Statistics

Director of undergraduate studies: Andrew Barron, 24 Hillhouse Ave., 432-0634, andrew.barron@yale.edu; statistics.yale.edu

FACULTY OF THE DEPARTMENT OF STATISTICS

Professors † Donald Andrews, Andrew Barron, Joseph Chang, John Hartigan (Emeritus), † Theodore Holford, † Peter Phillips, David Pollard, † Heping Zhang, † Hongyu Zhao, Huibin Zhou

Associate Professors John Emerson (Adjunct), † Sekhar Tatikonda

Assistant Professors Jessi Cisewski, Sahand Negahban

Senior Lecturer Jonathan Reuning-Scherer

† A joint appointment with primary affiliation in another department or school.

Statistics is the science and art of prediction and explanation. The mathematical foundation of statistics lies in the theory of probability, which is applied to problems of making inferences and decisions under uncertainty. Practical statistical analysis also uses a variety of computational techniques, methods of visualizing and exploring data, methods of seeking and establishing structure and trends in data, and a mode of questioning and reasoning that quantifies uncertainty.

The Statistics program at Yale is a blend of the mathematical theory of probability and statistical inference, the philosophy of inference under uncertainty, computational techniques, the practice of data analysis, and statistical analysis applied to economics, biology, medicine, engineering, and other areas. Statistical methods are widely used in the sciences, medicine, industry, business, and government; graduates can work in these areas or go on to graduate study.

The curriculum for the Statistics major is a synthesis of theory, methods, and applications. The requirements are designed to achieve balance and depth in each of the three directions of probability, statistics, and data analysis. Statistics can be taken either as a primary major or as one of two majors, in consultation with the director of undergraduate studies. Appropriate majors to combine with Statistics include programs in the social sciences, natural sciences, engineering, computer science, or mathematics. A Statistics concentration is also available within the Applied Mathematics major.

Prerequisites Multivariable calculus and linear algebra are required and should be taken before or during the sophomore year. This requirement may be satisfied by MATH 120 and MATH 222 or 225, or equivalents.

Requirements of the major for the B.A. degree program The B.A. degree program requires ten term courses beyond the prerequisites, including the senior project. Majors take two courses in the theory and applications of probability (STAT 241 and 251), two courses emphasizing the theory of statistical inference (STAT 242 and 312), and two courses in the methods and practice of data analysis, chosen from STAT 230, 361, and 363. STAT 238 may be substituted for STAT 241 with the permission of the director of undergraduate studies. All majors are also required to take a course in computing (ENAS 130 or CPSC 112). The two remaining courses are electives chosen from Statistics courses numbered above 200. Appropriate courses in other departments or in the Graduate School may count toward the major with permission of the director of undergraduate studies.

Requirements of the major for the B.S. degree program The B.S. degree program requires twelve term courses beyond the prerequisites. In addition to the courses indicated for the B.A. degree, the B.S. degree requires a course in mathematical analysis (MATH 260, 300, or 301) and an additional Statistics elective numbered above 200.

Senior requirement In the senior year, majors in both degree programs complete a research project in STAT 490. Students enrolled in this course work on a research project under the supervision of a faculty member, present and share their progress with each other during the seminar meetings, and write a final report.

Credit/D/Fail For students in the Class of 2017 and subsequent classes, a maximum of one course taken Credit/D/Fail may be counted toward the requirements of the major, with permission of the director of undergraduate studies.

REQUIREMENTS OF THE MAJOR

Prerequisites Both degrees — MATH 120 and MATH 222 or 225, or equivalents

Number of courses B.A. — 10 term courses beyond prereqs (incl senior project); B.S. — 12 term courses beyond prereqs (incl senior project)

Specific courses required B.A. — STAT 241, 242, 251, 312; 2 from STAT 230, 361, 363; ENAS 130 or CPSC 112; B.S. — same, plus MATH 260, 300, or 301

Distribution of courses B.A. — 2 Stat electives numbered above 200, as specified; B.S. — 3 Stat electives numbered above 200, as specified

Substitution permitted STAT 238 for STAT 241, with DUS permission; courses in other depts or grad courses, with DUS permission

Senior requirement Both degrees — Senior project (STAT 490)
* STAT 325a, Statistical Case Studies  John Emerson
Statistical analysis of a variety of statistical problems using real data. Emphasis on methods of choosing data, acquiring data, assessing
data quality, and the issues posed by extremely large data sets. Extensive computations using R. Prerequisite: STAT 230, 361, or equivalent. QR

STAT 101 – 106, Introduction to Statistics
A basic introduction to statistics, including numerical and graphical summaries of data, probability, hypothesis testing, confidence
intervals, and regression. Each course in this group focuses on applications to a particular field of study and is taught jointly by two
instructors, one specializing in statistics and the other in the relevant area of application. The first seven weeks of classes are attended by
all students in STAT 101–106 together, as general concepts and methods of statistics are developed. The remaining weeks are divided into
field-specific sections that develop the concepts with examples and applications. Computers are used for data analysis. These courses are
alternatives; they do not form a sequence and only one may be taken for credit. No prerequisites beyond high school algebra. May not be
taken after STAT 100 or 109.

Students enrolled in STAT 101–106 who wish to change to STAT 109, or those enrolled in STAT 109 who wish to change to STAT 101–
106, must submit a course change notice, signed by the instructor, to their residential college dean by Friday, October 2. The approval of
the Committee on Honors and Academic Standing is not required.

STAT 101a / E&EB 210a, Introduction to Statistics: Life Sciences  Jonathan Reuning-Scherer
Statistical and probabilistic analysis of biological problems, presented with a unified foundation in basic statistical theory. Problems are
drawn from genetics, ecology, epidemiology, and bioinformatics. QR

STAT 102a / EP&E 203a / PLSC 452a, Introduction to Statistics: Political Science  Jonathan Reuning-Scherer
Statistical analysis of politics, elections, and political psychology. Problems presented with reference to a wide array of examples: public
opinion, campaign finance, racially motivated crime, and public policy. QR

Descriptive and inferential statistics applied to analysis of data from the social sciences. Introduction of concepts and skills for
understanding and conducting quantitative research. QR

STAT 105a, Introduction to Statistics: Medicine  Jonathan Reuning-Scherer
Statistical methods used in medicine and medical research. Practice in reading medical literature competently and critically, as well as
practical experience performing statistical analysis of medical data. QR

[ STAT 106, Introduction to Statistics: Data Analysis ]

Courses in Statistics

STAT 100b, Introductory Statistics  Staff
An introduction to statistical reasoning. Topics include numerical and graphical summaries of data, data acquisition and experimental
design, probability, hypothesis testing, confidence intervals, correlation and regression. Application of statistical concepts to data; analysis
of real-world problems. May not be taken after STAT 101–106 or 109. QR

EP&E: Intro Statistics

General concepts and methods in statistics. Meets for the first half of the term only. May not be taken after STAT 100 or 101–106.
½ Course cr

STAT 230b, Introductory Data Analysis  Staff
Survey of statistical methods: plots, transformations, regression, analysis of variance, clustering, principal components, contingency
tables, and time series analysis. The R computing language and Web data sources are used. Prerequisite: a 100-level Statistics course or
equivalent, or with permission of instructor. QR

EP&E: Intro Statistics

STAT 238a, Probability and Statistics  Joseph Chang
Fundamental principles and techniques of probabilistic thinking, statistical modeling, and data analysis. Essentials of probability,
including conditional probability, random variables, distributions, law of large numbers, central limit theorem, and Markov chains.
Statistical inference with emphasis on the Bayesian approach: parameter estimation, likelihood, prior and posterior distributions,
Bayesian inference using Markov chain Monte Carlo. Introduction to regression and linear models. Computers are used for calculations,
simulations, and analysis of data. After MATH 118 or 120. QR

STAT 241a / MATH 241a, Probability Theory  Harrison Zhou
Introduction to probability theory. Topics include probability spaces, random variables, expectations and probabilities, conditional
probability, independence, discrete and continuous distributions, central limit theorem, Markov chains, and probabilistic modeling. After
or concurrently with MATH 120 or equivalent. QR
STAT 242b / MATH 242b, Theory of Statistics  Staff
Study of the principles of statistical analysis. Topics include maximum likelihood, sampling distributions, estimation, confidence intervals, tests of significance, regression, analysis of variance, and the method of least squares. Some statistical computing. After STAT 241 and concurrently with or after MATH 222 or 225, or equivalents.  QR

STAT 251b / MATH 251b, Stochastic Processes  Staff
Introduction to the study of random processes, including Markov chains, Markov random fields, martingales, random walks, Brownian motion, and diffusions. Techniques in probability, such as coupling and large deviations. Applications chosen from image reconstruction, Bayesian statistics, finance, probabilistic analysis of algorithms, and genetics and evolution. After STAT 241 or equivalent.  QR

STAT 312a, Linear Models  Taylor Arnold
The geometry of least squares; distribution theory for normal errors; regression, analysis of variance, and designed experiments; numerical algorithms, with particular reference to the R statistical language. After STAT 242 and MATH 222 or 225.  QR

STAT 330b / MATH 330b, Advanced Probability  Staff
Measure theoretic probability, conditioning, laws of large numbers, convergence in distribution, characteristic functions, central limit theorems, martingales. Some knowledge of real analysis assumed.  QR

STAT 361a / AMTH 361a, Data Analysis  Jessica Cisewski
Selected topics in statistics explored through analysis of data sets using the R statistical computing language. Topics include linear and nonlinear models, maximum likelihood, resampling methods, curve estimation, model selection, classification, and clustering. After STAT 242 and MATH 222 or 225, or equivalents.  QR

STAT 363b, Multivariate Statistics for Social Sciences  Staff
Introduction to the analysis of multivariate data as applied to examples from the social sciences. Topics include principal components analysis, factor analysis, cluster analysis (hierarchical clustering, k-means), discriminant analysis, multidimensional scaling, and structural equations modeling. Extensive computer work using either SAS or SPSS programming software. Prerequisites: knowledge of basic inferential procedures and experience with linear models.  QR

STAT 364b / AMTH 364b / EENG 454b, Information Theory  Staff
Foundations of information theory in communications, statistical inference, statistical mechanics, probability, and algorithmic complexity. Quantities of information and their properties: entropy, conditional entropy, divergence, redundancy, mutual information, channel capacity. Basic theorems of data compression, data summarization, and channel coding. Applications in statistics and finance. After STAT 241.  QR

STAT 365b, Data Mining and Machine Learning  Staff
Techniques for data mining and machine learning from both statistical and computational perspectives, including support vector machines, bagging, boosting, neural networks, and other nonlinear and nonparametric regression methods. Discussion includes the basic ideas and intuition behind these methods, a more formal understanding of how and why they work, and opportunities to experiment with machine learning algorithms and to apply them to data. After STAT 242.  QR

* STAT 480a or b, Individual Studies  Staff
Directed individual study for qualified students who wish to investigate an area of statistics not covered in regular courses. A student must be sponsored by a faculty member who sets the requirements and meets regularly with the student. Enrollment requires a written plan of study approved by the faculty adviser and the director of undergraduate studies.

* STAT 490b, Senior Seminar and Project  Staff
Under the supervision of a member of the faculty, each student works on an independent project. Students participate in seminar meetings at which they speak on the progress of their projects.

GRADUATE COURSES OF PARTICULAR INTEREST TO UNDERGRADUATES
Courses in the Graduate School are open to qualified undergraduates. Descriptions of graduate courses in Statistics are available on the departmental Web site (http://statistics.yale.edu). Permission of the instructor and of the director of graduate studies is required.