

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. Not open to students who have already completed any of: SDS 192, SDS 201, SDS 220, SDS 290 or SDS 291. Concurrent registration required in any of: SDS 192, SDS 201, SDS 220, SDS 290 or SDS 291. S/ U only. 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. Throughout the semester, we explore concepts in decision-making, human perception, color theory and storytelling as they apply to data-driven communication. Whether you're an aspiring data scientist or you just want to learn new ways of presenting information, this course helps you build a strong foundation in how to talk to people about data. {M}

Fall, Spring, Alternate Years

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. Corequisite: 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. Classes meet for lecture/discussion and a required laboratory that emphasizes the analysis of real data. 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. Normally, students receive credit for only one of the following introductory statistics courses: SDS 201, PSY 201, ECO 220, GOV 203, SDS 220 or SOC 204. Corequisite: SDS 100 required for students who have not completed SDS 192, SDS 220, SDS 290 or SDS 291. {M}

Fall, Spring, Annually

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. Normally students receive credit for only one of the following introductory statistics courses: SDS 201, PSY 201, GOV 203, ECO 220, SDS 220 or SOC 204. Exceptions may be allowed in special circumstances with adviser and instructor permission. Corequisite: SDS 100 required for students who have not completed SDS 192, SDS 201, SDS 290 or SDS 291. Prerequisite: MTH 111 or equivalent. 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 may elect to take this course as a programming intensive course, prerequisite: 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, Annually

SDS 261 SQL for Data Science (1 Credit)

A continuation of ideas learned in SDS 192, this course develops abilities for using SQL databases within the data science pipeline. The core of the course focuses on the why and the how associated with writing SELECT queries in SQL. Additional topics include subqueries, indexes, keys and regular expressions. Students learn how to run SQL queries from both the RStudio IDE as well as from a relational database management system client like MySQL Workbench or DBeaver. Prerequisite: SDS 192. S/U only. Enrollment limited to 20. {E}

Interterm, Variable

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, Annually

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

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. Corequisite: 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, Annually

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: one of the following: SDS 201, PSY 201, GOV 203, SDS 220, ECO 220 or equivalent or a score of 4 or 5 on the AP Statistics examination. Corequisite: 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 Modeling for Machine Learning (4 Credits)

In the era of “big data,” statistical models are becoming increasingly sophisticated. This course begins with linear regression models and introduces students to a variety of techniques for learning from data, as well as principled methods for assessing and comparing models. Topics include bias-variance trade-off, resampling and cross-validation, linear model selection and regularization, classification and regression trees, bagging, boosting, random forests, support vector machines, generalized additive models, principal component analysis, unsupervised learning and k-means clustering. Emphasis is placed on statistical computing in a high-level language (e.g. R or Python). Prerequisites: SDS 291 and MTH 211 (may be concurrent). {M}

Fall, Spring, Annually

SDS 300cr Seminar: Topics in Statistical & Data Sciences Applications-Data Science for Coral Reef Conservation (4 Credits)

Students develop the skills and tools needed to process, analyze and visualize data related to large-scale coral reef conservation and management in R. Specifically, students work to collate data from NGOs, governments and academic researchers to assess changes in coral cover and community structure across the Caribbean. Quantifying these changes across spatial scales within the basin is essential in planning and managing the coral reefs of today and those of the future. Students use statistical and meta-analytical approaches to seek patterns in the data and build toward a final synthesis and presentation of these data. Enrollment limited to 12. Juniors and seniors only. Instructor permission required. {M}

Fall, Spring, Variable

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

Students will 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. Enrollment limited to 12. Juniors and seniors only. Instructor permission required. {M}

Fall, Variable

SDS 300ed Seminar: Topics in Applications-Statistics in Education (4 Credits)

Students will learn educational measurement and assessment and apply this knowledge to a research project analyzing educational data. Discussions will cover sensitivity and specificity, reliability, validity, item response theory, logistic regression and the Rasch model. Students will use this knowledge to evaluate the effectiveness of a new curriculum on the performance of at-risk low-income students. Research will also be conducted on an additional dataset to analyze the relationship between student/family characteristics and educational outcomes. Enrollment limited to 12. Juniors and seniors only. Instructor permission required. {M}

Fall, Spring, Variable

SDS 320/ MTH 320 Seminar: 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 12. Juniors and seniors only. Instructor permission required. {M}

Spring, Alternate Years

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 congressmembers will vote with, social scientists are increasingly looking to political networks to recognize the inter-connectedness of the world around us. This course will overview the essentials of social network analysis and how they are applied to give us a better understanding of American politics. Prerequisites: SDS 220 or an equivalent introductory statistics course. Enrollment limited to 12. Juniors and seniors only. Instructor permission required. {S}

Fall, Spring, Variable

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 will be developed. Prerequisites: PSY 201, SDS 201, SDS 220 or equivalent and PSY 202. Enrollment limited to 12. Juniors and seniors only. Instructor permission required. {M}{N}{S}

Fall, Spring, Alternate Years

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 will be considered. Topics may include contingency table and chi-squared analyses, logistic, Poisson and negative-binomial regression models. R statistical software will be used. Prerequisites: SDS 291 or SDS 290 or equivalent.

Fall, Variable

SDS 390ef Topics in Statistical and Data Sciences-Ecological Forecasting (4 Credits)

Ecologists are asked to respond to unprecedented environmental challenges. How can they provide the best scientific information about what will happen in the future? The goal of this seminar is to bring together the concepts and tools needed to make ecology a more predictive science. Topics include Bayesian calibration and the complexities of real-world data; uncertainty quantification, partitioning, propagation and analysis; feedback from models to measurements; state-space models and data fusion; iterative forecasting and the forecast cycle; and decision support. A semester-long project will center on data from the Smithsonian Conservation Biology Institute (SCBI) forestry reserve. Prerequisites: SDS 192, SDS 291 and either MTH 112 or (MTH 111 and MTH 153.) (E)

Fall, Variable

SDS 400 Special Studies (1-4 Credits)

Admission by permission of the program, normally for juniors and seniors.

Fall, Spring

SDS 410 Seminar: 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. Enrollment limited to 12. Statistical and Data Science majors only. Juniors and seniors only. {M}

Fall, Spring

SDS 430D Honors Thesis (4 Credits)

Fall, Spring, Annually