STATISTICS

Emeriti: Theodore W. Anderson, Ingram Olkin, Charles Stein
Chair: Trevor J. Hastie
Professors: Thomas M. Cover, Amir Dembo, Persi Diaconis, David L. Donoho, Bradley Efron, Jerome H. Friedman, Trevor J. Hastie, Iain M. Johnstone, Tze L. Lai, Art Owen, Joseph P. Romano, David O. Siegmund, Paul Switzer, Robert J. Tibshirani, Wing H. Wong
Associate Professors: Jonathan Taylor, Guenther Walther
Assistant Professors: Andrea Montanari, Nancy Zhang
Professor (Teaching): Susan Holmes
Courtesy Professors: Philip W. Lavori, Richard A. Olshen
Courtesy Associate Professors: Simon Jackman, David Rogosa
Consulting Professors: John Chambers, Charles Chui, David G. Stork

Mail Code: 94305-4065
Phone: (650) 723-2620
Web Site: http://www-stat.stanford.edu

Courses offered by the Department of Statistics have the subject code STATS, and are listed in the "Statistics (STATS) Courses" section of this bulletin.

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 requirements for a degree in Statistics are flexible, depending on the needs and interests of the students. Some students may be interested in the theory of statistics and/or probability, whereas other students may wish to apply statistical and probabilistic methods to a substantive area. 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: Economics (Anderson), Education (Olkin, Rogosa), Electrical Engineering (Cover), Geological and Environmental Sciences (Switzer), Health Research and Policy (Efron, Hastie, Johnstone, Olshen, Tibshirani, Wong), Mathematics (Dembo, Diaconis), Political Science (Jackman), and the Stanford Linear Accelerator Center (Friedman). 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 majors, 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 of Statistics is well equipped for statistical applications and research in computational statistics. Computer facilities include several networked Unix servers and a PC lab for general research and teaching use. The Mathematical Sciences Library serves the department jointly with the departments of Mathematics and Computer Science.

The department has always drawn visitors from other countries and universities. As a consequence, there is usually a wide range of seminars offered by both the visitors and our own faculty.

UNDERGRADUATE PROGRAMS IN STATISTICS

MAJORING IN STATISTICS

Students wishing to build a concentration in probability and statistics are encouraged to consider declaring a major in Mathematical and Computational Science. This interdepartmental 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.

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 valued preparation for professional degree studies in postgraduate academic programs.

The minor consists of a minimum of six courses with a total of at least 20 units. There are two required courses (8 units) and four qualifying or elective courses (12 or more units). An overall 2.75 grade point average (GPA) is required for courses fulfilling the minor.

Required Courses: STATS 116 and 200.

1. *Qualifying Courses:* at most, one of these two courses may be counted toward the six course requirement for the minor: MATH 52; STATS 191.
2. *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. Examples of elective course sequences are:
   - STATS 202, 203, 204, emphasizing data analysis and applied statistics
   - STATS 205, 206, 207, emphasizing statistical methodology
   - STATS 206, ECON 160, 181, emphasizing economic optimization
   - STATS 206, PSYCH 156, 160, emphasizing psychology modeling and experiments
   - STATS 207, EE 264, 279, emphasizing signal processing
   - STATS 217, BIO 283, emphasizing genetic and ecologic modeling
   - STATS 217, 218, emphasizing probability and its applications
   - STATS 240, 250, emphasizing mathematical finance

GRADUATE PROGRAMS IN STATISTICS

University requirements for the M.S. and Ph.D. degrees are discussed in the "Graduate Degrees" section of this bulletin.

MASTER OF SCIENCE IN STATISTICS

The department requires that the student take 45 units of work from offerings in the Department of Statistics or from authorized courses in other departments. Ordinarily, four or five quarters are needed to complete all requirements.

Students must fulfill the following requirements for the M.S. degree:

STATS 116, 191, 200, and 217. All must be taken for a letter grade. Courses previously taken may be waived by the adviser, in which case they must be replaced by other graduate courses offered by the department.

3. One of MATH 103, 113, 115, 171; and one of CS 106A, 106X, 137, 138. Substitution of other courses in Mathematics and Computer Science may be made with consent of the adviser.
4. At least four additional Statistics courses from graduate offerings in the department (202-399). All must be taken for a letter grade. Consent of the adviser is required in order to take more than six units of STATS 260ABC, 298, 299, 390, or 399.
5. Additional elective units to complete the requirements may be chosen from the list available from the department web site. Other graduate courses (200 or above) may be authorized by the adviser if they provide skills relevant to statistics or deal primarily with an application of statistics or probability and do not overlap 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.

6. Courses below 200 level are generally not acceptable, with the following exceptions: STATS 116, 191; MATH 103, 113, 115, 171, 180; CS 106A, 106B, 106X, 137, 138. At most, one of these two courses may be counted: (1) MATH 151 or STATS 116, (2) MATH 103 or MATH 113.

Students with a strong mathematical background who may wish to go on to a Ph.D. in Statistics should consider applying to the Ph.D. program.

The eight Statistics courses required for the M.S. degree must be taken for letter grades, and an overall 2.75 grade point average (GPA) is required.

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. In particular, the department is expanding its research and educational activities towards computational biology, mathematical finance and information science, via a VIGRE program. The program normally takes four years to complete.

Program Summary—STATS 300A,B,C, 305, 306A,B, and 310A,B,C (first-year core program); pass two of three parts of the qualifying examinations (beginning of second year); breadth requirement (second or third year); University oral examination (end of third year or beginning of fourth year); dissertation (fourth year).

In addition, students are required to take 9 units of advanced topics courses offered by the department (including at least two of the following: 314, 317, 318, 315A, or 315B, but not including literature, research, or consulting), and 3 units of statistical consulting. All students who have passed the qualifying exam but have not yet passed the University oral examination must take 319 at least once per year.

First-Year Core Courses—STATS 300 systematically surveys the ideas of estimation and of hypothesis testing for parametric and nonparametric models involving small and large samples. 305 is concerned with linear regression and the analysis of variance. 306 surveys a large number of modeling techniques, related to but going beyond the linear models of 305. 310 is a measure-theoretic course 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 310 after their first year but need to have their first-year program approved by the Ph.D. program adviser.

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 may take two or three of these examinations and are expected to show acceptable performance in two examinations.

Breadth Requirement—Students are advised to choose an area of concentration in a specific scientific field of statistical applications; this can be realized by taking at least 15 units of course work approved by the Ph.D. program adviser.

Current areas with suggested course options include:

Computational Biology and Statistical Genomics—Students are expected to take 9 units of graduate courses in genetics or neurosciences (imaging), such as GENE 203/BIO 203, as well as 9 units of classes in Statistical Genetics or Bioinformatics, GENE 344A,B, STATS 345, STATS 366, STATS 367.

Machine Learning—Courses can be chosen from the following list:

Statistical Learning: STATS 315A and 315B
Data Bases: CS 245, 346, 347
Probabilistic Methods in AI: CS 221, 354
Statistical Learning Theory and Pattern Classification: CS 229

Applied Probability—Students are expected to take 15 units of graduate courses in some of the following areas:

Control and Stochastic Calculus: MS&E 322, 351, MATH 237, EE 363
Finance: STATS 250, FINANCE 622, MATH 236
Information Theory: EE 376A, 376B
Monte Carlo: STATS 318, 345, 362, MS&E 323
Queuing Theory: GSB 661, 663, MS&E 335
Stochastic Processes: STATS 317, MATH 234

Earth Science Statistics—Students are expected to take:

STATS 317, 318, 352
and three courses from the GES or Geophysics departments, such as GES 144 or GEOPHYS 210.

Social and Behavioral Sciences—Students are expected to take three advanced courses from the department with an applied orientation such as:

STATS 261/262, 324, 343, 354
and three advanced quantitative courses from departments such as Anthropology, Economics, Political Science, Psychology, and Sociology, and the schools of Education, Business, or Medicine.

University Oral Examination—The University oral examination is taken on the recommendation of the student's research adviser after the thesis problem has been well defined and some research progress has been made. Usually, this happens early in the student's fourth year. The oral examination consists of a 40-minute presentation on the thesis topic, followed by two question periods. The first question period relates directly to the student's presentation; the second is intended to explore the student's familiarity with broader statistical topics related to the thesis research.

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 strongly encouraged to apply for outside scholarships, fellowships, and other forms of financial support.

PH.D. MINOR IN STATISTICS

The minimum requirement for a Ph.D. minor is 20 units of course work at the 200 level, taken at Stanford. The Department of Statistics devises individual Ph.D. minor programs, but the department recommends that graduate students in other fields who wish to have a subspecialty in statistics study for an M.S. degree instead. The unit requirement for an M.S. degree is 45 units, whereas the number of units required for a minor averages around 30. This difference of 15 units can be made up by the student by including in the M.S. program courses from his or her own field which are related to statistics or applications of statistics. Contact the student services officer for further information.

STATISTICS (STATS) COURSES

For information on undergraduate and graduate programs in the Department of Statistics, see the "Statistics" section of this bulletin.

UNDERGRADUATE COURSES IN STATISTICS

STATS 46N. Experiments in Extrasensory Perception: A Critical Analysis
Stanford Introductory Seminar. Preference to freshmen. Old and new reports of ESP experiments at Stanford and elsewhere. Principles of experimental design, randomization, experimental control and confounding, response modeling, and probabilistic calculation. Design and execution of student ESP experiments, literature reviews, probability calculations, critiques, and oral and written presentations.
3 units, Aut (Switzer, P)

STATS 47N. Breaking the Code?
Stanford Introductory Seminar. Preference to freshmen. Cryptography and its counterpart, cryptanalysis or code breaking. How the earliest cryptanalysts used statistical tools to decrypt messages by uncovering recurring patterns. How such frequency-analysis tools have been used to analyze biblical texts to produce a Bible code, and to detect genes in the human genome. Overview of codes and ciphers. Statistical tools useful for code breaking. Students use simple computer programs to apply these tools to break codes and explore applications to various kinds of data. GER:DB-Math
3 units, Aut (Holmes, S)

STATS 50. Mathematics of Sports
(Same as MCS 100.) The use of mathematics, statistics, and probability in the analysis of sports performance, sports records, and strategy. Topics include mathematical analysis of the physics of sports and the determinations of optimal strategies. New diagnostic statistics and strategies for each sport. Corequisite: STATS 116. GER:DB-Math
3 units, not given this year

STATS 60. Introduction to Statistical Methods: Precalculus
(Same as PSYCH 10, STATS 160.) 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. GER:DB-Math
5 units, Aut (Thomas, E), Win (Walther, G), Spr (Boik, J), Sum (Staff)

STATS 110. Statistical Methods in Engineering and the Physical Sciences
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. GER:DB-Math
4-5 units, Aut (Staff), Sum (Staff)

STATS 116. Theory of Probability
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. GER:DB-Math
3-5 units, Aut (Ross, K), Spr (Staff), Sum (Staff)

STATS 141. Biostatistics
(Same as BIO 141.) 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. See http://www-stat.stanford.edu/~rag/stat141/. GER:DB-Math
4-5 units, Aut (Boik, J; Rogosa, D)

STATS 166. Computational Biology
(Same as BIOMEDIN 366, STATS 366.) Methods to understand sequence alignments and phylogenetic trees built from molecular data, and general genetic data. Phylogenetic trees, median networks, microarray analysis, Bayesian statistics. Binary labeled trees as combinatorial objects, graphs, and networks. Distances between trees. Multivariate methods (PCA, CA, multidimensional scaling). Combining data, nonparametric inference. Algorithms used: branch and bound, dynamic programming, Markov chain approach to combinatorial optimization (simulated annealing, Markov chain Monte Carlo, approximate counting, exact tests). Software such as Matlab, Phylip, Seq-gen, Arlequin, Puzzle, Splitstree, XGobi.
2-3 units, Spr (Wong, W)

STATS 191. Introduction to Applied Statistics
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. Recommended: 60, 110, or 141. GER:DB-Math
3-4 units, Win (Taylor, J)

STATS 199. Independent Study
For undergraduates.
1-15 units, Aut (Staff), Win (Staff), Spr (Staff), Sum (Staff)

GRADUATE COURSES IN STATISTICS

Primarily for graduate students; undergraduates may enroll with consent of instructor.

STATS 160. Introduction to Statistical Methods: Precalculus
(Same as PSYCH 10, STATS 60.) 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.
5 units, Aut (Thomas, E), Win (Walther, G), Spr (Boik, J), Sum (Staff)

STATS 200. Introduction to Statistical Inference
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: 116.
3 units, Win (Romano, J), Sum (Staff)

STATS 202. Data Mining and Analysis
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, neural networks, association rules, clustering, case based methods, and data visualization.
3 units, Aut (Walther, G)

STATS 203. Introduction to Regression Models and Analysis of Variance
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. Pre- or corequisite: 200.
3 units, Win (Zhang, N)

STATS 206. Applied Multivariate Analysis
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.
3 units, Aut (Khalessi, S), Sum (Staff)

STATS 208. Introduction to the Bootstrap
The bootstrap is a computer-based method for assigning measures of accuracy to statistical estimates. By substituting computation in place of mathematical formulas, it permits the statistical analysis of complicated estimators. Topics: nonparametric assessment of standard errors, biases, and confidence intervals; related resampling methods including the jackknife, cross-validation, and permutation tests. Theory and applications. Prerequisite: course in statistics or probability.
3 units, Spr (Holmes, S)

STATS 209. Understanding Statistical Models and their Social Science Applications
(Same as EDUC 260X, HRP 239.) Statistical modeling in experimental and non-experimental settings, including misconceptions in social science applications such as causal models. Text is Statistical Models: Theory and Practice, by David Freedman. See http://www-stat.stanford.edu/~rag/stat209. Prerequisite: intermediate-level statistical methods including multiple regression, logistic regression, and log-linear models.
3 units, Win (Rogosa, D)

STATS 211. Topics in Quantitative Methods: Meta-Analysis
Meta-analysis as a quantitative method for combining the results of independent studies enabling researchers to evaluate available evidence. Examples of meta-analysis in medicine, education, and social and behavioral sciences. Statistical methods include nonparametric methods, contingency tables, regression and analysis of variance, and Bayesian methods. Project involving an existing published meta-analysis. Prerequisite: basic sequence in statistics.
1-3 units, Win (Olkin, I)

STATS 212. Applied Statistics with SAS
Data analysis and implementation of statistical tools in SAS. Topics: reading in and describing data, categorical data, dates and longitudinal data, correlation and regression, nonparametric comparisons, ANOVA, multiple regression, multivariate data analysis, using arrays and macros in SAS. Prerequisite: statistical techniques at the level of STATS 191 or 203; knowledge of SAS not required.
3 units, Sum (Staff)

STATS 214. Randomness in the Physical World
(Same as APPPHYS 214.) Topics include: random numbers, and their generation and application; disordered systems, quenching, and annealing; percolation and fractal structures; universality, the renormalization group, and limit theorems; path integrals, partition functions, and Wiener measure; random matrices; and optical estimation. Prerequisite: introductory course in statistical mechanics or analysis.
3 units, Spr (Diaconis, P; Fisher, D; Holmes, S), alternate years, not given next year

STATS 215. Statistical Models in Biology
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.
3 units, Win (Zhang, N)

STATS 217. Introduction to Stochastic Processes
Discrete and continuous time Markov chains, point processes, random walks, branching processes, first passage times, recurrence and transience, stationary distributions. Prerequisite: STATS 116 or consent of instructor.
3 units, Win (Rajaratnam, B), Sum (Staff)

STATS 218. Introduction to Stochastic Processes
Renewal theory, Brownian motion, Gaussian processes, second order processes, martingales.
3 units, Spr (Staff), Sum (Staff)

STATS 219. Stochastic Processes
(Same as MATH 136.) 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.
3 units, Aut (Ross, K)

STATS 237. Time Series Modeling and Forecasting
Box-Jenkins and Bayesian approaches. State-space and change-point models. Application to revenue prediction, forecasting product demand, and other real world problems. Development and assessment of models and forecasts in practical applications. Hands-on experience with real data.
3 units, Sum (Staff)

STATS 239A. Workshop in Quantitative Finance
Topics of current interest.
1 unit, Aut (Lai, T)

STATS 239B. Workshop in Quantitative Finance
Topics of current interest. May be repeated for credit.
1 unit, Spr (Lai, T)

STATS 240. Statistical Methods in Finance
(SCPD students register for 240P.) Regression analysis and applications to pricing and investment models. Principal components and multivariate analysis. Parametric influence. Financial time series. Estimation and modeling of volatilities. Statistical methods for portfolio management. Hands-on experience with financial data.
3-4 units, Aut (Lai, T)

STATS 240P. Statistical Methods in Finance
For SCPD students; see 240.
3 units, Aut (Lai, T)

STATS 241. Statistical Modeling in Financial Markets
(SCPD students register for 241P.) Nonparametric regression and yield curve smoothing. Advanced time series modeling and forecasting. Market risk measures. Substantive and empirical modeling approaches in financial markets. Statistical trading strategies. Prerequisite: 240 or equivalent.
3-4 units, Spr (Lai, T)

STATS 241P. Statistical Modeling in Financial Markets
For SCPD students; see 241.
3 units, Spr (Lai, T)

STATS 243. Introduction to Mathematical Finance
Interest rate and discounted value. Financial derivatives, hedging, and risk management. Stochastic models of financial markets, introduction to Ito calculus and stochastic differential equations. Black-Scholes pricing of European options. Optimal stopping and American options. Prerequisites: MATH 53, STATS 116, or equivalents.
3-4 units, Sum (Staff)

STATS 250. Mathematical Finance
(Same as MATH 238.) 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.
3 units, Win (Papanicolaou, G)

STATS 252. Data Mining and Electronic Business
The Internet and related technologies have caused the cost of communication and transactions to plummet, and consequently the amount of potentially relevant data to explode. The underlying principles, statistical issues, and algorithmic approaches to data mining and e-business, with real world examples.
3 units, Spr (Weigend, A)

STATS 253. Spatial Statistics
(Same as STATS 352.) Statistical descriptions of spatial variability, spatial random functions, grid models, spatial partitions, spatial sampling, linear and nonlinear interpolation and smoothing with error estimation, Bayes methods and pattern simulation from posterior distributions, multivariate spatial statistics, spatial classification, nonstationary spatial statistics, space-time statistics and estimation of time trends from monitoring data, spatial point patterns, models of attraction and repulsion. Applications to earth and environmental sciences, meteorology, astronomy, remote-sensing, ecology, materials. GER:DB-Math
3 units, Spr (Taylor, J)

STATS 254. Correspondence Analysis and Related Methods
Use of correspondence analysis (CA) method for dimension-reduction based on the singular-value decomposition, aimed at frequency data or raw multivariate categorical observations. Comprehensive treatment of simple and multiple CA and related methods, using R packages and including 2- and 3-dimensional graphics.
3 units, Aut (Staff)

STATS 260A. Workshop in Biostatistics
(Same as HRP 260A.) Applications of statistical techniques to current problems in medical science.
1-2 units, Aut (Olshen, R)

STATS 260B. Workshop in Biostatistics
(Same as HRP 260B.) Applications of statistical techniques to current problems in medical science.
1-2 units, Win (Olshen, R)

STATS 260C. Workshop in Biostatistics
(Same as HRP 260C.) Applications of statistical techniques to current problems in medical science.
1-2 units, Spr (Olshen, R)

STATS 261. Intermediate Biostatistics: Analysis of Discrete Data
(Same as BIOMEDIN 233, 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. Special topics: cross-fold validation and bootstrap inference.
3 units, Win (Sainani, K)

STATS 262. Intermediate Biostatistics: Regression, Prediction, Survival Analysis
(Same as 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.
3 units, Spr (Sainani, K)

STATS 270. A Course in Bayesian Statistics
(Same as STATS 370.) Bayesian statistics including theory, applications, and computational tools. Topics: history of Bayesian methods, foundational problems (what is probability), subjective probability and coherence, exchangeability and deFinetti's theorem. Conjugate priors, Laplace approximations, Gibbs sampling, hierarchical and empirical Bayes, nonparametric methods, Dirichlet and Polya tree priors. Bayes robustness, asymptotic properties of Bayes procedures.
3 units, Win (Wong, W)

STATS 290. Paradigms for Computing with Data
For Statistics graduate students and others whose research involves data analysis and development of associated computational software. Programming and computing techniques to support projects in data analysis and related research. Prerequisites: CS 106, and STATS 110 or 141, or equivalent background.
3 units, Win (Narasimhan, B; Chambers, J)

STATS 297. Practical Training
For students in the M.S. program in Financial Mathematics only. Students obtain employment in a relevant industrial or research activity to enhance their professional experience. May be repeated for credit once. Prerequisite: consent of adviser.
1-3 units, Aut (Lai, T), Win (Lai, T), Spr (Lai, T), Sum (Lai, T)

STATS 298. Industrial Research for Statisticians
Masters-level research as in 299, but must be conducted for an off-campus employer. Final report required. Prerequisite: enrollment in Statistics M.S. or Ph.D. program, prior to candidacy.
1-9 units, Aut (Staff), Win (Staff), Spr (Staff), Sum (Staff)

STATS 299. Independent Study
For Statistics M.S. students only. Reading or research program under the supervision of a Statistics faculty member. May be repeated for credit.
1-10 units, Aut (Staff), Win (Staff), Spr (Staff), Sum (Staff)

STATS 300. Advanced Topics in Statistics
May be repeated for credit.
3 units, Sum (Staff)

STATS 300A. Theory of Statistics
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.
2-4 units, Aut (Walther, G)

STATS 300B. Theory of Statistics
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.
2-4 units, Win (Siegmund, D)

STATS 300C. Theory of Statistics
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.
2-4 units, Spr (Siegmund, D)

STATS 305. Introduction to Statistical Modeling
The linear model: simple linear regression, polynomial regression, multiple regression, anova models; and with some extensions, orthogonal series regression, wavelets, radial basis functions, and MARS. Topics: normal theory inference (tests, confidence intervals, power), related distributions (t, chi-square, F), numerical methods (QR, SVD), model selection/regularization (Cp, AIC, BIC), diagnostics of model inadequacy, and remedies including bootstrap inference, and cross-validation. Emphasis is on problem sets involving substantial computations with data sets, including developing extensions of existing methods. Prerequisites: consent of instructor, 116, 200, applied statistics course, CS 106A, MATH 114.
2-4 units, Aut (Owen, A)

STATS 306A. Methods for Applied Statistics
Extension of modeling techniques of 305: binary and discrete response data and nonlinear least squares. Topics include regression, Poisson loglinear models, classification methods, clustering. May be repeated for credit. Prerequisite: 305 or equivalent.
2-4 units, Win (Efron, B)

STATS 306B. Methods for Applied Statistics
Unsupervised learning techniques in statistics, machine learning, and data mining.
2-4 units, Spr (Hastie, T)

STATS 310A. Theory of Probability
(Same as MATH 230A.) Mathematical tools: asymptotics, metric spaces; measure and integration; Lp spaces; some Hilbert spaces theory. Probability: independence, Borel-Cantelli lemmas, almost sure and Lp convergence, weak and strong laws of large numbers. Weak convergence and characteristic functions; central limit theorems; local limit theorems; Poisson convergence. Prerequisites: 116, MATH 171.
2-4 units, Aut (Diaconis, P)

STATS 310B. Theory of Probability
(Same as MATH 230B.) Stopping times, 0-1 laws, Kolmogorov consistency theorem. Uniform integrability. Radon-Nikodym theorem, branching processes, conditional expectation, discrete time martingales. Exchangeability. Large deviations. Laws of the iterated logarithm. Birkhoff's and Kingman's ergodic theorems. Recurrence, entropy. Prerequisite: 310A or MATH 230A.
2-4 units, Win (Dembo, A)

STATS 310C. Theory of Probability
(Same as MATH 230C.) Infinitely divisible laws. Continuous time martingales, random walks and Brownian motion. Invariance principle. Markov and strong Markov property. Processes with stationary independent increments. Prerequisite: 310B or MATH 230B.
2-4 units, Spr (Dembo, A)

STATS 314. Advanced Statistical Methods
Topic this year is multiple hypothesis testing. The demand for new methodology for the simultaneous testing of many hypotheses as driven by modern applications in genomics, imaging, astronomy, and finance. High dimensionality: how tests of many hypotheses may be considered simultaneously. Classical techniques, and recent developments. Stepwise methods, generalized error rates such as the false discovery rate, and the role of resampling. May be repeated for credit.
2-3 units, not given this year

STATS 315A. Modern Applied Statistics: Learning
Topics: clustering, biclustering, and spectral clustering. Data analysis using the singular value decomposition, nonnegative decomposition, and generalizations. Plaid model, aspect model, and additive clustering. Correspondence analysis, Rasch model, and independent component analysis. Page rank, hubs, and authorities. Probabilistic latent semantic indexing. Recommender systems. Applications to genomics and information retrieval. Prerequisite: 315A,B, 305, 306A,B, or consent of instructor.
2-3 units, Aut (Tibshirani, R)

STATS 315B. Modern Applied Statistics: Data Mining
Three-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.
2-3 units, Win (Friedman, J)

STATS 315C. Modern Applied Statistics: Transposable data
Topics: clustering, biclustering, and spectral clustering. Data analysis using the singular value decomposition, nonnegative decomposition, and generalizations. Plaid model, aspect model, and additive clustering. Correspondence analysis, Rasch model, and independent component analysis. Page rank, hubs, and authorities. Probabilistic latent semantic indexing. Recommender systems. Applications to genomics and information retrieval. Prerequisite: 315A,B, 305/306A,B, or consent of instructor.
2-3 units, Spr (Owen, A)

STATS 316. Stochastic Processes on Graphs
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.
1-3 units, not given this year

STATS 317. Stochastic Processes
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.
2-3 units, Spr (Siegmund, D)

STATS 318. Modern Markov Chains
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-Cheger-Log-Sobolov inequalities. Comparison techniques, coupling, stationary times, Harris recurrence, central limit theorems, and large deviations.
2-3 units, not given this year

STATS 319. Literature of Statistics
Literature study of topics in statistics and probability culminating in oral and written reports. May be repeated for credit.
1-3 units, Aut (Taylor, J), Win (Montanari, A)

STATS 322. Function Estimation in White Noise
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.
2-3 units, Spr (Johnstone, I)

STATS 324. Multivariate Analysis
Classic multivariate statistics: properties of the multivariate normal distribution, determinants, volumes, projections, matrix square roots, the singular value decomposition; Wishart distributions, Hotelling's T-square; principal components, canonical correlations, Fisher's discriminant, the Cauchy projection formula.
2-3 units, not given this year

STATS 345. Computational Algorithms for Statistical Genetics
(Same as GENE 245.) Computational algorithms for human genetics research. Topics include: permutation, bootstrap, expectation maximization, hidden Markov model, and Markov chain Monte Carlo. Rationales and techniques illustrated with existing implementations commonly used in population genetics research, disease association studies, and genomics analysis. Prerequisite: GENE 244 or consent of instructor.

2-3 units, Spr (Tang, H; Zhang, N)

STATS 351A. An Introduction to Random Matrix Theory
(Same as MATH 231A.) Patterns in the eigenvalue distribution of typical large matrices, which also show up in physics (energy distribution in scattering experiments), combinatorics (length of longest increasing subsequence), first passage percolation and number theory (zeros of the zeta function). Classical compact ensembles (random orthogonal matrices). The tools of determinental point processes.

3 units, Aut (Diaconis, P)

STATS 352. Spatial Statistics
(Same as STATS 253.) Statistical descriptions of spatial variability, spatial random functions, grid models, spatial partitions, spatial sampling, linear and nonlinear interpolation and smoothing with error estimation, Bayes methods and pattern simulation from posterior distributions, multivariate spatial statistics, spatial classification, nonstationary spatial statistics, space-time statistics and estimation of time trends from monitoring data, spatial point patterns, models of attraction and repulsion. Applications to earth and environmental sciences, meteorology, astronomy, remote-sensing, ecology, materials.

3 units, Spr (Taylor, J)

STATS 362. Monte Carlo Sampling
Fundamentals of Monte Carlo methods. Generating uniform and nonuniform variables, random vectors and processes. Monte Carlo integration and variance reduction. Quasi-Monte Carlo sampling. Markov chain Monte Carlo, including Gibbs sampling and Metropolis-Hastings. Examples, problems and motivations from Bayesian statistics, computational finance, computer graphics, physics.

2-3 units, Aut (Owen, A)

STATS 366. Computational Biology
(Same as BIOMEDIN 366, STATS 166.) Methods to understand sequence alignments and phylogenetic trees built from molecular data, and general genetic data. Phylogenetic trees, median networks, microarray analysis, Bayesian statistics. Binary labeled trees as combinatorial objects, graphs, and networks. Distances between trees. Multivariate methods (PCA, CA, multidimensional scaling). Combining data, nonparametric inference. Algorithms used: branch and bound, dynamic programming, Markov chain approach to combinatorial optimization (simulated annealing, Markov chain Monte Carlo, approximate counting, exact tests). Software such as Matlab, Phylip, Seq-gen, Arlequin, Puzzle, Splitstree, XGobi.

2-3 units, Spr (Wong, W)

STATS 370. A Course in Bayesian Statistics
(Same as STATS 270.) Bayesian statistics including theory, applications, and computational tools. Topics: history of Bayesian methods, foundational problems (what is probability), subjective probability and coherence, exchangeability and deFinetti's theorem. Conjugate priors, Laplace approximations, Gibbs sampling, hierarchical and empirical Bayes, nonparametric methods, Dirichlet and Polya tree priors. Bayes robustness, asymptotic properties of Bayes procedures.

3 units, Win (Wong, W)

STATS 375. Inference in Graphical Models
Graphical models as a unifying framework for describing the statistical relationships between large sets of variables; computing the marginal distribution of one or a few such variables. Focus is on sparse graphical structures, low-complexity algorithms, and their analysis. Topics include: variational inference; message passing algorithms; belief propagation; generalized belief propagation; survey propagation. Analysis techniques: correlation decay; distributional recursions. Applications from engineering, computer science, and statistics. Prerequisite: EE 278, STATS 116, or CS 228. Recommended: EE 376A or STATS 217.

3 units, Win (Montanari, A)

STATS 390. Consulting Workshop
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.

1-3 units, Aut (Olshen, R), Win (Tibshirani, R), Spr (Owen, A), Sum (Staff)

STATS 398. Industrial Research for Statisticians
Doctoral research as in 298, but must be conducted for an off-campus employer. Final report required. May be repeated for credit. Prerequisite: Statistics Ph.D. candidate.

1-9 units, Aut (Staff), Win (Staff), Spr (Staff), Sum (Staff)

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

1-10 units, Aut (Staff), Win (Staff), Spr (Staff), Sum (Staff)