Data Science Courses (Statistics and Actuarial Science) (DATA)

DATA Courses

This is a list of courses with the subject code DATA. For more information, see Data Science and Statistics and Actuarial Science (College of Liberal Arts and Sciences) in the catalog.

DATA:1015 Introduction to Data Science 3 s.h.
In today's world, massive amounts of data are increasingly collected and leveraged for knowledge discovery, policy assessment, and decision-making across many fields, including business, natural sciences, social sciences, and humanities. Topics covered include data collection, visualization, and data wrangling; basics of probability and statistical inference; fundamentals of data learning, including regression, classification, prediction, and cross-validation; computing, learning, and reporting in the R environment; and literate programming and reproducible research. Requirements: one year of high school algebra or MATH:0100. GE: Quantitative or Formal Reasoning. Same as STAT:1015.

DATA:3120 Probability and Statistics 4 s.h.
Models, discrete and continuous random variables and their distributions, estimation of parameters, testing statistical hypotheses. Prerequisites: MATH:1560 or MATH:1860. Same as IGPI:3120, STAT:3120.

DATA:3200 Applied Linear Regression 3 s.h.
Regression analysis with focus on applications; model formulation, checking, and selection; interpretation and presentation of analysis results; simple and multiple linear regression; logistic regression; ANOVA; polynomial regression; tree models; bootstrapping; hands-on data analysis with computer software. Prerequisites: STAT:2020 or STAT:2010 or STAT:3120. Same as IGPI:3200, ISE:3760, STAT:3200.

DATA:4540 Statistical Learning 3 s.h.
Introduction to supervised and unsupervised statistical learning, with a focus on regression, classification, and clustering; methods will be applied to real data using appropriate software; supervised learning topics include linear and nonlinear (e.g., logistic) regression, linear discriminant analysis, cross-validation, bootstrapping, model selection, and regularization methods (e.g., ridge and lasso); generalized additive and spline models, tree-based methods, random forests and boosting, and support-vector machines; unsupervised learning topics include principal components and clustering. Requirements: an introductory statistics course and a regression course. Recommendations: prior exposure to programming and/or software, such as R, SAS, and Matlab. Same as BAIS:4540, IGPI:4540, STAT:4540.

DATA:4580 Data Visualization and Data Technologies 3 s.h.
Introduction to common techniques for visualizing univariate and multivariate data, data summaries, and modeling results; how to create and interpret these visualizations and assess effectiveness of different visualizations based on an understanding of human perception and statistical thinking; data technologies for obtaining and preparing data for visualization and further analysis; students learn how to present results in written reports and use version control to manage their work. Requirements: an introductory statistics course and a regression course. Recommendations: prior exposure to basic use of statistical programming software (e.g., R or SAS) as obtained from a regression course strongly recommended. Same as IGPI:4580, STAT:4580.

DATA:4600 Causal Inference for Data Science 3 s.h.
Introduce methods for reasoning about causes, effects, and bias when analyzing experimental and observational data. Topics include the potential outcomes framework, counterfactuals, confounding, and missing data; the identification and estimation of causal effects via propensity score methods, marginal structural models, instrumental variables, and directed acyclic graphs; as well as applications of machine learning and Bayesian methods to causal inference. Prerequisites: (DATA:3120 or STAT:3120) and (DATA:3200 or STAT:3200). Requirements: familiarity with the R programming. Same as STAT:4600.

DATA:4610 Data Acquisition and Management 3 s.h.
Introduction to common techniques for manipulating relational databases for data analysis; SQL and PostgreSQL fundamentals: querying, data manipulation and transformation, joins and subqueries, aggregation and grouping, data types and management; advanced topics: window functions, subqueries, common table expressions, indexing strategies, performance optimization techniques, security considerations; database building. Prerequisites: DATA:3200 or STAT:3200. Recommendations: Familiarity with basic programming logic, e.g., variables, loops, conditional statements.

DATA:4620 Text Data Analysis 3 s.h.
Introduction to text analytics techniques for real-world applications; Python fundamentals for text data exploration and manipulation; text processing via NLP libraries (NLTK, spaCy, Gensim); feature engineering; sentiment analysis; topic modeling; text summarization, machine translation, and deep learning applications. Prerequisites: (CS:1210 or DATA:5400) and DATA:4540. Recommendations: Basic knowledge of Python programming.

DATA:4750 Probabilistic Statistical Learning 3 s.h.
Essential machine learning and statistics ideas that are critical in analyzing modern complex and large data; supervised learning topics include linear models, deep neural networks, and nonparametric models; essential topics include nonlinear dimension reduction, clustering, and recommender systems. Prerequisites: (CS:1210 with a minimum grade of C- or ENGR:2730 with a minimum grade of C-) and (MATH:2700 or MATH:2550) and (STAT:2010 or STAT:2020 or STAT:4200) and STAT:4540. Same as STAT:4750.

DATA:4880 Data Science Creative Component 1 s.h.
Readings, group discussions, and short-term projects in area of data science; emphasis on communication of ideas learned in student's data science coursework, data ethics, and potential bias in algorithms.

DATA:4890 Data Science Practicum 3 s.h.
On- or off-campus internship or group-based consulting project that provides experience in a real-world setting; application of knowledge and techniques learned in coursework; practice in communicating results to others.

DATA:5400 Computing in Statistics 3 s.h.
R; database management; graphical techniques; importing graphics into word-processing documents (e.g., LaTeX); creating reports in LaTeX; SAS; simulation methods (Monte Carlo studies, bootstrap, etc.). Prerequisites: CS:1210 and STAT:3200 and (STAT:3120 or STAT:3101 or STAT:4101). Corequisites: STAT:5100 and STAT:5200 if not already completed. Same as IGPI:5400, STAT:5400.

DATA:5890 MS Data Science Practicum 2 s.h.
On- or off-campus internship or group-based consulting project that provides experience in a real-world setting; application of knowledge and techniques learned in coursework and practice communicating results to others.

DATA:6200 Predictive Analytics 3 s.h.
Linear mixed models; generalized linear mixed models; generalized additive models; applications of these models using associated R packages. Prerequisites: STAT:4560. Corequisites: STAT:4561. Requirements: comfort working with R software environment. Same as ACTS:6200, STAT:6200.

DATA:6220 Consulting and Communication With Data 3 s.h.
Realistic supervised data analysis experiences, including statistical packages, statistical graphics, writing statistical reports, dealing with complex or messy data. Offered spring semesters. Prerequisites: (STAT:3200 and STAT:3210) or (STAT:5201 and STAT:5200). Requirements: for undergraduate majors—major GPA of 3.00 or above, and grades of B or higher in STAT:3200 and STAT:3210. Same as STAT:6220.

DATA:7350 High-Dimensional Probability for Data Science 3 s.h.
Nonasymptotic probability with a view towards applications in data science; concentration inequalities for functions of independent variables, martingale inequalities, entropy method, random matrices, matrix inequalities, suprema of random processes, and sparse recovery. Prerequisites: STAT:5101. Requirements: linear algebra course and familiarity with R or Python.

DATA:7400 Computer Intensive Statistics 3 s.h.
Computer arithmetic, random variate generation, numerical optimization, numerical linear algebra, smoothing techniques, bootstrap methods, cross-validation, MCMC, EM and related algorithms; other topics per student/instructor interests. Prerequisites: (BIOS:5710 or STAT:5200) and STAT:3101 and STAT:5400. Requirements: proficiency in Fortran or C or C++ or Java. Same as IGPI:7400, STAT:7400.