STATISTICS

Emeriti: Theodore W. Anderson, Albert H. Bowker, M. Vernon Johns, Lincoln E. Moses, Herbert Solomon, Charles Stein
Chair: Tze L. Lai
Professors: Thomas M. Cover, Amir Dembo, Persi Diaconis, David L. Donoho, Bradley Efron, Jerome H. Friedman, Trevor J. Hastie, Iain M. Johnstone, T. L. Lai, Ingram Olkin, Art Owen, Joseph P. Romano, David O. Siegmund, Paul Switzer
Associate Professors: Jun Liu, Guenther Walther
Assistant Professor: Jonathan Taylor
Associate Professor (Teaching): Susan Holmes
Courtesy Professors: Richard A. Olshen, Neil Risch, Robert Tibshirani
Courtesy Associate Professor: David Rogosa

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

Courses given in Statistics have the subject code STATS. For a complete list of subject codes, see Appendix B.

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), Genetics (Risch), Health Research and Policy (Brown, Efron, Hastie, Johnstone, Moses, Olshen, Tibshirani), Mathematics (Dembo, Diaconis), and the Stanford Linear Accelerator (Friedman). The research activities of the department reflect an interest in both 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 SGI Challenge networked to approximately 40 X-terminals 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

MAJOR

Students wishing to build a concentration in probability and statistics are encouraged to consider declaring a major in Mathematical and Computational Sciences. This interdepartmental program is administered in the Department of Statistics (Bradley Efron, chair) and provides a 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.

MINORS

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.

1. Qualifying Courses: at most two of the following courses may be counted toward the six course requirement for the minor: MATH 52; STATS 191.
2. Required Courses: STATS 116 and 200.
3. 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, BIOSCI 283, emphasizing genetic and ecologic modeling
STATS 217, 218, emphasizing probability and its applications
STATS 240, 250, emphasizing mathematical finance

GRADUATE PROGRAMS

MASTER OF SCIENCE

The department requires that the student take 40 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.

Each student should fulfill the following requirements for the M.S. degree:

1. STATS 116, 191, 200, and 217. Courses previously taken may be waived by the adviser, in which case they must be replaced by other graduate courses offered by the department.
2. One of MATH 103, 113, 115, 171; and one of CS 106X (3 units), 137 (3 units), 138A. Substitution of other courses in Mathematics and Computer Science may be made with consent of the adviser.
3. At least four additional courses from graduate offerings in the department (202-399). Consent of the adviser is required in order to take more than 6 units of STATS 260, 390, or 399.
4. Additional units to complete the requirements may be chosen from the list available from the department. 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 social sciences.

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

All statistics courses required for the M.S. degree (116, 200, 217, and three additional statistics graduate courses) must be taken for letter grades, and an overall 2.75 grade point average (GPA) is required.

DOCTOR OF PHILOSOPHY

The department looks for motivated students who want 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 considered for admission. The program normally takes four years.

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 examination (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.

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 measure-theoretic probability theory, beginning with the basic concepts of analysis. 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 PhD program advisor.

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 given before the beginning of the Autumn Quarter of the student's second year. 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 PhD program advisor.

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/BIOSCI 203, as well as 9 units of classes in Statistical Genetics or Bioinformatics, GENE 344A and B, 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, 353, MS&E 323
Queuing Theory: MS&E 335, GSB 661, 663
Statistical Mechanics: STATS 365
Stochastic Processes: STATS 317, MATH 234

Earth Science Statistics—Students are expected to take:
STATS 317 (Stochastic Processes)
STATS 318 (Monte Carlo Markov Chains)
STATS 352 (Spatial Statistics)
and three courses from the GES or Geophysics departments, such as GES 290 (Geological Time Series), GES 296 (Geographic Information Systems), GEOPHYS 210 (Earth Imaging).

Social and Behavioral Sciences—Students are expected to take three advanced courses from the department with an applied orientation such as:
STATS 261/262 (Discrete Data and Survival Analysis)
STATS 324 (Multivariate Analysis)
STATS 343 (Time Series)
STATS 354 (Bootstrap and Resampling)
and three advanced quantitative courses from departments such as Psychology, Sociology, Political Science, Anthropology, Economics, and 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 relates directly to the student's presentation and 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 is continued for four 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

The Department of Statistics will devise individual Ph.D. minor programs, but the department urges all graduate students in other fields who wish to have a subspeciality in statistics to 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.

COURSES

INTRODUCTORY

Introductory courses for general students with an interest in the problems of descriptive statistics and statistical inferences are STATS 30, 40N, 50N, 60, 141. These courses have no mathematical prerequisites. STATS 30, 40N, 41N, 43N, 44N, and 141 are certified to meet the General Education Requirement in mathematics for undergraduates. STATS 40N, 41, 42N, 43N, 44N, 45N, and 50N are Stanford Introductory Seminars offering introductions to particular topics in a small group format with a preference to freshmen. STATS 60 and 141 explain the techniques and methods of statistical inference. STATS 60 emphasizes applications in the social sciences and STATS 141 applications in the biological sciences. STATS 60 and 141 can be followed by STATS 191, which explains more advanced methods and their applications.

STATS 110, 116, 200, 217-218 are introductory but have a calculus prerequisite. STATS 110 covers the most important techniques used in the analysis of experimental data in engineering and science. STATS 110 can be followed by STATS 191. STATS 116 provides a general introduction to the theory of probability. It may be followed by STATS 200, which deals with statistical theory, or by 217 and 218, which deal with stochastic processes. The sequence 116, 200 is a basic two quarter sequence in mathematical statistics; the sequence 116, 217, 218 is a basic one-year course in probability theory.

STATS 30. Statistical Thinking—Overview of statistical inference, presented with a minimum of mathematical formulation. Topics: comparisons and the randomized clinical trial, statistical significance, accuracy and the meaning of statistical error (plus or minus), correlation and regression to the mean, exploratory methods and data mining, life tables and survival analysis, and learning from experience (Bayesian inference). Lectures are supplemented with web-based statistical simulations. GER:2c

3 units (Staff) not given 2002-03

STATS 40N. Chance, Experiments, and Inference—Stanford Introductory Seminar. Preference to freshmen. The role of probability and statistics in understanding chance phenomena in an uncertain and unpredictable world. Goal: expose students to the range of real-world applications of probability and statistics, to read newspaper and journal articles with critical thinking, and to learn some simple "back of the envelope" calculations to interpret data. Applications: statistics in court cases, randomized clinical trials and assessing the efficacy of new drugs; chance and strategy in sports; paradoxes in probability and statistics; predicting the stock market and the random walk hypothesis; analysis of ESP experiments. GER:2c

3 units, not given 2002-03

STATS 41N. News and Numbers: Interpreting Information—Stanford Introductory Seminar. Preference to freshmen. Data reporting in newspaper and magazine accounts often leads to misinterpretations and erroneous conclusions. Goal is to introduce the basic statistical tools needed to critically interpret reported data. Applications from medicine, law, sports, parapsychology, business. GER:2c

3 units, not given 2002-03

STATS 43N. Displaying Data: Principles, Computer Graphics, and the Internet—Stanford Introductory Seminar. What is a good way to depict data and information and why? The principles for displaying data and envisioning information by reviewing historical graphical practice and recent scientific literature. How these principles, which have evolved with the printed medium in mind, apply to new media such as the Internet. Student project and presentation. No prerequisites. GER:2c

3 units, Win (Walther)

STATS 44N. The Pleasures of Counting—Stanford Introductory Seminar. Preference to freshmen; preference to students with either AP calculus or AP statistics, or equivalent. The interplay between applied mathematics and the world around it through a tour of celebrated topics in statistics and mathematics such as John Snow, graphic display and cholera, sorting algorithms, the census, and random matrices. Computational experimentation such as MATLAB. GER:2c

3 units, not given 2002-03

STATS 45N. Our Fractal World?—Stanford Introductory Seminar. Preference to freshmen. Over the last 30 years, mathematicians, physicists, and other scientists have claimed evidence of patterns such as fractals and multifractals throughout nature, and even in non-natural phenomena such as finance. Such patterns involving infinitely repeated geometric structures are beautiful to look at and fun to learn about, but the claims of the proponents of these ideas have invited backlash. The evidence for fractal-like behavior in different fields and why these ideas can be so attractive to proponents and yet invite reactions.

3 units (Staff) not given 2002-03

STATS 46N. Experiments in Extrasensory Perception: A Critical Analysis—Stanford Introductory Seminar. The history of ESP experimentation both at Stanford and elsewhere. Principles of experimental design, randomization, experimental control and confounding, response modeling, and calculations. Class designs and executes ESP experiments. Students prepare literature reviews, probability calculations, critiques, experiments, and oral and written presentations. GER:2c

3 units, Spr (Switzer)

STATS 47N. Breaking the Code?—Stanford Introductory Seminar. Cryptography has long been used for secret communications, and more recently to secure communication on the Internet. Its counterpart, cryptanalysis or code breaking, has been around almost as long as cryptography. Early cryptanalysts used statistical tools to decrypt messages by uncovering patterns. More recently, such frequency analysis tools have been used to analyze biblical texts and to detect genes in the human genome. Overview of codes and ciphers. Basic statistical tools useful in code breaking. Using computer programs, students apply these tools to break simple codes and explore applications to data. No prerequisites. GER:2c

3 units, Aut (Taylor)

STATS 50N. Mathematics in Sports—Stanford Introductory Seminar. Preference to freshmen. The mathematical and physical foundations of various sports are developed to provide new statistics, interpret old statistics, and suggest new physical and strategic approaches. Extremes are examined to find the optimum. Some game theory and assessment of odds. The extent to which all sports are equally exciting. Skill versus luck. The mathematics are followed as necessary to reach the desired conclusions. Recommended: mathematical aptitude.

3 units (Staff) not given 2002-03

STATS 60. Introduction to Statistical Methods: Precalculus—(Same as 160, PSYCH 10.) Emphasis is on 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:2c

5 units, Aut (Walther), Win (Thomas), Spr (Switzer), 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. Prerequisites: one year of calculus. GER:2c

4-5 units, Aut (Zerner), 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. Prerequisite: MATH 52 and some familiarity with infinite series, or equivalent. GER:2c

3-5 units, Aut (Taylor), Spr (Donoho), Sum (Staff)

STATS 141. Biostatistics—(Same as BIOSCI 141.) Introduction to the statistical analysis of biological data. Topics: discrete and continuous distributions, testing hypotheses and confidence procedures, fixed and random effects analysis of variance, regression, and correlation. Wilcoxon and other nonparametric procedures, inference on contingency tables and other data arising from counts. Tests of goodness of fit. Emphasis is on finding numerical solutions to biostatistical problems, and practical interpretations and their implications. GER:2c

4-5 units, Aut (Holmes)

STATS 160. Introduction to Statistical Methods: Precalculus—(Same as 60; see 60.)

5 units, Aut (Walther), Win (Thomas), Spr (Switzer), Sum (Staff)

STATS 166. Statistical Methods in Computational Genetics—The computational methods necessary for the construction and evaluation of sequence alignments and phylogenies built from molecular data and genetic data such as micro-arrays and data base searches. How to formulate biological problems in an algorithmic decomposed form, and building blocks common to many problems such as Markovian models, multivariate analyses. Some software covered in labs (Python, Biopython, XGobi, MrBayes, HMMER, Probe). Prerequisites: knowledge of probability equivalent to STATS 116, and one class in computing at the CS 106 level.

3 units, Aut (Holmes)

STATS 191. Introduction to Applied Statistics—Statistical tools for modern data analysis. Topics: regression and prediction, elements of the analysis of variance, bootstrap, and cross-validation. Emphasis is on conceptual rather than theoretical understanding. Student projects require extensive computer use. Recommended: 60, 110, or 141. GER:2c.
3-4 units, Spr (Taylor), Sum (Staff)

STATS 199. Independent Study—For undergraduates.
1-15 units (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)

CONTINUATION

Courses in this category have been designed for particular use in applications. Generally, they have introductory statistics or probability as prerequisites.

STATS 202. Data 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 data bases or through web mining. Topics: decision trees, neural networks, association rules, clustering, case based methods, and data visualization.
3 units, Aut (Friedman)

STATS 203. Introduction to Regression Models and Analysis of Variance—The most widely used statistical techniques; interpretation of observational data and empirical model building. Topics: simple and multiple linear regression, nonlinear regression, analysis of residuals and model selection, design of one-way and two-way factorial experiments, fixed and random effects models. Prerequisite or corequisite: 200.
3 units, Win (Olkin)

STATS 205. Introduction to Nonparametric Statistics—Nonparametric analogs of the one- and two-sample *t* tests and analysis of variance; the sign test, median test, Wilcoxon's tests, and the Kruskal-Wallis and Friedman tests, tests of independence. Nonparametric regression and nonparametric density estimation, modern nonparametric techniques, nonparametric confidence interval estimates.
3 units, Win (Romano)

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. Prerequisite: 200; concurrent registration in 200 is permitted.
3 units, not given 2002-03

STATS 207. Introduction to Time Series Analysis—Time series models used in economics, engineering, physics, geology, etc. Trend fitting, autoregressive schemes, moving average models, periodograms, second order stationary processes, spectral analysis. Prerequisites: 116 and a basic course in statistics at the level of 200.
3 units, not given 2002-03

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: at least one course in statistics or probability.
3 units, not given 2002-03

STATS 211. Statistical Methods in Meta-Analysis—Meta-analysis is a quantitative method for combining results of independent studies, and enables researchers to synthesize the results of related studies. Examples from the medical, behavioral, and social sciences. Topics: literature search, publication and selection bias, statistical methods (contingency tables, cumulative methods, sensitivity analyses, non-parametric methods). Project. Prerequisite: basic sequence in Statistics.
3 units, Win (Olkin)

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. No previous knowledge of SAS is required. Knowledge of statistical techniques at the level of Stats 191 or 203 is assumed.
3 units, Sum (Staff)

STATS 215. Statistical Methods in Computational Genetics—Computational methods necessary to understand and construct sequence alignments and phylogenies built from molecular data and the exploration of more general genetic data such as micro-arrays and database searches. How to formulate biological problems in an algorithmic decomposed form, and useful building blocks common to problems such as Markovian models and multivariate analyses. Some software covered in labs.
3 units, Spr (Siegmund)

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.
3 units, Win (Siegmund)

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

STATS 227. Statistical Computing—Numerical aspects of least squares, nonlinear, and robust regression. Eigenvector-eigenvalue computations and analyses. Monte Carlo methods: generation of uniformly distributed random numbers, generation of special distributions, variance reduction techniques. The complexity of algorithms used in statistics: sorting, computation of quantiles, nearest neighbor search, fast Fourier transform. Prerequisites: Statistics at the level of 200, matrix algebra, knowledge of a programming language.
3 units, not given 2002-03

STATS 230. Experimental Design—(Same as 340.) For graduate students in science, engineering, and statistics. Emphasis is on the how and why of doing experiments, and analyzing and presenting the results. Topics: control groups, anova, blocking and balance, factorial experiments, fractional factorials, screening designs, response surfaces, binary outcomes, Taguchi methods, computer experiments. Prerequisite: 116. Recommended: experience with experimentation or data analysis.
3 units, Spr (Staff)

STATS 240. Statistical Methods in Finance—Regression analysis and applications to the Capital Asset Pricing Model and multifactor pricing models. Smoothing techniques and estimation of yield curves. Classification and credit risk. Statistical analysis and econometric modeling of financial time series. Forecasting. Problem sets, hands-on experience with financial data.
3-4 units, Spr (Lai)

STATS 250. Mathematical Finance—(Same as MATH 241.) 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.
3 units, Win (Papanicolaou)

STATS 261. Intermediate Biostatistics: Analysis of Discrete Data—(Same as HRP 261, BIOMEDIN 233.) The 2x2 table. Chi-square test. Fisher's exact test. Odds ratios. Sampling plans; case control and cohort studies. Series of 2x2 tables. Mantel Hantzel. Other tests. *k x m* tables. Matched data logistic models. Conditional logistic analysis, application to case-control data. Log-linear models. Generalized estimating equations for longitudinal data. Cell phones and car crashes: the crossover design. Special topics: generalized additive models, classification trees, bootstrap inference.

3 units, Win (Olshen)

STATS 262. Intermediate Biostatistics: Regression, Prediction, Survival Analysis—(Same as HRP 262.) Linear and inherently nonlinear models. Prediction versus testing. Sample reuse methods. Analysis of variance. Components of variance. Introduction to multivariate analysis: the normal distribution. Principle components and *k*-means clustering. Survival analysis: the actuarial and Kaplan-Meier methods. The log-rank test. Weibull models. The Cox model, including estimation of baseline hazard.

3 units, Spr (Taylor)

STATS 270. A Course in Bayesian Statistics—(Same as 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 de-Finetti'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 (Diaconis)

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, Win, Spr, Sum (Staff)

PRIMARILY FOR DOCTORAL STUDENTS

Sequences 300A,B,C, 305, 306A,B, and 310A,B,C comprise the fundamental sequence which serves as a general introduction to and prerequisite for further work. Subsequent courses delve more deeply into special topics.

STATS 260A,B,C. Workshop in Biostatistics—(Same as HRP 260A,B,C.) Applications of statistical techniques to current problems in medical science. Enrollment for more than 2 units of credit involves extra reading or consulting and requires consent of the instructor.

260A: *1-5 units, Aut (Lazzeroni, Olshen, Bloch, Efron, Hastie, Lavori, Tibshirani)*

260B: *1-2 units, Win (Lazzeroni, Olshen, Bloch, Efron, Hastie, Lavori, Tibshirani)*

260C: *1-5 units, Spr (Lazzeroni, Olshen, Bloch, Efron, Hastie, Lavori, Tibshirani)*

STATS 300. Advanced Topics in Statistics

3 units, Sum (Staff)

STATS 300A,B,C. 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.

2-3 units, **A:** *Aut (Romano),* **B:** *Win (Donoho),* **C:** *Spr (Walther)*

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. Prerequisite: consent of instructor, 116, 200, one applied statistics course, Computer Science 106A, Mathematics 114.

2-4 units, Aut (Hastie)

STATS 306A,B. Methods for Applied Statistics—Survey of applied statistical methods, including computational methods. Topics: nonlinear least squares (including robust regression), generalized linear models, time series (autocorrelation, autoregression, periodiogram, spectrum), survey sampling (finite populations, stratification, clustering, ratio estimation), nonparametric regression (kernels, splines, projection pursuit, CART, MARS), survival analysis (Kaplan-Meier, Mantel-Haenszel, Cox model), design (factorial experiments, response surfaces), random number generation, numerical linear algebra, numerical optimization, sample reuse (bootstrap, jackknife, cross-validation, other Monte Carlo), matrix based multivariate statistics (canonical correlation, T-squared, factor analysis, principal components), and other topics briefly. Prerequisite: 305 or equivalent.

2-4 units, Win (Owen), Spr (Tibshirani)

STATS 310A. Theory of Probability—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.

3 units, Aut (Dembo)

STATS 310B. Theory of Probability—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.

2-3 units, Win (Siegmund)

STATS 310C. Theory of Probability—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.

2-3 units, Spr (Lai)

STATS 314. Advanced Statistical Methods—This year's topic is resampling and bootstrap methods.

3 units, Aut (Efron)

STATS 315A,B,C. Modern Applied Statistics: Learning—The rise in computing power has been accompanied by a rapid growth in statistical modeling and data analysis. New techniques have emerged for predictive and descriptive learning not possible 10 years ago using ideas that bridge the gaps among statistics, computer science, and artificial intelligence. Two-part series on these new methods emphasizes the 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 specific predictive goal. From a statistical perspective, it can be viewed as computer automated exploratory analysis of usually large complex data sets.

3 units, **A:** *Win (Hastie),* **B:** *Spr (Friedman),* **C:** *not given 2002-03*

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.

3 units, Aut (Diaconis)

STATS 319. Literature of Statistics—Literature study of topics in statistics and probability culminating in oral and written reports.

3 units, Aut (Staff), Win (Donoho), Spr (Walther)

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.

3 units, not given 2002-03

STATS 324. Multivariate Analysis—General theory of multivariate distributions; multivariate normal distribution and related distributions: Wishart distribution and Hotelling's T-squared. Tests for means, variances and covariances, multivariate analysis of variance, multiple regression, principal components, canonical correlations. General problems of classification and clustering of high dimensional data.

3 units, not given 2002-03

STATS 325. Analysis of Multivariate and Functional Data—Classical and modern methods in multivariate and functional settings: principal components analysis and linear modeling. Semiparametric models and smoothing. Regularization. Dimension reduction. Selective theoretical issues and applications and computing.

3 units, not given 2002-03

STATS 326. Sequential Experimentation—Sequential statistical decision problems, dynamic programming and optimal stopping, quality control and changepoint detection, dynamic allocation and bandit problems, stochastic approximation, sequential testing and estimation and boundary crossing probabilities.

3 units (Staff) not given 2002-03

STATS 332. Asymptotic Methods in Statistics—Concepts of efficiency, the asymptotic efficiency of maximum likelihood estimators, best asymptotically normal (BAN) estimators, asymptotic behavior of likelihood ratio tests, optimal designs, empirical Bayes methods.

3 units, not given 2002-03

STATS 340. Experimental Design—(Same as 230; see 230.)

3 units, Spr (Staff)

STATS 344A. Genetic Epidemiology—(Same as GENE 344A.) Methods for the design and analysis of studies in human genetics, focusing on the epidemiology of Mendelian disorders, and the genetic and environmental contributions to common, complex familial traits. Topics: study designs for assessing the importance of genetic factors (family, twin, and adoption studies); methods for determining modes of inheritance (segregation analysis); identification and mapping of major genes through linkage analysis and disease marker associations.

3 units (Risch) alternate years, given 2003-04

STATS 344B. Topics in Statistical Genetics—(Same as GENE 344B.) Statistical methods currently used in human genetic analysis. Topics depend on interests of the students and instructors: concepts of likelihood as used in the genetic context; measures of familial aggregation, including issues of censoring and age-dependent data; genetic modeling of quantitative traits; mode of inheritance analysis, including segregation analysis; analysis of extended pedigrees; parametric and nonparametric approaches to linkage analysis and gene mapping, including family studies; linkage disequilibrium; analysis of DNA profiles for individual identification.

3 units (Risch) alternate years, given 2003-04

STATS 350. Advanced Topics in Probability Theory—Current research topic(s) in probability theory, chosen to reflect the interests of the students and instructor. Possible topics: Brownian motion, course graining, concentration inequalities, discrete probability, Gibbs measures, interacting particle systems, large deviations, percolation, random matrices, Stein's method.

3 units, not given 2002-03

STATS 352. Spatial Statistics—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, not given 2002-03

STATS 354. Bootstrap and Resampling Methodology—Resampling methods in statistics: bootstrap, jackknife, cross-validation, permutation and randomization techniques. General asymptotic theory. Relevant mathematical tools, Edgeworth expansions and empirical processes; methods are developed. Survey of recent literature.

3 units, Aut (Efron)

STATS 362. Monte Carlo Sampling and Computer Experiments—Uniform random number generation and testing. Generating nonuniform random numbers, random rotations, random permutations, Brownian and geometric Brownian motions. Monte Carlo integration. Variance reduction by stratification, antithetic and cluster sampling, control variates, importance sampling, defensive mixture importance sampling. Output analysis. Latin hypercube sampling. Randomized orthogonal arrays. ANOVA decomposition of functions. Quasi-Monte Carlo methods. Lattice rules. (t,m,s)-nets. Koksma-Hlawka inequality. Discrepancy measures. Randomized QMC. Effective dimension and reduction thereof. Applications to finance. Computer experiments. Kriging. Choice of covariance. Design of input points. Numerical noise. Missing data. Regression and quasi-regressions methods. Visualization. Error estimation. Prerequisite: 116 or equivalent, and programming equivalent.

2-3 units, Aut (Owen)

STATS 365. Statistical Mechanics—Mathematical introduction to the basic findings of statistical mechanics and the development of tools and exponential families. Introduction to thermodynamics, Broadwell models, Boltzman's equation, entropy and the H-theorem. Gibbs states, Ising model, equivalence of ensembles, phase transitions, low density extensions, application to parameter estimation in high dimensional models.

3 units, not given 2002-03

STATS 366. Computational Biology—For biologists, applied mathematicians, statisticians, or computer scientists. The computational methods necessary to understand the construction and evaluation of sequence alignments and phylogenetic trees built from molecular data, and general genetic data such as micro-arrays and database search results. 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: most methods exist, in high level programming environments, e.g., Matlab. Specialized software: Phylip, Seq-gen, Arlequin, Puzzle, Splitstree, XGobi.

2-3 units, Win (Holmes)

STATS 367. Statistical Models in Genetics—Stochastic models and related statistical problems in linkage analysis of qualitative and quantitative traits in humans and experimental populations; sequence alignment and analysis; and population genetics/evolution, both classical (Wright-Fisher-Kimura) and modern (Kingman coalescent). Computational algorithms as applications of dynamic programming, Markov chain Monte Carlo, and hidden Markov models. Prerequisites: knowledge of probability through elementary stochastic processes and statistics through likelihood theory.
3 units, Spr (Siegmund)

STATS 370. A Course in Bayesian Statistics—(Same as 270; see 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 (Diaconis)

STATS 374. Large Deviations—(Same as MATH 234.) 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: 230A or STATS 310.
3 units, not given 2002-03

STATS 376A,B. Information Theory—(Enroll in EE 376A,B.)
3 units, **A:** *Win (Cover),* **B:** *Spr (Cover) alternate years, not given 2003-04*

STATS 390. Consulting Workshop—Provides the skills required of practicing statistical consultants, exposure to wide range of 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. Prerequisites: course work in applied statistics or data analysis, and consent of the instructor.
1-3 units, Aut (Owen), Win (Owen), Spr (Rogosa)

STATS 398. Industrial Research for Statisticians—Doctoral research as in 298, but must be conducted for an off-campus employer. Final report required. Prerequisite: Statistics Ph.D. candidate.
1-9 units, Aut, Win, Spr, Sum (Staff)

STATS 399. Research—Research work as distinguished from independent study of nonresearch character listed in 199.
1-10 units (Staff)

STATS 459. Frontiers in Interdisciplinary Biosciences—(Crosslisted in multiple departments in the schools of Humanities and Sciences, Engineering, and Medicine; students should enroll directly through their affiliated department, otherwise enroll in CHEMENG 459.) An introduction to cutting-edge research involving interdisciplinary approaches to bioscience and biotechnology; for specialists and non-specialists. Organized and sponsored by the Stanford BioX Program. Three seminars each quarter address a broad set of scientific and technical themes related to interdisciplinary approaches to important issues in bioengineering, medicine, and the chemical, physical, and biological sciences. Leading investigators from Stanford and throughout the world present the latest breakthroughs and endeavors that cut broadly across many core disciplines. Pre-seminars introduce basic concepts and provide background for non-experts. Registered students attend all pre-seminars in advance of the primary seminars, others welcome. Prerequisite: keen interest in all of science, engineering, and medicine with particular interest in life itself. Recommended: basic knowledge of mathematics, biology, chemistry, and physics.
1 unit, Aut, Win, Spr (Robertson)