

Statistics

*Departmental Counselor: Michael Stein, E 135, 702-8326 or 702-8333,
stein@galton.uchicago.edu*

Web: galton.uchicago.edu

Program of Study

The modern science of statistics involves the invention, study, and development of principles and methods for modeling uncertainty through mathematical probability; for designing experiments, surveys, and observational programs; and for analyzing and interpreting empirical data. Mathematics plays a major role in all statistical activity, whether of an abstract nature or dealing with specific techniques for analyzing data. Statistics is an excellent field for students with strong mathematical skills and an interest in applying these skills to problems in the natural and social sciences. A program leading to the B.A. degree in statistics offers coverage of the principles and methods of statistics in combination with a solid training in mathematics and some exposure to computing, which is essential to nearly all modern data analysis. In addition, there is considerable elective freedom enabling interested students to examine those areas of knowledge in the biological, physical, and social sciences that are often subjected to detailed statistical analysis. The major provides a base for graduate study in statistics or in other subjects with strong quantitative components. Students considering graduate study in statistics or related fields are encouraged to discuss their programs with the departmental counselor at an early stage, whether or not they plan to receive an undergraduate degree in statistics.

Statistics Courses for Students in Other Majors. Courses at the 20000 level are designed to provide instruction in statistics, probability, and statistical computation for students from all parts of the University. These courses differ in emphasis on theory or methods, on the mathematical level, and in the direction of applications. Most of the introductory courses make intensive use of computers to exemplify and explore statistical concepts and methods. The nature and extent of computer work varies according to the course and instructor. Statistics courses are not mathematics courses, but the mathematics prerequisites are a useful guide to the level of mathematical maturity assumed by a statistics course. Students are typically advised to take the course with the highest prerequisites that they can meet and, when possible, to take a two-quarter sequence rather than a one-quarter course. In particular, students who have taken calculus should not take STAT 20000 but, instead, should take 22000, 23400, 24400-24500, or 25100.

Explanations and comparisons of the various courses, both entry level and more advanced, are provided in the following sections. Students will also find the course descriptions to be helpful in choosing appropriate courses. Nonmajors should

contact Linda Collins (collins@galton.uchicago.edu) with remaining questions about appropriate courses for their backgrounds.

Introductory Courses and Sequences. STAT 22000, which is typically the statistics course taken first, is a general introduction to statistical concepts, techniques, and applications to data analysis and to problems in the design, analysis, and interpretation of experiments and observational programs. Computers are used throughout the course. A score of 4 or 5 on the AP test in statistics yields credit for STAT 22000, although this credit will not count toward the requirements for a major in statistics. STAT 23400 covers much of the same material as 22000, but at a somewhat higher mathematical level. STAT 23400 is a required course for students who are majoring in economics, but the class is a one-quarter introduction to statistics that is appropriate for any student with a good command of univariate calculus. For other students, STAT 20000 is an alternative with no calculus prerequisite that places less emphasis on statistical techniques. STAT 25100 is an introductory course in probability.

STAT 24400-24500 is recommended for students who wish to have a thorough introduction to statistical theory and methodology. STAT 24400-24500 is more mathematically demanding than either 22000 or 23400 and assumes some familiarity with multiple integration and with linear algebra. Students considering a major in statistics are strongly encouraged to take STAT 24400 rather than 23400. No prior training in statistics or probability is required for STAT 24400-24500. However, STAT 22000 would provide a helpful background. STAT 24600 is a follow-up to 24400-24500 that covers more advanced statistical methods. STAT 24400-24500 and 25100 form the core of the statistics major: this is recommended as a cognate sequence to students in the quantitative sciences and mathematics. Taking 25100 after 24400-24500 is recommended but not required.

For students interested in exploring methods and their applications, STAT 22200, 22400, and 22600 are recommended. These complementary second courses emphasize some class of methods for the analysis of data. They may be taken in any order. Each presumes a previous course in statistics (22000 or equivalent) and experience using computers in data analysis (as in STAT 22000). The emphasis is on linear models and experimental design in STAT 22200, multiple regression and least squares in 22400, and categorical data analysis in 22600. STAT 26100, which covers time dependent data, is appropriate for students with some knowledge of linear modeling (22400 or 24400-24500).

For students who have completed STAT 24500, many graduate courses in statistics offer opportunities for further study of statistical theory, methods, and applications. The introductory probability course (STAT 25100) may be taken separately from any statistics courses. STAT 25100 can be supplemented with more advanced probability courses, such as STAT 25200 (=31200), 31300, or 38100-38300. NOTE: College students may register for a number of other 30000-level courses in statistics. For details, consult the instructor or the departmental counselor, or visit galton.uchicago.edu.

Program Requirements

Degree Programs. Students should meet the general education requirements in the mathematical sciences with courses in calculus. The program includes four additional prescribed mathematics courses and four prescribed statistics courses; students should complete the four mathematics courses by the end of their third year. Additional requirements include one course in computer science and three approved elective courses in statistics. The four required statistics courses are STAT 24400-24500 and STAT 25100; and either 22400 or 34300. *If either STAT 22000 or STAT 23400 is included in the program (students may not include both), it must be taken before STAT 24400 is taken.* Although we allow STAT 23400 to count towards the degree if taken before STAT 24400, we do not recommend taking both STAT 23400 and 24400. Instead, we suggest taking STAT 24400 as a first course in statistics or, if a more elementary introduction is desired, taking STAT 22000 as preparation for STAT 24400. Candidates must obtain approval of their course program from the departmental counselor. NOTE: Students completing majors in both statistics and economics are encouraged to complete their mathematics requirements with MATH 20000-20100-25000 (or higher-level courses). Students who plan to complete both majors and have already taken MATH 19500-19600 should discuss with the departmental counselor how to best meet their mathematical requirements.

Summary of Requirements

<i>General Education</i>		MATH 13100-13200, 15100-15200, or 16100-16200*
<i>Major</i>	1	MATH 13300, 15300, or 16300*
	2	MATH 20000-20100, 20300-20400, or 20700-20800**
	1	MATH 25000 or 25500**
	4	STAT 24400, 24500, and 25100; and either 22400 or 34300
	1	CMSC 10500 or 10600 (10600 preferred), or 15100 or 16100
	<u>3</u>	approved elective courses in statistics***
	12	

* Credit may be granted by examination.

** Except for economics majors as noted in preceding section.

*** For example, STAT 22600 or 22700 (but not both), 22200, 24600, 25200, 26100, or 26700. Upon petition, one intermediate/advanced course in mathematics or computer science may be approved for this purpose by the statistics departmental counselor as relevant for a coherent degree program. The petition must include a documented strong case for the relevance.

Grading. Subject to College and divisional regulations, and with the consent of the instructor, *all students except majors in statistics* may register for quality grades or for *P/F* grades in any 20000-level statistics course. A grade of *P* is given only for work of *C-* quality or higher. Incompletes are allowed only in cases of serious emergency. A grade of at least *C-* must be earned in each of the twelve courses in the major; with the exception of STAT 29900, a grade of *P* is not acceptable for courses in the major.

Honors. The B.A. with honors is awarded to students with statistics as their primary major who have a GPA of 3.0 or higher overall and 3.25 or higher in the twelve required courses in the major, and who, in addition, also complete an approved honors paper (STAT 29900). This paper is typically based upon a structured research program that students undertake with faculty supervision in the first quarter of their fourth year. Eligible students who wish to be considered for honors should consult the departmental counselor before the end of their third year. The research paper or project used to meet this requirement may not be used to meet the B.A. paper or project requirement in another major. NOTE: Credit for STAT 29900 will not count towards the twelve courses required for a major in statistics.

Joint B.A./M.S. Program. This program enables qualified undergraduate students to complete an M.S. in statistics along with a B.A. during their four years at the College. Although a student may receive a B.A. in any field, a program of study other than statistics is recommended.

Participants must be admitted to the M.S. program in statistics. Students must submit applications by June 1 of their third year for admission to candidacy for an M.S. in statistics during their fourth year. Interested students are strongly encouraged to consult the departmental counselor early in their third year.

Participants in the joint B.A./M.S. program must meet the same requirements as students in the M.S. program in statistics. Of the nine courses that are required at the appropriate level, up to three may also meet the requirements of an undergraduate program. For example, STAT 24400-24500 and 24600, which are required for the M.S. in statistics, could also be used to meet part of the requirements of a B.A. or B.S. program in mathematics for courses outside of mathematics.

Other requirements include a master's paper and participation in the consulting program of the Department of Statistics. For details, visit <http://galton.uchicago.edu/admissions/master.html>.

Faculty

Y. Amit, L. Collins, M. Coram, M. Drton, S. Lalley, P. McCullagh, M. S. McPeck, P. Mykland, D. Nicolae, P. Niyogi, M. Stein, S. Stigler, R. Thisted, M. Wang, M. Wichura, K. Wilder, W. Wu

Courses: Statistics (STAT)

20000. Basic Concepts in Statistics. *This course meets one of the general education requirements in the mathematical sciences. NOTE: STAT 20000 may not be used in the statistics major.* This course is an introduction to statistical concepts and methods for the collection, presentation, analysis, and interpretation of data. Elements of sampling, simple techniques for analysis of means, proportions, and linear association are used to illustrate both effective and fallacious uses of statistics. *Autumn, Winter, Spring.*

22000. Introductory Statistics with Applications. *PQ: MATH 15200 or equivalent.* This course is an introduction to statistical techniques and methods of data analysis, including the use of computers. Examples are drawn from the biological, physical, and social sciences. Students are required to apply the techniques discussed to data drawn from actual research. Topics include data description, graphical techniques, exploratory data analyses, random variation and sampling, one- and two-sample problems, the analysis of variance, linear regression, and analysis of discrete data. *Summer, Autumn, Winter, Spring.*

22200. Linear Models and Experimental Design. *PQ: STAT 22000 or equivalent.* This course covers principles and techniques for the analysis of experimental data and the planning of the statistical aspects of experiments, surveys, and observational programs. Topics include linear and nonlinear models; analysis of variance and response surface analysis; randomization, blocking, and factorial designs; fractional replication and confounding; incorporation of covariate information; sample surveys; designs subject to constraints; and split-plot and nested experiments. *Spring.*

22400. Applied Regression Analysis. *PQ: STAT 22000 or equivalent.* This course is an introduction to the methods and applications of fitting and interpreting multiple regression models. The primary emphasis is on the method of least squares and its many varieties. Topics include the examination of residuals, the transformation of data, strategies and criteria for the selection of a regression equation, the use of dummy variables, tests of fit, nonlinear models, biases due to excluded variables and measurement error, and the use and interpretation of computer package regression programs. The techniques discussed are illustrated by many real examples involving data from both the physical and social sciences. Matrix notation is introduced as needed. *Autumn.*

22600. Analysis of Categorical Data. *PQ: STAT 22000 or equivalent.* This course covers statistical methods for the analysis of structured, counted data. Topics may include Poisson, multinomial, and product-multinomial sampling models; chi-square and likelihood ratio tests; log-linear models for cross-classified counted data, including models for data with ordinal categories and log-multiplicative models; logistic regression and logit linear models; and measures of association. Applications in the social and biological sciences are considered,

and the interpretation of models and fits, rather than mathematical details of computational procedures, is emphasized. *Winter*.

22700. Biostatistical Methods. (=HSTD 32700) *PQ: STAT 22000 or equivalent, or consent of instructor.* This course is designed to provide students with tools for analyzing categorical, count, and time-to-event data frequently encountered in medicine, public health, and related biological and social sciences. This course emphasizes application of the methodology rather than statistical theory (e.g., recognition of the appropriate methods, interpretation and presentation of results). Methods covered include contingency table analysis, Kaplan-Meier survival analysis, Cox proportional-hazards survival analysis, logistic regression, and Poisson regression. *Winter*.

23400. Statistical Models and Methods. *PQ: MATH 13300, 15300, or 16300.* This course presents basic ideas of probability theory and statistics, and is recommended for students throughout the natural and social sciences who want a broad background in statistical methodology and exposure to probability models and the statistical concepts underlying the methodology. Probability is developed for the purpose of modeling outcomes of random phenomena. Random variables and their expectations are studied; including means and variances of linear combinations and an introduction to conditional expectation. Binomial, Poisson, normal and other standard probability distributions are considered. Some probability models are studied mathematically and others via simulation on a computer. Sampling distributions and related statistical methods are explored mathematically, studied via simulation and illustrated on data. Statistical methods for describing data and making inferences based on samples from populations are presented. Methods include, but are not limited to, inference for means and variances for one- and two-sample problems, correlation and simple linear regression. Graphical and numerical data description are used for exploration, communication of results, and comparing mathematical consequences of probability models and data. Mathematics employed is to the level of univariate calculus, but is less demanding than that required by STAT 24400. *Autumn, Winter, Spring*.

24400-24500. Statistical Theory and Methods I, II. *PQ: Calculus, including some familiarity with multiple integration, and some familiarity with linear algebra (e.g., MATH 19600 or 20100 or 20400, or equivalent).* Some previous experience with statistics recommended but not required. This course is a systematic introduction to the principles and techniques of statistics, with emphasis on the analysis of experimental data. The first quarter covers tools from probability and the elements of statistical theory. Topics include the definitions of probability and random variables, binomial and other discrete probability distributions, normal and other continuous probability distributions, joint probability distributions and the transformation of random variables, principles of inference (including Bayesian inference), maximum likelihood estimation, hypothesis testing and confidence intervals, likelihood ratio tests, multinomial distributions, and chi-square tests. Examples are drawn from the social, physical, and biological sciences.

The coverage of topics in probability is limited and brief so those who have taken a course in probability will find reinforcement rather than redundancy. The second quarter covers statistical methodology, including the analysis of variance, regression, correlation, and some multivariate analysis. Some principles of data analysis are introduced, and an attempt is made to present the analysis of variance and regression in a unified framework. The computer is used in the second quarter. *Autumn, Winter; Winter, Spring.*

24600. Complex Statistical Problems. *PQ: STAT 24400-24500 or equivalent. Knowledge of probability distributions, random variables, and estimation techniques, such as maximum likelihood from STAT 24400-24500.* Topics vary from year to year. The course recently has treated the impact of missing data on statistical analyses (e.g., probability models and methods of estimation and inference); algorithms for iterative maximum likelihood estimation (e.g., the Expectation-Maximization [EM] and Newton-Raphson algorithms); and Bayesian computation (e.g., Data Augmentation and Monte Carlo Markov Chain methods). *Spring.*

25100. Introduction to Mathematical Probability. *PQ: MATH 20000 or 20300, or consent of instructor.* This course covers fundamentals and axioms; combinatorial probability; conditional probability and independence; binomial, Poisson, and normal distributions; the law of large numbers and the central limit theorem; and random variables and generating functions. *Spring.*

25200/31200. Introduction to Stochastic Processes I. *PQ: STAT 25100, and MATH 20100 or 20400.* This course introduces stochastic processes not requiring measure theory. Topics include branching processes, recurrent events, renewal theory, random walks, Markov chains, Poisson, and birth-and-death processes. *Winter.*

26100. Time Dependent Data. *PQ: STAT 24400-24500 or 22400, or consent of instructor.* This course considers the modeling and analysis of data that are ordered in time. The main focus is on quantitative observations taken at evenly spaced intervals and includes both time-domain and spectral approaches. Time permitting, statistical approaches to other data types (e.g., categorical observations, point processes) are considered. *Spring.*

26700/36700. History of Statistics. (=CHSS 32900, HIPS 25600) *PQ: Prior statistics course.* This course covers topics in the history of statistics, from the eleventh century to the middle of the twentieth century. We focus on the period from 1650 to 1950, with an emphasis on the mathematical developments in the theory of probability and how they came to be used in the sciences. Our goals are both to quantify uncertainty in observational data and to develop a conceptual framework for scientific theories. This course includes broad views of the development of the subject, and closer looks at specific people and investigations, including reanalyses of historical data. *S. Stigler. Spring.*

29700. Undergraduate Research. *PQ: Consent of faculty adviser and departmental counselor. Students are required to submit the College Reading and Research Course Form. Open to statistics majors and nonmajors. May be taken either for quality grades or for P/F grades. However, a grade of P will not count toward the requirements for a major in statistics. This course consists of reading and research in an area of statistics or probability under the guidance of a faculty member. A written report must be submitted at the end of the quarter. Autumn, Winter, Spring.*

29900. Bachelor's Paper. *PQ: Consent of faculty adviser and departmental counselor. Students are required to submit the College Reading and Research Course Form. Open only to statistics majors. May be taken P/F. This course consists of reading and research in an area of statistics or probability under the guidance of a faculty member, leading to a bachelor's paper. The paper must be submitted at the end of the quarter. Credit for STAT 29900 will not count toward the twelve courses required for a major in statistics. Autumn, Winter, Spring.*

The following list of 30000-level courses may be of interest to advanced undergraduate majors. For more information, consult the departmental counselor. For a complete listing and updates, visit galton.uchicago.edu/.

30100-30200. Mathematical Statistics. *PQ: STAT 30400 or consent of instructor. This course surveys the mathematical structure of modern statistics. Topics include statistical models, methods for parameter estimation, comparison of estimators, large sample theory, efficiency, confidence sets, theory of hypothesis tests, elements of linear hypothesis theory, analysis of discrete data, and an introduction to Bayesian analysis. Winter, Spring.*

30400. Distribution Theory. *PQ: STAT 24500 and MATH 20500, or consent of instructor. This course is a systematic introduction to random variables and probability distributions. Topics include standard distributions (i.e., uniform, normal, beta, gamma, F, t, Cauchy, Poisson, binomial, and hypergeometric), moments and cumulants, characteristic functions, exponential families, modes of convergence, central limit theorem, and Laplace's method. Autumn.*

30700. Numerical Computation for Statistics. (=CMSC 37800) *PQ: STAT 34300 or consent of instructor. This course covers topics in numerical methods and computation that are useful in statistical research (e.g., simulation, random number generation, Monte Carlo methods, quadrature, optimization, and matrix methods). Autumn.*

31200. Introduction to Stochastic Processes I. *PQ: STAT 25100 and MATH 20400. This course introduces stochastic processes not requiring measure theory. Topics include branching processes, recurrent events, renewal theory, random walks, Markov chains, Poisson, and birth-and-death processes. Winter.*

31300. Introduction to Stochastic Processes II. *PQ: STAT 31200 or consent of instructor. This course is a sequel to STAT 31200. Topics include continuous-time Markov chains, Markov chain Monte Carlo, discrete-time martingales, and*

Brownian motion and diffusions. The emphasis is on defining the processes and calculating or approximating various related probabilities. The measure theoretic aspects of these processes are not covered rigorously. *Spring*.

33100. Sample Surveys. *PQ: Consent of instructor.* This course covers random sampling methods; stratification, cluster sampling, and ratio estimation; and methods for dealing with non-response and partial response. *Autumn*.

33800. Statistical Inference for Financial Data. *PQ: STAT 30100 and 30200, or consent of instructor. Primarily intended for second-year graduate students in statistics, this course is also open to students with similar backgrounds in econometrics or finance.* Financial data is commonly modeled by diffusion, jump-diffusion, and related models, and it is usually supposed that observation is discrete. This course is concerned with inference in such settings. *Spring*.

34300. Applied Linear Statistical Methods. *PQ: STAT 24500 and MATH 25000, or equivalents.* This course introduces the theory, methods, and applications of fitting and interpreting multiple regression models. Topics include the examination of residuals, the transformation of data, strategies and criteria for the selection of a regression equation, nonlinear models, biases due to excluded variables and measurement error, and the use and interpretation of computer package regression programs. The theoretical basis of the methods, the relation to linear algebra, and the effects of violations of assumptions are studied. Techniques discussed are illustrated by examples involving both physical and social sciences data. *Autumn*.

34500. Design and Analysis of Experiments. *PQ: STAT 34300.* This course introduces the methodology and application of linear models in experimental design. We emphasize the basic principles of experimental design (e.g., blocking, randomization, incomplete layouts). Many of the standard designs (e.g., fractional factorial, incomplete block, split unit designs) are studied within this context. The analysis of these experiments is developed as well, with particular emphasis on the role of fixed and random effects. Time permitting, additional topics may include response surface analysis, the use of covariates in the analysis of designed experiments, and spatial analysis of field trials. *Winter*.

34700. Generalized Linear Models. *PQ: STAT 34300 or knowledge of linear models at the level of the texts by Draper and Smith or Weisberg.* This applied course covers factors, variates, contrasts, and interactions; exponential-family models (i.e., variance function); definition of a generalized linear model (i.e., link functions); specific examples of GLMs; logistic and probit regression; cumulative logistic models; log-linear models and contingency tables; inverse linear models; Quasi-likelihood and least squares; estimating functions; and partially linear models. *Spring*.

35000. Principles of Epidemiology. (=ENST 27400, HSTD 30900, PPHA 36400) *Prior course in statistics recommended.* Epidemiology is the study of the distribution and determinants of health and disease in human populations. This

course introduces the basic principles of epidemiologic study design, analysis, and interpretation through lectures, assignments, and critical appraisal of classic and contemporary research articles. *Autumn.*

37800. Statistical Computing. *PQ: Consent of instructor.* Students are introduced to and gain experience in using a variety of computational tools that are useful for large statistical computing projects. These include HTML, Unix shell programming, python, and C++, as well as utilities such as the gdb debugger, the cvs source control system, and make. We emphasize choosing and being able to use the right tools for projects that involve execution speed, memory usage, coding ease, portability, and other issues. Of particular interest are projects for which favorite tools such as R and Matlab may not be sufficient. *Autumn.*

38100. Measure-Theoretic Probability I. *PQ: STAT 31300 or consent of instructor.* This course provides a detailed, rigorous treatment of probability from the point of view of measure theory, as well as existence theorems, integration and expected values, characteristic functions, moment problems, limit laws, Radon-Nikodym derivatives, and conditional probabilities. *Autumn.*

38300. Measure-Theoretic Probability II. *PQ: STAT 38100.* This course is a continuation of STAT 38100. Lp spaces, Radon-Nikodym theorem, conditional expectation, and martingale theory are discussed. *Winter.*

38500. Measure-Theoretic Probability III. *PQ: STAT 38300.* This course is a continuation of STAT 38300. Continuous-parameter martingales, Brownian motion, and Skorohod embedding are discussed; and the Ito calculus is introduced. *Spring.*