STATISTICS

Courses offered by the Department of Statistics are listed under the subject code STATS on the Stanford Bulletin's ExploreCourses (http://explorecourses.stanford.edu/CourseSearch/search/?view=catalog&catalog=&page=0&q=STATS&filter-catalognumber-STATS=on) web site.

The department's goals are to acquaint students with the role played in science and technology by probabilistic and statistical ideas and methods, to provide instruction in the theory and application of techniques that have been found to be commonly useful, and to train research workers in probability and statistics. There are courses for general students as well as those who plan careers in statistics in business, government, industry, and teaching.

The department has long recognized the relation of statistical theory to applications. It has fostered this by encouraging a liaison with other departments in the form of joint and courtesy faculty appointments, as well as membership in various interdisciplinary programs: Biomedical Data Science, Bio-X, Center for Computational, Evolutionary and Human Genomics, Computer Science, Economics, Education, Electrical Engineering, Environmental Earth System Science, Genetics, Mathematics, Mathematical and Computational Finance, and Medicine. The research activities of the department reflect an interest in applied and theoretical statistics and probability. There are workshops in biology/medicine and in environmental factors in health.

In addition to courses for Statistics students, the department offers a number of service courses designed for students in other departments. These tend to emphasize the application of statistical techniques rather than their theoretical development.

The department has always drawn visitors from other countries and universities, and as a result there are a wide range of seminars offered by both the visitors and the department's own faculty.

Undergraduate Programs in Statistics

The department offers a minor in Statistics and in Data Science (https://statistics.stanford.edu/academics/undergraduate-programs/). Program details can be found under the Minor section.

Undergraduates Interested in Statistics

Students wishing to build a concentration in probability and statistics are encouraged to consider declaring a major in Mathematical and Computational Science (https://mcs.stanford.edu/). This interdisciplinary program is administered in the Department of Statistics and provides core training in computing, mathematics, operations research, and statistics, with opportunities for further elective work and specialization. See the "Mathematical and Computational Science" section of this bulletin.

Graduate Programs in Statistics

University requirements for the M.S. and Ph.D. degrees are discussed in the "Graduate Degrees (http://exploredegrees.stanford.edu/graduatedegrees/)" section of this bulletin.

Learning Outcomes (Graduate)

The purpose of the master's program is to further develop knowledge and skills in Statistics and to prepare students for a professional career or doctoral studies. This is achieved through completion of courses, in the primary field as well as related areas, and experience with independent work and specialization.

The Ph.D. is conferred upon candidates who have demonstrated substantial scholarship and the ability to conduct independent research and analysis in Statistics. Through completion of advanced course work and rigorous skills training, the doctoral program prepares students to make original contributions to the knowledge of Statistics and to interpret and present the results of such research.

The Department of Statistics offers two minor programs for undergraduates, a minor in Data Science (p. 1) and a minor in Statistics (p. 2). To declare either minor for a degree program, visit the Statistics website (https://statistics.stanford.edu/academic-programs/undergraduate-programs/) and submit the appropriate form to the department.

Minor in Data Science

The undergraduate Data Science minor has been designed for majors in the humanities and social sciences who want to gain practical knowledge of statistical data analytic methods as it relates to their field of interest. The minor:

- provides students with the knowledge of exploratory and confirmatory data analyses of diverse data types such as text, numbers, images, graphs, trees, and binary input)
- strengthens social research by teaching students how to correctly apply data analysis tools and the techniques of data visualization to convey their conclusions.

No previous programming or statistical background is assumed.

Learning Outcomes

Students are expected to:

1. be able to connect data to underlying phenomena and to think critically about conclusions drawn from data analysis.
2. be knowledgeable about programming abstractions so that they can later design their own computational inferential procedures

All courses for the minor must be taken for a letter grade, with the exception of the Data Mining requirement.

Seven courses are required, 22 units minimum. An overall 2.75 grade point average (GPA) is required for courses fulfilling the minor.

Requirements

Linear Algebra

		Units
One of the following:		
MATH 51	Linear Algebra, Multivariable Calculus, and Modern Applications	5
CME 100	Vector Calculus for Engineers	5

Programming

		Units
CS 106A	Programming Methodology	3-5

Programming in *R*

		Units
One of the following:		
THINK 3		4
STATS 32	Introduction to R for Undergraduates	1
STATS 48N	Riding the Data Wave	3
STATS 195	Introduction to R	1
Or other course that teaches proficiency in R programming.		

Data Science

		Units
STATS 101	Data Science 101	5
STATS 191	Introduction to Applied Statistics	3
CS 102		
MS&E 226	Fundamentals of Data Science: Prediction, Inference, Causality	3

Statistics

		Units
One of the following:		
ECON 102A	Introduction to Statistical Methods (Postcalculus) for Social Scientists	5
PHIL 166	Probability: Ten Great Ideas About Chance	4
STATS 48N	Riding the Data Wave	3
STATS 141	Biostatistics	5
STATS 191	Introduction to Applied Statistics	3
STATS 211	Meta-research: Appraising Research Findings, Bias, and Meta-analysis	3

Data Mining and Analysis

STATS 202	Data Mining and Analysis (may be taken CR/NC)	3
STATS 216	Introduction to Statistical Learning	3

Elective Course

		Units
One course fulfilling Data Science methodology from cognate field of interest. Suggested courses:		
CS 224W	Machine Learning with Graphs	3-4
ECON 291	Social and Economic Networks	3-5
ENGLISH 184E	Literary Text Mining	5
LINGUIST 275	Probability and Statistics for linguists	2-4
MS&E 135	Networks	3
PHIL 166	Probability: Ten Great Ideas About Chance	4
POLISCI 150B	Machine Learning for Social Scientists	5
POLISCI 450A	Political Methodology I: Regression	5
PSYCH 109	An introduction to computation and cognition	4
PUBLPOL 105	Empirical Methods in Public Policy	4-5
SOC 126	Introduction to Social Networks	4
SOC 180A	Foundations of Social Research	4
or SOC 180B	Introduction to Data Analysis	

*STATS 191 cannot count for both requirements.

return to top (p. 1)

Minor in Statistics

The undergraduate minor in Statistics is designed to complement major degree programs primarily in the social and natural sciences. Students with an undergraduate Statistics minor should find broadened possibilities for employment. The Statistics minor provides valuable preparation for professional degree studies in postgraduate academic programs.

The minor consists of a minimum of six courses with a total of at least 19 units. There are two required courses (8 units) and four qualifying or elective courses (12 or more units). All courses for the minor must be taken for a letter grade. An overall 2.75 grade point average (GPA) is required for courses fulfilling the minor.

Required Courses

		Units
Both:		
STATS 116	Theory of Probability	4
STATS 200	Introduction to Statistical Inference	4

Qualifying Courses

At most, one of these two courses may be counted toward the six course requirement for the minor.

		Units
Choose one from the following:		
MATH 52	Integral Calculus of Several Variables	5
STATS 191	Introduction to Applied Statistics	3

Three Elective Courses

At least one of the elective courses should be a STATS 200-level course. The remaining two elective courses may also be 200-level courses. Alternatively, one or two elective courses may be approved courses in other departments. Special topics courses and seminars for undergraduates are offered from time to time by the department, and these may be counted toward the course requirement. Students may not count any Statistics courses below the 100 level toward the minor.

Examples of elective course sequences are:

		Units
Data Analysis and Applied Statistics		
STATS 202	Data Mining and Analysis	3
STATS 203	Introduction to Regression Models and Analysis of Variance	3
Statistical Methodology		
STATS 205	Introduction to Nonparametric Statistics	3
STATS 206	Applied Multivariate Analysis	3
STATS 207	Introduction to Time Series Analysis	3
Economic Optimization		
STATS 206	Applied Multivariate Analysis	3
ECON 160	Game Theory and Economic Applications	5
Psychology Modeling and Experiments		
STATS 206	Applied Multivariate Analysis	3
Signal Processing		
STATS 207	Introduction to Time Series Analysis	3
EE 264	Digital Signal Processing	3
EE 279	Introduction to Digital Communication	3
Genetic and Ecologic Modeling		
STATS 217	Introduction to Stochastic Processes I	3
BIO 283	Theoretical Population Genetics	3
Probability and Applications		
STATS 217	Introduction to Stochastic Processes I	3
STATS 218	Introduction to Stochastic Processes II	3
Mathematical Finances		
STATS 240	Statistical Methods in Finance	3
STATS 243	Risk Analytics and Management in Finance and Insurance	3
STATS 250	Mathematical Finance	3

return to top (p. 1)

Master of Science in Statistics

The University's basic requirements for the M.S. degree are discussed in the "Graduate Degrees" (http://exploredegrees.stanford.edu/graduatedegrees/) section of this bulletin. The following are specific departmental requirements.

The M.S. in Statistics and the M.S. in Statistics, Data Science track, are intended as terminal degree programs and do not lead to the Ph.D. program in Statistics. Students interested in pursuing doctoral study in Statistics should apply directly to the Ph.D. program.

Admission

Prospective applicants should consult the Graduate Admissions (https://gradadmissions.stanford.edu/) and the Statistics Department admissions webpages (https://statistics.stanford.edu/admissions/graduate-application-information-and-instructions/) for complete information on admission requirements and deadlines.

Recommended preparatory courses include advanced undergraduate level courses in linear algebra, statistics/probability and proficiency in programming.

Stanford students interested in the Data Science track (subplan) in Statistics must apply as external candidates. Visit Graduate Admissions (https://gradadmissions.stanford.edu/) to start an application.

Coterminal Master's Program

Stanford undergraduates who want to apply for the coterminal master's degree in Statistics must submit a complete application to the department by the deadline published on the department's coterminal admissions webpage. (https://statistics.stanford.edu/admissions/ms/statistics-coterm-eligibility/)

Applications are accepted twice a year in autumn and winter quarters for winter and spring quarter start, respectively. The general GRE is not required of coterminal applicants.

Students pursuing the Statistics coterminal master's degree must follow the same curriculum requirements stated in the Requirements for the Master of Science in Statistics section.

University Coterminal Requirements

Coterminal master's degree candidates are expected to complete all master's degree requirements as described in this bulletin. University requirements for the coterminal master's degree are described in the "Coterminal Master's Program (http://exploredegrees.stanford.edu/cotermdegrees/)" section. University requirements for the master's degree are described in the "Graduate Degrees (http://exploredegrees.stanford.edu/graduatedegrees/#masterstext)" section of this bulletin.

After accepting admission to this coterminal master's degree program, students may request transfer of courses from the undergraduate to the graduate career to satisfy requirements for the master's degree. Transfer of courses to the graduate career requires review and approval of both the undergraduate and graduate programs on a case by case basis.

In this master's program, courses taken during or after the first quarter of the sophomore year are eligible for consideration for transfer to the graduate career; the timing of the first graduate quarter is not a factor. No courses taken prior to the first quarter of the sophomore year may be used to meet master's degree requirements.

Course transfers are not possible after the bachelor's degree has been conferred.

The University requires that the graduate advisor be assigned in the student's first graduate quarter even though the undergraduate career may still be open. The University also requires that the Master's Degree Program Proposal be completed by the student and approved by the department by the end of the student's first graduate quarter.

Master of Science in Statistics

Curriculum and Degree Requirements

The department requires that a master's student take 45 units of work from offerings in the Department of Statistics (http://explorecourses.stanford.edu/search/?view=catalog&filter-coursestatus-Active=on&page=0&catalog=&academicYear=&q=STATS&collapse=) or from authorized courses in other departments. With the advice of the master's program advisors, each student selects his or her own set of electives.

All requirements for a master's degree, including the coterminal master's degree, must be completed within three years after the student's first term of enrollment in the master's program. Ordinarily, four or five quarters are needed to complete all requirements. Honors Cooperative students must finish within five years.

Units for a given course may not be counted to meet the requirements of more than one degree, with the exception that up to 45 units of a Stanford M.A. or M.S. degree may be applied to the residency requirement for the Ph.D., D.M.A. or Engineer degrees. See the "Residency Policy for Graduate Students (http://exploredegrees.stanford.edu/graduatedegrees/#residencytext)" section of this Bulletin for University rules.

As defined in the general graduate student requirements, students must maintain a grade point average (GPA) of 3.0 (or better) for courses used to fulfill degree requirements and classes must be taken at the 200 level or higher.

Master's Degree Program Proposal

The Statistics Master's Degree Program Proposal form (https://statistics.stanford.edu/academic-programs/graduate-programs/ms-program-forms/) must be signed and approved by the department's student services administrator before submission to the student's program advisor. This form is due no later than the end of the first quarter of enrollment in the program.

A revised program proposal must be submitted if degree plans change.

There is no thesis requirement.

For further information about the Statistics master's degree program requirements, see the program's webpage (https://statistics.stanford.edu/academics/ms-statistics/).

1. Statistics Core Courses (must complete all four courses):

		Units
Probability		
STATS 116	Theory of Probability [1]	4
Applied Statistics		
STATS 203	Introduction to Regression Models and Analysis of Variance	3
or STATS 305A	Applied Statistics I	
or STATS 191	Introduction to Applied Statistics	
Theoretical Statistics		
STATS 200	Introduction to Statistical Inference	3-4
or STATS 300A	Theory of Statistics I	
or STATS 370	A Course in Bayesian Statistics	
Stochastic Processes		
STATS 217	Introduction to Stochastic Processes I [1,2]	3

or STATS 218	Introduction to Stochastic Processes II	
or STATS 219	Stochastic Processes	
or STATS 318	Modern Markov Chains	

Students with prior background may replace each course with a more advanced course from the same area, or a more advanced course offered by the department, with consent of the adviser. All must be taken for a letter grade.

2. Statistics Depth:

Five additional Statistics courses must be taken from graduate offerings in the department (at or above the 200-level). During the 2020-21 academic year, three of five courses must be taken for a letter grade (with the exception of courses that may only be offered satisfactory(S)/credit (CR) only).

The following courses that may only be used to fulfill elective credit [3]: STATS 260A Workshop in Biostatistics series, STATS 299 Independent Study, STATS 298 Industrial Research for Statisticians, and STATS 390 Consulting Workshop (see list of electives below).

		Units
Courses which may be offered by the department:		
STATS 202	Data Mining and Analysis	3
STATS 203	Introduction to Regression Models and Analysis of Variance (STATS 203V)	3
STATS 204	Sampling	3
STATS 205	Introduction to Nonparametric Statistics	3
STATS 206	Applied Multivariate Analysis	3
STATS 207	Introduction to Time Series Analysis	3
STATS 208	Bootstrap, Cross-Validation, and Sample Re-use	3
STATS 209A	Topics in Causal Inference	3
STATS 211	Meta-research: Appraising Research Findings, Bias, and Meta-analysis	3
STATS 215	Statistical Models in Biology	3
STATS 216	Introduction to Statistical Learning	3
STATS 222	Statistical Methods for Longitudinal Research	2-3
STATS 229	Machine Learning	3-4
or CS 229	Machine Learning	
STATS 237	Investment Portfolios, Derivative Securities, and Risk Measures	3
STATS 240	Statistical Methods in Finance	3
STATS 241	Data-driven Financial Econometrics	3
STATS 244	Quantitative Trading: Algorithms, Data, and Optimization	2-4
STATS 245	Data, Models and Applications to Healthcare Analytics	3
STATS 250	Mathematical Finance	3
STATS 263	Design of Experiments	3
STATS 266	Advanced Statistical Methods for Observational Studies	2-3
STATS 270	A Course in Bayesian Statistics	3
or STATS 370	A Course in Bayesian Statistics	
STATS 271	Applied Bayesian Statistics	3
or STATS 371	Applied Bayesian Statistics	
STATS 285	Massive Computational Experiments, Painlessly	2
STATS 290	Computing for Data Science	3
STATS 300A	Theory of Statistics I	3
or STATS 300B	Theory of Statistics II	
or STATS 300C	Theory of Statistics III	
STATS 305A	Applied Statistics I	3
or STATS 305B	Applied Statistics II: Generalized Linear Models, Survival Analysis, and Exponential Families	
or STATS 305C	Applied Statistics III	
STATS 310A	Theory of Probability I	3
or STATS 310B	Theory of Probability II	
or STATS 310C	Theory of Probability III	
STATS 311	Information Theory and Statistics	3
or EE 377	Information Theory and Statistics	
STATS 314A	Advanced Statistical Theory	3
STATS 315A	Modern Applied Statistics: Learning	3
STATS 315B	Modern Applied Statistics: Data Mining	3
STATS 317	Stochastic Processes	3
STATS 318	Modern Markov Chains	3
STATS 319	Literature of Statistics	1
STATS 322	Function Estimation in White Noise	3
STATS 325	Multivariate Analysis and Random Matrices in Statistics	3
STATS 334	Mathematics and Statistics of Gambling	3
STATS 359	Topics in Mathematical Physics	3
or MATH 273	Topics in Mathematical Physics	
STATS 361	Causal Inference ((NEW))	3
STATS 364	Theory and Applications of Selective Inference ((NEW))	3
STATS 363	Design of Experiments	3
STATS 366	Modern Statistics for Modern Biology	3
STATS 374	Large Deviations Theory	3
or MATH 234	Large Deviations Theory	
STATS 368	Empirical Process Theory and its Applications	3
STATS 369	Methods from Statistical Physics	3
STATS 376A	Information Theory	3
STATS 376B	Topics in Information Theory and Its Applications	3
or EE 376B	Topics in Information Theory and Its Applications	
STATS 385	Analyses of Deep Learning	1

3. Linear Algebra Requirement:

		Units
Must be taken for a letter grade, with the exception of courses offered satisfactory/no credit only.		
Select one of the following:		
MATH 104	Applied Matrix Theory	3
MATH 113	Linear Algebra and Matrix Theory	3
MATH 115	Functions of a Real Variable	3
MATH 171	Fundamental Concepts of Analysis	3
CME 302	Numerical Linear Algebra	3
CME 364A	Convex Optimization I	3
or CME 364B	Convex Optimization II	

Substitution of more advanced courses in Mathematics, that provide similar skills, may be made with consent of the adviser.

4. Programming Requirement:

		Units
2020-21: May be taken for a letter grade or CR.		
Select one of the following:		
CS 106A	Programming Methodology	3

CS 106B	Programming Abstractions	3
CS 106X	Programming Abstractions	3
CS 107	Computer Organization and Systems	3-5
CME 108	Introduction to Scientific Computing	3
Substitution more advanced courses in Computer Science (140 - 181), that provide similar skills, may be made with consent of the adviser.		

5. Breadth/Elective Courses:

Courses that provide breadth to the degree may be chosen as elective units to complete the degree requirements. List of suggested of courses available from the program's webpage (https://statistics.stanford.edu/academics/statistics-ms-electives/). Other graduate courses (200 or above) may be authorized by the advisor if they provide skills relevant to degree requirements or deal primarily with an application of statistics or probability and do not significantly overlap (repeat) courses in the student's program.

There is sufficient flexibility to accommodate students with interests in applications to business, computing, economics, engineering, health, operations research, and biological and social sciences.

Courses that fulfill elective units may be taken concerning 'CR' (credit) or 'S' (satisfactory).

Students may enroll in up to 6 units of the following workfshops and training seminars to fulfill elective coursework: [3]		
STATS 242	NeuroTech Training Seminar	1
STATS 260A	Workshop in Biostatistics	1-2
STATS 260B	Workshop in Biostatistics	1-2
STATS 260C	Workshop in Biostatistics	1-2
STATS 298	Industrial Research for Statisticians	1
STATS 299	Independent Study	1-5
STATS 390	Consulting Workshop	1

Courses below 200 level are not acceptable with the following exceptions; however, students are strongly advised to avoid redundancy in coursework:

		Units
STATS 191	Introduction to Applied Statistics	3
MATH 115	Functions of a Real Variable	3
MATH 171	Fundamental Concepts of Analysis	3
CS 106A	Programming Methodology	3-5
CS 106B	Programming Abstractions	3-5
CS 106X	Programming Abstractions	3-5
CS 140	Operating Systems and Systems Programming	3-4
CS 142	Web Applications	3
CS 143	Compilers	3-4
CS 144	Introduction to Computer Networking	3-4
CS 145	Data Management and Data Systems	3-4
CS 147	Introduction to Human-Computer Interaction Design	3-5
CS 148	Introduction to Computer Graphics and Imaging	3-4
CS 149	Parallel Computing	3-4
CS 154	Introduction to the Theory of Computation	3-4
CS 155	Computer and Network Security	3
CS 157	Computational Logic	3
CS 161	Design and Analysis of Algorithms	3-5
CS 170	Stanford Laptop Orchestra: Composition, Coding, and Performance	1-5
CS 181	Computers, Ethics, and Public Policy [4]	4
And at most, one of these courses ay be counted as an elective. [4]		
MATH 104	Applied Matrix Theory	3
MATH 113	Linear Algebra and Matrix Theory	3
STATS 116	Theory of Probability	4

1. Students who replace STATS 116 with STATS 217 must take a second course in Stochastic Processes or Probability.
2. Enrollment in STATS 116 after successful completion of STATS 217, 218, and/or 219, may not be used to fulfill degree requirements, including as an elective.
3. Students admitted to the Statistics M.S. program prior to academic year 2018-19 fulfill the requirements in effect at the time of their admission.
4. Enrollment in a course that provides redundant coursework cannot be used to fulfill the M.S. degree requirements.

Master of Science in Statistics, Data Science Track

The Data Science track[5] develops strong mathematical, statistical, and computational and programming skills through the general master's core and programming requirements. In addition, it provides a fundamental data science education through general and focused electives requirement from courses in data sciences and related areas. Course choices are limited to predefined courses from the data sciences and related courses group. The final requirement is a practical component to be completed through capstone project, data science clinic, or other courses that have strong hands-on or practical component, such as statistical consulting.

Admission

Prospective applicants should consult the Graduate Admissions (https://studentaffairs.stanford.edu/gradadmissions/) and the Statistics Department admissions webpages (https://statistics.stanford.edu/admissions/graduate-application-information-and-instructions/) for complete information on admission requirements and deadlines.

Applicants apply to the Master of Science degree program in Statistics and subsequently declare their preference for the Data Science track (subplan) within the graduate application ("Department Specialization" option).

Prerequisites

Recommended preparatory courses include advanced undergraduate level courses in linear algebra and probability, and introductory courses in stochastic processes, numerical methods and proficiency in programming (Basic usage of the Python and C/C++ programming languages).

Curriculum and Degree Requirements

As defined in the general graduate student requirements, students must maintain a grade point average (GPA) of 3.0 or better and classes must be taken at the 200 level or higher. Students must complete 45 units of required coursework in Data Science.

Master's Degree Program Proposal

The Statistics (Data Science) Master's Degree Program Proposal form (https://statistics.stanford.edu/academic-programs/graduate-programs/ms-program-forms/) must be signed and approved by the department's student services administrator before submission to the student's program advisor. This form is due no later than the end of the first quarter of enrollment in the program.

A revised program proposal must be submitted if degree plans change.

There is no thesis requirement.

The Data Science track (subplan) is printed on the student transcript and diploma.

Mathematical and Statistical Foundations (15 units)

Students must demonstrate foundational knowledge in the field by completing the following courses. Courses in this area must be taken for letter grades.

		Units
STATS 200	Introduction to Statistical Inference	3
or STATS 300A	Theory of Statistics I	
STATS 203	Introduction to Regression Models and Analysis of Variance	3
or STATS 305A	Applied Statistics I	
STATS 315A	Modern Applied Statistics: Learning	3
or STATS/CS 229	Machine Learning	
CME 302	Numerical Linear Algebra	3
CME 308	Stochastic Methods in Engineering	3

Experimentation (3 units)

Experimental method and causal considerations are fundamental to data science. The course chosen from this area must be taken for letter grades.

		Units
STATS 263	Design of Experiments	3
ECON 271	Intermediate Econometrics II	3-5
MS&E 327	Topics in Causal Inference	3

Software Development & Scientific Computing (6 - 9 units)

To ensure that students have a strong foundation in programming, 3 units of scientific software development (CME212) is required.

Software Development: (3 units)

Minimum of 3 units in scientific computing. (Additional 3 units for those who need to take CME211.)

ICME offers a placement test Summer Quarter. Students who pass this placement test are not required to take CME 211. Courses in this area must be taken for letter grades.

		Units
CME 212	Advanced Software Development for Scientists and Engineers (prerequisite: CME 211)	3
Programming proficiency at the level of CME 211 is a hard prerequisite for CME 212. can be waived with placement exam (summer).		

Scientific Computing Foundations and Methods (minimum 3 units)

		Units
CME 213	Introduction to parallel computing using MPI, openMP, and CUDA	3
CME 305	Discrete Mathematics and Algorithms	3
CME 307	Optimization	3
CME 323	Distributed Algorithms and Optimization	3
CME 364A	Convex Optimization I	3
CS 246	Mining Massive Data Sets	3-4

Students may take 6 units as CR/S in Scientific Computing or Machine Learning for the 2020-21 academic year.

Machine Learning Methods and Applications (6 - 9 units)

Ordinarily, courses in machine learning should be taken for letter grades. Students may take two courses as 'CR' (credit) or 'S' (satisfactory) for academic year 2020-21.

		Units
STATS 315B	Modern Applied Statistics: Data Mining	3
CS 221	Artificial Intelligence: Principles and Techniques	3-4
CS 224N	Natural Language Processing with Deep Learning	3-4
CS 230	Deep Learning	3-4
CS 231N	Convolutional Neural Networks for Visual Recognition	3-4
CS 234	Reinforcement Learning	3
CS 236	Deep Generative Models	3
Students may take 6 units as CR/S in Scientific Computing or Machine Learning for the 2020-21 academic year.		

Practical Component (3 units)

A Capstone project, supervised by a faculty member and approved by the student's advisor. The capstone project should ideally build on the work done in the student's coursework. Students should submit a one-page proposal, supported by the faculty member and sent to the student's Data Science advisor for approval (at least one quarter prior to start of project).

Students are required to take 3 units of practical component that may include any combination of:

		Units
CME 217	Analytics Accelerator (Real-world project-based research; Application required; Autumn quarter commitment, winter quarter optional.)	3
ENGR 150	Data Challenge Lab (https://datalab.stanford.edu/challenge-lab)	3-5
ENGR 350	Data Impact Lab (https://datalab.stanford.edu/impact-lab)	1-6
STATS 299	Independent Study	1-5
or CME 291	Master's Research	
STATS 390	Consulting Workshop (repeatable)	1

Electives (6 - 9 units)

Courses in data science, machine learning, statistics, advanced programming or practical components, chosen in consultation with the student's course advisor.

Doctor of Philosophy in Statistics

The department looks for students who wish to prepare for research careers in statistics or probability, either applied or theoretical. Advanced undergraduate or master's level work in mathematics and statistics provides a good background for the doctoral program. Quantitatively oriented students with degrees in other scientific fields are also encouraged to apply for admission. The program normally takes five years to complete.

Program Summary

		Units
First-year core program		
STATS 300A	Theory of Statistics I	3
STATS 300B	Theory of Statistics II	3
STATS 300C	Theory of Statistics III	3
STATS 305A	Applied Statistics I	3
STATS 305B	Applied Statistics II: Generalized Linear Models, Survival Analysis, and Exponential Families	3
STATS 305C	Applied Statistics III	3
STATS 310A	Theory of Probability I	3
STATS 310B	Theory of Probability II	3
STATS 310C	Theory of Probability III	3
STATS 302	Qualifying Exams Workshop	5-10

- Pass two of three parts of the qualifying examinations (end of first year); breadth requirement (second, third and fourth year); successfully complete the dissertation proposal meeting (early spring quarter of third year); pass the University oral examination (fourth or fifth year); dissertation (fifth year).
- In addition, students are required to complete a 'depth' requirement consisting of a minimum of three courses (nine units) of advanced topics courses offered by the department. Courses for the depth and breadth (see below) requirements must equal a combined minimum of 24 units. Recommended advanced topics courses include the following:

		Units
STATS 311	Information Theory and Statistics	3
STATS 314A	Advanced Statistical Theory	3
STATS 315A	Modern Applied Statistics: Learning	3
STATS 315B	Modern Applied Statistics: Data Mining	3
STATS 317	Stochastic Processes	3
STATS 318	Modern Markov Chains	3
STATS 322	Function Estimation in White Noise	3
STATS 325	Multivariate Analysis and Random Matrices in Statistics	3
STATS 350	Topics in Probability Theory	3
STATS 359	Topics in Mathematical Physics	3
STATS 362	Topic: Monte Carlo	3
STATS 367	Statistical Models in Genetics	3
STATS 370	A Course in Bayesian Statistics	3
EE 364A	Convex Optimization I	3
EE 364B	Convex Optimization II	3

- Take STATS 390 Consulting Workshop at least twice in years two and three.
- Take STATS 319 Literature of Statistics once per year after passing the Qualifying Exam until the year after passing the dissertation proposal meeting.

First-Year Core Courses

- STATS 300A Theory of Statistics I, STATS 300B Theory of Statistics II and STATS 300C Theory of Statistics III systematically survey the ideas of estimation and of hypothesis testing for parametric and nonparametric models involving small and large samples.
- STATS 305A Applied Statistics I is concerned with linear regression and the analysis of variance.
- STATS 305B Applied Statistics II: Generalized Linear Models, Survival Analysis, and Exponential Families and STATS 305C Applied Statistics III survey a large number of modeling techniques, related to but going beyond the linear models of STATS 305A Applied Statistics I.
- STATS 310A Theory of Probability I, STATS 310B Theory of Probability II, and STATS 310C Theory of Probability III are measure-theoretic courses in probability theory, beginning with basic concepts of the law of large numbers and martingale theory.

Students who do not have enough mathematics background can take STATS 310 A,B,C after their first year but need to have their first-year program approved by the Director of Graduate Studies.

Qualifying Examinations

These are intended to test the student's level of knowledge when the first-year program, common to all students, has been completed. There are separate examinations in the three core subjects of statistical theory and methods, applied statistics, and probability theory, and all are typically taken during the summer between the student's first and second years. Students are expected to show acceptable performance in two examinations. Letter grades are not given. After passing the qualifying exams students file for Ph.D. candidacy, a University milestone.

Breadth Requirement

Students are required to take a minimum of three courses (nine units) outside of the department and are advised to choose an area of concentration in a specific scientific field of statistical applications approved by their Ph.D. program adviser. Courses for the depth and breadth requirements must equal a combined minimum of 24 units.

Popular areas include: Computational Biology and Statistical Genomics, Machine Learning, Applied Probability, Earth Science Statistics, and Social and Behavioral Sciences.

Dissertation Reading Committee, Dissertation Proposal Meeting and University Oral Examinations

The dissertation reading committee consists of the student's adviser plus two faculty readers, all of whom are responsible for reading and approving the full dissertation.

The dissertation proposal meeting is intended to demonstrate students' depth in some areas of statistics, and to examine the general plan for their research. It also confirms that students have chosen a Ph.D. faculty adviser and have started to work with that adviser on a research topic. In the meeting, the student will give a 60-minute presentation and discuss their ideas for completing a Ph.D. thesis, with a committee typically consisting of the members of the dissertation reading committee. The meeting must be successfully completed by early spring quarter of the third year. "Successful completion" means that the general research plan is sound and has a reasonable chance of success. If the student does not pass, the meeting must be repeated. Repeated failure by the end of Year 3 can lead to a loss of financial support.

The oral examination/dissertation defense is scheduled when the student has finished their dissertation and is in the process of completing their final draft. The oral exam consists of a 60-minute presentation on the dissertation topic, followed by a question and answer period attended only by members of the examining committee. The questions relate both to the student's presentation and also explore the student's familiarity with broader statistical topics related to the thesis research. The oral examination is normally completed within the last few months of the student's Ph.D. period. The examining committee usually consists of at least five members: four examiners including the three members of the Dissertation Reading Committee, plus an outside chair who serves as an impartial representative of the academic standards of the University. Four out of five passing votes are required and no grades are given. Nearly all students can expect to pass this examination, although it is common for specific recommendations to be made regarding completion of the written dissertation.

For further information on University oral examinations and committees, see the Graduate Academic Policies and Procedures (GAP) Handbook, section 4.7 (http://gap.stanford.edu/4-7.html) or the "University Oral Examination (http://exploredegrees.stanford.edu/graduatedegrees/#doctoraltext)" section of this bulletin.

Doctoral and Research Advisers

From the student's arrival until the selection of a research adviser, the student's academic progress is monitored by the department's Director of Graduate Studies. Each student should meet at least once a quarter with the Doctoral Adviser to discuss their academic plans and their progress towards choosing a dissertation adviser. See Graduate Advising Expectations section for more information.

Financial Support

Students accepted to the Ph.D. program are offered financial support. All tuition expenses are paid and there is a fixed monthly stipend determined to be sufficient to pay living expenses. Financial support can be continued for five years, department resources permitting, for students in good standing. The resources for student financial support derive from funds made available for student teaching and research assistantships. Students receive both a teaching and research assignment each quarter which, together, do not exceed 20 hours. Students are encouraged to apply for outside scholarships, fellowships, and other forms of financial support.

Ph.D. Minor in Statistics

Students must complete a total of 30 units for the Ph.D. minor. 20 units must be from Statistics courses numbered 300 and above and taken for a letter grade (minimum grade of B for each course). The remaining 10 units can be from Statistics courses numbered 200 and above, and may be taken for a letter grade or credit. Students may not include more than one unit of Stats 390, Consulting Workshop, towards the 30 units. The selection of courses must be approved by the Statistics Department and the *Application for the Ph.D. Minor* form must be approved by both the student's Ph.D. department and the Statistics department.

For further information about the Statistics Ph.D. degree program requirements, see the department web site (https://statistics.stanford.edu/academics/doctoral-program/).

COVID-19 Policies

On July 30, the Academic Senate adopted grading policies effective for all undergraduate and graduate programs, excepting the professional Graduate School of Business, School of Law, and the School of Medicine M.D. Program. For a complete list of those and other academic policies relating to the pandemic, see the "COVID-19 and Academic Continuity (http://exploredegrees.stanford.edu/covid-19-policy-changes/#tempdepttemplatetabtext)" section of this bulletin.

The Senate decided that all undergraduate and graduate courses offered for a letter grade must also offer students the option of taking the course for a "credit" or "no credit" grade and recommended that deans, departments, and programs consider adopting local policies to count courses taken for a "credit" or "satisfactory" grade toward the fulfillment of degree-program requirements and/or alter program requirements as appropriate.

Undergraduate Degree Requirements

Grading

The Statistics department counts all courses taken in academic year 2020-21 with a grade of 'CR' (credit) or 'S' (satisfactory) towards satisfaction of undergraduate minor requirements that otherwise require a letter grade.

Graduate Degree Requirements

Grading

The Statistics department's M.S. program has modified its policy concerning 'CR' (credit) or 'S' (satisfactory) grades in degree requirements requiring a letter grade for academic year 2020-21 as follows: Letter grade is required of the core statistics courses (4), three of the five (5) statistics depth courses and linear algebra requirement. The programming requirement may be taken 'CR' (credit) or 'S' (satisfactory).

The Statistics department's M.S. program in Data Science has modified its policy concerning 'CR' (credit) or 'S' (satisfactory) grades in degree requirements requiring a letter grade for academic year 2020-21 as follows: Students may take two courses as 'CR' (credit) or 'S' (satisfactory) in Machine Learning and/or Scientific Computing Foundations (up to 6 units).

The Statistics department's Ph.D. program counts all courses taken in academic year 2020-21 with a grade of 'CR' (credit) or 'S' (satisfactory) towards satisfaction of Ph.D. degree requirements that otherwise require a letter grade, though first year Statistics Ph.D. students are strongly encouraged to take the first year required courses for a letter grade.

Graduate Advising Expectations

The Department of Statistics is committed to providing academic advising in support of graduate student scholarly and professional development. When most effective, this advising relationship entails collaborative and sustained engagement by both the adviser and the advisee. As a best practice, advising expectations should be periodically discussed and reviewed to ensure mutual understanding. Both the adviser and the advisee are expected to maintain professionalism and integrity.

Faculty advisers guide students in key areas such as selecting courses, designing and conducting research, developing of teaching pedagogy, navigating policies and degree requirements, and exploring academic opportunities and professional pathways.

Graduate students are active contributors to the advising relationship, proactively seeking academic and professional guidance and taking responsibility for informing themselves of policies and degree requirements for their graduate program.

For a statement of University policy on graduate advising, see the "Graduate Advising (http://exploredegrees.stanford.edu/graduatedegrees/#advisingandcredentialstext)" section of this bulletin.

M.S. in Statistics and Data Science

Master's students are assigned an academic adviser for the duration of their tenure in the program. The adviser serves as a key resource for the purposes of course placement and approval of elective coursework as it relates to fulfilling degree requirements. Since the majority of MS students choose employment in the field of industry (tech/programming), the program adviser may provide assistance with regards to internships and general professional opportunities. Those planning to apply to doctoral programs are also able to receive feedback on research opportunities.

Ph.D. in Statistics

First and second year students are advised on course selection and other academic matters by the Director of Graduate Studies who is available by appointment to consult with students about any graduate student related matter, including degree progress. The DGS also leads cohort-specific workshops addressing topics such as qualifying exams, adviser selection, oral exams and post-graduation placement.

By the final study list deadline of Spring Quarter of the second year students are expected to have selected a research adviser who later serves as their principal dissertation adviser. The dissertation adviser must be a member of the Academic Council, and may be from outside the department. Students may also opt to have two co-advisers rather than one principal adviser, which may include one from outside the department.

The adviser-student mentorship takes many different forms, including, but not limited to programmatic consultation and degree progress, and support and collaboration relating to research, conferences, publications, and academic and professional opportunities.

It is the responsibility of the student to meet with their adviser at least once per quarter during the academic year to discuss academic standing and graduate degree progress. In addition, the Director of Graduate Studies is always available to Ph.D. students for consultation.

Program requirements and milestones, as well as more detailed descriptions of the program's expectations of advisers and students, are listed in the Stats Ph.D. Handbook, available on the department website. (https://statistics.sites.stanford.edu/stats-phd-handbook-2018-19/)

Faculty

Emeriti: (Professors) Jerome H. Friedman, Paul Switzer

Chair: Art Owen

Director of Graduate Studies: Joseph P. Romano

Director of Undergraduate Studies: Guenther Walther

Professors: Emmanuel Candès, Sourav Chatterjee, Amir Dembo, Persi Diaconis, David L. Donoho, Bradley Efron, Trevor J. Hastie, Susan P. Holmes, Iain M. Johnstone, Tze L. Lai, Andrea Montanari, Art Owen, Joseph P. Romano, Chiara Sabatti, David O. Siegmund, Jonathan Taylor, Robert J. Tibshirani, Guenther Walther, Wing H. Wong

Assistant Professors: Guillaume Basse, John Duchi, Scott Linderman, Tengyu Ma, Julia Palacios, Dominik Rothenhäusler, Tselil Schramm

Courtesy Professors: John Ioannidis, Hua Tang

Courtesy Associate Professors: David Rogosa, Lu Tian

Courtesy Assistant Professors: Mike Baiocchi, Percy Shuo Liang, Stefan Wager

Stein Fellows: Paromita Dubey, Vishesh Jain

Courses

STATS 32. Introduction to R for Undergraduates. 1 Unit.
This short course runs for weeks one through five of the quarter. It is recommended for undergraduate students who want to use R in the humanities or social sciences and for students who want to learn the basics of R programming. The goal of the short course is to familiarize students with R's tools for data analysis. Lectures will be interactive with a focus on learning by example, and assignments will be application-driven. No prior programming experience is needed. Topics covered include basic data structures, File I/O, data transformation and visualization, simple statistical tests, etc, and some useful packages in R. Prerequisite: undergraduate student. Priority given to non-engineering students. Laptops necessary for use in class.

STATS 48N. Riding the Data Wave. 3 Units.
Imagine collecting a bit of your saliva and sending it in to one of the personalized genomics company: for very little money you will get back information about hundreds of thousands of variable sites in your genome. Records of exposure to a variety of chemicals in the areas you have lived are only a few clicks away on the web; as are thousands of studies and informal reports on the effects of different diets, to which you can compare your own. What does this all mean for you? Never before in history humans have recorded so much information about themselves and the world that surrounds them. Nor has this data been so readily available to the lay person. Expression as "data deluge" are used to describe such wealth as well as the loss of proper bearings that it often generates. How to summarize all this information in a useful way? How to boil down millions of numbers to just a meaningful few? How to convey the gist of the story in a picture without misleading oversimplifications? To answer these questions we need to consider the use of the data, appreciate the diversity that they represent, and understand how people instinctively interpret numbers and pictures. During each week, we will consider a different data set to be summarized with a different goal. We will review analysis of similar problems carried out in the past and explore if and how the same tools can be useful today. We will pay attention to contemporary media (newspapers, blogs, etc.) to identify settings similar to the ones we are examining and critique the displays and summaries there documented. Taking an experimental approach, we will evaluate the effectiveness of different data summaries in conveying the desired information by testing them on subsets of the enrolled students.
Same as: BIODS 48N

STATS 60. Introduction to Statistical Methods: Precalculus. 5 Units.
Techniques for organizing data, computing, and interpreting measures of central tendency, variability, and association. Estimation, confidence intervals, tests of hypotheses, t-tests, correlation, and regression. Possible topics: analysis of variance and chi-square tests, computer statistical packages.
Same as: PSYCH 10, STATS 160

STATS 100. Mathematics of Sports. 3 Units.
This course will teach you how statistics and probability can be applied in sports, in order to evaluate team and individual performance, build optimal in-game strategies and ensure fairness between participants. Topics will include examples drawn from multiple sports such as basketball, baseball, soccer, football and tennis. The course is intended to focus on data-based applications, and will involve computations in R with real data sets via tutorial sessions and homework assignments. Prereqs: No statistical or programming background is assumed, but introductory courses, e.g, Stats 60,101 or 116, are recommended. A prior knowledge of Linear Algebra (e.g., Math 51) and basic probability is strongly recommended.

STATS 101. Data Science 101. 5 Units.
https://statweb.stanford.edu/~tibs/stat101.html This course will provide a hands-on introduction to statistics and data science. Students will engage with the fundamental ideas in inferential and computational thinking. Each week, we will explore a core topic comprising three lectures and two labs (a module), in which students will manipulate real-world data and learn about statistical and computational tools. Students will engage in statistical computing and visualization with current data analytic software (Jupyter, R). The objectives of this course are to have students (1) be able to connect data to underlying phenomena and to think critically about conclusions drawn from data analysis, and (2) be knowledgeable about programming abstractions so that they can later design their own computational inferential procedures. No programming or statistical background is assumed. Freshmen and sophomores interested in data science, computing and statistics are encouraged to attend. Open to graduates as well.

STATS 110. Statistical Methods in Engineering and the Physical Sciences. 5 Units.
Introduction to statistics for engineers and physical scientists. Topics: descriptive statistics, probability, interval estimation, tests of hypotheses, nonparametric methods, linear regression, analysis of variance, elementary experimental design. Prerequisite: one year of calculus.

STATS 116. Theory of Probability. 4 Units.
Probability spaces as models for phenomena with statistical regularity. Discrete spaces (binomial, hypergeometric, Poisson). Continuous spaces (normal, exponential) and densities. Random variables, expectation, independence, conditional probability. Introduction to the laws of large numbers and central limit theorem. Prerequisites: MATH 52 and familiarity with infinite series, or equivalent.

STATS 141. Biostatistics. 5 Units.
Introductory statistical methods for biological data: describing data (numerical and graphical summaries); introduction to probability; and statistical inference (hypothesis tests and confidence intervals). Intermediate statistical methods: comparing groups (analysis of variance); analyzing associations (linear and logistic regression); and methods for categorical data (contingency tables and odds ratio). Course content integrated with statistical computing in R.
Same as: BIO 141

STATS 155. Modern Statistics for Modern Biology. 3 Units.
Application based course in nonparametric statistics. Modern toolbox of visualization and statistical methods for the analysis of data, examples drawn from immunology, microbiology, cancer research and ecology. Methods covered include multivariate methods (PCA and extensions), sparse representations (trees, networks, contingency tables) as well as nonparametric testing (Bootstrap, permutation and Monte Carlo methods). Hands on, use R and cover many Bioconductor packages. Prerequisite: Working knowledge of R and two core Biology courses. Note that the 155 offering is a writing intensive course for undergraduates only and requires instructor consent. (WIM). See https://web.stanford.edu/class/bios221/index.html.
Same as: BIOS 221, STATS 256, STATS 366

STATS 160. Introduction to Statistical Methods: Precalculus. 5 Units.
Techniques for organizing data, computing, and interpreting measures of central tendency, variability, and association. Estimation, confidence intervals, tests of hypotheses, t-tests, correlation, and regression. Possible topics: analysis of variance and chi-square tests, computer statistical packages.
Same as: PSYCH 10, STATS 60

STATS 167. Probability: Ten Great Ideas About Chance. 4 Units.
Foundational approaches to thinking about chance in matters such as gambling, the law, and everyday affairs. Topics include: chance and decisions; the mathematics of chance; frequencies, symmetry, and chance; Bayes great idea; chance and psychology; misuses of chance; and harnessing chance. Emphasis is on the philosophical underpinnings and problems. Prerequisite: exposure to probability or a first course in statistics at the level of STATS 60 or 116.
Same as: PHIL 166, PHIL 266, STATS 267

STATS 191. Introduction to Applied Statistics. 3 Units.
Statistical tools for modern data analysis. Topics include regression and prediction, elements of the analysis of variance, bootstrap, and cross-validation. Emphasis is on conceptual rather than theoretical understanding. Applications to social/biological sciences. Student assignments/projects require use of the software package R. Prerequisite: introductory statistical methods course. Recommended: 60, 110, or 141.

STATS 195. Introduction to R. 1 Unit.
This short course runs for four weeks and is offered in fall and spring. It is recommended for students who want to use R in statistics, science or engineering courses, and for students who want to learn the basics of data science with R. The goal of the short course is to familiarize students with some of the most important R tools for data analysis. Lectures will focus on learning by example and assignments will be application-driven. No prior programming experience is assumed.
Same as: CME 195

STATS 196A. Multilevel Modeling Using R. 1 Unit.
See http://rogosateaching.com/stat196/ . Multilevel data analysis examples using R. Topics include: two-level nested data, growth curve modeling, generalized linear models for counts and categorical data, nonlinear models, three-level analyses.
Same as: EDUC 401D

STATS 199. Independent Study. 1-15 Unit.
For undergraduates.

STATS 200. Introduction to Statistical Inference. 4 Units.
Modern statistical concepts and procedures derived from a mathematical framework. Statistical inference, decision theory; point and interval estimation, tests of hypotheses; Neyman-Pearson theory. Bayesian analysis; maximum likelihood, large sample theory. Prerequisite: STATS 116. Please note that students must enroll in one section of their choosing in addition to the main lecture.

STATS 202. Data Mining and Analysis. 3 Units.
Data mining is used to discover patterns and relationships in data. Emphasis is on large complex data sets such as those in very large databases or through web mining. Topics: decision trees, association rules, clustering, case based methods, and data visualization. Prereqs: Introductory courses in statistics or probability (e.g., Stats 60), linear algebra (e.g., Math 51), and computer programming (e.g., CS 105).

STATS 203. Introduction to Regression Models and Analysis of Variance. 3 Units.
Modeling and interpretation of observational and experimental data using linear and nonlinear regression methods. Model building and selection methods. Multivariable analysis. Fixed and random effects models. Experimental design. Prerequisites: A post-calculus introductory probability course, e.g. STATS 116, basic computer programming knowledge, some familiarity with matrix algebra, and a pre- or co-requisite post-calculus mathematical statistics course, e.g. STATS 200.

STATS 203V. Introduction to Regression Models and Analysis of Variance. 3 Units.
Modeling and interpretation of observational and experimental data using linear and nonlinear regression methods. Model building and selection methods. Multivariable analysis. Fixed and random effects models. Experimental design. This course is offered remotely only via video segments (MOOC style). TAs will host remote weekly office hours using an online platform such as Zoom. Prerequisites: A post-calculus introductory probability course, e.g. STATS 116, basic computer programming knowledge, some familiarity with matrix algebra, and a pre- or co-requisite post-calculus mathematical statistics course, e.g. STATS 200.

STATS 204. Sampling. 3 Units.
How best to take data and where to sample it. Examples include surveys and sampling from data warehouses. Emphasis is on methods for finite populations. Topics: simple random sampling, stratified sampling, cluster sampling, ratio and regression estimators, two stage sampling.

STATS 205. Introduction to Nonparametric Statistics. 3 Units.
Nonparametric regression and nonparametric density estimation, modern nonparametric techniques, nonparametric confidence interval estimates, nearest neighbor algorithms (with non-linear features), wavelet, bootstrap. Nonparametric analogs of the one- and two-sample t-tests and analysis of variance.

STATS 206. Applied Multivariate Analysis. 3 Units.
Introduction to the statistical analysis of several quantitative measurements on each observational unit. Emphasis is on concepts, computer-intensive methods. Examples from economics, education, geology, psychology. Topics: multiple regression, multivariate analysis of variance, principal components, factor analysis, canonical correlations, multidimensional scaling, clustering. Pre- or corequisite: 200.

STATS 207. Introduction to Time Series Analysis. 3 Units.
Time series models used in economics and engineering. Trend fitting, autoregressive and moving average models and spectral analysis, Kalman filtering, and state-space models. Seasonality, transformations, and introduction to financial time series. Prerequisite: basic course in Statistics at the level of 200.

STATS 208. Bootstrap, Cross-Validation, and Sample Re-use. 3 Units.
By re-using the sample data, sometimes in ingenious ways, we can evaluate the accuracy of predictions, test the significance of a conclusion, place confidence bounds on an unknown parameter, select the best prediction architecture, and develop more accurate predictors. In this course, we will describe the many ways that samples get reused to achieve these goals, including the bootstrap, the parametric bootstrap, cross-validation, conformal prediction, random forests, and sample splitting. We also develop basic theory justifying such methods. Prerequisite: course in statistics or probability.

STATS 209A. Topics in Causal Inference. 3 Units.
This course introduces the fundamental ideas and methods in causal inference, and surveys a broad range of problems and applications. Emphasis will be on framing causal problems and identifying causal effects in both randomized experiments and observational studies. Topics will include: the potential outcomes framework; randomization-based inference and covariate adjustment; matching, and IPW; instrumental variables, regression discontinuity and synthetic ncontrols. Examples and applications will be taken from the fields of education, political science, economics, public health and digital marketing.
Same as: MS&E 327

STATS 209B. Applications of Causal Inference Methods. 2 Units.
See http://rogosateaching.com/stat209/. Application of potential outcomes formulation for causal inference to research settings including: mediation, compliance adjustments, time-1 time-2 designs, encouragement designs, heterogeneous treatment effects, aggregated data, instrumental variables, analysis of covariance regression adjustments, and implementations of matching methods. Prerequisite: STATS 209A/MSE 327 or other introduction to causal inference methods. (Formerly HRP 239).
Same as: EDUC 260A, EPI 239

STATS 211. Meta-research: Appraising Research Findings, Bias, and Meta-analysis. 3 Units.
Open to graduate, medical, and undergraduate students. Appraisal of the quality and credibility of research findings; evaluation of sources of bias. Meta-analysis as a quantitative (statistical) method for combining results of independent studies. Examples from medicine, epidemiology, genomics, ecology, social/behavioral sciences, education. Collaborative analyses. Project involving generation of a meta-research project or reworking and evaluation of an existing published meta-analysis. Prerequisite: knowledge of basic statistics.
Same as: CHPR 206, EPI 206, MED 206

STATS 214. Machine Learning Theory. 3 Units.
How do we use mathematical thinking to design better machine learning methods? This course focuses on developing mathematical tools for answering these questions. This course will cover fundamental concepts and principled algorithms in machine learning. We have a special focus on modern large-scale non-linear models such as matrix factorization models and deep neural networks. In particular, we will cover concepts and phenomenon such as uniform convergence, double descent phenomenon, implicit regularization, and problems such as matrix completion, bandits, and online learning (and generally sequential decision making under uncertainty). Prerequisites: linear algebra (MATH 51 or CS 205), probability theory (STATS 116, MATH 151 or CS 109), and machine learning (CS 229, STATS 229, or STATS 315A).
Same as: CS 229M

STATS 215. Statistical Models in Biology. 3 Units.
Poisson and renewal processes, Markov chains in discrete and continuous time, branching processes, diffusion. Applications to models of nucleotide evolution, recombination, the Wright-Fisher process, coalescence, genetic mapping, sequence analysis. Theoretical material approximately the same as in STATS 217, but emphasis is on examples drawn from applications in biology, especially genetics. Prerequisite: 116 or equivalent.

STATS 216. Introduction to Statistical Learning. 3 Units.
Overview of supervised learning, with a focus on regression and classification methods. Syllabus includes: linear and polynomial regression, logistic regression and linear discriminant analysis;cross-validation and the bootstrap, model selection and regularization methods (ridge and lasso); nonlinear models, splines and generalized additive models; tree-based methods, random forests and boosting; support-vector machines; Some unsupervised learning: principal components and clustering (k-means and hierarchical). Computing is done in R, through tutorial sessions and homework assignments. This math-light course is offered via video segments (MOOC style), and in-class problem solving sessions. Prereqs: Introductory courses in statistics or probability (e.g., Stats 60 or Stats 101), linear algebra (e.g., Math 51), and computer programming (e.g., CS 105).

STATS 216V. Introduction to Statistical Learning. 3 Units.
Overview of supervised learning, with a focus on regression and classification methods. Syllabus includes: linear and polynomial regression, logistic regression and linear discriminant analysis; cross-validation and the bootstrap, model selection and regularization methods (ridge and lasso); nonlinear models, splines and generalized additive models; tree-based methods, random forests and boosting; support-vector machines; Some unsupervised learning: principal components and clustering (k-means and hierarchical). Computing is done in R, through tutorial sessions and homework assignments. This math-light course is offered remotely only via video segments (MOOC style). TAs will host remote weekly office hours using an online platform such as Zoom. There are four homework assignments, a midterm, and a final exam, all of which are administered remotely. Prereqs: Introductory courses in statistics or probability (e.g., Stats 60 or Stats 101), linear algebra (e.g., Math 51), and computer programming (e.g., CS 105).

STATS 217. Introduction to Stochastic Processes I. 3 Units.
Discrete and continuous time Markov chains, poisson processes, random walks, branching processes, first passage times, recurrence and transience, stationary distributions. Non-Statistics masters students may want to consider taking STATS 215 instead. Prerequisite: a post-calculus introductory probability course e.g. STATS 116.

STATS 218. Introduction to Stochastic Processes II. 3 Units.
Renewal theory, Brownian motion, Gaussian processes, second order processes, martingales.

STATS 219. Stochastic Processes. 3 Units.
Introduction to measure theory, Lp spaces and Hilbert spaces. Random variables, expectation, conditional expectation, conditional distribution. Uniform integrability, almost sure and Lp convergence. Stochastic processes: definition, stationarity, sample path continuity. Examples: random walk, Markov chains, Gaussian processes, Poisson processes, Martingales. Construction and basic properties of Brownian motion. Prerequisite: STATS 116 or MATH 151 or equivalent. Recommended: MATH 115 or equivalent. http://statweb.stanford.edu/~adembo/math-136/.
Same as: MATH 136

STATS 220. Machine Learning Methods for Neural Data Analysis. 3 Units.
With modern high-density electrodes and optical imaging techniques, neuroscientists routinely measure the activity of hundreds, if not thousands, of cells simultaneously. Coupled with high-resolution behavioral measurements, genetic sequencing, and connectomics, these datasets offer unprecedented opportunities to learn how neural circuits function. This course will study statistical machine learning methods for analysing such datasets, including: spike sorting, calcium deconvolution, and voltage smoothing techniques for extracting relevant signals from raw data; markerless tracking methods for estimating animal pose in behavioral videos; network models for connectomics and fMRI data; state space models for analysis of high-dimensional neural and behavioral time-series; point process models of neural spike trains; and deep learning methods for neural encoding and decoding. We will develop the theory behind these models and algorithms and then apply them to real datasets in the homeworks and final project.This course is similar to STATS215: Statistical Models in Biology and STATS366: Modern Statistics for Modern Biology, but it is specifically focused on statistical machine learning methods for neuroscience data. Prerequisites: Students should be comfortable with basic probability (STATS 116) and statistics (at the level of STATS 200). This course will place a heavy emphasis on implementing models and algorithms, so coding proficiency is required.
Same as: CS 339N, NBIO 220, STATS 320

STATS 221. Random Processes on Graphs and Lattices. 3 Units.
Covering modern topics in the study of random processes on graphs and lattices. Specifically, a subset of: Random walks, electrical networks and flows. Uniform spanning trees. Percolation and self-avoiding walks. Contact process, voter model and the exclusion process. Ising, Potts, and Random-Cluster model. Random graphs. Prerequisites: MATH 115 (or equivalent), STAT 217 (or equivalent).

STATS 222. Statistical Methods for Longitudinal Research. 2 Units.
See http://rogosateaching.com/stat222/. Research designs and statistical procedures for time-ordered (repeated-measures) data. The analysis of longitudinal panel data is central to empirical research on learning, development, aging, and the effects of interventions. Topics include: measurement of change, growth curve models, analysis of durations including survival analysis, experimental and non-experimental group comparisons, reciprocal effects, stability. Prerequisite: intermediate statistical methods.
Same as: EDUC 351A

STATS 229. Machine Learning. 3-4 Units.
Topics: statistical pattern recognition, linear and non-linear regression, non-parametric methods, exponential family, GLMs, support vector machines, kernel methods, deep learning, model/feature selection, learning theory, ML advice, clustering, density estimation, EM, dimensionality reduction, ICA, PCA, reinforcement learning and adaptive control, Markov decision processes, approximate dynamic programming, and policy search. Prerequisites: knowledge of basic computer science principles and skills at a level sufficient to write a reasonably non-trivial computer program in Python/numpy, familiarity with probability theory to the equivalency of CS109 or STATS116, and familiarity with multivariable calculus and linear algebra to the equivalency of MATH51.
Same as: CS 229

STATS 237. Investment Portfolios, Derivative Securities, and Risk Measures. 3 Units.
Asset returns and their volatilities. Markowitz portfolio theory, capital asset pricing model, multifactor pricing models. Measures of market risk and statistical models and methods for their estimation and backtesting. Financial derivatives and hedging. Black-Scholes pricing of European options and implied volatilities. Prerequisite: STATS 116 or equivalent.

STATS 237P. Investment Portfolios, Derivative Securities, and Risk Measures. 3 Units.
For SCPD students; see STATS237.

STATS 240. Statistical Methods in Finance. 3 Units.
(SCPD students register for 240P.) Regression analysis and applications to investment models. Principal components and multivariate analysis. Likelihood inference and Bayesian methods. Financial time series. Estimation and modeling of volatilities. Statistical methods for portfolio management. Prerequisite: STATS 200 or equivalent.

STATS 240P. Statistical Methods in Finance. 3 Units.
For SCPD students; see 240.

STATS 241. Data-driven Financial Econometrics. 3 Units.
(SCPD students register for 241P) Approximate dynamic programming and time series approaches in options, interest rate, and credit markets. Nonlinear least squares, nonparametric regression and model selection. Behavioral finance and efficient markets. Economic capital, risk measures, and regulatory supervision. Quantile regression, extreme value theory, and applications to market risk analytics. Empirical Bayes approach to pricing insurance contracts. Corporate bonds, bond ratings, and corporate default analytics. Prerequisite or corequisite: STATS 240 or equivalent.

STATS 241P. Data-driven Financial Econometrics. 3 Units.
For SCPD students; see STATS241.

STATS 242. NeuroTech Training Seminar. 1 Unit.
This is a required course for students in the NeuroTech training program, and is also open to other graduate students interested in learning the skills necessary for neurotechnology careers in academia or industry. Over the academic year, topics will include: emerging research in neurotechnology, communication skills, team science, leadership and management, intellectual property, entrepreneurship and more.
Same as: NSUR 239

STATS 243. Risk Analytics and Management in Finance and Insurance. 2-4 Units.
Market risk and credit risk, credit markets. Back testing, stress testing and Monte Carlo methods. Logistic regression, generalized linear models and generalized mixed models. Loan prepayment and default as competing risks. Survival and hazard functions, correlated default intensities, frailty and contagion. Risk surveillance, early warning and adaptive control methodologies. Banking and bank regulation, asset and liability management. Prerequisite: STATS 240 or equivalent.
Same as: CME 243

STATS 243P. Risk Analytics and Management in Finance and Insurance. 3 Units.
For SCPD students; see STATS243.

STATS 244. Quantitative Trading: Algorithms, Data, and Optimization. 2-4 Units.
Statistical trading rules and performances evaluation. Active portfolio management and dynamic investment strategies. Data analytics and models of transactions data. Limit order book dynamics in electronic exchanges. Algorithmic trading, informatics, and optimal execution. Market making and inventory control. Risk management and regulatory issues. Prerequisites: STATS 240 or equivalent.

STATS 244P. Quantitative Trading: Algorithms, Data and Optimization. 3 Units.
For SCPD students; see 244.

STATS 245. Data, Models and Applications to Healthcare Analytics. 3 Units.
Topics on fundamentals of data science, biological and statistical models, application to medical product safety evaluation, health risk models and their evaluation, benefit-risk assessment and multi-criteria decision analytics. Applications to environmental health, nutritional epidemiology, wellness and prevention will also be discussed. Prerequisite: Graduate students - STATS 202 or 216, or CS 229; Undergraduate students - consent of instructor.

STATS 245P. Data, Models, and Applications to Healthcare Analytics. 3 Units.
For SCPD students; see STATS245.

STATS 248. Clinical Trial Design in the Age of Precision Medicine and Health. 3 Units.
Overview of requirements, designs, and statistical foundations for traditional Phase I, II, and III clinical trials for medical product approval and Phase IV postmarketing studies for safety evaluation. As these methods cost too much and take too much time in the era of precision medicine and precision health, this course then introduces innovative designs that have been developed for affordable clinical trials, which can be completed within reasonable time constraints and which have been encouraged by regulatory agencies. Prerequisites: Working knowledge of statistics and R.
Same as: BIODS 248, BIODS 248P, BIOMEDIN 248

STATS 249. Experimental Immersion in Neuroscience. 1 Unit.
This course provides students from technical backgrounds (e.g., physics, applied physics, electrical or chemical engineering, bioengineering, computer science, statistics) the opportunity to learn how they can apply their expertise to advancing experimental research in the neurosciences. Students will visit one neuroscience lab per week to watch experiments, understand the technical apparatus and animal models being used, discuss the questions being addressed, and interact with students and others conducting the research. This course is strongly encouraged for students who wish to apply to the NeuroTech graduate training program.
Same as: NSUR 249

STATS 250. Mathematical Finance. 3 Units.
Stochastic models of financial markets. Forward and futures contracts. European options and equivalent martingale measures. Hedging strategies and management of risk. Term structure models and interest rate derivatives. Optimal stopping and American options. Corequisites: MATH 236 and 227 or equivalent.nnNOTE: Undergraduates require instructor permission to enroll. Undergraduates interested in taking the course should contact the instructor for permission, providing information about relevant background such as performance in prior coursework, reading, etc.
Same as: MATH 238

STATS 253. Analysis of Spatial and Temporal Data. 3 Units.
A unified treatment of methods for spatial data, time series, and other correlated data from the perspective of regression with correlated errors. Two main paradigms for dealing with autocorrelation: covariance modeling (kriging) and autoregressive processes. Bayesian methods. Prerequisites: applied linear algebra (MATH 103 or equivalent), statistical estimation (STATS 200 or CS 229), and linear regression (STATS 203 or equivalent).

STATS 256. Modern Statistics for Modern Biology. 3 Units.
Application based course in nonparametric statistics. Modern toolbox of visualization and statistical methods for the analysis of data, examples drawn from immunology, microbiology, cancer research and ecology. Methods covered include multivariate methods (PCA and extensions), sparse representations (trees, networks, contingency tables) as well as nonparametric testing (Bootstrap, permutation and Monte Carlo methods). Hands on, use R and cover many Bioconductor packages. Prerequisite: Working knowledge of R and two core Biology courses. Note that the 155 offering is a writing intensive course for undergraduates only and requires instructor consent. (WIM). See https://web.stanford.edu/class/bios221/index.html.
Same as: BIOS 221, STATS 155, STATS 366

STATS 260A. Workshop in Biostatistics. 1-2 Unit.
Applications of statistical techniques to current problems in medical science. To receive credit for one or two units, a student must attend every workshop. To receive two units, in addition to attending every workshop, the student is required to write an acceptable one page summary of two of the workshops, with choices made by the student.
Same as: BIODS 260A

STATS 260B. Workshop in Biostatistics. 1-2 Unit.
Applications of statistical techniques to current problems in medical science. To receive credit for one or two units, a student must attend every workshop. To receive two units, in addition to attending every workshop, the student is required to write an acceptable one page summary of two of the workshops, with choices made by the student.
Same as: BIODS 260B

STATS 260C. Workshop in Biostatistics. 1-2 Unit.
Applications of statistical techniques to current problems in medical science. To receive credit for one or two units, a student must attend every workshop. To receive two units, in addition to attending every workshop, the student is required to write an acceptable one page summary of two of the workshops, with choices made by the student.
Same as: BIODS 260C

STATS 261. Intermediate Biostatistics: Analysis of Discrete Data. 3 Units.
(Formerly HRP 261) Methods for analyzing data from case-control and cross-sectional studies: the 2x2 table, chi-square test, Fisher's exact test, odds ratios, Mantel-Haenzel methods, stratification, tests for matched data, logistic regression, conditional logistic regression. Emphasis is on data analysis in SAS or R. Special topics: cross-fold validation and bootstrap inference.
Same as: BIOMEDIN 233, EPI 261

STATS 262. Intermediate Biostatistics: Regression, Prediction, Survival Analysis. 3 Units.
(Formerly HRP 262) Methods for analyzing longitudinal data. Topics include Kaplan-Meier methods, Cox regression, hazard ratios, time-dependent variables, longitudinal data structures, profile plots, missing data, modeling change, MANOVA, repeated-measures ANOVA, GEE, and mixed models. Emphasis is on practical applications. Prerequisites: basic ANOVA and linear regression.
Same as: EPI 262

STATS 263. Design of Experiments. 3 Units.
Experiments vs observation. Confounding. Randomization. ANOVA.Blocking. Latin squares. Factorials and fractional factorials. Split plot. Response surfaces. Mixture designs. Optimal design. Central composite. Box-Behnken. Taguchi methods. Computer experiments and space filling designs. Prerequisites: probability at STATS 116 level or higher, and at least one course in linear models.
Same as: STATS 363

STATS 264. Foundations of Statistical and Scientific Inference. 1 Unit.
(Formerly HRP 264) The course will consist of readings and discussion of foundational papers and book sections in the domains of statistical and scientific inference. Topics to be covered include philosophy of science, interpretations of probability, Bayesian and frequentist approaches to statistical inference and current controversies about the proper use of p-values and research reproducibility. Recommended preparation: At least 2 quarters of biostatistics and one of epidemiology. Intended for second year Masters students or PhD students with at least 1 year of preceding graduate training.
Same as: EPI 264

STATS 266. Advanced Statistical Methods for Observational Studies. 2-3 Units.
Design principles and statistical methods for observational studies. Topics include: matching methods, sensitivity analysis, and instrumental variables. 3 unit registration requires a small project and presentation. Computing is in R. Pre-requisites: EPI 261 and 262 or STATS 209 (EPI 239), or equivalent. See http://rogosateaching.com/somgen290/.
Same as: CHPR 266, EDUC 260B, EPI 292

STATS 267. Probability: Ten Great Ideas About Chance. 4 Units.
Foundational approaches to thinking about chance in matters such as gambling, the law, and everyday affairs. Topics include: chance and decisions; the mathematics of chance; frequencies, symmetry, and chance; Bayes great idea; chance and psychology; misuses of chance; and harnessing chance. Emphasis is on the philosophical underpinnings and problems. Prerequisite: exposure to probability or a first course in statistics at the level of STATS 60 or 116.
Same as: PHIL 166, PHIL 266, STATS 167

STATS 270. A Course in Bayesian Statistics. 3 Units.
This course will treat Bayesian statistics at a relatively advanced level. Assuming familiarity with standard probability and multivariate distribution theory, we will provide a discussion of the mathematical and theoretical foundation for Bayesian inferential procedures. In particular, we will examine the construction of priors and the asymptotic properties of likelihoods and posterior distributions. The discussion will include but will not be limited to the case of finite dimensional parameter space. There will also be some discussions on the computational algorithms useful for Bayesian inference. Prerequisites: Stats 116 or equivalent probability course, plus basic programming knowledge; basic calculus, analysis and linear algebra strongly recommended; Stats 200 or equivalent statistical theory course desirable.
Same as: STATS 370

STATS 271. Applied Bayesian Statistics. 3 Units.
This course is a modern treatment of applied Bayesian statistics with a focus on high-dimensional problems. We will study a collection of canonical methods that see heavy use in applications, including high-dimensional linear and generalized linear models, hierarchical/random effects models, Gaussian processes, variable-dimension and Dirichlet process mixtures, graphical models, and methods used in Bayesian inverse problems. Each method will be accompanied by one or more motivating datasets. Through these examples the course will cover: (1) Bayesian hypothesis testing, multiplicity correction, selection, shrinkage, and model averaging; (2) prior choice; (3) Frequentist properties of Bayesian procedures in high dimensions; and (4) computation by Markov chain Monte Carlo, including constructing efficient Gibbs, Metropolis, and more exotic samplers, empirical convergence analysis, strategies for scaling computation to high dimensions (approximations, divide-and-conquer, minibatching, et cetera), and the theory of convergence rates.
Same as: STATS 371

STATS 281. Statistical Analysis of Fine Art. 3 Units.
This course presents the application of rigorous statistical analysis, machine learning, and data analysis to problems in the history and interpretation of fine art paintings, drawings, and other two-dimensional artworks. The course focuses on the aspects of these problems that are unlike those addressed widely elsewhere in statistical image analysis, such as applied to photographs, videos, and medical images. These novel problems include statistical analysis of brushstrokes and marks, medium, inferring artists¿ working methods, compositional principles, stylometry (quantification of style), the tracing of artistic influence, and art attribution and authentication. The course revisits classic problems, such as image-based object recognition and scene description, but in the environment of highly non-realistic, stylized artworks.

STATS 285. Massive Computational Experiments, Painlessly. 2 Units.
Ambitious Data Science requires massive computational experimentation; the entry ticket for a solid PhD in some fields is now to conduct experiments involving 1 Million CPU hours. Recently several groups have created efficient computational environments that make it painless to run such massive experiments. This course reviews state-of-the-art practices for doing massive computational experiments on compute clusters in a painless and reproducible manner. Students will learn how to automate their computing experiments first of all using nuts-and-bolts tools such as Perl and Bash, and later using available comprehensive frameworks such as ClusterJob and CodaLab, which enables them to take on ambitious Data Science projects. The course also features few guest lectures by renowned scientists in the field of Data Science. Students should have a familiarity with computational experiments and be facile in some high-level computer language such as R, Matlab, or Python.

STATS 290. Computing for Data Science. 3 Units.
Programming and computing techniques for the requirements of data science: acquisition and organization of data; visualization, modelling and inference for scientific applications; presentation and interactive communication of results. Emphasis on computing for substantial projects. Software development with emphasis on R, plus other key software tools. Prerequisites: Programming experience including familiarity with R; computing at least at the level of CS 106; statistics at the level of STATS 110 or 141.

STATS 298. Industrial Research for Statisticians. 1 Unit.
Masters-level research as in 299, but with the approval and supervision of a faculty adviser, it must be conducted for an off-campus employer. Students must submit a written final report upon completion of the internship in order to receive credit. Repeatable for credit. Prerequisite: enrollment in Statistics M.S. program.

STATS 299. Independent Study. 1-5 Unit.
For Statistics M.S. students only. Reading or research program under the supervision of a Statistics faculty member. May be repeated for credit.

STATS 300A. Theory of Statistics I. 3 Units.
Finite sample optimality of statistical procedures; Decision theory: loss, risk, admissibility; Principles of data reduction: sufficiency, ancillarity, completeness; Statistical models: exponential families, group families, nonparametric families; Point estimation: optimal unbiased and equivariant estimation, Bayes estimation, minimax estimation; Hypothesis testing and confidence intervals: uniformly most powerful tests, uniformly most accurate confidence intervals, optimal unbiased and invariant tests. Prerequisites: Real analysis, introductory probability (at the level of STATS 116), and introductory statistics.

STATS 300B. Theory of Statistics II. 3 Units.
Elementary decision theory; loss and risk functions, Bayes estimation; UMVU estimator, minimax estimators, shrinkage estimators. Hypothesis testing and confidence intervals: Neyman-Pearson theory; UMP tests and uniformly most accurate confidence intervals; use of unbiasedness and invariance to eliminate nuisance parameters. Large sample theory: basic convergence concepts; robustness; efficiency; contiguity, locally asymptotically normal experiments; convolution theorem; asymptotically UMP and maximin tests. Asymptotic theory of likelihood ratio and score tests. Rank permutation and randomization tests; jackknife, bootstrap, subsampling and other resampling methods. Further topics: sequential analysis, optimal experimental design, empirical processes with applications to statistics, Edgeworth expansions, density estimation, time series.

STATS 300C. Theory of Statistics III. 3 Units.
Decision theory formulation of statistical problems. Minimax, admissible procedures. Complete class theorems ("all" minimax or admissible procedures are "Bayes"), Bayes procedures, conjugate priors, hierarchical models. Bayesian non parametrics: diaichlet, tail free, polya trees, bayesian sieves. Inconsistency of bayes rules.

STATS 302. Qualifying Exams Workshop. 5-10 Units.
Prepares Statistics Ph.D. students for the qualifying exams by reviewing relevant course topics and problem solving strategies.

STATS 303. Statistics Faculty Research Presentations. 1 Unit.
For Statistics first and second year PhD students only. Discussion of statistics topics and research areas; consultation with PhD advisors.

STATS 305A. Applied Statistics I. 3 Units.
Statistics of real valued responses. Review of multivariate normal distribution theory. Univariate regression. Multiple regression. Constructing features from predictors. Geometry and algebra of least squares: subspaces, projections, normal equations, orthogonality, rank deficiency, Gauss-Markov. Gram-Schmidt, the QR decomposition and the SVD. Interpreting coefficients. Collinearity. Dependence and heteroscedasticity. Fits and the hat matrix. Model diagnostics. Model selection, Cp/AIC and crossvalidation, stepwise, lasso. Multiple comparisons. ANOVA, fixed and random effects. Use of bootstrap and permutations. Emphasis on problem sets involving substantive computations with data sets. Prerequisites: consent of instructor, 116, 200, applied statistics course, CS 106A, MATH 114.

STATS 305B. Applied Statistics II: Generalized Linear Models, Survival Analysis, and Exponential Families. 3 Units.
This course uses exponential family structure to motivate generalized linear models and other useful applied techniques including survival analysis methods and Bayes and empirical Bayes analyses. The lectures are based on a forthcoming book whose notes will be distributed. Prerequisites: 305A or consent of the instructor.

STATS 305C. Applied Statistics III. 3 Units.
Methods for multivariate responses. Theory, computation and practice for multivariate statistical tools. Multivariate Gaussian and undirected graphical models, graphical displays. Hotelling's T-squared, principal components, canonical correlations, linear discriminant analysis, correspondence analysis, and recent variants of these. Hierarchical and k-means clustering. Bi-clustering. Factor analysis and independent component analysis. Topic modeling. Multidimensional scaling and variants (e.g., Isomap, spectral clustering, t-SNE). Matrix completion. Extensive work with data involving programming, ideally in R. Prerequisites: Stats 305A and Stats 305B or consent of the instructor.

STATS 310A. Theory of Probability I. 3 Units.
Mathematical tools: sigma algebras, measure theory, connections between coin tossing and Lebesgue measure, basic convergence theorems. Probability: independence, Borel-Cantelli lemmas, almost sure and Lp convergence, weak and strong laws of large numbers. Large deviations. Weak convergence; central limit theorems; Poisson convergence; Stein's method. Prerequisites: STATS 116, MATH 171.
Same as: MATH 230A

STATS 310B. Theory of Probability II. 3 Units.
Conditional expectations, discrete time martingales, stopping times, uniform integrability, applications to 0-1 laws, Radon-Nikodym Theorem, ruin problems, etc. Other topics as time allows selected from (i) local limit theorems, (ii) renewal theory, (iii) discrete time Markov chains, (iv) random walk theory,n(v) ergodic theory. http://statweb.stanford.edu/~adembo/stat-310b. Prerequisite: 310A or MATH 230A.
Same as: MATH 230B

STATS 310C. Theory of Probability III. 3 Units.
Continuous time stochastic processes: martingales, Brownian motion, stationary independent increments, Markov jump processes and Gaussian processes. Invariance principle, random walks, LIL and functional CLT. Markov and strong Markov property. Infinitely divisible laws. Some ergodic theory. Prerequisite: 310B or MATH 230B. http://statweb.stanford.edu/~adembo/stat-310c/.
Same as: MATH 230C

STATS 311. Information Theory and Statistics. 3 Units.
Information theoretic techniques in probability and statistics. Fano, Assouad,nand Le Cam methods for optimality guarantees in estimation. Large deviationsnand concentration inequalities (Sanov's theorem, hypothesis testing, thenentropy method, concentration of measure). Approximation of (Bayes) optimalnprocedures, surrogate risks, f-divergences. Penalized estimators and minimumndescription length. Online game playing, gambling, no-regret learning. Prerequisites: EE 276 (or equivalent) or STATS 300A.
Same as: EE 377

STATS 314A. Advanced Statistical Theory. 3 Units.
Covers a range of topics, including sequential experimentation, self-normalized statistics, multi-stage testing of multiple hypotheses, and their recent far-reaching multidisciplinary applications.

STATS 315A. Modern Applied Statistics: Learning. 3 Units.
Overview of supervised learning. Linear regression and related methods. Model selection, least angle regression and the lasso, stepwise methods. Classification. Linear discriminant analysis, logistic regression, and support vector machines (SVMs). Basis expansions, splines and regularization. Kernel methods. Generalized additive models. Kernel smoothing. Gaussian mixtures and the EM algorithm. Model assessment and selection: crossvalidation and the bootstrap. Pathwise coordinate descent. Sparse graphical models. Prerequisites: STATS 305A, 305B, 305C or consent of instructor.

STATS 315B. Modern Applied Statistics: Data Mining. 3 Units.
Two-part sequence. New techniques for predictive and descriptive learning using ideas that bridge gaps among statistics, computer science, and artificial intelligence. Emphasis is on statistical aspects of their application and integration with more standard statistical methodology. Predictive learning refers to estimating models from data with the goal of predicting future outcomes, in particular, regression and classification models. Descriptive learning is used to discover general patterns and relationships in data without a predictive goal, viewed from a statistical perspective as computer automated exploratory analysis of large complex data sets.

STATS 316. Stochastic Processes on Graphs. 1-3 Unit.
Local weak convergence, Gibbs measures on trees, cavity method, and replica symmetry breaking. Examples include random k-satisfiability, the assignment problem, spin glasses, and neural networks. Prerequisite: 310A or equivalent. https://web.stanford.edu/~montanar/TEACHING/Stat316/stat316.html.

STATS 317. Stochastic Processes. 3 Units.
Semimartingales, stochastic integration, Ito's formula, Girsanov's theorem. Gaussian and related processes. Stationary/isotropic processes. Integral geometry and geometric probability. Maxima of random fields and applications to spatial statistics and imaging.

STATS 318. Modern Markov Chains. 3 Units.
Tools for understanding Markov chains as they arise in applications. Random walk on graphs, reversible Markov chains, Metropolis algorithm, Gibbs sampler, hybrid Monte Carlo, auxiliary variables, hit and run, Swedson-Wong algorithms, geometric theory, Poincare-Nash-Cheeger-Log-Sobolov inequalities. Comparison techniques, coupling, stationary times, Harris recurrence, central limit theorems, and large deviations.

STATS 319. Literature of Statistics. 1 Unit.
Literature study of topics in statistics and probability culminating in oral and written reports. May be repeated for credit.

STATS 320. Machine Learning Methods for Neural Data Analysis. 3 Units.
With modern high-density electrodes and optical imaging techniques, neuroscientists routinely measure the activity of hundreds, if not thousands, of cells simultaneously. Coupled with high-resolution behavioral measurements, genetic sequencing, and connectomics, these datasets offer unprecedented opportunities to learn how neural circuits function. This course will study statistical machine learning methods for analysing such datasets, including: spike sorting, calcium deconvolution, and voltage smoothing techniques for extracting relevant signals from raw data; markerless tracking methods for estimating animal pose in behavioral videos; network models for connectomics and fMRI data; state space models for analysis of high-dimensional neural and behavioral time-series; point process models of neural spike trains; and deep learning methods for neural encoding and decoding. We will develop the theory behind these models and algorithms and then apply them to real datasets in the homeworks and final project.This course is similar to STATS215: Statistical Models in Biology and STATS366: Modern Statistics for Modern Biology, but it is specifically focused on statistical machine learning methods for neuroscience data. Prerequisites: Students should be comfortable with basic probability (STATS 116) and statistics (at the level of STATS 200). This course will place a heavy emphasis on implementing models and algorithms, so coding proficiency is required.
Same as: CS 339N, NBIO 220, STATS 220

STATS 322. Function Estimation in White Noise. 3 Units.
Gaussian white noise model sequence space form. Hyperrectangles, quadratic convexity, and Pinsker's theorem. Minimax estimation on Lp balls and Besov spaces. Role of wavelets and unconditional bases. Linear and threshold estimators. Oracle inequalities. Optimal recovery and universal thresholding. Stein's unbiased risk estimator and threshold choice. Complexity penalized model selection. Connecting fast wavelet algorithms and theory. Beyond orthogonal bases.

STATS 325. Multivariate Analysis and Random Matrices in Statistics. 3 Units.
Topics on Multivariate Analysis and Random Matrices in Statistics (full description TBA).

STATS 334. Mathematics and Statistics of Gambling. 3 Units.
Probability and statistics are founded on the study of games of chance. Nowadays, gambling (in casinos, sports and the Internet) is a huge business. This course addresses practical and theoretical aspects. Topics covered: mathematics of basic random phenomena (physics of coin tossing and roulette, analysis of various methods of shuffling cards), odds in popular games, card counting, optimal tournament play, practical problems of random number generation. Prerequisites: Statistics 116 and 200.
Same as: MATH 231

STATS 345. Statistical and Machine Learning Methods for Genomics. 3 Units.
Introduction to statistical and computational methods for genomics. Sample topics include: expectation maximization, hidden Markov model, Markov chain Monte Carlo, ensemble learning, probabilistic graphical models, kernel methods and other modern machine learning paradigms. Rationales and techniques illustrated with existing implementations used in population genetics, disease association, and functional regulatory genomics studies. Instruction includes lectures and discussion of readings from primary literature. Homework and projects require implementing some of the algorithms and using existing toolkits for analysis of genomic datasets.
Same as: BIO 268, BIOMEDIN 245, CS 373

STATS 350. Topics in Probability Theory. 3 Units.
See http://statweb.stanford.edu/~adembo/stat-350/concentration/ Selected topics of contemporary research interest in probability theory. May be repeated once for credit. Prerequisite: 310A or equivalent.

STATS 352. Topics in Computing for Data Science. 3 Units.
A seminar-style course jointly supported by the Statistics department and Stanford Data Science, and suitable for doctoral students engaged in either research on data science techniques (statistical or computational, for example) or research in scientific fields relying on advanced data science to achieve its goals. Seminars will usually consist of a student presentation of a relevant technical topic followed by discussion of the topic by all. Topics will be assigned to individuals to combine relevance for the course and suitability to the individual student's background and research interests. Prerequisites: Competence in the basic data science needed for the student's research goals plus preparation for presenting a suitable topic. Before enrolling, participants should have a topic approved as prescribed on the website https://stat352.stanford.edu.

STATS 359. Topics in Mathematical Physics. 3 Units.
Covers a list of topics in mathematical physics. The specific topics may vary from year to year, depending on the instructor's discretion. Background in graduate level probability theory and analysis is desirable.
Same as: MATH 273

STATS 360. Advanced Statistical Methods for Earth System Analysis. 3 Units.
Introduction for graduate students to important issues in data analysis relevant to earth system studies. Emphasis on methodology, concepts and implementation (in R), rather than formal proofs. Likely topics include the bootstrap, non-parametric methods, regression in the presence of spatial and temporal correlation, extreme value analysis, time-series analysis, high-dimensional regressions and change-point models. Topics subject to change each year. Prerequisites: STATS 110 or equivalent.
Same as: ESS 260

STATS 361. Causal Inference. 3 Units.
This course covers statistical underpinnings of causal inference, with a focus on experimental design and data-driven decision making. Topics include randomization, potential outcomes, observational studies, propensity score methods, matching, double robustness, semiparametric efficiency, treatment heterogeneity, structural models, instrumental variables, principal stratification, mediation, regression discontinuities, synthetic controls, interference, sensitivity analysis, policy learning, dynamic treatment rules, invariant prediction, graphical models, and structure learning. We will also discuss the relevance of optimization and machine learning tools to causal inference. Prerequisite: STATS 300A, or equivalent graduate-level coursework on the theory of statistics.

STATS 362. Topic: Monte Carlo. 3 Units.
Random numbers and vectors: inversion, acceptance-rejection, copulas. Variance reduction: antithetics, stratification, control variates, importance sampling. MCMC: Markov chains, detailed balance, Metropolis-Hastings, random walk Metropolis,nnindependence sampler, Gibbs sampling, slice sampler, hybrids of Gibbs and Metropolis, tempering. Sequential Monte Carlo. Quasi-Monte Carlo. Randomized quasi-Monte Carlo. Examples, problems and motivation from Bayesian statistics,nnmachine learning, computational finance and graphics. May be repeat for credit.

STATS 363. Design of Experiments. 3 Units.
Experiments vs observation. Confounding. Randomization. ANOVA.Blocking. Latin squares. Factorials and fractional factorials. Split plot. Response surfaces. Mixture designs. Optimal design. Central composite. Box-Behnken. Taguchi methods. Computer experiments and space filling designs. Prerequisites: probability at STATS 116 level or higher, and at least one course in linear models.
Same as: STATS 263

STATS 364. Theory and Applications of Selective Inference. 3 Units.
This course focuses on the problem of inference under the presence of multiplicity or selection. Topics covered include classical topics multiple comparisons (FWER, FDR, FCR) as well as newer methods such as knockoffs. We will also cover inference when targeted parameters are determined only after inspection of the data, considering both conditional and simultaneous approaches. Both theoretical and computational considerations will be stressed throughout the course. Prerequisite: STATS 200 or equivalent.

STATS 366. Modern Statistics for Modern Biology. 3 Units.
Application based course in nonparametric statistics. Modern toolbox of visualization and statistical methods for the analysis of data, examples drawn from immunology, microbiology, cancer research and ecology. Methods covered include multivariate methods (PCA and extensions), sparse representations (trees, networks, contingency tables) as well as nonparametric testing (Bootstrap, permutation and Monte Carlo methods). Hands on, use R and cover many Bioconductor packages. Prerequisite: Working knowledge of R and two core Biology courses. Note that the 155 offering is a writing intensive course for undergraduates only and requires instructor consent. (WIM). See https://web.stanford.edu/class/bios221/index.html.
Same as: BIOS 221, STATS 155, STATS 256

STATS 367. Statistical Models in Genetics. 3 Units.
This course will cover statistical problems in population genetics and molecular evolution with an emphasis on coalescent theory. Special attention will be paid to current research topics, illustrating the challenges presented by genomic data obtained via high-throughput technologies. No prior knowledge of genomics is necessary. Familiarity with the R statistical package or other computing language is needed for homework assignments. Prerequisites: knowledge of probability through elementary stochastic processes and statistics through likelihood theory.

STATS 368. Empirical Process Theory and its Applications. 3 Units.
This course is on the theory of empirical processes. In the course we will focus on weak convergence of stochastic processes, M-estimation and empirical risk minimization. The course will cover topics like covering numbers and bracketing numbers, maximal inequalities, chaining and symmetrization, uniform law of large numbers and uniform central limit theorems, rates of convergence of MLEs and (penalized) least squares estimators, and concentration inequalities.

STATS 369. Methods from Statistical Physics. 3 Units.
Mathematical techniques from statistical physics have been applied with increasing success on problems form combinatorics, computer science, machine learning. These methods are non-rigorous, but in several cases they were proved to yield correct predictions. This course provides a working knowledge of these methods for non-physicists. Specific topics: the Sherrington-Kirkpatrick model; sparse regression with random designs;.

STATS 370. A Course in Bayesian Statistics. 3 Units.
This course will treat Bayesian statistics at a relatively advanced level. Assuming familiarity with standard probability and multivariate distribution theory, we will provide a discussion of the mathematical and theoretical foundation for Bayesian inferential procedures. In particular, we will examine the construction of priors and the asymptotic properties of likelihoods and posterior distributions. The discussion will include but will not be limited to the case of finite dimensional parameter space. There will also be some discussions on the computational algorithms useful for Bayesian inference. Prerequisites: Stats 116 or equivalent probability course, plus basic programming knowledge; basic calculus, analysis and linear algebra strongly recommended; Stats 200 or equivalent statistical theory course desirable.
Same as: STATS 270

STATS 371. Applied Bayesian Statistics. 3 Units.
This course is a modern treatment of applied Bayesian statistics with a focus on high-dimensional problems. We will study a collection of canonical methods that see heavy use in applications, including high-dimensional linear and generalized linear models, hierarchical/random effects models, Gaussian processes, variable-dimension and Dirichlet process mixtures, graphical models, and methods used in Bayesian inverse problems. Each method will be accompanied by one or more motivating datasets. Through these examples the course will cover: (1) Bayesian hypothesis testing, multiplicity correction, selection, shrinkage, and model averaging; (2) prior choice; (3) Frequentist properties of Bayesian procedures in high dimensions; and (4) computation by Markov chain Monte Carlo, including constructing efficient Gibbs, Metropolis, and more exotic samplers, empirical convergence analysis, strategies for scaling computation to high dimensions (approximations, divide-and-conquer, minibatching, et cetera), and the theory of convergence rates.
Same as: STATS 271

STATS 374. Large Deviations Theory. 3 Units.
Combinatorial estimates and the method of types. Large deviation probabilities for partial sums and for empirical distributions, Cramer's and Sanov's theorems and their Markov extensions. Applications in statistics, information theory, and statistical mechanics. Prerequisite: MATH 230A or STATS 310. Offered every 2-3 years. http://statweb.stanford.edu/~adembo/large-deviations/.
Same as: MATH 234

STATS 375. Mathematical problems in Machine Learning. 3 Units.
Mathematical tools to understand modern machine learning systems. Generalization in machine learning, the classical view: uniform convergence, Radamacher complexity. Generalization from stability. Implicit (algorithmic) regularization. Infinite-dimensional models: reproducing kernel Hilbert spaces. Random features approximations to kernel methods. Connections to neural networks, and neural tangent kernel. Nonparametric regression. Asymptotic behavior of wide neural networks. Properties of convolutionalnnetworks. Prerequisites: EE364A or equivalent; Stat310A or equivalent.
Same as: EE 375

STATS 376A. Information Theory. 3 Units.
(Formerly EE 376A.) Project-based course about how to measure, represent, and communicate information effectively. Why bits have become the universal currency for information exchange. How information theory bears on the design and operation of modern-day systems such as smartphones and the Internet. The role of entropy and mutual information in data compression, communication, and inference. Practical compressors and error correcting codes. The information theoretic way of thinking. Relations and applications to probability, statistics, machine learning, biological and artificial neural networks, genomics, quantum information, and blockchains. Prerequisite: a first undergraduate course in probability.
Same as: EE 276

STATS 376B. Topics in Information Theory and Its Applications. 3 Units.
Information theory establishes the fundamental limits on compression and communication over networks. The tools of information theory have also found applications in many other fields, including probability and statistics, computer science and physics. The course will cover selected topics from these applications, including communication networks, through regular lectures and student projects. Prerequisites: EE276 (Formerly EE376A).
Same as: EE 376B

STATS 385. Analyses of Deep Learning. 1 Unit.
Deep learning is a transformative technology that has delivered impressive improvements in image classification and speech recognition. Many researchers are trying to better understand how to improve prediction performance and also how to improve training methods. Some researchers use experimental techniques; others use theoretical approaches. In this course we will review both experimental and theoretical analyses of deep learning. We will have 8-10 guest lecturers as well as graded projects for those who take the course for credit.

STATS 390. Consulting Workshop. 1 Unit.
Skills required of practicing statistical consultants, including exposure to statistical applications. Students participate as consultants in the department's drop-in consulting service, analyze client data, and prepare formal written reports. Seminar provides supervised experience in short term consulting. May be repeated for credit. Prerequisites: course work in applied statistics or data analysis, and consent of instructor.

STATS 397. PhD Oral Exam Workshop. 1 Unit.
For Statistics PhD students defending their dissertation.

STATS 398. Industrial Research for Statisticians. 1 Unit.
Doctoral research as in 399, but must be conducted for an off-campus employer. A final report acceptable to the advisor outlining work activity, problems investigated, key results, and any follow-up projects they expect to perform is required. The report is due at the end of the quarter in which the course is taken. May be repeated for credit. Prerequisite: Statistics Ph.D. candidate.

STATS 399. Research. 1-10 Unit.
Research work as distinguished from independent study of nonresearch character listed in 199. May be repeated for credit.

STATS 801. TGR Project. 0 Units.
.

STATS 802. TGR Dissertation. 0 Units.
.