Statistics

Departmental Counselor: Stephen M. Stigler, E 102, 702-8328 or 702-8333, stigler@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. 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 concentration provides a base for graduate study in statistics or in other subjects with strong quantitative components. An honors program is available. 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 Concentrations. 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 serious 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 provide a useful guide to the level of mathematical maturity assumed by a statistics course. In general, students are 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, rather, should take STAT 22000, 23400-23500, 24400-24500, or 25100.

Introductory Courses and Sequences. STAT 22000 is the usual first course in statistics, providing 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. One or two sections of STAT 22000 in the Autumn, Winter, and Spring Quarters use examples drawn from economics and business and a selection of texts and topics that are more appropriate for concentrators in economics. A score of 4 or 5 on the

AP test in statistics yields credit for STAT 22000. STAT 20000 is an alternative that has no calculus prerequisite and places less emphasis on exploring statistical techniques. STAT 25100 is an introductory course in probability.

STAT 23400-23500 or 24400-24500 is recommended for students who want a thorough introduction to statistical theory and methodology. The two sequences differ primarily in the level of mathematics employed; both make consistent use of calculus, but 24400-24500 is more demanding and assumes some familiarity with multiple integration and with linear algebra. Normally students would not receive credit for both sequences, so it is important that they plan their schedules carefully in light of their level of mathematical training. No prior training in statistics or probability is required for STAT 23400-23500 or 24400-24500. However, STAT 20000 or 22000 would provide a helpful background; students who have taken one of these are encouraged to take STAT 23400-23500 or 24400-24500 if they want more extensive training in the basis of statistical methods. STAT 24600 is offered as a supplement to either sequence.

STAT 24400-24500 and 25100 form the core of the statistics concentration: this is recommended as a cognate sequence to concentrators in the quantitative sciences and mathematics. It would be preferable, but not mandatory, to take STAT 25100 after 24400-24500; accordingly, 25100 is now offered in the Spring Quarter to permit the completion of this cognate sequence in one year. Students who take STAT 23400-23500 and then decide to concentrate in statistics are urged to consult with the departmental counselor. In no case will credit be given for 23400-23500 if taken after 24400-24500.

For students more interested in exploring methods and their applications, STAT 22200, 22400, and 22600 are recommended. These are complementary second courses that emphasize some class of methods for the analysis of data. They may be taken in any order. Each presumes a previous course in statistics (STAT 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 STAT 22400, and categorical data analysis in STAT 22600.

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 31200, 31300, or 38100-38300. NOTE: College students may register for a number of other 30000-level courses in statistics. For further information, see the instructor, the departmental counselor, or *galton.uchicago.edu/.*

Program Requirements

Degree Programs. Concentrators should meet the general education requirements in the mathematical sciences with courses in calculus. Concentration requirements include four additional prescribed mathematics courses and five prescribed statistics courses; the four mathematics courses should be completed by the end of the third year. Additional requirements

include one course in computer science and two more courses in mathematics, statistics, or computer science. The five required statistics courses must include STAT 24400-24500 and STAT 25100; and either 22400 or 34300. The fifth required statistics course may be either STAT 22000 or another course such as STAT 22200, 22600, 24600, 30100, 31200, or 32100. *If STAT 22000 is included as part of the program, it should be taken before STAT 24400 is taken.* Candidates must obtain approval of their course program from the departmental counselor. NOTE: Students completing concentrations in both statistics and economics may replace MATH 20000-20100 and MATH 25000/25500 with MATH 19500-19600 and MATH 20300. NOTE: STAT 23400-23500 is new in 2002-2003, and its role in the department's degree requirements is still under consideration. Students who wish to take STAT 23400-23500 and are considering a statistics concentration should consult first with the departmental counselor.

Summary of Requirements

General Education		MATH 13100-13200, 15100-15200, or 16100-16200†
Concentration	1	MATH 13300, 15300, or 16300†
	2	MATH 20000-20100, 20300-20400, or 20700-20800
	1	MATH 25000 or 25500
	5	STAT 24400, 24500, 25100, and 22400 or 34300, and one other approved statistics course
	1	CMSC 10500 or 11500
	2	approved courses in statistics, mathematics, or computer science*
	12	

† *Credit may be granted by examination.*

* *For example, STAT 22200, 22600, 24600, 26700, or 32900; or an intermediate/advanced course in mathematics or computer science approved by the statistics departmental counselor as relevant for a coherent degree program*

Grading. Subject to College and divisional regulations, and with the consent of the instructor, *all students except concentrators in statistics* may register for regular letter grades or *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. To meet the concentration requirement in statistics, a grade of at least *C-* must be earned in each of the twelve courses; a grade of *P* is not acceptable for meeting these concentration requirements.

Honors. The B.A. with honors is awarded to students who have a GPA of 3.0 or higher overall and 3.25 or higher in the twelve required courses in the concentration and who, in addition to these courses, complete an approved honors paper (STAT 29900). This paper is usually based upon a structured research program undertaken with faculty supervision in the first quarter of the student's fourth year. Interested students who meet the program requirements should see the departmental counselor before the end of their third year in the College.

Faculty

Y. Amit, Z. Chi, S. Lalley, M. Larsen, P. McCullagh, M. S. McPeek, P. Mykland, D. Nicolae, P. Niyogi, M. Stein, S. Stigler, R. Thisted, M. Wang, M. Wichura, W. Wu

Courses

12500. Quantitative Methods in Environmental Science. (=ENST 12500, NTSC 12500) *PQ: NTSC 12400 or consent of instructor.* This course studies mathematical, statistical, and computational approaches to scientific issues raised previously in this sequence. Three principal tools are: differential equations as a way to model a changing world, probability theory as a way to quantify uncertainty, and the application of computer simulations to understanding environmental processes. *M. Stein. Winter.*

20000. Basic Concepts in Statistics. *PQ: MATH 10600, placement into 13100 or higher, or satisfactory performance on a special elementary diagnostic mathematics examination. This course meets one of the general education requirements in the mathematical sciences. NOTE: STAT 20000 does not count toward the concentration requirement.* 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. Except in the Summer Quarter, one or more sections of STAT 22000 use examples drawn from economics and business and a selection of texts and topics that are more appropriate for concentrators in economics. *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 may 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; design and analysis of sample surveys; designs subject to constraints; split-plot and nested experiments; and components of variance. *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 discussed 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.*

23400-23500. Statistical Models and Methods I, II. *PQ: MATH 15300 or equivalent. Some previous experience with statistics recommended but not required.* 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 the probability models and statistical concepts underlying the methodology. Probability is developed for the purpose of modeling outcomes of random phenomena. Some models are studied mathematically and others via simulation on a computer. Binomial, Poisson, normal and other standard probability distributions are considered. Statistical methods for describing data and making inferences based on samples from populations are presented. Methods are illustrated on examples and studied via simulation. Topics include Bayesian inference, maximum likelihood estimation, and repeated-sampling frequentist inference. Methods for one- and two-sample problems, analysis of variance, analysis of counted data, and correlation and regression are studied. Graphical and numerical data description are used for exploration, communication of results, and comparing mathematical consequences of probability models and data. Mathematics is employed to the level of univariate calculus but is less demanding than that required by STAT 24400-24500. Other than the mathematical level, the content of the two sequences are similar. *Autumn, Winter.*

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.* A systematic introduction to the principles and techniques of statistics, with emphasis on the analysis of experimental data. The first quarter will cover 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, chi-square tests. Examples

will be drawn from the social, physical, and biological sciences. The coverage of topics in probability will be limited and brief, so that those who have taken a course in probability will find reinforcement rather than redundancy. The second quarter will cover statistical methodology, including the analysis of variance, regression, correlation, and some multivariate analysis. Some principles of data analysis will be introduced, and an attempt will be made to present the analysis of variance and regression in a unified framework. The computer will be used in the second quarter. *Autumn, Winter.*

24600. Complex Statistical Problems. *PQ: STAT 23400-23500 or 24400-24500, or equivalent. Knowledge of probability distributions, random variables, and estimation techniques, such as maximum likelihood from STAT 23400-23500 or 24400-24500.* In this course, the impact of missing data on statistical analyses is considered. Probability models, methods of estimation and inference, and applications to data are studied for various examples. Algorithms for iterative maximum likelihood estimation, such as the Expectation-Maximization (EM) and Newton-Raphson algorithms, and for Bayesian computation, such as Data Augmentation and Monte Carlo Markov Chain methods, are introduced. *Spring.*

24700/31000. Statistical Methods for the Neurosciences II: Introduction to Probability Models. (=BIOS 29406) *PQ: Students must have completed the equivalent of one year of college calculus and at least one introductory statistics course such as STAT 22000, preferably 23400 or 24400 or 25100.* This course is for students in computational neuroscience and other biological or physical sciences interested in stochastic modeling. The course offers an overview of basic concepts in probability and information theory, discrete Markov chains, continuous time Markov chains and diffusions. The emphasis is on modeling and computation, with particular attention to applications in computational neuroscience. *Winter.*

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.*

26700/36700. History of Statistics. (=CFSC 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. The emphasis is on the period from 1650 to 1950, and on the mathematical developments in the theory of probability and how they came to be used in the sciences, both to quantify uncertainty in observational data and as a conceptual framework for scientific theories. The 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 both concentrators and nonconcentrators. May be taken either for a* P/F *grade or for a quality grade.* 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 concentrators. May be taken* P/N *or* 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. *Autumn, Winter, Spring.*

For more information on 30000-level courses, consult the departmental counselor. For updates, see 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, 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 covers methods of deriving, characterizing, displaying, approximating, and comparing distributions. Topics include algebra by computer (Maple and Macsyma), standard distributions (uniform, normal, beta, gamma, F, t, Cauchy, Poisson, binomial, and hypergeometric), moments and cumulants, characteristic functions, exponential families, the Pearson system, Edgeworth and saddlepoint approximations, and Laplace's method. *Autumn.*

31200. Introduction to Stochastic Processes I. *PQ: STAT 25100, and MATH 20100 or 20400.* This course is an introduction to 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 covered include continuous time Markov chains: birth-and-death processes and queues, introduction to discrete time martingales, and Brownian motion and diffusions. Stochastic ordering and Poisson approximations may also be discussed. 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.*

32900. Applied Multivariate Analysis. (=GSBC 42400) *PQ: STAT 22400 or equivalent.* This course is an introduction to multivariate analysis. Topics include principal component analysis, multidimensional scaling, discriminant analysis, canonical correlation analysis, and cluster analysis. *Spring.*

34300. Applied Linear Statistical Methods. *PQ: STAT 24500 and MATH 25000, or equivalents.* This course is an introduction to 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.*

35000-35100. Epidemiology. (=HSTD 31000-31100) *PQ: Consent of instructor.* The topic of this course is the quantitative study of the spread of diseases in a population. *Autumn, Spring.*

35600. Introduction to Survival Analysis. (=HSTD 33100) *PQ: Consent of instructor.* This course covers the analysis of longitudinal data on patients. *Winter.*

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.*