Department of Statistics and Data Science Courses

About Course Numbers:

Each Carnegie Mellon course number begins with a two-digit prefix that designates the department offering the course (i.e., 76-xxx courses are offered by the Department of English). Although each department maintains its own course numbering practices, typically, the first digit after the prefix indicates the class level: xx-1xx courses are freshmen-level, xx-2xx courses are sophomore level, etc. Depending on the department, xx-6xx courses may be either undergraduate senior-level or graduate-level, and xx-7xx courses and higher are graduate-level. Consult the Schedule of Classes (https://enr-apps.as.cmu.edu/open/SOC/SOCServlet/) *each semester for course offerings and for any necessary pre-requisites or co-requisites.*

36-198 Research Training: Writing in Statistics

Intermittent
TBD
Prerequisite: 36-200

36-200 Reasoning with Data

All Semesters: 9 units
This course is an introduction to learning how to make statistical decisions and now to reason with data. The approach will emphasize the thinking-through of empirical problems from beginning to end and using statistical tools to look for evidence for/against explicit arguments/hypotheses. Types of data will include continuous and categorical variables, images, text, networks, and repeated measures over time. Applications will largely drawn from interdisciplinary case studies spanning the humanities, social sciences, and related fields. Methodological topics will include basic exploratory data analysis, elementary probability, significance tests, and empirical research methods. There will be once-weekly computer lab for additional hands-on practice using an interactive software platform that allows student-driven inquiry.

36-202 Methods for Statistics & Data Science

All Semesters: 9 units
This course builds on the principles and methods of statistical reasoning developed in 36-200 (or its equivalents). The course covers simple and multiple regression, basic analysis of variance methods, logistic regression, and introduction to data mining including classification and clustering. Students will also learn the principles of overfitting, training vs testing, ensemble methods, variable selection, and bootstrapping. Course objectives include applying the basic principles and methods that underlie statistical practice and empirical research to real data sets and interdisciplinary problems. Learning the Data Analysis Pipeline is strongly emphasized through structured coding and data analysis projects. In addition to three lectures a week, students attend a computer lab once a week for "hands-on" practice of the material covered in lecture. There is no programming language pre-requisite. Students will learn the basics of R Markdown and related analytics tools.
Prerequisites: 36-200 or 36-220 or 36-247 or 36-207 or 70-207

36-204 Discovering the Data Universe

Intermittent: 3 units
Every day we wake up in the data universe, we use the information around us to make decisions. We are constantly evaluating and interpreting data from our environment, in everything from spreadsheets to Instagram posts. At the same time, our own personal data are being observed and recorded and #8212;through websites we visit online, our smart devices, and even our interactions with other students and faculty at CMU. Navigating this data universe requires knowledge of what data is and how to use it responsibly. For example, can a plant be a data set? Discovering the truth behind a piece of data, including who made it, what it looks like, and what we can learn from it, is a critical skill. Understanding data can be the difference between being able to distinguish truth from lies; and the key to identifying your data footprint and succeeding in research and in your career. In this course, we will explore the data universe from multiple angles and across several types of data. We will define, find, and analyze data, and most importantly, identify narratives within data to tell stories about the world around us. We will examine data using the following questions: How can we tell multiple stories from the same dataset? What biases can exist in data? And, who creates or decides what data matters enough to collect, preserve, and share? NOTE: There will be one in person and one virtual pre-recorded lecture each week.

36-218 Probability Theory for Computer Scientists

Fall and Spring: 9 units
Probability theory is the mathematical foundation for the study of both statistics and of random systems. This course is an intensive introduction to probability,from the foundations and mechanics to its application in statistical methods and modeling of random processes. Special topics and many examples are drawn from areas and problems that are of interest to computer scientists and that should prepare computer science students for the probabilistic and statistical ideas they encounter in downstream courses and research. A grade of C or better is required in order to use this course as a pre-requisite for 36-226, 36-326, and 36-410. If you hold a Statistics primary/additional major or minor you will be required to complete 36-226. For those who do not have a major or minor in Statistics, and receive at least a B in 36-218, you will be eligible to move directly onto 36-401.
Prerequisites: (21-111 and 21-112) or 21-120 or 21-256 or 21-259

Course Website: http://www.stat.cmu.edu/academics/courselist (http://www.stat.cmu.edu/academics/courselist/)

36-219 Probability Theory and Random Processes

All Semesters: 9 units
This course provides an introduction to probability theory. It is designed for students in electrical and computer engineering. Topics include elementary probability theory, conditional probability and independence, random variables, distribution functions, joint and conditional distributions, limit theorems, and an introduction to random processes. Some elementary ideas in spectral analysis and information theory will be given. A grade of C or better is required in order to use this course as a pre-requisite for 36-226 and 36-410.
Prerequisites: (21-111 and 21-112) or 21-120 or 21-256 or 21-259

36-220 Engineering Statistics and Quality Control

Fall and Spring: 9 units
This is a course in introductory statistics for engineers with emphasis on modern product improvement techniques. Besides exploratory data analysis, basic probability, distribution theory and statistical inference, special topics include experimental design, regression, control charts and acceptance sampling.
Prerequisites: 21-120 or 21-112

36-225 Introduction to Probability Theory

Fall and Summer: 9 units
This course is the first half of a year-long course which provides an introduction to probability and mathematical statistics for students in the data sciences. Topics include elementary probability theory, conditional probability and independence, random variables, distribution functions, joint and conditional distributions, law of large numbers, and the central limit theorem.
Prerequisites: (21-112 and 21-111) or 21-120 or 21-256 or 21-259

Course Website: http://coursecatalog.web.cmu.edu/schools-colleges/dietrichcollegeofhumanitiesandsocialsciences/depar (http://coursecatalog.web.cmu.edu/schools-colleges/dietrichcollegeofhumanitiesandsocialsciences/depar/)

36-226 Introduction to Statistical Inference

Spring and Summer: 9 units
This course is the second half of a year-long course in probability and mathematical statistics. Topics include maximum likelihood estimation, confidence intervals, hypothesis testing, and properties of estimators, such as unbiasedness and consistency. If time permits there will also be a discussion of linear regression and the analysis of variance. A grade of C or better is required in order to advance to 36-401, 36-402 or any 36-46x course. Not open to students who have received credit for 36-626.
Prerequisites: 21-325 Min. grade C or 36-219 Min. grade C or 36-225 Min. grade C or 15-259 Min. grade C or 36-218 Min. grade C or 36-217 Min. grade C

36-235 Probability and Statistical Inference I

Fall: 9 units
This class is the first half of a two-semester, calculus-based course sequence that introduces theoretical aspects of probability and statistical inference to students. The material in this course and in 36-236 (Probability and Statistical Inference II) is organized so as to provide repeated exposure to essential concepts: the courses cover specific probability distributions and their inferential applications one after another, starting with the normal distribution and continuing with the binomial and Poisson distributions, etc. Topics specifically covered in 36-235 include basic probability, random variables, univariate and multivariate distribution functions, point and interval estimation, hypothesis testing, and regression, with the discussion being supplemented with computer-based examples and exercises (e.g., visualization and simulation). Given its organization, the course is only appropriate for those taking the full two-semester sequence, and thus it is currently open only to statistics majors (primary, additional, dual) and minors. (Check with the statistics advisors for the exact declaration deadline.) Non-majors/minors requiring a probability course are directed to take 36-225 or one of its analogues. A grade of C or better in 36-235 is required in order to advance to 36-236 (or 36-226) and/or 36-410. This course is not open to students who have received credit for 36-217, 36-218, 36-219, or 36-700, or for 21-325 or 15-259.
Prerequisites: (21-112 and 21-111) or 21-256 or 21-259 or 21-120

36-236 Probability and Statistical Inference II

Spring: 9 units
This class is the second half of a two-semester, calculus-based course sequence that introduces theoretical aspects of probability and statistical inference to students. The material in this course and in 36-235 (Probability and Statistical Inference I) is organized so as to provide repeated exposure to essential concepts: the courses cover specific probability distributions and their inferential applications one after another, starting with the normal distribution and continuing with the binomial and Poisson distributions, etc. Topics specifically covered in 36-236 include the binomial and related distributions, the Poisson and related distributions, and the uniform distribution, and how they are used in point and interval estimation, hypothesis testing, and regression. Also covered in 36-236 are topics related to multivariate distributions: marginal and conditional distributions, covariance, and conditional distribution moments. All discussion is supplemented with computer-based examples and exercises (e.g., visualization and simulation). Given its organization, the course is only appropriate for those who first take 36-235, and thus it is currently open only to statistics majors (primary, additional, dual) and minors, as well as to CS majors using both 36-235 and 36-236 to complete their probability requirement. All others are directed to take 36-226. A grade of C or better in 36-236 is required in order to advance to 36-401.
Prerequisite: 36-235 Min. grade C

36-290 Introduction to Statistical Research Methodology

Fall: 9 units
This is a first course in statistical practice, targeted to first-semester sophomores. It is designed as a high-level introduction to the ways by which statisticians go about approaching and analyzing quantitative observational data, thus preparing students for future work in capstone classes. Students in the course are taught the basic concepts of statistical learning and #8212;inference vs.prediction, supervised vs. unsupervised learning, regression vs. classification, etc. and #8212;and will reinforce this knowledge by applying, e.g., linear regression, random forest, principal components analysis, and/or hierarchical clustering and more to datasets provided by the instructor. Students will also practice disseminating the results of their analyses via oral presentations and posters. Analyses will be carried out using the R programming language.
Prerequisites: 36-220 or 70-207 or 36-207 or 36-200 or 36-247

Course Website: http://coursecatalog.web.cmu.edu/schools-colleges/dietrichcollegeofhumanitiesandsocialsciences/depar (http://coursecatalog.web.cmu.edu/schools-colleges/dietrichcollegeofhumanitiesandsocialsciences/depar/)

36-297 Early Undergraduate Research

Fall and Spring: 6 units
This course is designed to give early undergraduate students (those who have not yet taken 36-401) experience navigating real data science research problems. Small groups of students are matched with clients and do supervised research for a semester. From an academic perspective, the course presents an opportunity for students to gain skills in, e.g., data acquisition and cleaning, exploratory data analysis, and basic statistical modeling; which skills are practiced is project-dependent. Additionally, the course will help students develop the professional skills necessary for successfully navigating team-based project delivery roles. Programming will be performed in R and/or Python; previous programming experience is not required.

36-300 Statistics & Data Science Internship

Summer: 3 units
The Department of Statistics and amp; Data Science considers experiential learning as an integral part of our program. One such option is through an internship. If a student has an internship, they dont have to register for this class unless they want it listed on their official transcripts. This process should be used by international students interested in Curricular Practical Training (CPT) and should also be authorized by the Office of International Education (OIE). More information regarding CPT is available on OIE's website. This course will be taken as Pass/Fail, and students will be charged tuition for 3 units. There is an approval process in order to register for this course. Please contact your advisor the Department of Statistics and amp; Data Science for more details.

36-301 Documenting Human Rights

Intermittent: 9 units
This course will teach students about the origins of modern human rights and the evolution of methods to document the extent to which these rights are being upheld or violated. The need to understand and document human rights issues is at the center of the most pressing current events. From threats to democracy and civil rights to work holding perpetrators of mass harm accountable in legal proceedings to efforts to quantify and advance economic, social, cultural, and environmental rights, making human rights violations visible is fundamental to achieving a more just world. We will begin with an overview of the history of human rights, the main philosophical and political debates in the field, and the most relevant organizations, institutions, and agreements. We will then delve into specific cases that highlight methodological opportunities and challenges, including: the identification of mass atrocity victims, the disappeared, and missing migrants; efforts to estimate civilian casualties in war; the documentation of police brutality and other human rights violations with smartphones; as well as the use of satellite imagery and drone footage for the documentation of genocide, environmental rights, and war crimes. We will critically assess the technical challenges that arise in each context and how the human rights and scientific communities have responded. After reviewing these cases, we will conclude by reflection on why the documentation of human rights actually matters and what happens to evidence once it is gathered. Students will then take what they've learned and do two multidisciplinary group projects, one involving the document of a rights violation in Western Pennsylvania and the other involving an international situation. Assignments include an essay, a data analysis assignment, and a group project that include a written component, quantitative and/or qualitative data analysis, and a presentation.

36-303 Sampling, Survey and Society

Spring: 9 units
This course will revolve around the role of sampling and sample surveys in the context of U.S. society and its institutions. We will examine the evolution of survey taking in the United States in the context of its economic, social and political uses. This will eventually lead to discussions about the accuracy and relevance of survey responses, especially in light of various kinds of nonsampling error. Students will be required to design, implement and analyze a survey sample.
Prerequisites: 70-208 or 36-236 or 36-218 Min. grade B or 36-208 or 36-202 or 36-309 or 36-220 or 36-226 or 36-326

36-309 Experimental Design for Behavioral & Social Sciences

Fall and Summer: 9 units
This course focuses on the statistical aspects of the design and analysis stages of planned experiments. The design stage focuses on determining how experimental factors are allocated, the sample size necessary to achieve adequate statistical power, and how subjects/variables are measured. The analysis stage focuses on how data are collected and which statistical models are most appropriate to answer the research questions of interest. Although students will have to do some computer programming to implement these statistical techniques, the most important aspect of the course will be on interpreting analyses' results (e.g., whether a given analysis is appropriate, to what extent that analysis can answer research questions of interest, and the broader implications of an analysis within the context of the experiment). In addition to a weekly lecture, students will attend a computer lab once a week to get guidance and hands-on practice implementing statistical techniques we learn in class.
Prerequisites: 36-218 or 70-207 or 36-326 or 36-226 or 36-220 or 15-260 or 36-247 or 36-200 or 36-236

Course Website: http://www.stat.cmu.edu/academics/courselist (http://www.stat.cmu.edu/academics/courselist/)

36-311 Statistical Analysis of Networks

Intermittent: 9 units
Networks are omnipresent. In this course, students will get an introduction to network science, mainly focusing on social network analysis. The course will start with some empirical background, and an overview of concepts used when measuring and describing networks. We will also discuss network visualization. Most traditional models cannot be applied straightforwardly to social network data, because of their complex dependence structure. We will discuss random graph models and statistical network models, that have been developed for the study of network structure and growth. We will also cover models of how networks impact individual behavior.
Prerequisite: 36-226

36-313 Statistics of Inequality and Discrimination

Intermittent: 9 units
Many social questions about inequality, injustice and unfairness are, in part, questions about evidence, data, and statistics. This class lays out the statistical methods which let us answer questions like "Does this employer discriminate against members of that group?", "Is this standardized test biased against that group?", "Is this decision-making algorithm biased, and what does that even mean?" and "Did this policy which was supposed to reduce this inequality actually help?" We will also look at inequality within groups, and at different ideas about how to explain inequalities between groups. The class will interweave discussion of concrete social issues with the relevant statistical concepts.
Prerequisite: 36-202

36-315 Statistical Graphics and Visualization

All Semesters: 9 units
Graphical displays of quantitative information take on many forms, and they help us understand data and statistical methods by (hopefully) clearly communicating arguments, results, and ideas. This course introduces students to the most common forms of graphical displays and their uses and misuses. Ideally, graphs are designed according to three key elements: The data structure, the graph's audience, and the designer's intended message. Students will learn how to create well-designed graphs and understand them from a statistical perspective. Furthermore, the course will consider complex data structures that are becoming increasingly common in data visualizations (temporal, spatial, and text data); we will discuss common ways to process these data that make them easy to visualize. As time permits, we may also consider more advanced graphical methods (e.g., interactive graphics and computer-generated animations). In addition to two weekly lectures, there will be weekly computer labs and homework assignments where students use R to visualize and analyze real datasets. Along the way, students also make monthly Piazza posts discussing the strengths and weaknesses of a graph they found online, thereby critiquing real graphical designs found in the wild. The course culminates in a group final project, where students make public-facing data visualizations and analyses for a real dataset. All assignments will be in R; although this is not a programming class, using programming-based statistical software like R is essential to create modern-day graphics, and this class will give you practice using this kind of software. Throughout, communication skills (usually written or visual, but sometimes spoken) will play an important role. Indeed, if it's true that "a picture speaks a thousand words," then ideally the one thousand words you are communicating with your graphics are statistically correct, clear, and compelling.
Prerequisites: 36-309 or 36-225 or 36-218 or 70-208 or 36-202 or 36-219 or 36-235 or 36-208 or 15-259 or 21-325

36-318 Introduction to Causal Inference

Intermittent: 9 units
Many social science and scientific inquiries can be framed as causal questions. Does a new cancer treatment cause a reduction in mortality? Do financial grants cause students to do better in college? Does a new public policy cause an increase in voter turnout? When tackling these questions, we frequently come across the phrase "correlation does not imply causation." If that's the case, then what does imply causation? In this course, we will discuss causal inference methods for measuring causal effects of different interventions (e.g., drug treatments, financial grants, and public policies). First, we will discuss how experiments and #8212;-where interventions are randomized among subjects and #8212;-can imply causation when an appropriate experimental design and statistical analysis is used. Then, we will discuss how observational studies and #8212;-where interventions are not randomized and #8212;-can also imply causation when approaches like propensity score methods, matching, and doubly robust estimation are employed. Finally, we will discuss instrumental variables and regression discontinuity designs and #8212;-which are frequently used in medicine and public policy for establishing causal inferences. Throughout we will use R to conduct causal analyses. A working knowledge of regression is encouraged, but regression will also be discussed and taught during much of the course.
Prerequisites: 15-259 Min. grade C or 36-225 Min. grade C or 36-219 Min. grade C or 36-218 Min. grade C or 36-235 Min. grade C or 21-325 Min. grade C

36-326 Mathematical Statistics (Honors)

Spring: 9 units
This course is a rigorous introduction to the mathematical theory of statistics. A good working knowledge of calculus and probability theory is required. Topics include maximum likelihood estimation, confidence intervals, hypothesis testing, Bayesian methods, and regression. A grade of C or better is required in order to advance to 36-401, 36-402 or any 36-46x course. Not open to students who have received credit for 36-625.
Prerequisites: 15-359 or 21-325 or 36-217 or 36-225 with a grade of A AND advisor approval. Students interested in the course should add themselves to the waitlist pending review.
Prerequisites: 36-218 Min. grade A or 21-325 Min. grade A or 36-217 Min. grade A or 36-225 Min. grade A or 15-359 Min. grade A

36-350 Statistical Computing

All Semesters: 9 units
Statistical Computing is a one-semester course that will introduce you to the fundamentals of computational data analysis, as carried out in the R programming language, and to the fundamentals of working with relational databases, such as SQLite. No previous knowledge of either is required.
Prerequisites: 21-325 Min. grade C or 36-218 Min. grade C or 36-219 Min. grade C or 36-225 Min. grade C or 36-217 Min. grade C or 15-259 Min. grade C or 36-235 Min. grade C

36-390 Study Abroad Experience in Statistics and Data Science

Summer: 9 units
Statistics and Data Science at the Monteverde Institute in Costa Rica. This is a five-week study abroad experience in which students will directly engage with, and will process, visualize, and/or analyze data collected by, researchers at the institute. Students will also have the opportunity to participate in data collection, as appropriate. The mission of the institute is to promote sustainable practices that benefit both the local community and local wildlife, and the data that students can examine include, but are not limited to, ecological data on bats, birds, reforestation, and stream beds, as well as data arising from community surveys. This course does not require prior knowledge of, or exposure to, data processing, visualization, or analysis techniques beyond what is covered in the prerequisite classes, and necessary techniques and methods will be introduced and discussed in daily classes. Project goals will be modified for students with more advanced backgrounds (e.g., students who have completed 36-401 and 36-402). The 2024 class is limited to six students overall.

36-400 Introduction to Statistical Modeling and Learning

Spring: 9 units
This course is a high-level introduction both to fundamental concepts of probability and statistics and to the ways by which statisticians go about approaching and analyzing data. The course will cover data processing, exploratory data analysis, parameter estimation and hypothesis testing, clustering, and common regression and classification models. Students will carry out work using the R and Python programming languages. This course is open only to students not majoring in Stat and amp; DS who have taken the prerequisite courses.
Prerequisites: 36-200 and (36-309 or 36-202 or 36-290)

36-401 Modern Regression

Fall: 9 units
This course is an introduction to the real world of statistics and data analysis using linear regression modeling. We will explore real data sets, examine various models for the data, assess the validity of their assumptions, and determine which conclusions we can make (if any). We will use the R programming language to implement our analyses and produce graphs and tables of results. Data analysis is a bit of an art; there may be several valid approaches. We will strongly emphasize the importance of critical thinking about the data and the question of interest. Our overall goal is to use data and a basic set of modeling tools to answer substantive questions, and to present the results in a scientific report.
Prerequisites: (36-236 Min. grade C or 36-326 Min. grade C or 36-226 Min. grade C or 36-218 Min. grade B) and (21-242 or 21-240 or 21-241)

36-402 Advanced Methods for Data Analysis

Spring: 9 units
This course introduces modern methods of data analysis, building on the theory and application of linear models from 36-401. Topics include nonlinear regression, nonparametric smoothing, density estimation, generalized linear and generalized additive models, simulation and predictive model-checking, cross-validation, bootstrap uncertainty estimation, multivariate methods including factor analysis and mixture models, and graphical models and causal inference. Students will analyze real-world data from a range of fields, coding small programs and writing reports.
Prerequisite: 36-401 Min. grade C

36-410 Introduction to Probability Modeling

Spring: 9 units
An introductory-level course in stochastic processes. Topics typically include Poisson processes, Markov chains, birth and death processes, random walks, recurrent events, and renewal theory. Examples are drawn from reliability theory, queuing theory, inventory theory, and various applications in the social and physical sciences.
Prerequisites: 21-325 or 15-259 or 36-225 or 36-235 or 36-217

36-460 Special Topics: Sports Analytics

Spring: 9 units
This course introduces students to fundamental topics in sports analytics and the relevant statistical methods for tackling problems in this growing area. The first half of the course will cover foundational topics in sports analytics including models for the expected value of game states, win probability, team ratings, and hierarchical models for player evaluation. The second half of the course will focus on spatio-temporal methods appropriate for modeling complex player-tracking data. The focus is on understanding the foundations of the considered methods and introducing software for implementation. Students will develop their own sports analytics project using techniques covered in the course for their final assessment.
Prerequisite: 36-401 Min. grade C

36-461 Special Topics: Statistical Methods in Epidemiology

Intermittent: 9 units
Epidemiology is concerned with understanding factors that cause, prevent, and reduce diseases by studying associations between disease outcomes and their suspected determinants in human populations. Epidemiologic research requires an understanding of statistical methods and design. Epidemiologic data is typically discrete, i.e., data that arise whenever counts are made instead of measurements. In this course, methods for the analysis of categorical data are discussed with the purpose of learning how to apply them to data. The central statistical themes are building models, assessing fit and interpreting results. There is a special emphasis on generating and evaluating evidence from observational studies. Case studies and examples will be primarily from the public health sciences.
Prerequisite: 36-401 Min. grade C

Course Website: http://coursecatalog.web.cmu.edu/schools-colleges/dietrichcollegeofhumanitiesandsocialsciences/depar (http://coursecatalog.web.cmu.edu/schools-colleges/dietrichcollegeofhumanitiesandsocialsciences/depar/)

36-462 Special Topics: Statistical Machine Learning

Intermittent: 9 units
Data mining is the science of discovering patterns and learning structure in large data sets. Covered topics include information retrieval, clustering, dimension reduction, regression, classification, and decision trees.
Prerequisite: 36-401 Min. grade C

Course Website: http://www.stat.cmu.edu/academics/courselist (http://www.stat.cmu.edu/academics/courselist/)

36-463 Special Topics: Multilevel and Hierarchical Models

Intermittent: 9 units
Multilevel and hierarchical models are among the most broadly applied "sophisticated" statistical models, especially in the social and biological sciences. They apply to situations in which the data "cluster" naturally into groups of units that are more related to each other than they are the rest of the data. In the first part of the course we will review linear and generalized linear models. In the second part we will see how to generalize these to multilevel and hierarchical models and relate them to other areas of statistics, and in the third part of the course we will learn how Bayesian statistical methods can help us to build, estimate and diagnose problems with these models using a variety of data sets and examples.
Prerequisite: 36-401 Min. grade C

Course Website: http://www.stat.cmu.edu/academics/courselist (http://www.stat.cmu.edu/academics/courselist/)

36-464 Special Topics: Psychometrics: A Statistical Modeling Approach

Intermittent: 9 units
Much of the social, educational, policy, and professional worlds involve measuring the skills, abilities, attitudes, decision-making, etc. of people and #8212; from SAT's and GRE's for school, to 360-evaluations in business. This is the field of modern psychometrics, and it involves (at least) two kinds of craft: designing good sets of questions, and designing and fitting statistical models that extract the information we want from the responses to those questions. In this course we will touch on both kinds of craft, but we will concentrate on the second: what do statistical models for psychometric data look like, and how can we design, fit, and use them in practice? We will look at these models from a variety of statistical perspectives, but we will concentrate on the applied Bayesian point of view.
Prerequisite: 36-401 Min. grade C

Course Website: http://www.stat.cmu.edu/academics/courselist (http://www.stat.cmu.edu/academics/courselist/)

36-465 Special Topics: Conceptual Foundations of Statistical Learning

Intermittent: 9 units
This class is an introduction to the foundations of statistical learning theory, and its uses in designing and analyzing machine-learning systems. Statistical learning theory studies how to fit predictive models to training data, usually by solving an optimization problem, in such a way that the model will predict well, on average, on new data. The course will focus on the key concepts and theoretical tools, at a level accessible to students who have taken 36-401 and its pre-requisites. The course will also illustrate those concepts and tools by applying them to carefully selected kinds of machine learning systems (such as kernel machines). Students wanting exposure to a broad range of algorithms and applications would be better served by 36-462/662 ("Data Mining"). This class is for those who want a deeper understanding of the principles underlying all machine learning methods.
Prerequisite: 36-401 Min. grade C

36-466 Special Topics: Statistical Methods in Finance

Intermittent: 9 units
Financial econometrics is the interdisciplinary area where we use statistical methods and economic theory to address a wide variety of quantitative problems in finance. These include building financial models, testing financial economics theory, simulating financial systems, volatility estimation, risk management, capital asset pricing, derivative pricing, portfolio allocation, proprietary trading, portfolio and derivative hedging, and so on and so forth. Financial econometrics is an active field of integration of finance, economics, probability, statistics, and applied mathematics. Financial activities generate many new problems and products, economics provides useful theoretical foundation and guidance, and quantitative methods such as statistics, probability and applied mathematics are essential tools to solve quantitative problems in finance. Professionals in finance now routinely use sophisticated statistical techniques and modern computation power in portfolio management, proprietary trading, derivative pricing, financial consulting, securities regulation, and risk management.
Prerequisite: 36-401

36-467 Special Topics: Data over Space & Time

Intermittent: 9 units
This course is an introduction to the opportunities and challenges of analyzing data from processes unfolding over space and time. It will cover basic descriptive statistics for spatial and temporal patterns; linear methods for interpolating, extrapolating, and smoothing spatio-temporal data; basic nonlinear modeling; and statistical inference with dependent observations. Class work will combine practical exercises in R, a little mathematics on the underlying theory, and case studies analyzing real problems from various fields (economics, history, meteorology, ecology, etc.). Depending on available time and class interest, additional topics may include: statistics of Markov and hidden-Markov (state-space) models; statistics of point processes; simulation and simulation-based inference; agent-based modeling; dynamical systems theory.
Prerequisite: 36-401 Min. grade C

Course Website: http://coursecatalog.web.cmu.edu/schools-colleges/dietrichcollegeofhumanitiesandsocialsciences/depar (http://coursecatalog.web.cmu.edu/schools-colleges/dietrichcollegeofhumanitiesandsocialsciences/depar/)

36-468 Special Topics: Text Analysis

Intermittent: 9 units
The analysis of language is concerned with how variables relate to people (their gender, age, and location, for example), how variables relate to use (such as writing in different academic disciplines), and how variables change over time. While we are surrounded by data that might potentially shed light on many of these questions, working with real-world linguistic data can present some unique challenges in sampling, in the distribution of features, and in their high dimensionality. In this course, we work through some of these issues, paying particular attention to the aligning of the statistical questions we want to investigate with the choice of statistical models, as well as focusing on the interpretation of results. Analysis will be carried out in R and students will develop a suite of tools as they work through their course projects.
Prerequisites: 36-218 Min. grade B or 36-226 Min. grade C or 36-236 Min. grade C

36-469 Special Topics: Statistical Genomics and High Dimensional Inference

Intermittent: 9 units
The field of computational and statistical genomics focuses on developing and applying computationally efficient and statistically robust methods to sort through increasingly rich and massive genome wide data sets to identify complex genetic patterns, gene interactions, and disease associations. Because the genome is vast, analytical approaches require high dimensional statistical approaches such as multiple testing, dimension reduction techniques, regularization and high dimensional regression analysis, best linear unbiased prediction models, networks and graphical models. In this course, we will motivate these topics using data obtained from the human genetic and genomic literature. No prior knowledge in biology is required.
Prerequisite: 36-401 Min. grade C

36-470 Special Topics: Statistical Methods in Health Sciences

Intermittent: 9 units
As the volume of health and clinical data continues to expand, the integration of statistical and machine learning methods becomes increasingly important for enhancing healthcare efficiency. However, there are challenges in modeling health data, for example, annotated data is often limited or subject to incompleteness. In this course, we will introduce statistical methods that address these challenges, including survival analysis, latent variable models, clustering, semi-supervised learning, and so on. An emphasis will put on understanding methodological foundations and how to appropriately apply methods to health data. Through homework assignments, labs, paper presentations, and a final project, students will gain hands-on-experience in applying statistical methods to solve problems arising from health sciences.
Prerequisite: 36-401 Min. grade C

36-471 Special Topics: Time Series

Fall: 9 units
This course covers time series analysis from fundamentals to advanced models in both time and frequency domains. The focus is on practical execution and interpretation of time series analyses with realistic real-world data.
Prerequisite: 36-401

36-490 Undergraduate Research

Fall and Spring: 9 units
This course is designed to give undergraduate students experience using statistics in real research problems. Small groups of students are matched with clients and do supervised research for a semester. From an academic perspective, the course presents an opportunity for students to gain skills in approaching a research problem, critical thinking, and statistical analyses. Additionally, the course will help students develop the professional skills necessary for successfully navigating team-based project delivery roles. Client-facing and collaborative skills will be emphasized within a team setting, and students will learn leading practices for engaging stakeholders as well as gain a conceptual understanding of leading practices for project delivery.

36-497 Corporate Capstone Project

Fall and Spring: 9 units
This course is designed to give undergraduate students experience applying statistics data science methodology to real industry projects. Small groups of students will be matched with industry clients and do supervised projects for a semester. From an academic perspective, the course presents an opportunity for students to gain skills in approaching a research problem, critical thinking, and statistical analyses. Additionally, the course will help students develop the professional skills necessary for successfully navigating team-based project delivery roles. Client-facing and collaborative skills will be emphasized within a team setting, and students will learn leading practices for engaging stakeholders as well as gain a conceptual understanding of leading practices for project delivery. The industry clients will change and rotate each semester; available projects will be advertised prior to the first week of class. The course size is limited; students apply the previous semester and placed on the course waitlist until project matching is performed. Students with skill sets matching project needs will be given priority. We will also take into consideration whether or not a student has had a recent prior corporate capstone experience with the goal of providing experiences to a broad group of qualified students. Note that there is no guarantee a waitlisted student will be matched to a project in any given semester.

36-498 Corporate Capstone II

Fall and Spring
This course allows students to continue work on projects begun as part of 36-497, Corporate Capstone Project. Enrollment is at the discretion of the external advisor for the 36-497 project and the Department of Statistics and amp; Data Science.

36-680 Quantitative Financial Analytics and Algorithmic Trading

Fall and Spring: 12 units
Algorithmic trading serves as a practical application of software engineering and data science methodologies and quantitative analysis techniques within the context of financial markets. This project-based course offers an introduction to algorithmic trading and the principles behind it, while emphasizing universally applicable engineering concepts and data-driven methodologies. Students will gain an understanding of the fundamentals of financial markets and trading systems, learn how to manage data, generate signals, backtest strategies, and use APIs to execute trades. Additionally, they will apply risk management principles, position sizing, and software development best practices such as unit testing in Python. Most importantly, the course will teach students specific thinking patterns and data science methodologies that can be applied across various engineering and data analysis fields. Students will be equipped with a toolbox needed to continue researching trading strategies, predictive analytics, or other data science-related topics independently. Following condensed lecture videos, the course will emulate a professional environment through a series of individual assignments culminating in a functional project. Delivery of the project will be guided by direct instruction, Q and amp;A calls, and an online chat group with the lecturers, similar to a real workplace. Students will deliver a functional project in Python, according to a specification, while also taking exams on the theoretical materials covered in the lectures. Student progress is assessed through the delivery of practical projects according to a specification and evaluation criteria. While there are no prerequisites for this course, an understanding of statistics, probabilities, hypothesis testing, measures of spread, confidence intervals, and related topics is assumed.

36-700 Probability and Mathematical Statistics

Fall: 12 units
This is a one-semester course covering the basics of statistics. We will first provide a quick introduction to probability theory, and then cover fundamental topics in mathematical statistics such as point estimation, hypothesis testing, asymptotic theory, and Bayesian inference. If time permits, we will also cover more advanced and useful topics including nonparametric inference, regression and classification. Prerequisites: one- and two-variable calculus and matrix algebra. Graduate students in degree-seeking programs are given priority.