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, 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: Amir Dembo, Jun Liu, Guenther Walther
Assistant Professor: Jonathan Taylor
Courtesy Professors: Richard A. Olshen, Neil Risch, Robert Tibshirani
Courtesy Associate Professor: David Rogosa
Senior Lecturer: Susan Holmes

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. Furthermore, 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: Mathematics. 52; Statistics 191.
2. Required Courses: Statistics 116 and 200.
3. Elective Courses: at least one of the elective courses should be a Statistics 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:

Statistics 202, 203 (204), emphasizing data analysis and applied statistics
Statistics 205, 206 (207), emphasizing statistical methodology
Statistics 206; Economics 160 (181), emphasizing economic optimization
Statistics 206; Psychology 156 (160), emphasizing psychology modeling and experiments
Statistics 207; Electrical Engineering 264 (279), emphasizing signal processing
Statistics 217; Biological Sciences 283, emphasizing genetic and ecologic modeling
Statistics 217, 218, emphasizing probability and its applications
Statistics 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. Statistics 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 Mathematics 103, 113, 115, 171; and one of Computer Science 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 Statistics 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—Statistics 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 (courses numbered above 310, not including literature, research, or consulting), and 3 units of statistical consulting.

First-Year Core Courses—Statistics 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 probability theory, beginning with the basic concepts of analysis.

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 methods, mathematical statistics, and probability theory, and all are given at 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—In order to appreciate scientific problems, students are required to take 9 units of course work at a graduate or advanced undergraduate level in some other department. These units must be in courses higher than 200. Students with a graduate degree in a scientific area that is not essentially mathematics or statistics are exempted from this requirement.

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 Statistics 30, 40N, 50N, 60, 141. These courses have no mathematical prerequisites. Statistics 30, 40N, 41N, 43N, 44N, and 141 are certified to meet the General Education Requirement in mathematics for undergraduates. Statistics 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. Statistics 60 and 141 explain the techniques and methods of statistical inference. Statistics 60 emphasizes applications in the social sciences and Statistics 141 applications in the biological sciences. Statistics 60 and 141 can be followed by Statistics 191, which explains more advanced methods and their applications.

Statistics 110, 116, 200, 217-218 are introductory but have a calculus prerequisite. Statistics 110 covers the most important techniques used in the analysis of experimental data in engineering and science. Statistics 110 can be followed by Statistics 191. Statistics 116 provides a general introduction to the theory of probability. It may be followed by Statistics 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.

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, not given 2001-02

40N. Stanford Introductory Seminar: Chance, Experiments, and Interface—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 2001-02

41N. Stanford Introductory Seminar: News and Numbers—Interpreting Information—Preference to freshmen. Data reporting in newspaper and magazine accounts often leads to misinterpretations and erroneous conclusions. Goal: introduce the basic statistical tools needed to critically interpret reported data. Applications from medicine, law, sports, parapsychology, business, etc. GER:2c
3 units, not given 2001-02

42N. Stanford Introductory Seminar: Monte Carlo Methods—From Molecular Structure to Dynamic Systems—Preference to freshmen. Computer simulation techniques (Monte Carlo methods) appeared early (1945-55) in electronic computing. Statistical physicists introduced the Markov Chain-based, dynamic Monte Carlo method for the simulation of simple fluids. 1980s statisticians and computer scientists developed the method for varieties of Bayesian computation tasks. In the '90s, the method played a role in computational biology, in sequence motif identification, and the analysis of complex pedigrees. Application areas of Monte Carlo methods include biology, chemistry, computer science, economics and finance, engineering, material science, physics, and statistics. The basic principle of the Monte Carlo simulation method and the power of the method; applications from polymer simulations to target tracking.
3 units, not given 2001-02

43N. Stanford Introductory Seminar: Displaying Data—Principles, Computer Graphics, and the Internet—Preference to freshmen. Based on literature and historical examples. Principles for displaying data and envisioning information. What is a good way to depict data and information, and why? How do these principles apply to or have to be modified for media such as computer graphics and the Internet? Student project. GER:2c
3 units, Win (Walther)

44N. Stanford Introductory Seminar: The Pleasures of Counting—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, e.g., John Snow, graphic display and cholera; sorting algorithms; the census, random matrices, etc. Computational experimentation, e.g., MATLAB, is encouraged. GER:2c

3 units, not given 2001-02

45N. Stanford Introductory Seminar: Our Fractal World?—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 like 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. Books for the nonmathematician audience explain concepts like fractals power laws, and self-similarity. 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, Aut (Donoho)

50N. Stanford Introductory Seminar: Mathematics in Sports—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 vs. luck. The mathematics are followed as necessary to reach the desired conclusions. Recommended: mathematical aptitude.

3 units, not given 2001-02

60. Introduction to Statistical Methods: Precalculus—(Graduate students register for 160; same as Psychology 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)

105Q. Stanford Introductory Seminar: Statistical Design of Experiments in Extrasensory Perception—Preference to sophomores. The principle of good and bad experimental designs and how experimental data is used to reach conclusions. Randomization, experimental control, efficient design, elementary probability, statistical inference, quantification of uncertainty.

3 units, not given 2001-02

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)

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: Mathematics 52 and some familiarity with infinite series, or equivalent. GER:2c

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

141. Biostatistics—(Same as Biological Sciences 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 (Johnstone)
Win (Feldman)

160. Introduction to Statistical Methods: Precalculus—See 60. For graduate students.

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

191. Introduction to Regression Analysis and 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 use of a computer. Recommended: Statistics 60, 110, or 141. GER:2c

3-4 units, Win, Spr (Taylor)

199. Independent Study—For undergraduates.

(Staff)

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.

3-4 units, Win (Siegmund)

CONTINUATION

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

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 data bases or through web mining. Topics: decision trees, neural networks, association rules, clustering, case based methods, and data visualization.

3 units, Aut (Friedman)

203. Introduction to Regression Models and the 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 effects and random effects models. Prerequisite: 200.

3 units, Spr (Walther)

204. Sampling—The issue in sampling is how best to take data, when you can choose where to sample it. Motivating examples include surveys, and sampling from data warehouses. Emphasis is on methods for finite populations. Topics: simple random sampling, stratified sampling, cluster sampling, rato and regression estimators, two stage sampling. Computing required. Prerequisite: 61, 110, or 200.

3 units, not given 2001-02

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 (Owen)

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, Win (Donoho)

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 2001-02

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, Win (Holmes)

211. Statistical Methods for Meta-Analysis—(Same as Education 493B, Health Research and Policy 206.) Meta-analysis is a quantitative method for combining results of independent studies, and enables researchers to synthesize the results of related studies so that the combined weight of evidence can be considered and applied. 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)

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)

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

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 or 201, matrix algebra, knowledge of a programming language.
3 units, Spr (Owen)

228. Probabilistic Models in Artificial Intelligence—(Enroll in Computer Science 228.)
3 units, Win (Koller)

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 real data.
3 units, Spr (Lai)

245. Computation and Simulation in Finance—(Enroll in Mathematics 240.)
3 units, Spr (Lee)

250. Mathematical Finance—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)

261. Intermediate Biostatistics: Analysis of Discrete Data—(Same as Health Research and Policy 261.) 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 (Tibshirani)

262. Regression, Prediction, Survival Analysis—(Same as Health Research and Policy 262.) Linear and inherently nonlinear models. Prediction vs. 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 (Owen)

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).
3 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.

260A,B,C. Workshop in Biostatistics—(Same as Health Research and Policy 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 (Olshen, Bloch, Efron, Hastie, Johnstone, Lavori, Lazzeroni, Tibshirani)*
260B. *1-5 units, Win (Olshen, Bloch, Efron, Hastie, Johnstone, Lavori, Lazzeroni, Tibshirani)*
260C. *1-5 units, Spr (Olshen, Bloch, Efron, Hastie, Johnstone, Lavori, Lazzeroni, Tibshirani)*

300A,B,C. Theory of Statistics—Elementary finite sample theory of point estimation: statistical models; sufficiency; applications to exponential families, group families, and nonparametric families; minimum risk unbiased estimation; minimum risk equivariant estimation; Cramér-Rao inequality. Elementary decision theory: loss and risk functions, Bayes estimation; minimax estimation; shrinkage estimators. Large sample estimation theory: asymptotic efficiency, maximum likelihood estimation, delta method, asymptotic distribution of quantiles and trimmed means, differentiability of statistical functionals robustness and influence. Hypothesis testing and confidence intervals: Neyman-Pearson theory; uniformly most powerful tests and uniformly most accurate confidence intervals for distributions with monotone likelihood ratio; systematic use of sufficiency and conditioning to eliminate nuisance parameters in exponential families; use of invariance to eliminate nuisance parameters in group families; asymptotic theory of likelihood ratio test; Pitman asymptotic efficiency; rank, permutation, and randomization tests; jackknife, bootstrap, and sample reuse methods. Density estimation: kernel density estimation; bias vs. variance tradeoff; choice of bandwidth and kernel. Time series: first- and second-order autoregressive processes; conditions for stationarity; use of maximum likelihood in time series with asymptotic theory. Possible topics: sequential analysis, optimal experimental design, empirical processes with applications to statistics, Edgeworth expansions with applications to statistics.

300A. *3-4 units, Aut (Romano)*
300B. *3-4 units, Win (Romano)*
300C. *3-4 units, Spr (Romano)*

305. Linear Models—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.

3-4 units, Aut (Hastie)

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.

306A. *3-4 units, Win (Donoho)*
306B. *3-4 units, Spr (Efron)*

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, Mathematics 171.

3-4 units, Aut (Diaconis)

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 Mathematics 230A.

3-4 units, Win (Dembo)

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 Mathematics 230B.

3-4 units, Spr (Lai)

315A,B. Modern Applied Statistics: Elements of Statistical Learning—The rise in computing power has been accompanied by a rapid growth in the areas of statistical modeling and data analysis. New techniques have emerged for predictive and descriptive learning not possible 10 years ago, using ideas that bridge the gap between statistics, computer science, and artificial intelligence. Two-part series on these new methods emphasizes the statistical aspects of their application and their 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. Regression topics: linear regression with recent advances to deal with large numbers of variables, smoothing techniques, additive models, projection pursuit, MARS, local regression, and neural networks. Classification topics: discriminant analysis, logistic regression, support vector machines, generalized additive models, decision trees, naive Bayes, mixture models, boosting, and nearest neighbor methods. 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. Topics: association rules, clustering, principal components, self-organizing maps, principal curves and surfaces, independent component analysis, multidimensional scaling, and visualization techniques.

315A. *3 units, Win (Hastie)*
315B. *3 units, Spr (Friedman)*

317. Point and Spatial Processes—History via life tables and renewal theory. Poisson processes on the line and general spaces, stationary point processes and the Palm-Kinchin equations. Finite point processes: generating functions and combinatorics. Introduction to random measures, cluster processes, infinite divisibility, and Cox processes. Spectral theory, Palm theory, applications to eigenvalues of random matrices.

3 units, not given 2001-02

318. Modern Markov Chains—Examples: the Gibbs sampler, metropolis, algorithm and hybrid Monte Carlo. Irreducibility, recurrence periodicity, Doeblins techniques, C-sets, Harris recurrence. Coupling, spectral techniques, bounds on convergence from bounds on the spectrum. Iterated function systems with application to image compression recurrence; Foster's criteria and applications to statistical decision theory. Introduction to control theory and nonlinear time series analysis. Prerequisite: graduate probability course.

3 units, not given 2001-02

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

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

322. Function Estimation in White Noise—Gaussian white noise model sequence space form. Hyperrectangles, quadratic convexity, and Pinsker's theorem. Minimax estimation on l_p 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 2001-02

323. Wavelets and Beyond, with Applications—New representations have been proposed for signals and images, e.g., wavelets, wavelet packets, and cosine packets. How these representations are constructed, comparing their special features. The range of potential applications (data compression, feature extraction, pattern classification), and newer representation tools, e.g., noiselets, brushlets, ridgelets, etc.

3 units, not given 2001-02

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, Win (Olkin)

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 2001-02

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, not given 2001-02

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 2001-02

338. Analysis of Lifetime Data—Survival distributions. Types of censored data. Parametric, non-parametric, and semiparametric estimation in various survival models. Life tables. The Kaplan-Meier and related estimators. Rank tests and comparison of two or more survival distributions. The proportional hazards and accelerated lifetime models for covariate data. Regression analysis of censored lifetime data. Applications.
3 units, not given 2001-02

340. Experimental Design—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, not given 2001-02

343. Time Series Analysis—Autoregressive and moving average models; estimation of parameters and tests. Serial correlation. Stationary stochastic processes, spectral densities, prediction. Estimation of spectral densities and covariance sequences, regression analysis.
3 units, not given 2001-02

344A. Genetic Epidemiology—(Same as Genetics 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. Applications to birth defects, coronary heart disease, psychiatry, neurology, cancer, and immunology.
3 units, Win (Risch) alternate years, not given 2002-03

344B. Topics in Statistical Genetics—(Same as Genetics 344B.) In-depth discussion of 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, radiation hybrid data, sperm typing, and DNA contig mapping; linkage disequilibrium; analysis of DNA profiles for individual identification; DNA sequence analysis.
3 units, Spr (Risch) alternate years, not given 2002-03

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 2001-02

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 2001-02

353. Monte Carlo Methods—Recent advances in iterative and noniterative Monte Carlo sampling methods involved in scientific and statistical computations. Basic iterative methods: the Metropolis-Hastings algorithm and the Gibbs sampler; noniterative methods mainly cover variations of importance and sampling. Topics: molecular dynamics, hybrid Monte Carlo (and other Hamiltonian-based methods), Langevin diffusion, simulated and parallel tempering, adaptive directional sampling, multigrid Monte Carlo and generalized Gibbs sampling, multiple-try Metropolis, conjugate-gradient MC, growth Monte Carlo, and Monte Carlo filtering methods. Applications to Bayesian computations, computational biology, computer chip design, digital signal processing, polymer simulation, statistical physics, and target tracking.
3 units, not given 2001-02

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, not given 2001-02

362. Monte Carlo Methods for Integration and Approximation—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.
3 units, Aut (Owen)

365. Statistical Mechanics for Statisticians and Probabilists—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, Win (Diaconis)

366. Computational Biology—For biologists and for applied mathematicians, statisticians, or computer scientists interested in computational biology. The computational methods necessary to understand the construction and evaluation of sequence alignments and phylogenetic trees built from molecular data, and general genetic data, e.g., micro-arrays and database search results. Topics: 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.
3 units, Aut (Holmes)

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 (classical, i.e., Wright-Fisher-Kimura and modern. i.e., 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)

371. Bayesian Modeling and Computations—Bayesian methods treat unknowns as random variables and are coherent and flexible. Basic Bayesian models, whose answers often appear similar to classical answers. Complicated hierarchical and mixture models with nonstandard solutions. Methods for model checking, sensitivity analysis, and predictions. Emphasis is on drawing inferences via computer simulation. Mathematical analysis discussion.
3 units, not given 2001-02

372. Graphical Models and Bayesian Network—Using graphs to represent statistical models (the associational and casual relations between random variables under study) enables concise representations, easy interpretation, and computationally feasible incorporation of new information. Emphasis is on Bayesian inference using these graphical models, e.g., Bayesian networks. Topics: basic graph theory, graphical Gaussian models, log linear models, expert systems, missing data problems, and genetic modeling. Computational issues and their applications to artificial intelligence.
3 units, not given 2001-02

374. Large Deviations—(Same as Mathematics 234.) Combinatorial estimates and the method of types. Large deviation probabilities for partial sums and for empirical distributions, Cramêr's and Sanov's theorems and their Markov extensions. Application in statistics, information theory, and statistical mechanics. Prerequisite: 310 or Mathematics 230A.
3 units, not given 2001-02

376A. Information Theory—(Same as Electrical Engineering 376A.) Information theory and statistics. The extreme points of communication theory: data compression to the entropy limit, and communication at the channel capacity limit. Kolmogorov complexity, Shannon entropy. Rate distortion theory. Huffman coding and random coding. Unified treatment based on the asymptotic equipartition theorem. Prerequisite: 116, or Electrical Engineering 278, or equivalent.
3 units, Win (Gill)

376B. Information Theory—(Enroll in Electrical Engineering 376B.)
3 units, alternate years, given 2002-03

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.
3 units, Aut (Holmes)
Win (Switzer)
Spr (Johnstone)

398. Industrial Research for Statisticians—Doctoral research as in 199, 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)

399. Research—Research work as distinguished from independent study of nonresearch character listed in 199.
(Staff)

This file has been excerpted from the *Stanford Bulletin, 2001-02*, pages 631-636. Every effort has been made to ensure accuracy; late changes (after print publication of the bulletin) may have been made here. Contact the editor of the *Stanford Bulletin* via email at arod@stanford.edu with changes, corrections, updates, etc.