STATISTICAL AND DATA SCIENCES

SDS 100 Laboratory: Reproducible Scientific Computing with Data (1 Credit)

The practice of data science rests upon computing environments that foster responsible uses of data and reproducible scientific inquiries. This course develops students' ability to engage in data science work using modern workflows, open-source tools and ethical practices. Students learn how to author a scientific report written in a lightweight markup language (e.g., markdown) that includes code (e.g., R), data, graphics, text and other media. Students also learn to reason about ethical practices in data science. S/U only. Concurrent registration required in any of: SDS 192, SDS 201, SDS 220, SDS 290 or SDS 291. Restrictions: Not open to students who have already completed any of: SDS 192, SDS 201, SDS 220, SDS 290 or SDS 291. Enrollment limited to 30. Students not registered for a corequisite course will be dropped without notification. Fall, Spring

SDS 109/ CSC 109 Communicating with Data (4 Credits)

Offered as SDS 109 and CSC 109. The world is growing increasingly reliant on collecting and analyzing information to help people make decisions. Because of this, the ability to communicate effectively about data is an important component of future job prospects across nearly all disciplines. In this course, students learn the foundations of information visualization and sharpen their skills in communicating using data. This course explores concepts in decision-making, human perception, color theory and storytelling as they apply to data-driven communication. This course helps students build a strong foundation in how to talk to people about data, for both aspiring data scientists and students who want to learn new ways of presenting information. Enrollment limited to 40. {M} Fall, Spring

SDS 192 Introduction to Data Science (4 Credits)

An introduction to data science using Python, R and SQL. Students learn how to scrape, process and clean data from the web; manipulate data in a variety of formats; contextualize variation in data; construct point and interval estimates using resampling techniques; visualize multidimensional data; design accurate, clear and appropriate data graphics; create data maps and perform basic spatial analysis; and query large relational databases. Prerequisite: concurrent registration in SDS 100 required for students who have not previously completed SDS 201, SDS 220, SDS 290 or SDS 291. {M}

Fall, Spring

SDS 201 Statistical Methods for Undergraduates (4 Credits)

(Formerly MTH 201/ PSY 201). An overview of the statistical methods needed for undergraduate research, emphasizing methods for data collection, data description and statistical inference, including an introduction to study design, confidence intervals, testing hypotheses, analysis of variance and regression analysis. Techniques for analyzing both quantitative and categorical data are discussed. Applications are emphasized and students use R for data analysis. This course satisfies the basic requirement for the psychology major. Students who have taken MTH 111 or equivalent should take SDS 220, which also satisfies the basic requirement. Prerequisite: concurrent registration in SDS 100 required for students who have not completed SDS 192, SDS 220, SDS 290 or SDS 291. Restrictions: Students do not normally earn credit for more than one course on this list: ECO 220, GOV 203, MTH 220, PSY 201, SDS 201, SDS 220 or SOC 204. Enrollment limited to 40. {M}

Fall, Spring

SDS 220 Introduction to Probability and Statistics (4 Credits)

(Formerly MTH 220/SDS 220). An application-oriented introduction to modern statistical inference: study design, descriptive statistics, random variables, probability and sampling distributions, point and interval estimates, hypothesis tests, resampling procedures, and multiple regression. A wide variety of applications from the natural and social sciences are used. This course satisfies the basic requirement for biological science, engineering, environmental science, neuroscience, and psychology. Prerequisite: MTH 111, or equivalent; SDS 100 must be taken concurrently for students who have not completed SDS 192, SDS 201, SDS 290 or SDS 291. Restrictions: Students do not normally earn credit for more than one course on this list: ECO 220, GOV 203, MTH 220, PSY 201, SDS 201, SDS 220 or SOC 204. Enrollment limited to 40. {M}

Fall, Spring

SDS 235/ CSC 235 Visual Analytics (4 Credits)

Offered as CSC 235 and SDS 235. Visual analytics techniques can help people to derive insight from massive, dynamic, ambiguous and often conflicting data. During this course, students learn the foundations of the emerging, multidisciplinary field of visual analytics and apply these techniques toward a focused research problem in a domain of personal interest. Students who elect to take this course as a programming intensive course should have previously taken CSC 212. In this track, students learn to use R, Python and HTML5/JavaScript to develop custom visual analytic tools. Students preferring a non-programming intensive track may elect to use existing visual analytic software, such as Tableau or Plotly. Designations: Theory, Programming. Prerequisite: CSC 120 or equivalent. {M}

Fall, Spring, Variable

SDS 236 Data Journalism (4 Credits)

Data journalism is the practice of telling stories with data. This course will focus on journalistic practices, interviewing data as a source, and interpreting results in context. We will discuss the importance of audience in a journalistic context, and will focus on statistical ideas of variation and bias. The course will include hands-on work with data, using appropriate computational tools such as R, Python, and data APIs. In addition, we will explore the use of visualization and storytelling tools such as Tableau, plot.ly, and D3. No prior experience with programming or journalism is required. Prerequisites: An introductory statistics course (including SDS 220, SOC 204, GOV 203, ECO 220, PSY 201). Enrollment limited to 20. WI {M}

Fall, Spring, Variable

SDS 237 Data Ethnography (4 Credits)

This course introduces the theory and practice of data ethnography, demonstrating how qualitative data collection and analysis can be used to study data settings and artifacts. Students will learn techniques in field-note writing, participant observation, in-depth interviewing, documentary analysis and archival research and how they may be used to contextualize the cultural underpinnings of datasets. Students will learn how to visualize datasets in ways that foreground their sociopolitical provenance in R. Students will also learn how ethnographic methods can be leveraged to improve data documentation and communication. The course will introduce debates regarding the politics of technoscientific fieldwork. Recommended prerequisite: SDS 192. Enrollment limited to 40. {S}

Fall, Spring

SDS 238 Community-Based Data Science (4 Credits)

This course introduces concepts in human-centered design and design justice, considering how their principles can be applied in the context of community-based data science work. Students learn how to define social problems, engage stakeholders, design data science solutions, and evaluate social impact. Students also learn techniques in collaborative data science project planning and execution, engaging best practices (e.g. version control and code review) in the context of a communitybased data science project. Strategies for effectively communicating project approach, outcomes, and impact are addressed throughout the course. Enrollment limited to 24.

Fall, Spring, Alternate Years

SDS 270 Programming for Data Science in R (4 Credits)

This course is not about data analysis—rather, students learn the R programming language at a deep level. Topics may include data structures, control flow, regular expressions, functions, environments, functional programming, object-oriented programming, debugging, testing, version control, documentation, literate programming, code review and package development. The major goal for the course is to contribute to a viable, collaborative, open-source, publishable R package. Prerequisites: SDS 192 and CSC 110, or equivalent. Enrollment limited to 40. {M}

Fall, Spring

SDS 271 Programming for Data Science in Python (4 Credits)

This course covers the skills and tools needed to process, analyze and visualize data in Python and work on collaborative projects. Topics include functional and object oriented programming in Python, data wrangling in Pandas, visualization in Matplotlib in seaborn, as well as creating a reproducible workflow: debugging, testing and documenting programs, and effectively using version control. The major goal for the course is to create a viable, open-source Python package like those in the Python Package Index (PyPI). Prerequisites: SDS 192 and CSC 110. Enrollment limited to 40. (E) {M}

Fall, Spring, Variable

SDS 290 Research Design and Analysis (4 Credits)

(Formerly MTH/SDS 290). A survey of statistical methods needed for scientific research, including planning data collection and data analyses that provide evidence about a research hypothesis. The course can include coverage of analyses of variance, interactions, contrasts, multiple comparisons, multiple regression, factor analysis, causal inference for observational and randomized studies and graphical methods for displaying data. Special attention is given to analysis of data from student projects such as theses and special studies. Statistical software is used for data analysis. Prerequisites: One of the following: PSY 201, SDS 201, GOV 203, ECO 220, SDS 220 or a score of 4 or 5 on the AP Statistics examination or the equivalent; concurrent registration in SDS 100 required for students who have not completed SDS 192, SDS 201, SDS 220 or SDS 291. Enrollment limited to 40. {M} Fall, Spring

SDS 291 Multiple Regression (4 Credits)

(Formerly MTH 291/ SDS 291). Theory and applications of regression techniques: linear and nonlinear multiple regression models, residual and influence analysis, correlation, covariance analysis, indicator variables and time series analysis. This course includes methods for choosing, fitting, evaluating and comparing statistical models and analyzes data sets taken from the natural, physical and social sciences. Prerequisite: SDS 201, PSY 201, GOV 203, SDS 220, ECO 220 or equivalent or a score of 4 or 5 on the AP Statistics examination; concurrent registration in SDS 100 required for students who have not completed SDS 192, 201, 220 or 290. Enrollment limited to 40. {M}{N} Fall, Spring

SDS 293/ CSC 293 Machine Learning (4 Credits)

Offered as CSC 293 and SDS 293. The field of statistical learning encompasses a variety of computational tools for modeling and understanding complex data. In this introductory course, we will explore many of the most popular of these tools, such as sparse regression, classification trees, boosting and support vector machines. In addition to unpacking the mathematics underlying the computational methods, students will also gain hands-on experience in applying these techniques to real datasets using R. Prerequisite: SDS 201, SDS 220 or CSC 210, or equivalent intro statistics course. Enrollment limited to 60. {M} Fall, Spring, Annually

SDS 300di Seminar: Topics in the Applications of Statistics and Data Science-Disability Inclusion and Data Analytics (4 Credits)

Students learn the social model of disability and critical disability theory as well as research design and process, and work on a research project analyzing disability inclusion public data. The statistical methods covered in this course may include logistic regression, multivariate analysis, factor analysis, etc. Students are expected to submit their final projects to a journal, conference or competition by the end of the semester. Prerequisite: SDS 201, SDS 220 or ECO 220. Restrictions: Juniors and seniors only. Enrollment limited to 15. Instructor permission required. {M}

Fall, Spring, Variable

SDS 300fs Seminar. Topics in the Applications of Statistics and Data Science-Understanding Food Systems through Engaging the Data (4 Credits)

This course examines the global Food System, with a focus on the US, through the examination of established data used to study the system and recommend food policy. In the United States, the US Department of Agriculture (USDA) oversees much of the food system, both promoting food products and regulating health impacts. In general, the operation of these systems generates a vast amount of data, much of which is in large open or semi-open online databases. Researchers and policy makers draw on these databases to aid in their decision making. This course aims to familiarize students with the data and its uses. Prerequisite: SDS 201 or SDS 220. Restrictions: Juniors and seniors only. Enrollment limited to 15. Instructor permission required. {M}{S} Fall, Spring, Alternate Years

SDS 320/ MTH 320 Mathematical Statistics (4 Credits)

Offered as MTH 320 and SDS 320. An introduction to the mathematical theory of statistics and to the application of that theory to the real world. Discussions include functions of random variables, estimation, likelihood and Bayesian methods, hypothesis testing and linear models. Prerequisites: a course in introductory statistics, MTH 212 and MTH 246, or equivalent. Enrollment limited to 20. {M} Spring

SDS 338/ GOV 338 Research Seminar in Political Networks (4 Credits)

Offered as GOV 338 and SDS 338. How does the behavior of a state, politician or interest group affect the behavior of others? Does Massachusetts's decision to legalize recreational marijuana influence Vermont's marijuana policies? From declarations of war to the decision of who congress members' voting alignments, social scientists are increasingly looking to political networks to recognize the interconnectedness of the world. This course presents the essentials of social network analysis and how they can be applied to American politics. Prerequisites: SDS 220 or an equivalent introductory statistics course. Restrictions: Juniors and seniors only. Enrollment limited to 12. Instructor permission required. {S} Fall, Spring, Alternate Years

SDS 364/ PSY 364 Research Seminar. Intergroup Relationships (4 Credits)

Offered as PSY 364 and SDS 364. Research on intergroup relationships and an exploration of theoretical and statistical models used to study mixed interpersonal interactions. Example research projects include examining the consequences of sexual objectification for both women and men, empathetic accuracy in interracial interactions and gender inequality in household labor. A variety of skills including, but not limited to, literature review, research design, data collection, measurement evaluation, advanced data analysis and scientific writing are developed. Prerequisites: PSY 201, SDS 201, SDS 220 or equivalent; and PSY 202. Restrictions: Juniors and seniors only. Enrollment limited to 12. Instructor permission required. {M}{N}S} Fall, Spring, Alternate Years

SDS 390be Topics in Statistical and Data Sciences-Methods in Biostatistics and Epidemiology (4 Credits)

Epidemiology concerns the distribution and determinants of disease in human populations, while biostatistics focuses on the development and application of statistical methods to a wide range of topics in biology, medicine and public health. This course focuses on foundational concepts in epidemiology, including measures of association and common epidemiological study designs, and statistical methods for public health data. Discussions include categorical data analysis (contingency table analysis, multinomial regression, ordinal regression and Poisson regression) and survival analysis (Kaplan-Meier estimators and Cox proportional hazards models). No background in biology is expected or required. Prerequisites: SDS 291 and [MTH 112 or (MTH 111 and MTH 153)]. Restrictions: SDS 390 may be taken a total of 3 times with different topics. Enrollment limited to 30. {M} Fall, Spring, Variable

SDS 390cd Topics in Statistical and Data Sciences-Categorical Data Analysis (4 Credits)

Theory and applications of statistical methods for the analysis of categorical data. The course includes an overview of statistical methods for analyzing discrete data including binary, multinomial and count response variables. Nominal and ordinal responses are considered. Discussions may include contingency table and chi-squared analyses, logistic, Poisson and negative-binomial regression models. R statistical software is used. Prerequisites: SDS 291 or SDS 290 or equivalent. Restrictions: SDS 390 may be taken a total of 3 times with different topics. Enrollment limited to 30.

Fall, Spring, Variable

SDS 400 Special Studies (1-4 Credits)

Normally for juniors and seniors. Instructor permission required. Fall, Spring

SDS 410 Capstone in Statistical & Data Sciences (4 Credits)

This one-semester course leverages students' previous coursework to address a real-world data analysis problem. Students collaborate in teams on projects sponsored by academia, government or industry. Professional skills developed include: ethics, project management, collaborative software development, documentation and consulting. Regular team meetings, weekly progress reports, interim and final reports, and multiple presentations are required. Open only to Statistical and Data Science majors. Prerequisites: SDS 192, SDS 291 and CSC 111. Restrictions: Statistical and Data Science majors only. Enrollment limited to 20. Instructor permission required. {M} Fall, Spring

SDS 430D Honors Thesis (4 Credits) Department permission required. Fall, Spring