement, model selection, cross-validation. The probabilistic basis of all methods will be emphasized.
Prerequisite: Statistics 110 and Math 21a and 21b or equivalent.

Statistics 149. Generalized Linear Models
Sequel to Statistics 139, emphasizing common methods for analyzing continuous non-normal and categorical data. Topics include logistic regression, log-linear models, multinomial logit models, proportional odds models for ordinal data, Gamma and inverse-Gaussian models, over-dispersion, analysis of deviance, model selection and criticism, model diagnostics, and an introduction to non-parametric regression methods.
Note: Examples will be drawn from several fields, particularly from biology and social sciences.
Prerequisite: Statistics 139 or with permission of instructor.

Statistics 244. Linear and Generalized Linear Models
The theory and application of linear and generalized linear models, including linear models for normal responses, logistic models for binary and multinomial data, loglinear models for count data, overdispersion and quasi likelihood methods, and models and methods for clustered (e.g., repeated measurement) correlated data.
Prerequisite: Strong statistics background required (at the second-year graduate level), Statistics 210a may be taken concurrently, Statistics 211a desirable.

Economics 1123. Introduction to Econometrics
An introduction to multiple regression techniques with focus on economic applications. Discusses extensions to discrete response, panel data, and time series models, as well as issues such as omitted variables, missing data, sample selection, randomized and quasi-experiments, and instrumental variables. Also develops the ability to apply econometric and statistical methods using computer packages.
Prerequisite: Statistics 100 or 104.

Economics 2110. Econometrics I
(listed above under Statistical Inference)

Economics 2115. Econometric Methods II
(listed above under Statistical Inference)

Government 2001. Quantitative Social Science Methods I
(listed above under Statistical Inference)

Government 2002. Quantitative Social Science Methods II
This course provides a rigorous foundation necessary for quantitative research in the social sciences. After reviewing the basic probability theory, we offer a systematic introduction to the linear model and its variants -- the workhorse models for social scientists. We cover the classic linear regression model, least squares estimation and projection, fixed and random effects models, principal components analysis, instrumental variables, flexible regression models, and regularization for high dimensional data. In covering these topics, we deepen our knowledge of fundamental concepts in statistical inference while also demonstrating how these methods are applied in political science.
Prerequisite: Gov 2001 or permission of instructor.

BST 211. Regression and Analysis of Variance in Experimental Research
Covers analysis of variance and regression, including details of data-analytic techniques and implications for study design. Also included are probability models and computing. Students learn to formulate a scientific question in

terms of a statistical model, leading to objective and quantitative answers.
Course Note: Lab or section times to be announced at first meeting.
Prerequisite: ID201 or BST201 or (BST202 & BST203) or [BST206 & (BST207 or BST208)]. Concurrent enrollment allowed.

BST 213. Applied Regression for Clinical Research
This course will introduce students involved with clinical research to the practical application of multiple regression analysis. Linear regression, logistic regression and proportional hazards survival models will be covered, as well as general concepts in model selection, goodness-of-fit, and testing procedures. Each lecture will be accompanied by a data analysis using SAS and a classroom discussion of the results. The course will introduce, but will not attempt to develop the underlying likelihood theory. Background in SAS programming ability required.
Note: Lab or section times to be announced at first meeting.
Prerequisite: ID201 or BST201 or (BST202 & BST203) or [BST206 & (BST207 or BST208)]. Concurrent enrollment allowed.

Biostatistics 232 / BST 232. Methods I
Introductory course in the analysis of Gaussian and categorical data. The general linear regression model, ANOVA, robust alternatives based on permutations, model building, resampling methods (bootstrap and jackknife), contingency tables, exact methods, logistic regression.

Biostatistics 235 / BST 235. Advanced Regression and Statistical Learning
An advanced course in linear models, including both classical theory and methods for high dimensional data. Topics include theory of estimation and hypothesis testing, multiple testing problems and false discovery rates, cross validation and model selection, regularization and the LASSO, principal components and dimension reduction, and classification methods.
Prerequisite: Biostatistics 231 AND Biostatistics 233. Background in matrix algebra and linear regression required.

GHP 525. Econometrics for Health Policy
This is a course in applied econometrics for doctoral and advanced master level students. The course has two primary objectives: (1) to develop skills in linking economic behavioral models and quantitative analysis, in a way that students can use in their own research; (2) to develop students' abilities to understand and evaluate critically other peoples' econometric studies.
The course focuses on developing the theoretical basis and practical application of the most common empirical models used in health policy research. In particular, it pays special attention to a class of models identifying causal effects in observational data, including instrumental variable estimation, simultaneous equations and two-stage-least-squares, quasi-experiments and difference-in-difference method, sample selection, treatment effect models and propensity score methods. Lectures will be complemented with computer exercises building on public domain data sets commonly used in health research. The statistical package recommended for the exercises is Stata.
Course Activities: Optional review and computer lab sessions will be held.
Note: Students are expected to be familiar with probability theory (density and distribution functions) as well as the concepts underlying basic ordinary least square (OLS) estimation.
Prerequisites: BST210 or BST213; or equivalent course taken at Harvard Chan or HGSE with instructor permission.

API 210. Advanced Quantitative Methods II: Econometric Methods
Intended as a continuation of API-209, Advanced Quantitative Methods I, this course focuses on developing the theoretical basis and practical application of the most common tools of empirical analysis including non-linear models, instrumental variables, and panel data. Foundations of analysis will be coupled with hands-on examples and assignments involving the analysis of data sets.
Note: This course is open to non-MPA/ID students only by permission of the instructor. May not be taken for credit with API-202.
Prerequisite: API-209 or permission of instructor.

Impact Evaluation (1 credit)

GHP 228. Econometric Methods in Impact Evaluation
The objective of this course is to provide students with a set of theoretical, econometric and reasoning skills to estimate the causal impact of one variable on another. Examples from the readings explore the causal effect of policies, laws, programs and natural experiments derived from pension programs to television shows to natural disasters. We will go beyond estimating causal effects to analyze the channels through which the causal impact was

likely achieved. This will require that the students are familiar with microeconomic theories of incentives, institutions, social networks, etc.

The course will introduce students to a variety of econometric techniques in impact evaluation and a set of reasoning skills intended to help them become both a consumer and producer of applied empirical research. Students will learn to critically analyze evaluation research and to gauge how convincing the research is in identifying a causal impact. They will use these skills to develop an evaluation plan for a topic of their own, with the aim of stimulating ideas for dissertation research. This is a methods class that relies heavily on familiarity with econometrics and microeconomics. These are pre-requisites for the course without exception. The course is intended for doctoral students who are finishing their course work and aims to help them transition into independent research.

The aim of this course is to prepare doctoral students in the health systems track of the Global Health and Population department for the dissertation phase of their research and thus they will be given priority in enrollment. The course is also open to other GHP doctoral students, other GHP masters students and students from other departments, conditional on having adequate training in economics and the course having enough space.

Prerequisites: A course in econometrics and a course in intermediate microeconomics are required. While students can get by with just these two subjects, some previous experience with regression analysis and applied economic research will be a huge advantage. Students seeing applied regression analysis for the first time in this course will most likely struggle with the reading.

Course Note: Students interested in taking this course must email the course instructor (cohenj@hsph.harvard.edu) by January 4, 2019. Students will be notified of their status by January 18, 2019. This essay should include the following information: name, academic department, degree program and year, previous courses taken in economics (specify if these were beginner, intermediate or advanced) and econometrics, any previous experience with impact evaluation, and the reason you want to take the course.

MPR Reading Course (1 credit)

Health Policy 3080A&B. Graduate Reading Course: Methods for Policy Research
Graduate reading course covering major topics in study design and quantitative research methods for health policy research. This course is designed to help students in the Methods for Policy Research track of the Health Policy PhD program prepare for their concentration exam.

Statistics 393. Design of Experimental and Non-Experimental Studies
This is a seminar course on causal inference. We will discuss new and old papers on the design (and analysis) of experimental and non-experimental (observational) studies. Occasionally, leading scholars in the field will visit us and present their work. This course is geared towards methodological research in causal inference.
Recommended Prep: Statistics 211 or equivalent.

Methods and Analysis (4 credits); must be from at least 3 of the following categories

➢ **Statistics and Econometrics**

Statistics 109A / Computer Science 109A. Introduction to Data Science
Data Science 1 is the first half of a one-year introduction to data science. The course will focus on the analysis of messy, real life data to perform predictions using statistical and machine learning methods. Material covered will integrate the five key facets of an investigation using data: (1) data collection - data wrangling, cleaning, and sampling to get a suitable data set; (2) data management - accessing data quickly and reliably; (3) exploratory data analysis – generating hypotheses and building intuition; (4) prediction or statistical learning; and (5) communication – summarizing results through visualization, stories, and interpretable summaries. Part one of a two part series. The curriculum for this course builds throughout the academic year. Students are strongly encouraged to enroll in both the fall and spring course within the same academic year.
Course Notes: Only one of the following can be taken for credit: Stat 109A, Stat 121A, CS109A, AC 209A.
Recommended Prep: Programming knowledge at the level of CS 50 or above, and statistics knowledge at the level of Stat 100 or above (Stat 110 recommended.)

Statistics 109B / Computer Science 109B. Introduction to Data Science
Data Science 2 is the second half of a one-year introduction to data science. Building upon the material in Data Science 1, the course introduces advanced methods for statistical modeling, representation, and prediction. Topics include multiple deep learning architectures such as CNNs, RNNs, transformers, language models, autoencoders,

and generative models as well as basic Bayesian methods, and unsupervised learning. Students are strongly encouraged to enroll in both the fall and spring course within the same academic year. Part two of a two-part series. *Course Notes:* Can only be taken after successful completion of CS 109a, AC 209a, Stat 109a, or Stat 121a, or equivalent. *Prerequisite:* CS 109a, AC 209a, Stat 109a, or Stat 121a required.

Statistics 120. Introduction to Bayesian Inference and Applications
Provides students a comprehensive understanding to the questions as of what is, how and why Bayesian. Introduction to classic Bayesian models, basic computational algorithms/methods for Bayesian inference, as well as their applications in various domain fields, and comparisons with classic Frequentist methods. As Bayesian inference finds its roots and merits particularly in application, this course puts great emphasis on enhancing students' hands-on skills in statistical computation (mostly with R) and data analysis.
Prerequisite: STAT110, STAT111 and basics of R programming are required.

Statistics 131. Time Series and Prediction
Introduction to time series models and forecasting. Introduction to classical time series model: autoregressive, moving average, ARIMA models. Some concepts from stochastic processes: martingales, stationarity, Gaussian processes, Brownian motions, ergodic theorems. Some aspects of advanced time series: hidden Markov models, state space models, filtering, smoothing, Kalman filters, sequential Monte Carlo methods.
Prerequisite: Statistics 111 and 139 or equivalent.

Statistics 149. Generalized Linear Models
(listed above under Regression)

Statistics 151 / EDU S-043. Multilevel or Longitudinal Models
Data often have structure that needs to be modeled explicitly. For example, when investigating students' outcomes we need to account for the fact that students are nested inside classes that are in turn nested inside schools. If we are watching students develop over time, we need to account for the dependence of measurements across time. If we do not, our inferences will tend to be overly optimistic and wrong. The course provides an overall framework, the multilevel and generalized multilevel (hierarchical) model, for thinking about and analyzing these forms of data. We will focus on specific versions of these tools for the most common forms of longitudinal and clustered data. This course will focus on applied work, using real data sets and the statistical software R. R will be specifically taught and supported. While the primary focus will be on the linear model with continuous outcomes (i.e., the classic regression framework) we will also discuss binary, categorical, and ordinal outcomes. We will emphasize how to think about the applicability of these methods, how they might fail, and what one might do to protect oneself in such circumstances. Applications of hierarchical (multi-level) models will include the canonical specific cases of random-slope, random-intercept, mixed effect, crossed effect, marginal, and growth-curve models.
Prerequisite: Permission of instructor required. S-052, Stat 139, or an equivalent. Jointly-offered in the Graduate School of Education.

Statistics 160. Introduction to Survey Sampling and Estimation
Provides students with the tools to engage with complex survey data and survey weights. The course will cover various probability sampling techniques, such as clustering and stratification, and the impacts of these designs and survey weights on estimation. Students will gain experience with common survey estimators and variance estimation techniques. Additional topics will include methods for handling nonresponse, small area estimation, and nonprobability samples. Students will learn the methods presented through engagement with real-world, complex survey data.
Recommended Prep: Stat 110, Stat 111 (may be taken concurrently); Math 21A or equivalent, Math 21B or equivalent; basics of R programming.

Statistics 220. Bayesian Data Analysis
Basic Bayesian models, followed by more complicated hierarchical and mixture models with nonstandard solutions. Includes methods for monitoring adequacy of models and examining sensitivity of models.
Note: Emphasis throughout term on drawing inferences via computer simulation rather than mathematical analysis.
Prerequisite: Statistics 110 and 111.

Economics 2140. Econometric Methods
This course continues the first year sequence in econometrics and covers a variety of topics and ideas that are important for pursuing and interpreting empirical research in economics. The first half of the course covers core econometric approaches that are important for a wide range of applications, including identification analysis, asymptotic approximations, large sample theory for estimation and hypothesis testing, and the bootstrap. The second part of the course examines a range of complementary topics and new developments, including reasons why

canonical econometric methods may be unreliable (such as model misspecification, identification failure, and the incidental parameters problem) and extensions of and alternatives to the traditional econometric paradigm (such as partial identification, Bayesian inference, nonparametrics, and machine learning). Economic applications will be discussed throughout. Enrollment limited.some topics in structural estimation, in particular moment inequalities, demand analysis, and other models.
Prerequisite: Economics 2120 or equivalent.

Sociology 2211. Analysis of Longitudinal Data: Seminar
This course takes an applied approach to the analysis of longitudinal data. Lectures will provide an overview of a variety of techniques, including fixed effects models, multilevel models, and duration models. Students will develop their own empirical projects and receive support as they begin to work with longitudinal datasets.
Note: Primarily for graduate students in sociology.

BST 210. Applied Regression Analysis
Topics include model interpretation, model building, and model assessment for linear regression with continuous outcomes, logistic regression with binary outcomes, and proportional hazards regression with survival time outcomes. Specific topics include regression diagnostics, confounding and effect modification, goodness of fit, data transformations, splines and additive models, ordinal, multinomial, and conditional logistic regression, generalized linear models, overdispersion, Poisson regression for rate outcomes, hazard functions, and missing data. The course will provide students with the skills necessary to perform regression analyses and to critically interpret statistical issues related to regression applications in the public health literature.
Prerequisite: ID201 or BST201 or (BST202 and BST203) or (BST206 and (BST207 or BST208)) or permission of instructor.

BST 223. Applied Survival Analysis
Topics will include types of censoring, hazard, survivor, and cumulative hazard functions, Kaplan-Meier and actuarial estimation of the survival distribution, comparison of survival using log rank and other tests, regression models including the Cox proportional hazards model and the accelerated failure time model, adjustment for time-varying covariates, and the use of parametric distributions (exponential, Weibull) in survival analysis. Methods for recurrent survival outcomes and competing risks will also be discussed, as well as design of studies with survival outcomes. Class material will include presentation of statistical methods for estimation and testing along with current software (SAS, Stata) for implementing analyses of survival data. Applications to real data will be emphasized.
Prerequisite: BST210 or BST213 or BST232 or BST260 or PHS2000A.

BST 226. Applied Longitudinal Analysis
This course covers modern methods for the analysis of repeated measures, correlated outcomes and longitudinal data, including the unbalanced and incomplete data sets characteristic of biomedical research. Topics include an introduction to the analysis of correlated data, analysis of response profiles, fitting parametric curves, covariance pattern models, random effects and growth curve models, and generalized linear models for longitudinal data, including generalized estimating equations (GEE) and generalized linear mixed effects models (GLMMs).
Course Activities: Homework assignments will focus on data analysis in SAS using PROC GLM, PROC MIXED, PROC GENMOD, and PROC GLIMMIX.
Note: Lab or section times will be announced at first meeting.
Prerequisite: BST210 or BST213 or BST232 or BST260 or PHS2000A.

BST 228. Applied Bayesian Analysis
This course is a practical introduction to the Bayesian analysis of biomedical data. It is an intermediate Master's level course in the philosophy, analytic strategies, implementation, and interpretation of Bayesian data analysis. Specific topics that will be covered include: the Bayesian paradigm; Bayesian analysis of basic models; Bayesian computing: Markov Chain Monte Carlo; STAN R software package for Bayesian data analysis; linear regression; hierarchical regression models; generalized linear models; meta-analysis; models for missing data.
Programming and case studies will be used throughout the course to provide hands-on training in these concepts.
Prerequisite: **(**BST210 or PHS 2000A&B) and BST222, or permission of the instructor.

Biostatistics 232 / BST 232. Methods I
(listed above under Regression)

Biostatistics 245 / BST 245. Analysis of Multivariate and Longitudinal Data

The multivariate normal distribution, Hotelling's T2, MANOVA, repeated measures, the multivariate linear model, random effects and growth curve models, generalized estimating equations, multivariate categorical outcomes, missing data, computational issues for traditional and new methodologies.
Prerequisite: Biostatistics 231 and Biostatistics 235.

Biostatistics 249 / BST 249. Bayesian Methodology in Biostatistics

General principles of the Bayesian approach, prior distributions, hierarchical models and modeling techniques, approximate inference, Markov chain Monte Carlo methods, model assessment and comparison. Bayesian approaches to GLMMs, multiple testing, nonparametrics, clinical trials, survival analysis.
Prerequisite: Biostatistics 231 and Biostatistics 232, or signature of instructor required.

SBS 263. Multilevel Statistical Methods: Concept and Application

This course is designed to provide doctoral students with a training experience in the concept and application of multilevel statistical modeling. Students will be motivated to think about correlated and dependent data structures that arise due to sampling design and/or are inherent in the population (such as pupils nested within schools; patients nested within clinics; individuals nested within neighborhoods and so on). The substantive motivation for analyzing such complex data structures would be to make quantitative assessments about the role of contexts (e.g., schools, clinics, neighborhoods) in predicting individual outcomes. In particular, the principles of recognizing and modeling the underlying heterogeneity in average relationships would be emphasized. Linear, non-linear, and multivariate multilevel models will be covered. Upon completion, students should be able to conceptualize multilevel modeling strategies and to undertake empirical, quantitative multilevel research. The course will be lecture-based with substantial hands-on component.
Course Activities: Data management, modeling and analysis; individual assignments; project submission and class participation.
Notes: This course is a requirement for all SBS doctoral students. Required lab.

EDU S052. Applied Data Analysis

This course is designed for those who want to extend their data analytic skills beyond a basic knowledge of multiple regression analysis and who want to communicate their findings clearly to audiences of researchers, scholars, and policymakers. S-052 contributes directly to the diverse data analytic toolkit that the well-equipped empirical researcher must possess in order to perform sensible analyses of complex educational, psychological, and social data. The course begins with general linear models and continues with generalized linear models, survival analysis, multilevel models, multivariate methods, causal inference, and measurement. Specific methods exemplifying each of these topics include regression, discrete-time survival analysis, fixed- and random-effects models, principal components analysis, instrumental variables, and reliability, respectively. S-052 is an applied course. It offers conceptual explanations of statistical techniques and provides many opportunities to examine, implement, and practice these techniques using real data. Students will lean to produce readable and sensible code to enable others to replicate and extend their analyses. Attendance at weekly sections is required.
*Prerequisite: S*uccessful completion of S-040 (B+ or better allowed, A- or A recommended) or an equivalent course or courses that include 12 or more full hours of class time on multiple regression and its direct extensions. Students who have not passed S-40 must discuss their previous training before or at the first class meeting. Students who do not meet the prerequisite should consider S-030. See the syllabus at the instructor's website, https://scholar.harvard.edu/andrewho/classes, for more details.

➢ **Survey Research Methods**

Statistics 160. Introduction to Survey Sampling and Estimation
(listed above under Statistics and Econometrics)

BST 212. Survey Research Methods In Community Health

Covers research design, sample selection, questionnaire construction, interviewing techniques, the reduction and interpretation of data, and related facets of population survey investigations. Focuses primarily on the application of survey methods to problems of health program planning and evaluation. Treatment of methodology is sufficiently broad to be suitable for students who are concerned with epidemiological, nutritional, or other types of survey research.

BST 239. Health Survey Samples

To comprehend and monitor the public health, one needs to measure it. To this end, surveys are indispensable, as evidenced by the National Center for Health Statistics, within the CDC, and the surveys they carry out; the ACS, as carried out by the Census Bureau; and the powerful DHS and any number of ubiquitous health related surveys. This

course is meant to expose the student to sample surveys: their theoretical underpinnings, how to design them, how to collect the data, how to analyze the data, how to judge and interpret the data, and how to make inference about the public health from the survey. It is targeted at graduate students in the biostatistics department, but qualified students from the School are welcome, as sample surveys form the basis for a large fraction of public health data.
Course Notes: Lab or section times announced at first meeting. This course requires an understanding of statistical inference and the value of a random sample. Plus a facility with programming in R, or Stata.
Prerequisite: BST 230 / BIOSTAT 230

➢ **Causal Inference**

Economics 2115. Econometric Methods II
(listed above under Statistical Inference)

Government 2003. Causal Inference with Applications
(listed above under Statistical Inference)

Statistics 186. Causal Inference
(listed above under Statistical Inference)

Statistics 293. Design of Experimental and Non-Experimental Studies
This is a seminar course on causal inference. We will discuss new and old papers on the design (and analysis) of experimental and non-experimental (observational) studies. Occasionally, leading scholars in the field will visit us and present their work. This course is geared towards methodological research in causal inference.
Recommended Prep: Statistics 211 or equivalent.

EPI 289. Epidemologic Methods III: Models for Causal Inference
Causal Inference is a fundamental component of epidemiologic research. EPI289 describes models for causal inference, their application to epidemiologic data, and the assumptions required to endow the parameter estimates with a causal interpretation. The course introduces outcome regression, propensity score methods, the parametric g-formula, inverse probability weighting of marginal structural models, g-estimation of nested structural models, and instrumental variable methods. Each week students are asked to analyze the same data using a different method. EPI289 is designed to be taken after EPI201/EPI202 and before EPI204 and EPI207. Epidemiologic concepts and methods studied in EPI201/202 will be reformulated within a modeling framework in EPI289. This is the first course in the sequence of EPI core courses on modeling (EPI289, EPI204, EPI207). EPI289 focuses on time-fixed dichotomous exposures and time-fixed dichotomous and continuous outcomes. Continuous exposures and failure time outcomes (survival analysis) will be discussed in EPI204, and time-varying exposures in EPI207. Familiarity with either SAS or R language is strongly recommended.
Prerequisite: EPI201 and EPI202; may not be taken concurrently.

➢ **Decision Sciences**

RDS 280. Decision Analysis for Health and Medical Practices
This course is designed to introduce the student to the methods and growing range of applications of decision analysis and cost-effectiveness analysis in health technology assessment, medical and public health decision making, and health resource allocation. The objectives of the course are: (1) to provide a basic technical understanding of the methods used, (2) to give the student an appreciation of the practical problems in applying these methods to the evaluation of clinical interventions and public health policies, and (3) to give the student an appreciation of the uses and limitations of these methods in decision making at the individual, organizational, and policy level both in developed and developing countries.
Prerequisite: BST201 or BST202&203 or BST206&207 or BST206&208 (all courses may be taken concurrently). Introductory economics is recommended but not required.

RDS 282. Economic Evaluation of Health Policy and Program Management
This course features case studies in the application of health decision science to policymaking and program management at various levels of the health system. Both developed and developing country contexts will be covered. Topics include: [1] theoretical foundations of cost-effectiveness analysis (CEA); [2] controversies and limitations of CEA in practice; [3] design and implementation of tools and protocols for measurement and valuation of cost and benefit of health programs; [4] integration of evidence of economic value into strategic planning and resource allocation decisions, performance monitoring and program evaluation; [5] the role of evidence of economic value in the context of other stakeholder criteria and political motivations.
Prerequisite: Students must have taken RDS280 or RDS286. Prior coursework in Microeconomics is recommended.

RDS 284. Decision Theory

Introduces the standard model of decision-making under uncertainty, its conceptual foundations, challenges, alternatives, and methodological issues arising from the application of these techniques to health issues. Topics include von Neumann-Morgenstern and multi-attribute utility theory, Bayesian statistical decision theory, stochastic dominance, the value of information, judgment under uncertainty and alternative models of probability and decision making (regret theory, prospect theory, generalized expected utility). Applications are to preferences for health and aggregation of preferences over time and across individuals.

RDS 285. Decision Analysis Methods in Public Health and Medicine

An intermediate-level course on methods and health applications of decision analysis modeling techniques. Topics include Markov models, microsimulation models, life expectancy estimation, deterministic and probabilistic sensitivity analysis, ROC analysis and diagnostic technology assessment, and cost-effectiveness analysis.
Note: Familiarity with matrix algebra and elementary calculus may be helpful but not required; lab or section times to be announced at first meeting.
Prerequisite: (BST201 or ID201) and (RDS280 or RDS286).

API 302 / ECON 1415. Analytic Frameworks for Policy

This course develops abilities in using analytic frameworks in the formulation and assessment of public policies. It considers a variety of analytic techniques, particularly those directed toward uncertainty and interactive decision problems. It emphasizes the application of techniques to policy analysis, not formal derivations. Students encounter case studies, methodological readings, modeling of current events, the computer, a final exam, and challenging problem sets.
Prerequisite: An understanding of intermediate-level microeconomic theory and introductory techniques of optimization and decision analysis; API-101, API-102, or equivalent.

➢ **Epidemiology**

EPI 202. Epidemiologic Methods II: Elements of Epidemiologic Research

Introduces elements of study design, data analysis and inference in epidemiologic research. Principles and methods are illustrated with examples, and reviewed through homework and in-class exercises. May serve as an introduction to more advanced study or as a concluding course for those desiring a working knowledge of epidemiologic methods. EPI 202 extends the concepts of study design, data analysis, and inference introduced in the introductory epidemiology courses.
Prerequisite: EPI201 AND (BST201 or BST202&203 or BST206&207/8 or ID207 or PHS2000) (all courses may be taken concurrently).

EPI 203. Study Design in Epidemiologic Research

Beginning with the randomized clinical trial as a paradigm, this course examines common problems in the design, analysis, and interpretation of observational studies. Cohort and case-control studies are the focus of the discussion, but not to the exclusion of other designs. Problems of exposure and disease definitions, time-dependent effects, confounding, and misclassification are considered in the light of data sources typically available. Relevant statistical methods are introduced but not developed in detail.
Prerequisite: EPI202 and (BIO200 or ID200 or BST201 or BST202&203 or BST206&207/8/9 or PHS 2000A). All prerequisites may be taken concurrently.

EPI 204. Analysis of Case-Control, Cohort, and Other Epidemiologic Data

This course will examine, through practical examples, the use of regression methods for analyses of epidemiologic data, primarily case-control and cohort studies. Methods used will include linear, logistic, Poisson, conditional logistic and Cox regression models. The lectures will focus on the principle ideas and issues underlying the regression analyses, and the computer labs will provide practical experience applying those methods, using SAS software. Issues to be dealt with include dose-response, confounding, influence, and interaction. It will emphasize analysis and interpretation of results in the context of the study design. Familiarity with basic SAS is required, as this will be used in the labs. This can be met through BIO 113 (Introduction to Data Management and Programming in SAS) or other significant SAS experience.
Course Activities: Written group projects, class discussion, quizzes, homework.
Note: Computer lab is required, please sign up for one lab session when registering.
Prerequisite: (BST210 (concurrent enrollment allowed) or BST213 or PHS2000A/B) and (EPI200 or EPI201 or EPI208 or EPI505) and EPI202.

EPI 207. Advanced Epidemiologic Methods

Provides an in-depth investigation of statistical methods for drawing causal inferences from observational studies. Informal epidemiologic concepts such as confounding, selection bias, overall effects, direct effects, and intermediate variables will be formally defined within the context of a counterfactual causal model and with the help of causal diagrams. Methods for the analysis of the causal effects of time-varying exposures in the presence of time dependent covariates that are simultaneously confounders and intermediate variables will be emphasized. These methods include g-computation algorithm estimators, inverse probability weighted estimators of marginal structural models, g-estimation of structural nested models. As a practicum, students will reanalyze data sets using the above methods.
Course Activities: Class discussion, homework, practicum and final examination.
Note: Familiarity with logistic regression and survival analysis is expected; lab time will be announced at first meeting.
Prerequisite: EPI204 or (BST 210 and EPI289) or BIO233

- ➤ **Machine Learning**

API 222. Machine Learning and Big Data Analytics

In the last couple of decades, the amount of data available to organizations has significantly increased. Individuals who can use this data together with appropriate analytical techniques can discover new facts and provide new solutions to various existing problems. This course provides an introduction to the theory and applications of some of the most popular machine learning techniques. It is designed for students interested in using machine learning and related analytical techniques to make better decisions in order to solve policy and societal level problems.

We will cover various recent techniques and their applications from both supervised and unsupervised learning. In addition, students will get the chance to work with some data sets using software and apply their knowledge to a variety of examples from a broad array of industries and policy domains. Some of the intended course topics (time permitting) include: *K-Nearest Neighbors, Naive Bayes, Logistic Regression, Linear and Quadratic Discriminant Analysis, Model Selection (Cross Validation, Bootstrapping), Support Vector Machines, Smoothing Splines, Generalized Additive Models, Shrinkage Methods (Lasso, Ridge), Dimension Reduction Methods (Principal Component Regression, Partial Least Squares), Decision Trees, Bagging, Boosting, Random Forest, K-Means Clustering, Hierarchical Clustering, Neural Networks, Deep Learning, and Reinforcement Learning.*
Prerequisite: An understanding of intermediate-level statistics and probability theory (e.g., API-201, API-202, or equivalent courses).
Note: API 222A and API 222B are the exact same course, simply offered on different days.

BST 263: Statistical Learning

Statistical learning is a collection of flexible tools and techniques for using data to construct prediction algorithms and perform exploratory analysis. This course will introduce students to the theory and application of methods for supervised learning (classification and regression) and unsupervised learning (dimension reduction and clustering). Students will learn the mathematical foundations underlying the methods, as well as how and when to apply different methods. Topics will include the bias-variance tradeoff, cross-validation, linear regression, logistic regression, KNN, LDA/QDA, variable selection, penalized regression, generalized additive models, CART, random forests, gradient boosting, kernels, SVMs, PCA, and K-means. Homework will involve mathematical and programming exercises, and exams will contain conceptual and mathematical problems. Programming in R will be used throughout the course to provide hands-on training and practical examples.
Prerequisite: BST 260 or BST 210 or BST 232.

MIT 6.7910J/9.520J. Statistical Learning Theory and Applications

Provides students with the knowledge needed to use and develop advanced machine learning solutions to challenging problems. Covers foundations and recent advances of machine learning in the framework of statistical learning theory. Focuses on regularization techniques key to high-dimensional supervised learning. Starting from classical methods such as regularization networks and support vector machines, addresses state-of-the-art techniques based on principles such as geometry or sparsity, and discusses a variety of algorithms for supervised learning, feature selection, structured prediction, and multitask learning. Also focuses on unsupervised learning of data representations, with an emphasis on hierarchical (deep) architectures.

MIT 14.388. Inference on Causal and Structural Parameters Using ML and AI

Provides an applied treatment of modern causal inference with high-dimensional data, focusing on empirical economic problems encountered in academic research and the tech industry. Formulates problems in the languages of structural equation modeling and potential outcomes. Presents state-of-the-art approaches for inference on causal and structural parameters, including de-biased machine learning, synthetic control methods, and reinforcement learning. Introduces tools from machine learning and deep learning developed for prediction purposes, and discusses

how to adapt them to learn causal parameters. Emphasizes the applied and practical perspectives. Requires knowledge of mathematical statistics and regression analysis and programming experience in R or Python.

GOV 2004. Introduction to Machine Learning

This course serves as a graduate-level introduction to machine/statistical learning for social scientists. It will cover some, but not all, common techniques to collect, analyze and utilize large and unstructured data for social science questions. The general goal of this course is to introduce students to modern machine learning techniques and provide the skills necessary to apply the methods widely. Note, this course does not utilize Python or Julia. All computation work is conducted in R.

➢ **Network Analysis**

BST 267. Introduction to Social and Biological Networks

Many systems of scientific and societal interest consist of a large number of interacting components. The structure of these systems can be represented as networks where network nodes represent the components and network edges the interactions between the components. Network analysis can be used to study how pathogens, behaviors and information spread in social networks, having important implications for our understanding of epidemics and the planning of effective interventions. In a biological context, at a molecular level, network analysis can be applied to gene regulation networks, signal transduction networks, protein interaction networks, and more. This introductory course covers some basic network measures, models, and processes that unfold on networks. The covered material applies to a wide range of networks, but we will focus on social and biological networks. To analyze and model networks, we will learn the basics of the Python programming language and its NetworkX module.
Note: The course contains a number of hands-on computer lab sessions. There are five homework assignments and four reading assignments that will be discussed in class. In addition, each student will complete a final project that applies network analysis techniques to study a public health problem.
Prerequisite: BST201 or ID201 or (BST202 & 203) or [BST206 & (BST207 or 208)].

➢ **Program Evaluation**

API 211 / EDU S164. Program Evaluation

As school districts and state agencies accumulate quantitative student outcome data, demand for evidence of impact will grow. All of us must learn to be critical consumers of quantitative evidence of impact. The key challenge when evaluating the impact of an education policy or program is to identify what would have happened if that policy or program had not been implemented. There are a number of different approaches to constructing a plausible estimate of what would have happened, using experimental or quasi-experimental techniques. In this course, we will have three goals: to gain insight into the strengths and weaknesses of different evaluation designs, including experimental and quasi-experimental techniques; to develop the skills required to be a critical reader of impact evaluations; and to develop the ability to more clearly recognize opportunities for impact evaluations in education and to implement policies in a manner that would be amenable to evaluation. During the course, we will read and critique a number of impact evaluations, replicate the results of several evaluations, and design evaluations of educational programs. The course will focus on quantitative impact evaluations, as opposed to qualitative or process evaluations.
Prerequisite: Successful completion of S-030 or S-040, or prior equivalent training in multiple regression. Also offered by the Graduate School of Education as A-164. Permission of the instructor required. Enrollment procedure will be posted on the HGSE course website.

HPM 543. Quantitative Methods in Program Evaluation

This course will give students the tools that they need to evaluate policy interventions, social programs, and health initiatives. Did the program achieve its goals? Did it reach its target audience? Could it have been more effective? In order to answer these questions, students will develop a flexible set of analytical tools, including both the ability to design an evaluation study and the ability to evaluate existing studies critically.
By the end of the course students will be able to construct a well-designed study to answer well-posed questions, gauge the adequacy of available data, implement an econometric analysis, interpret the results of such studies, and draw policy implications. The course will focus on health policies and programs such as public insurance expansions and public health campaigns, but the techniques will be broadly applicable to other realms such as welfare or education.
Note: The material in this course is inherently quantitative, and builds on a base of statistics fundamentals. The prerequisite is a course in basic statistics and probability, such as BIO 200, BIO 201, BIO 202/203, ID 538, ID201 or equivalent. This includes knowledge of confidence intervals and hypothesis testing. It also includes familiarity with the statistical package of your choice- ideally STATA, but SAS or SPSS are fine. During the course students will be given data sets to analyze, but there will be no instruction on the mechanics of opening and manipulating the

data with a statistical software package. Students should contact instructor if they are uncertain about whether they have adequate preparation for the class.
Prerequisite: BIO200 or BIO201 or BIO202&203, or BIO206 & (207 or 208 or 209) or ID538 or ID201 or equivalent.

➢ **Qualitative Research**

GHP 504. Introduction to Qualitative Research for Global Health
This course introduces students to qualitative research design and methods applied in global health. Contrasts will be drawn with quantities and mixed-methods approaches in order to consider the place of qualitative research in global health. The course aims to provide students with an understanding of when to use qualitative research approaches, explores the philosophical debates around qualitative research and the theories that underpin qualitative research designs to consider which qualitative methodologies is appropriate for the research, and critically appraise the quality and credibility of qualitative research.
This practical-oriented course will equip students with the knowledge and skills to appropriately design, plan and appraise qualitative research. The course topics will include the application of qualitative research approaches in global health, qualitative research designs, qualitative methods, ethics and critical appraisal of qualitative research.

SBS 288. Qualitative Research Methods in Public Health
What students can expect from this course: Qualitative research can be used alone or in combination with quantitative research to investigate public health questions. This introductory course will provide students with an overview of the range of important conceptual and practical issues associated with qualitative research, including providing general familiarity with the design of qualitative studies and conduct of commonly-used qualitative methods. The course begins by examining the variety of potential uses of qualitative methods in public health research and diverse qualitative research approaches. The course then explores specific topics, including: developing research questions; ethics in qualitative research; "entering" the community to conduct qualitative research; role of theory; ensuring study rigor; selecting and implementing qualitative data collection methods (participant observation, semi-structured interviews and focus groups); writing open-ended questions; sampling; data management and analysis; publishing results; writing research proposals; and considerations for choosing qualitative methods for mixed-methods qualitative or mixed-methods qualitative/quantitative studies.
Students should come to class prepared to apply concepts from readings and lectures through participation in class discussions and small group activities that will occur during every class period. In addition, students will demonstrate application of concepts through completion of written assignments.
What this course is not: As this is an introductory course on qualitative research that provides an overview of all pertinent topics to foster familiarity with this research approach as a whole, the course cannot dwell deeply on any one topic. Students who are looking for in-depth training on a particular step in qualitative research, such as how to analyze their own qualitative dataset, or how to use qualitative coding software, are advised to select a different course.

SBS 521. Qualitative Data Analysis for Public Health
This intermediate seminar focuses on the analysis of qualitative data in public health. The course coversa range of approaches to the analysis of textual data from interview and focus group transcripts, including thematic analysis, grounded theory, framework analysis and hybrid and general approaches to qualitative data analysis. Students will learn practical skills involved in qualitative data analysis, including identifying themes using inductive and deductive strategies, developing a codebook, coding transcripts using computer software, creating matrices and networks, writing memos, and building conceptual models. The course also addresses issues of ethics in the context of qualitative data analysis and research with marginalized populations and discusses team-based and participatory approaches to qualitative data analysis. Class sessions focus on active learning and include lectures, large and small group discussions, coding practice activities, article critiques, student presentations, and guest lectures by qualitative researchers. This course is open to masters and doctoral students who meet the prerequisites.
Prerequisites: SBS288 or GHP504 or DRPH200 or S-063 or another qualitative research methods course/prior qualitative research experience (by permission of instructor)

EDU S510F. Qualitative Research Methods in Practice
What does the process of conducting qualitative research look like? How can qualitative research tools and approaches be incorporated into everyday practice? How does one craft research questions, collect and analyze data, and make appropriate claims about findings? This introductory module is designed for those who are interested in incorporating qualitative research techniques or dispositions into their work as practitioners, as well as

for students considering future careers in research. It does not presuppose prior experience or familiarity with qualitative research methods.

Prerequisite: none.

Learning Goals: (1) Gain a sense of some qualitative research methods, including tools and approaches that can be incorporated into everyday professional practice.

(2) Begin to develop an understanding of the overall process of conducting a qualitative research study, including the iterative nature of doing qualitative research.

(3) Start to develop specific qualitative research skills related to crafting research questions, collecting and analyzing data, and making appropriate claims about findings.

(4) Consider some of the ethical, epistemological, and relational issues that thoughtful researchers take into account when doing their work, and develop their own capacity for reflexivity.

EDU S510H. Developing a Research Proposal Informed by Qualitative Methods

The primary purpose of this module, which builds on S510F, is to support students to develop an original, coherent research proposal around a topic of their choice. While the emphasis will be on qualitative methods, there is scope to develop a mixed methods proposal—albeit with a focus on the qualitative components. The module will follow a seminar-style format, with students learning both with and from one another as they develop individual proposals that are informed by preliminary data gathering and analysis. The course will also allow students to deepen their understanding of key issues in qualitative research, including epistemological, ethical, and practical concerns.

Prerequisite: Students must have completed S510F.

Note: Limited enrollment and instructor permission required. Priority will be given to doctoral students.

EDU S513. Introduction to Quality Program and Policy Evaluation

Why is an intervention that was successful in a pilot program faltering in the broader rollout? What are the enablers of and barriers to implementing a new statewide policy? How do students and teachers perceive the effectiveness of a targeted program? Qualitative research, either as a sole methodology or in tandem with quantitative research, often plays a central role in education program evaluation and policy development. While quantitative research can show whether there is a relationship between programs and outcomes, it cannot answer why or how, or offer insight into the process of implementation. This course will provide students with a comprehensive introduction to qualitative program and policy evaluation, and is appropriate both for consumers and producers of research. Class time will include a mix of discussion, group activities and short presentations of key points, by the instructor and members of the class. A key element of the course will be conducting an authentic evaluation of a policy or program.

Prerequisite: none.

> ### ➤ Other

SBS 245. Social and Behavioral Research Methods

Provides a broad overview of social and behavioral research methodology, including experimental, quasi-experimental and non-experimental research design, measurement, sampling, data collection, and testing causal theories. By case studies, methodological readings, discussion, written assignments, and data analytic homeworks students learn to conduct social and behavioral research and more applied program evaluations. Homework includes analytic work with observational and experimental studies and development of new measures.

Course Activities: Assigned readings, class participation, homeworks, reflections, two papers.

Note: a multivariate statistics course strongly recommended; course primarily for doctoral students.

Prerequisite: BST210 or BST211 or BSt213.

HBSDOC 4070. Design of Field Research Methods

Field research involves collecting original data (qualitative or quantitative) in field sites. The course will combine informal lecture and discussion with practical sessions designed to build specific skills for conducting field research in organizations. Readings include books and papers about research methodology, as well as articles that provide exemplars of field research, including both theory driven and phenomenon driven work. Specific topics covered include variance versus process models, blending qualitative and quantitative data (in one paper, one study, or one career), collecting and analyzing different kinds of data (observation interview, survey, archival), levels of analysis, construct development, and writing up field research for publication. A core aim of the course is to help students understand the contingent relationship between the nature of the research question and the field research methods used to answer it, and to use this understanding to design and carry out original field research. Course requirements include several short assignments assessing readings and a final paper designed to help students' further their own field research goals. This seminar fulfills a requirement for HBS Organizational Behavior and Management students.

Prerequisite: Students are required to be in or beyond their second year of study.