

STATISTICS

The Statistics Department Office:

1005 School of Social Work (1255 Amsterdam Avenue); 212-851-2132
<http://www.stat.columbia.edu>

Statistics Major and Concentration Advising:

Ronald Neath, 612 Watson; 212-853-1398; rcn2112@columbia.edu

Data Science Major Advising:

Computer Science: Tim Roughgarden, 410 Mudd; 212-853-8474;
tr@cs.columbia.edu (cannon@cs.columbia.edu)

Statistics: Ronald Neath, 612 Watson;
 212-853-1398; rcn2112@columbia.edu

Economics - Statistics Major Advising:

Economics: Susan Elmes, 1006 IAB; 212-854-9124; se5@columbia.edu

Statistics: Ronald Neath, 612 Watson;
 212-853-1398; rcn2112@columbia.edu

Mathematics - Statistics Major Advising:

Mathematics: Julien Dubedat, 601 Mathematics;
 212-854-8806; jd2653@columbia.edu

Statistics: Ronald Neath, 612 Watson;
 212-853-1398; rcn2112@columbia.edu

Political Science - Statistic Major Advising:

Political Science: Naoki Egami, 734 IAB; 212-854-3623;
naoki.egami@columbia.edu (rys3@columbia.edu)

Statistics: Ronald Neath, 612 Watson;
 212-853-1398; rcn2112@columbia.edu

Department Administrator:

Dood Kalicharan, 1003 School of Social Work;
 212-851-2130; dk@stat.columbia.edu

The Department offers several introductory courses. Students interested in statistical concepts, who plan on consuming, but not creating statistics, should take STAT UN1001 INTRO TO STATISTICAL REASONING. The course is designed for students who have taken a pre-calculus course, and the focus is on general principles. It is suitable for students seeking to satisfy the Barnard quantitative reasoning requirements. Students seeking an introduction to applied statistics should take STAT UN1101 INTRODUCTION TO STATISTICS. The course is designed for students who have some mathematical maturity, but who may not have taken a course in calculus, and the focus is on the elements of data analysis. It is recommended for pre-med students, and students contemplating the concentration in statistics. Students seeking a foundation for further study of probability theory and statistical theory and methods should take STAT UN1201 CALC-BASED INTRO TO STATISTICS. The course is designed for students who have taken a semester of college calculus or the equivalent, and the focus is on preparation for a mathematical study of probability and statistics. It is recommended for students seeking to complete the prerequisite for econometrics, and for students contemplating the major in statistics. Students seeking a one-semester calculus-based survey of probability theory and statistical theory and methods should take STAT GU4001 INTRODUCTION TO PROBABILITY AND STATISTICS. This course is designed for students who have taken calculus, and is meant as a terminal course. It provides a somewhat abridged version of the more demanding sequence STAT GU4203 PROBABILITY THEORY and STAT GU4204 STATISTICAL INFERENCE. While some mathematically mature students take the more demanding sequence as

an introduction to the field, it is generally recommended that students prepare for the sequence by taking STAT UN1201 CALC-BASED INTRO TO STATISTICS.

The Department offers the Major in Statistics, the Concentration in Statistics, and interdisciplinary majors with Computer Science, Economics, Mathematics, and Political Science. The concentration is suitable for students preparing for work or study where substantial skills in data analysis are valued and may be taken without mathematical prerequisites. The concentration consists of a sequence of six courses in applied statistics, but students may substitute statistics electives numbered 4203 or above with permission of the concentration advisors. The major consists of mathematical and computational prerequisites, an introductory course, and five core courses in probability theory and theoretical and applied statistics together with three electives. The training in the undergraduate major is comparable to a masters degree in statistics.

Students may wish to consult the following guidelines when undertaking course planning. It is advisable to take STAT UN1101 INTRODUCTION TO STATISTICS and STAT UN2102 Applied Statistical Computing before taking any of the more advanced concentration courses, STAT UN2103 APPLIED LINEAR REG ANALYSIS, STAT UN2104 APPL CATEGORICAL DATA ANALYSIS, STAT UN3105 APPLIED STATISTICAL METHODS, and STAT UN3106 APPLIED MACHINE LEARNING. It is advisable to take STAT UN1201 CALC-BASED INTRO TO STATISTICS, STAT GU4203 PROBABILITY THEORY, STAT GU4204 STATISTICAL INFERENCE, and STAT GU4205 LINEAR REGRESSION MODELS in sequence.

Courses in stochastic analysis should be preceded by STAT GU4203 PROBABILITY THEORY, and for many students, it is advisable to take STAT GU4207 ELEMENTARY STOCHASTIC PROCESS before embarking on STAT GU4262 Stochastic Processes for Finance, STAT GU4264 STOCHASTIC PROCESSES-APPLICANTS I, or STAT GU4265 STOCHASTIC METHODS IN FINANCE. Most of the statistics courses numbered from 4221 to 4234 are best preceded by STAT GU4205 LINEAR REGRESSION MODELS. The data science courses STAT GU4206 STAT COMP # INTRO DATA SCIENCE, STAT GU4241 STATISTICAL MACHINE LEARNING, and STAT GU4242 Advanced Machine Learning should be taken in sequence.

Advanced Placement

The Department offers three points of advanced credit for a score of 5 on the AP statistics exam. Students who are required to take an introductory statistics course for their major should check with their major advisor to determine whether this credit provides exemption from their requirement.

Departmental Honors

Students are considered for department honors on the basis of GPA and the comprehensiveness and difficulty of their course work in the Department. The Department is generally permitted to nominate one tenth of graduating students for departmental honors.

Undergraduate Research in Statistics and the Summer Internship

Matriculated students who will be undergraduates at Columbia College, Barnard College, the School of General Studies, or the School of Engineering and Applied Sciences may apply to the Department's summer internship program. The internship provides summer housing and a stipend. Students work with Statistics Department faculty mentors. Applicants should send a brief statement of interest and

a copy of their transcript to Ms. Dood Kalicharan in the Statistics Department office by the end of March to be considered. If summer project descriptions are posted on the Department's website, please indicate in the statement of interest which project is of interest. Students seeking research opportunities with Statistics Department faculty during the academic year are advised to be entrepreneurial and proactive: identify congenial faculty whose research is appealing, request an opportunity to meet, and provide some indication of previous course work when asking for a project.

Professors

David Blei (with Computer Science)
 John Cunningham
 Richard R. Davis
 Victor H. de la Peña
 Andrew Gelman (with Political Science)
 Ioannis Karatzas (with Mathematics)
 Jingchen Liu
 Shaw-Hwa Lo
 Marcel Nutz (with Mathematics)
 Liam Paninski
 Philip Protter
 Daniel Rabinowitz
 Bodhisattva Sen
 Michael Sobel
 Simon Tavaré (with Biological Sciences)
 Zhiliang Ying
 Ming Yuan
 Tian Zheng (Chair)

Associate Professors

Samory Kpotufe
 Arian Maleki
 Sumit Mukherjee

Assistant Professors

Marco Avella
 Yuqi Gu
 Cynthia Rush
 Anne van Delft

Term Assistant Professors

Carsten Chong
 Gokce Dayanikli
 Yongchen Kwon
 Johannes Wiesel
 Chenyang Zhong

Adjunct Faculty

Demissie Alemayehu
 Mark Brown
 Guy Cohen
 Regina Dolgoarshinnykh
 Hammou El Barmi
 Tat Sang Fung

Xiaofu He
 Ying Liu
 Ka-Yi Ng
 Ha Nguyen
 Cristian Pasarica
 Kamiar Rahnama Rad
 Ori Shental
 Haiyuan Wang
 Rongning Wu

Lecturers in Discipline

Banu Baydil
 Anthony Donoghue
 Wayne Lee
 Dobrin Marchev
 Ronald Neath
 Alex Pijyan
 David Rios
 Joyce Robbins
 Gabriel Young

Major in Statistics

The requirements for this program were modified in March 2016. Students who declared this program before this date should contact the director of undergraduate studies for the department in order to confirm their options for major requirements.

The major should be planned with the director of undergraduate studies. Courses taken for a grade of Pass/D/Fail, or in which the grade of D has been received, do not count toward the major. The requirements for the major are as follows:

Mathematics and Computer Science Prerequisites

MATH UN1101	CALCULUS I
MATH UN1102	CALCULUS II
MATH UN1201	CALCULUS III
MATH UN2010	LINEAR ALGEBRA

One of the following five courses

COMS W1007	
ENGI E1006	INTRO TO COMP FOR ENG/APP SCI
COMS W1005	Introduction to Computer Science and Programming in MATLAB
STAT UN2102	Applied Statistical Computing
COMS W1004	Introduction to Computer Science and Programming in Java

Core courses in probability and statistics

STAT UN1201	CALC-BASED INTRO TO STATISTICS
STAT GU4203	PROBABILITY THEORY
STAT GU4204	STATISTICAL INFERENCE
STAT GU4205	LINEAR REGRESSION MODELS
STAT GU4206	STAT COMP # INTRO DATA SCIENCE
STAT GU4207	ELEMENTARY STOCHASTIC PROCESS

Three approved electives in statistics or, with permission, a cognate field.

- Students preparing for a career in actuarial science are encouraged to replace STAT GU4205 LINEAR REGRESSION MODELS with

STAT GU4282 Linear Regression and Time Series Methods, and should take as one of their electives STAT GU4281 Theory of Interest.

- Students preparing for graduate study in statistics are encouraged to replace two electives with MATH GU4061 INTRO MODERN ANALYSIS I and MATH GU4062 INTRO MODERN ANALYSIS II .

STAT GU4205	LINEAR REGRESSION MODELS
-------------	--------------------------

And select one of the following courses:

STAT GU4207	ELEMENTARY STOCHASTIC PROCESS
STAT GU4262	Stochastic Processes for Finance
STAT GU4264	STOCHASTC PROCSESSES-APPLICTNS I
STAT GU4265	STOCHASTIC METHODS IN FINANCE

Computer Science

Select one of the following courses:

COMS W1004	Introduction to Computer Science and Programming in Java
COMS W1005	Introduction to Computer Science and Programming in MATLAB
ENGI E1006	INTRO TO COMP FOR ENG/APP SCI
COMS W1007	or an advanced Computer Science offering in programming

Electives

An approved selection of three advanced courses in mathematics, statistics, applied mathematics, industrial engineering and operations research, computer science, or approved mathematical methods courses in a quantitative discipline. At least one elective must be a Mathematics Department course numbered 3000 or above.

- Students interested in modeling applications are recommended to take MATH UN3027 Ordinary Differential Equations and MATH UN3028 PARTIAL DIFFERENTIAL EQUATIONS
- Students interested in finance are recommended to include among their electives, MATH GR5010 INTRO TO THE MATH OF FINANCE, STAT GU4261 STATISTICAL METHODS IN FINANCE, and STAT GU4221 TIME SERIES ANALYSIS.
- Students interested in graduate study in mathematics or in statistics are recommended to take MATH GU4061 INTRO MODERN ANALYSIS I and MATH GU4062 INTRO MODERN ANALYSIS II.
- Students preparing for a career in actuarial science are encouraged to replace STAT GU4205 LINEAR REGRESSION MODELS with STAT GU4282 Linear Regression and Time Series Methods, and to take among their electives STAT GU4281 Theory of Interest.

Minor in Statistics

Courses taken for a grade of Pass/D/Fail, or in which the grade of D, has been received do not count towards the minor. The requirements for the minor are as follows.

STAT UN1101	INTRODUCTION TO STATISTICS
STAT UN2102	Applied Statistical Computing
STAT UN2103	APPLIED LINEAR REG ANALYSIS
STAT UN2104	APPL CATEGORICAL DATA ANALYSIS
STAT UN3105	APPLIED STATISTICAL METHODS
STAT UN3106	APPLIED MACHINE LEARNING

- Students may replace courses required for the minor by approved Statistics Department courses.

Major in Mathematics-Statistics

The program is designed to prepare the student for: (1) a career in industries such as finance and insurance that require a high level of mathematical sophistication and a substantial knowledge of probability and statistics; and (2) graduate study in quantitative disciplines. Students choose electives in finance, actuarial science, operations research, or other quantitative fields to complement requirements in mathematics, statistics, and computer science.

Courses taken for a grade of Pass/D/Fail, or in which the grade of D has been received, do not count toward the major. The requirements for the major are as follows:

Mathematics

Select one of the following sequences:

MATH UN1101	CALCULUS I
MATH UN1102	CALCULUS II
MATH UN1201	CALCULUS III
MATH UN2010	LINEAR ALGEBRA
MATH UN2500	ANALYSIS AND OPTIMIZATION
OR	
MATH UN1101	CALCULUS I
MATH UN1102	CALCULUS II
MATH UN1205	ACCELERATED MULTIVARIABLE CALC
MATH UN2010	LINEAR ALGEBRA
MATH UN2500	ANALYSIS AND OPTIMIZATION
OR	
MATH UN1207	HONORS MATHEMATICS A
MATH UN1208	HONORS MATHEMATICS B
MATH UN2500	ANALYSIS AND OPTIMIZATION

Statistics required courses

STAT UN1201	CALC-BASED INTRO TO STATISTICS
STAT GU4203	PROBABILITY THEORY
STAT GU4204	STATISTICAL INFERENCE

Introductory Courses

Students interested in statistical concepts, but who do not anticipate undertaking statistical analyses, should take STAT UN1001 Introduction to Statistical Reasoning. Students seeking an introduction to applied statistics or preparing for the concentration should take STAT UN1101 Introduction to Statistics (without calculus). Students seeking a foundation for further study of probability theory and statistical theory and methods should take STAT UN1201 Calculus-based Introduction to Statistics. Students seeking a one-semester calculus-based survey should take STAT GU4001 Introduction to Probability and Statistics. The undergraduate seminar STAT UN1202 features faculty lectures prepared with undergraduates in mind; students may attend without registering.

STAT UN1001 INTRO TO STATISTICAL REASONING. 3.00 points.

A friendly introduction to statistical concepts and reasoning with emphasis on developing statistical intuition rather than on mathematical rigor. Topics include design of experiments, descriptive statistics, correlation and regression, probability, chance variability, sampling, chance models, and tests of significance

Spring 2024: STAT UN1001

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 1001	001/13610	M W 2:40pm - 3:55pm 602 Hamilton Hall	Ronald Neath	3.00	75/86
STAT 1001	002/13674	M W 10:10am - 11:25am 903 School Of Social Work	Shaw-Hwa Lo	3.00	33/50
STAT 1001	003/13611	T Th 6:10pm - 7:25pm 602 Hamilton Hall	Victor de la Pena	3.00	66/86

STAT UN1010 Statistical Thinking For Data Science. 4.00 points.

CC/GS: Partial Fulfillment of Science Requirement

The advent of large scale data collection and the computer power to analyze the data has led to the emergence of a new discipline known as Data Science. Data Scientists in all sectors analyze data to derive business insights, find solutions to societal challenges, and predict outcomes with potentially high impact. The goal of this course is to provide the student with a rigorous understanding of the statistical thinking behind the fundamental techniques of statistical analysis used by data scientists. The student will learn how to apply these techniques to data, understand why they work and how to use the analysis results to make informed decisions. The student will gain this understanding in the classroom and through the analysis of real-world data in the lab using the programming language Python. The student will learn the fundamentals of Python and how to write and run code to apply the statistical concepts taught in the classroom

Spring 2024: STAT UN1010

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 1010	001/13612	M W 1:10pm - 2:25pm 516 Hamilton Hall	Anthony Donoghue	4.00	27/86
STAT 1010	001/13612	W 2:40pm - 3:55pm 516 Hamilton Hall	Anthony Donoghue	4.00	27/86

STAT UN1101 INTRODUCTION TO STATISTICS. 3.00 points.

Prerequisites: intermediate high school algebra. Designed for students in fields that emphasize quantitative methods. Graphical and numerical summaries, probability, theory of sampling distributions, linear regression, analysis of variance, confidence intervals and hypothesis testing. Quantitative reasoning and data analysis. Practical experience with statistical software. Illustrations are taken from a variety of fields. Data-collection/analysis project with emphasis on study designs is part of the coursework requirement

Spring 2024: STAT UN1101

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 1101	001/13613	M W 8:40am - 9:55am 517 Hamilton Hall	Alexander Clark	3.00	75/86
STAT 1101	002/13614	T Th 10:10am - 11:25am 602 Hamilton Hall	David Rios	3.00	70/86
STAT 1101	003/13615	M W 6:10pm - 7:25pm 602 Hamilton Hall	Banu Baydil	3.00	71/86

STAT UN1201 CALC-BASED INTRO TO STATISTICS. 3.00 points.

Prerequisites: one semester of calculus. Designed for students who desire a strong grounding in statistical concepts with a greater degree of mathematical rigor than in STAT W1111. Random variables, probability distributions, pdf, cdf, mean, variance, correlation, conditional distribution, conditional mean and conditional variance, law of iterated expectations, normal, chi-square, F and t distributions, law of large numbers, central limit theorem, parameter estimation, unbiasedness, consistency, efficiency, hypothesis testing, p-value, confidence intervals, maximum likelihood estimation. Serves as the pre-requisite for ECON W3412

Spring 2024: STAT UN1201

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 1201	001/13616	M W 10:10am - 11:25am 517 Hamilton Hall	Pratyay Datta	3.00	81/86
STAT 1201	002/13617	M W 8:40am - 9:55am 602 Hamilton Hall	Joyce Robbins	3.00	79/85
STAT 1201	003/13618	T Th 10:10am - 11:25am 702 Hamilton Hall	Joyce Robbins	3.00	90/86
STAT 1201	004/13619	M W 6:10pm - 7:25pm 702 Hamilton Hall	Sheela Kolluri	3.00	71/86

STAT UN1202 UNDERGRADUATE SEM/STATISTICS. 1.00 point.

Prerequisites: Previous or concurrent enrollment in a course in statistics would make the talks more accessible. Prepared with undergraduates majoring in quantitative disciplines in mind, the presentations in this colloquium focus on the interface between data analysis, computation, and theory in interdisciplinary research. Meetings are open to all undergraduates, whether registered or not. Presenters are drawn from the faculty of department in Arts and Sciences, Engineering, Public Health and Medicine

STAT GU4001 INTRODUCTION TO PROBABILITY AND STATISTICS. 3.00 points.

Prerequisites: Calculus through multiple integration and infinite sums. A calculus-based tour of the fundamentals of probability theory and statistical inference. Probability models, random variables, useful distributions, conditioning, expectations, law of large numbers, central limit theorem, point and confidence interval estimation, hypothesis tests, linear regression. This course replaces SIEO 4150

Spring 2024: STAT GU4001

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 4001	001/13625	M 6:10pm - 8:40pm 142 Uris Hall	Pratyay Datta	3.00	78/100
STAT 4001	002/13626	M W 1:10pm - 2:25pm 602 Hamilton Hall	Hammou El Barmi	3.00	68/86

Applied Statistics Concentration Courses

The applied statistics sequence, together with an introductory course, forms the concentration in applied statistics. STAT UN2102 Applied statistical computing may be used to satisfy the computing requirement for the major, and the other concentration courses may be used to satisfy the elective requirements for the major. (Students who sat STAT GU4205 Linear Regression for the major would find that they have covered essentially all of the material in STAT UN2103 Applied Linear Regression Analysis.

STAT UN2102 Applied Statistical Computing. 3.00 points.

Corequisites: An introductory course in statistic (STAT UN1101 is recommended).

Corequisites: An introductory course in statistic (STAT UN1101 is recommended). This course is an introduction to R programming. After learning basic programming component, such as defining variables and vectors, and learning different data structures in R, students will, via project-based assignments, study more advanced topics, such as conditionals, modular programming, and data visualization. Students will also learn the fundamental concepts in computational complexity, and will practice writing reports based on their data analyses

Spring 2024: STAT UN2102

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 2102	001/13620	T Th 4:10pm - 5:25pm 428 Pupin Laboratories	Alex Pijyan	3.00	79/120

STAT UN2103 APPLIED LINEAR REG ANALYSIS. 3.00 points.

Prerequisites: An introductory course in statistics (STAT UN1101 is recommended). Students without programming experience in R might find STAT UN2102 very helpful. Develops critical thinking and data analysis skills for regression analysis in science and policy settings. Simple and multiple linear regression, non-linear and logistic models, random-effects models. Implementation in a statistical package. Emphasis on real-world examples and on planning, proposing, implementing, and reporting

Spring 2024: STAT UN2103

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 2103	001/13621	M W 6:10pm - 7:25pm 717 Hamilton Hall	Daniel Rabinowitz	3.00	24/84

STAT UN2104 APPL CATEGORICAL DATA ANALYSIS. 3.00 points.

Prerequisites: STAT UN2103 is strongly recommended. Students without programming experience in R might find STAT UN2102 very helpful.

Prerequisites: STAT UN2103 is strongly recommended. Students without programming experience in R might find STAT UN2102 very helpful.

This course covers statistical models and methods for analyzing and drawing inferences for problems involving categorical data. The goals are familiarity and understanding of a substantial and integrated body of statistical methods that are used for such problems, experience in analyzing data using these methods, and proficiency in communicating the results of such methods, and the ability to critically evaluate the use of such methods. Topics include binomial proportions, two-way and three-way contingency tables, logistic regression, log-linear models for large multi-way contingency tables, graphical methods. The statistical package R will be used

Spring 2024: STAT UN2104

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 2104	001/13622	M W 8:40am - 9:55am 702 Hamilton Hall	Ronald Neath	3.00	40/86

STAT UN3105 APPLIED STATISTICAL METHODS. 3.00 points.

Prerequisites: At least one, and preferably both, of STAT UN2103 and UN2104 are strongly recommended. Students without programming experience in R might find STAT UN2102 very helpful.

Prerequisites: At least one, and preferably both, of STAT UN2103 and UN2104 are strongly recommended. Students without programming experience in R might find STAT UN2102 very helpful. This course is intended to give students practical experience with statistical methods beyond linear regression and categorical data analysis. The focus will be on understanding the uses and limitations of models, not the mathematical foundations for the methods. Topics that may be covered include random and mixed-effects models, classical non-parametric techniques, the statistical theory causality, sample survey design, multi-level models, generalized linear regression, generalized estimating equations and over-dispersion, survival analysis including the Kaplan-Meier estimator, log-rank statistics, and the Cox proportional hazards regression model. Power calculations and proposal and report writing will be discussed

STAT UN3106 APPLIED MACHINE LEARNING. 3.00 points.

Prerequisites: STAT UN2103. Students without programming experience in R might find STAT UN2102 very helpful.

Prerequisites: STAT UN2103. Students without programming experience in R might find STAT UN2102 very helpful. This course is a machine learning class from an application perspective. We will cover topics including data-based prediction, classification, specific classification methods (such as logistic regression and random forests), and basics of neural networks. Programming in homeworks will require R

Spring 2024: STAT UN3106

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 3106	001/13623	T Th 2:40pm - 3:55pm 332 Uris Hall	Alex Pijyan	3.00	51/50

Foundation Courses

The calculus-based foundation courses for the core of the statistics major. These courses are GU4203 Probability Theory, GU4204 Statistical Inference, GU4205 Linear Regression, GU4206 Statistical Computing and Introduction to Data Science, and GU4207 Elementary Stochastic processes. Ideally, students would take Probability theory or the equivalent before taking either Statistical Inference or Elementary Stochastic Processes, and would have taken Statistical Inference before, or at least concurrently with taking Linear Regression Analysis, and would have taken Linear Regression analysis before, or at least concurrently, with taking the computing and data science course. A semester of calculus should be taken before Probability, additional semesters of calculus are recommended before Statistical Inference, and a course in linear algebra before Linear Regression is strongly recommended. For the more advanced electives in stochastic processes, Probability Theory is an essential prerequisite, and many students would benefit from taking Elementary Stochastic Processes, too. Linear Regression and the computing and data science course should be taken before the advanced electives in machine learning and data science. Linear Regression is a strongly recommended prerequisite, or at least co-requisite, for the remaining advanced statistical electives.

STAT GU4203	PROBABILITY THEORY
STAT GU4204	STATISTICAL INFERENCE
STAT GU4205	LINEAR REGRESSION MODELS
STAT GU4206	STAT COMP # INTRO DATA SCIENCE
STAT GU4207	ELEMENTARY STOCHASTIC PROCESS

Advanced Statistics Courses

Advanced statistics courses combine theory with methods and practical experience in data analysis. Undergraduates enrolling in advanced statistics courses would be well-advised to have completed STAT GU4203 (Probability Theory), GU4204 (Statistical Inference), and GU4205 (Linear Regression).

STAT GU4221 TIME SERIES ANALYSIS. 3.00 points.

Prerequisites: STAT GU4205 or the equivalent. Least squares smoothing and prediction, linear systems, Fourier analysis, and spectral estimation. Impulse response and transfer function. Fourier series, the fast Fourier transform, autocorrelation function, and spectral density. Univariate Box-Jenkins modeling and forecasting. Emphasis on applications. Examples from the physical sciences, social sciences, and business. Computing is an integral part of the course

Spring 2024: STAT GU4221

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 4221	001/13633	Sa 10:10am - 12:40pm 301 Uris Hall	Franz Rembart	3.00	6/25

STAT GU4222 NONPARAMETRIC STATISTICS. 3.00 points.

CC/GS: Partial Fulfillment of Science Requirement

Prerequisites: STAT GU4204 or the equivalent.

Prerequisites: STAT GU4204 or the equivalent. Statistical inference without parametric model assumption. Hypothesis testing using ranks, permutations, and order statistics. Nonparametric analogs of analysis of variance. Non-parametric regression, smoothing and model selection

Spring 2024: STAT GU4222

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 4222	001/13678	M W 10:10am - 11:25am 501 Schermerhorn Hall	Alberto Gonzalez Sanz	3.00	0/25

STAT GU4223 MULTIVARIATE STAT INFERENCE. 3.00 points.

Prerequisites: STAT GU4205 or the equivalent.

Prerequisites: STAT GU4205 or the equivalent. Multivariate normal distribution, multivariate regression and classification; canonical correlation; graphical models and Bayesian networks; principal components and other models for factor analysis; SVD; discriminant analysis; cluster analysis

STAT GU4224 BAYESIAN STATISTICS. 3.00 points.

Prerequisites: STAT GU4204 or the equivalent.

This course introduces the Bayesian paradigm for statistical inference. Topics covered include prior and posterior distributions: conjugate priors, informative and non-informative priors; one- and two-sample problems; models for normal data, models for binary data, Bayesian linear models; Bayesian computation: MCMC algorithms, the Gibbs sampler; hierarchical models; hypothesis testing, Bayes factors, model selection; use of statistical software. Prerequisites: A course in the theory of statistical inference, such as STAT GU4204 or a course in statistical modeling and data analysis, such as STAT GU4205

Spring 2024: STAT GU4224

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 4224	001/13634	T Th 7:40pm - 8:55pm 501 Schermerhorn Hall	Dobrin Marchev	3.00	18/25

STAT GU4231 SURVIVAL ANALYSIS. 3.00 points.

Prerequisites: STAT GU4205 or the equivalent.

Prerequisites: STAT GU4205 or the equivalent. Survival distributions, types of censored data, estimation for various survival models, nonparametric estimation of survival distributions, the proportional hazard and accelerated lifetime models for regression analysis with failure-time data. Extensive use of the computer

STAT GU4232 GENERALIZED LINEAR MODELS. 3.00 points.

CC/GS: Partial Fulfillment of Science Requirement

Prerequisites: STAT GU4205 or the equivalent.

Prerequisites: STAT GU4205 or the equivalent. Statistical methods for rates and proportions, ordered and nominal categorical responses, contingency tables, odds-ratios, exact inference, logistic regression, Poisson regression, generalized linear models

STAT GU4233 Multilevel Models. 3 points.

Prerequisites: STAT GU4205 or the equivalent.

Theory and practice, including model-checking, for random and mixed-effects models (also called hierarchical, multi-level models). Extensive use of the computer to analyse data.

STAT GU4234 SAMPLE SURVEYS. 3.00 points.

Prerequisites: STAT GU4204 or the equivalent. Introductory course on the design and analysis of sample surveys. How sample surveys are conducted, why the designs are used, how to analyze survey results, and how to derive from first principles the standard results and their generalizations. Examples from public health, social work, opinion polling, and other topics of interest

Spring 2024: STAT GU4234

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 4234	001/13635	T Th 2:40pm - 3:55pm 312 Mathematics Building	Rongning Wu	3.00	2/7

STAT GU4241 STATISTICAL MACHINE LEARNING. 3.00 points.

Prerequisites: STAT GU4206.

Prerequisites: STAT GU4206. The course will provide an introduction to Machine Learning and its core models and algorithms. The aim of the course is to provide students of statistics with detailed knowledge of how Machine Learning methods work and how statistical models can be brought to bear in computer systems - not only to analyze large data sets, but to let computers perform tasks that traditional methods of computer science are unable to address. Examples range from speech recognition and text analysis through bioinformatics and medical diagnosis. This course provides a first introduction to the statistical methods and mathematical concepts which make such technologies possible

Spring 2024: STAT GU4241

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 4241	001/13636	M W 10:10am - 11:25am 503 Hamilton Hall	Samory Kpotufe	3.00	15/50

STAT GU4261 STATISTICAL METHODS IN FINANCE. 3.00 points.

Prerequisites: STAT GU4205 or the equivalent. A fast-paced introduction to statistical methods used in quantitative finance. Financial applications and statistical methodologies are intertwined in all lectures. Topics include regression analysis and applications to the Capital Asset Pricing Model and multifactor pricing models, principal components and multivariate analysis, smoothing techniques and estimation of yield curves statistical methods for financial time series, value at risk, term structure models and fixed income research, and estimation and modeling of volatilities. Hands-on experience with financial data

Spring 2024: STAT GU4261

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 4261	001/13638	Sa 10:10am - 12:40pm 501 Schermerhorn Hall	Zhiliang Ying	3.00	23/25

STAT GU4263 STAT INF/TIME-SERIES MODELLING. 3.00 points.

Prerequisites: STAT GU4204 or the equivalent. STAT GU4205 is recommended. Modeling and inference for random processes, from natural sciences to finance and economics. ARMA, ARCH, GARCH and nonlinear models, parameter estimation, prediction and filtering. This is a core course in the MS program in mathematical finance

STAT GU4291 ADVANCED DATA ANALYSIS. 3.00 points.

Prerequisites: STAT GU4205 and at least one statistics course numbered between GU4221 and GU4261. This is a course on getting the most out of data. The emphasis will be on hands-on experience, involving case studies with real data and using common statistical packages. The course covers, at a very high level, exploratory data analysis, model formulation, goodness of fit testing, and other standard and non-standard statistical procedures, including linear regression, analysis of variance, nonlinear regression, generalized linear models, survival analysis, time series analysis, and modern regression methods. Students will be expected to propose a data set of their choice for use as case study material

Spring 2024: STAT GU4291

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 4291	001/13640	F 10:10am - 12:40pm 301 Uris Hall	Gabriel Young	3.00	5/25

Actuarial Sciences Courses

Only students preparing for a career in actuarial sciences should consider the courses in this section. Such students may also be interested in courses offered through the School of Professional Studies M.S. Program in Actuarial Science, but must check with the academic advisors in their schools to know whether they are allowed to register for those courses. Students majoring in statistics and preparing for a career in actuarial science may take STAT GU4282 (Regression and Time Series Analysis) in place of the major requirement STAT GU4205 (Linear Regression Analysis).

STAT GU4281	Theory of Interest
STAT GU4282	Linear Regression and Time Series Methods

Advanced Data Science Courses

In response to the ever growing importance of “big data” in scientific and policy endeavors, the last few years have seen an explosive growth in theory, methods, and applications at the interface between computer science and statistics. The Department offers a sequence that begins with the core course STAT GU4206 (Statistical Computing and Introduction to Data Science) and continues with the advanced electives

GU4241 (Statistical Machine Learning) and GU4242 (Advanced Machine Learning), and also the advanced elective STAT GU4243 (Applied Data Science). Undergraduate students without experience in programming would likely benefit from taking the statistical computing and data science course before attempting GU4241, GU4242, or GU4243.

STAT GU4241	STATISTICAL MACHINE LEARNING
STAT GU4242	Advanced Machine Learning
STAT GU4243	APPLIED DATA SCIENCE
STAT GU4702	Exploratory Data Analysis and Visualization

Advanced Stochastic Processes Courses

The stochastic processes electives in this section have STAT GU4203 (Probability Theory) or the equivalent as prerequisites Most students would also benefit from taking STAT GU4207 (Elementary Stochastic Processes) before embarking on the more advanced stochastic processes electives.

STAT GU4262	Stochastic Processes for Finance
STAT GU4264	STOCHASTC PROCSSSES-APPLICTNS I
STAT GU4265	STOCHASTIC METHODS IN FINANCE