The course 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 Calculus-Based Introduction 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 Regression Analysis, STAT UN2104 Applied Categorical Data Analysis, STAT UN3105 Applied Statistical Methods, and STAT UN3106 Applied Data Mining. It is advisable to take STAT UN1201 Calculus-Based Introduction 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 Processes before embarking on STAT GU4262 Stochastic Processes for Finance, STAT GU4264 STOCHASTIC PROCESSES-APPLIC, 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 Statistical Computing and Introduction to Data Science, STAT GU4241, 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 departmental 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)
Richard R. Davis
Victor H. de la Peña
Andrew Gelman (with Political Science)
Ioannis Karatzas (with Mathematics)
Jingchen Liu
Shaw-Hwa Lo
David Madigan
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

John Cunningham
Samory Kpotufe
Arian Maleki
Sumit Mukherjee

Assistant Professors

Cynthia Rush
Anne van Delft

Term Assistant Professors

Marco Avella
Carsten Chong
Haoran Li
Xiaofei Shi
Thibault Vatter

Adjunct Faculty

Demissie Alemayehu
Flavio Bartmann
Mark Brown
Guy Cohen
Regina Dolgoarshinnykh
Anthony Donoghue
Hammou El Barmi
Tat Sang Fung
Xiaofu He
Margaret Holen
Irene Hueter
Ying Liu
Ka-Yi Nb
Ha Nguyen
Cristian Pasarica
David Rios
Ori Shental
Haiyuan Wang
Larry Wright
Rongning Wu

Lecturers in Discipline

Banu Baydil
Wayne Lee
Ronald Neath
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

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATH UN1101</td>
<td>CALCULUS I</td>
</tr>
<tr>
<td>MATH UN1102</td>
<td>Calculus II</td>
</tr>
<tr>
<td>MATH UN1201</td>
<td>Calculus III</td>
</tr>
<tr>
<td>MATH UN2010</td>
<td>Linear Algebra</td>
</tr>
</tbody>
</table>

One of the following five courses

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMS W1007</td>
<td>Honors Introduction to Computer Science</td>
</tr>
<tr>
<td>ENGI E1006</td>
<td>Introduction to Computing for Engineers and Applied Scientists</td>
</tr>
<tr>
<td>COMS W1005</td>
<td>Introduction to Computer Science and Programming in MATLAB</td>
</tr>
<tr>
<td>STAT UN2102</td>
<td>Applied Statistical Computing</td>
</tr>
<tr>
<td>COMS W1004</td>
<td>Introduction to Computer Science and Programming in Java</td>
</tr>
</tbody>
</table>
### Core courses in probability and statistics

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT UN1201</td>
<td>Calculus-Based Introduction to Statistics</td>
</tr>
<tr>
<td>STAT GU4203</td>
<td>PROBABILITY THEORY</td>
</tr>
<tr>
<td>STAT GU4204</td>
<td>Statistical Inference</td>
</tr>
<tr>
<td>STAT GU4205</td>
<td>Linear Regression Models</td>
</tr>
<tr>
<td>STAT GU4206</td>
<td>Statistical Computing and Introduction to Data Science</td>
</tr>
<tr>
<td>STAT GU4207</td>
<td>Elementary Stochastic Processes</td>
</tr>
</tbody>
</table>

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, 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 Introduction To Modern Analysis II.

### Concentration 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 concentration. The requirements for the concentration are as follows.

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT UN1101</td>
<td>Introduction to Statistics</td>
</tr>
<tr>
<td>STAT UN2102</td>
<td>Applied Statistical Computing</td>
</tr>
<tr>
<td>STAT UN2103</td>
<td>Applied Linear Regression Analysis</td>
</tr>
<tr>
<td>STAT UN2104</td>
<td>Applied Categorical Data Analysis</td>
</tr>
<tr>
<td>STAT UN3105</td>
<td>Applied Statistical Methods</td>
</tr>
<tr>
<td>STAT UN3106</td>
<td>Applied Data Mining</td>
</tr>
</tbody>
</table>

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

### Major in Data Science

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 Statistics Department and the Department of Computer Science have responded with a joint-major that emphasizes the interface between the disciplines.

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:

**Mathematical Prerequisites**

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATH UN1101</td>
<td>CALCULUS I</td>
</tr>
<tr>
<td>MATH UN1102</td>
<td>Calculus II</td>
</tr>
<tr>
<td>MATH UN1201</td>
<td>Calculus III</td>
</tr>
<tr>
<td>MATH UN2010</td>
<td>Linear Algebra</td>
</tr>
</tbody>
</table>

**Statistics Required Courses**

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT UN1201</td>
<td>Calculus-Based Introduction to Statistics</td>
</tr>
<tr>
<td>STAT GU4203</td>
<td>PROBABILITY THEORY</td>
</tr>
<tr>
<td>STAT GU4204</td>
<td>Statistical Inference</td>
</tr>
<tr>
<td>STAT GU4205</td>
<td>Linear Regression Models</td>
</tr>
<tr>
<td>STAT GU4241</td>
<td>or COMS W4771 Machine Learning</td>
</tr>
</tbody>
</table>

**Computer Science Introductory Courses**

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 Introduction to Computing for Engineers and Applied Scientists
- COMS W1007 Honors Introduction to Computer Science

### Statistics Electives

Select two of the following courses:

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT UN3106</td>
<td>Applied Data Mining</td>
</tr>
<tr>
<td>STAT GU4206</td>
<td>Statistical Computing and Introduction to Data Science</td>
</tr>
<tr>
<td>STAT GU4243</td>
<td>Applied Data Science</td>
</tr>
<tr>
<td>STAT GU224</td>
<td>BAYESIAN STATISTICS</td>
</tr>
<tr>
<td>STAT GU4242</td>
<td>Advanced Machine Learning</td>
</tr>
</tbody>
</table>

### Computer Science Required Courses

Select three of the following courses:

- COMS W3203 Discrete Mathematics: Introduction to Combinatorics and Graph Theory
- CSOR W4231 Analysis of Algorithms I
- COMS W3261 Computer Science Theory
- COMS W4236 Introduction to Computational Complexity
- COMS W4252 Introduction to Computational Learning Theory
- COMS W4111 INTRODUCTION TO DATABASES
- COMS W4130 Principles and Practice of Parallel Programming

Any COMS W47xx course EXCEPT W4771

### Major in Economics-Statistics

Please read Requirements for all Economics Majors, Concentrators, and Interdepartmental Majors in the Economics (http://bulletin.columbia.edu/general-studies/undergraduates/majors-concentrations/economics/) section of this Bulletin.

The major in Economics-Statistics provides students with a grounding in economic theory comparable to that of the general economics major, but also exposes students to a more rigorous and extensive statistics training. This program is recommended for students with strong quantitative skills and for those contemplating graduate studies in economics.

Two advisers are assigned for the interdepartmental major, one in the Department of Economics and one in the Department of Statistics. The economics adviser can only advise on economics requirements and the statistics adviser can only advise on statistics requirements.
Students should be aware of the rules regarding the use of the Pass/D/Fail option. Courses in which a grade of D has been received do not count toward the major requirements.

The economics-statistics major requires a total of 59 points: 29 in economics, 15 points in statistics, 12 points in mathematics, and 3 points in computer science, as follows:

**Economics Core Courses**
Complete the Economics core courses.

**Economics Electives**
Select three electives at the 3000-level or above, of which no more than one may be a Barnard course.

**Mathematics**
Select one of the following sequences:

- MATH UN1101 - MATH UN1102 - MATH UN1201 - MATH UN2010
  - CALCULUS I and Calculus II and Calculus III and Linear Algebra

- MATH UN1101 - MATH UN1102 - MATH UN1205 - MATH UN2010
  - CALCULUS I and Calculus II and Accelerated Multivariable Calculus and Linear Algebra

- MATH UN1207 - MATH UN1208
  - Honors Mathematics A and Honors Mathematics B

**Statistics**
Select one of the following sequences:

- STAT UN1201
  - Calculus-Based Introduction to Statistics

- STAT GU4203
  - PROBABILITY THEORY

- STAT GU4204
  - Statistical Inference

- STAT GU4205
  - Linear Regression Models

One elective from among courses numbered STAT GU4206 through GU4266.

**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

- COMS W1007
  - Honors Introduction to Computer Science

- ENGI E1006
  - Introduction to Computing for Engineers and Applied Scientists

- STAT UN2102
  - Applied Statistical Computing

**Seminar**

- ECON GU4918
  - Seminar In Econometrics

**Students who declared before Spring 2014:**
The requirements for this program were modified in 2014. Students who declared this program before Spring 2014 should contact the director of undergraduate studies for the department in order to confirm their options for major requirements.

**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

**Statistics required courses**

- STAT UN1201
  - Calculus-Based Introduction to Statistics

- STAT GU4203
  - PROBABILITY THEORY

- STAT GU4204
  - Statistical Inference

- STAT GU4205
  - Linear Regression Models

**And select one of the following courses:**

- STAT GU4207
  - Elementary Stochastic Processes

- STAT GU4262
  - Stochastic Processes for Finance

- STAT GU4264
  - Stochastic Processes-Applic

- 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
  - Introduction to Computing for Engineers and Applied Scientists

- COMS W1007
  - Honors Introduction to Computer Science

- 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 Introduction to the Mathematics 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 Introduction To Modern Analysis II.
• Students preparing for a career in actuarial science are encouraged to replace STAT GU4205 Linear Regression Models with STAT GU4282, and to take among their electives STAT GU4281 Theory of Interest.

Major in Political Science–Statistics

The interdepartmental major of political science–statistics is designed for students who desire an understanding of political science to pursue advanced study in this field and who also wish to have at their command a broad range of sophisticated statistical tools to analyze data related to social science and public policy research.

Students should be aware of the rules regarding the use of the Pass/D/Fail option. Courses in which a grade of D has been received do not count toward the major requirements.

Political science–statistics students are eligible for all prizes reserved for political science majors.

The political science-statistics major requires a minimum of 15 courses in political science, statistics, and mathematics, to be distributed as follows:

Political Science

Students must choose a primary subfield to study. Within the subfield, students must take a minimum of three courses, including the subfield’s introductory course. The subfields and their corresponding introductory courses are as follows:

American Politics:
POLS UN1201 Introduction To American Government and Politics

Comparative Politics:
POLS UN1501 Introduction to Comparative Politics

International Relations:
POLS UN1601 INTERNATIONAL POLITICS

Political Theory:
POLS UN1101 Political Theory I

Additionally, students must take a 4-point seminar in their primary subfield.

Research Methods

Students must take the following two research methods courses:
POLS GU4710 PRINC OF QUANT POL RESEARCH 1
POLS UN3704 RESEARCH DESIGN: DATA ANALYSIS
POLS GU4712 PRINC OF QUANT POL RESEARCH 2

Statistics

Select one of the following two sequences.

Sequence recommended for students preparing for graduate study in statistics.

MATH UN1101 CALCULUS I
MATH UN1102 Calculus II
MATH UN2010 Linear Algebra
STAT UN1201 Calculus-Based Introduction to Statistics
STAT GU4203 PROBABILITY THEORY
STAT GU4204 Statistical Inference
STAT GU4205 Linear Regression Models
STAT GU4206 Statistical Computing and Introduction to Data Science

Students taking the first track may replace the Mathematics prerequisites with both of MATH UN1207 and MATH UN1208

Sequence recommended for students preparing to apply statistical methods in the social sciences.

STAT UN1101 Introduction to Statistics
STAT UN2102 Applied Statistical Computing
STAT UN2103 Applied Linear Regression Analysis
STAT UN2104 Applied Categorical Data Analysis
STAT UN3105 Applied Statistical Methods
STAT UN3106 Applied Data Mining

Statistics elective:

Students must take an approved elective in a statistics or a quantitatively oriented course in a social science.

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 Introduction to Statistical Reasoning. 3 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.

Fall 2020: STAT UN1001
Course Number 001/12833 Times/Location Instructor Points Enrollment
STAT 1001 T Th 10:10am - 11:26am Guy Cohen 3 150/150
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.

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.

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.

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.

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.

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 points.
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 recursion, conditionals, modular programming, and data visualization. Students will also learn the fundamental concepts in computational complexity, and will practice writing reports based on their statistical analyses.

STAT UN2103 Applied Linear Regression Analysis. 3 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.

Fall 2020: STAT UN2103
<table>
<thead>
<tr>
<th>Course Number</th>
<th>Section/Call Number</th>
<th>Times/Location</th>
<th>Instructor</th>
<th>Points</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2103</td>
<td>001/12883</td>
<td>M W 6:10pm - 7:25pm</td>
<td>Daniel</td>
<td>3</td>
<td>57/60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Online Only</td>
<td>Rabinowitz</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

STAT UN2104 Applied Categorical Data Analysis. 3 points.
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.

STAT UN3105 Applied Statistical Methods. 3 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.
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.

Fall 2020: STAT UN3105
<table>
<thead>
<tr>
<th>Course Number</th>
<th>Section/Call Number</th>
<th>Times/Location</th>
<th>Instructor</th>
<th>Points</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN3105</td>
<td>001/12884</td>
<td>T Th 11:40am - 12:55pm</td>
<td>Wayne Lee</td>
<td>3</td>
<td>27/60</td>
</tr>
</tbody>
</table>

STAT UN3106 Applied Data Mining. 3 points.
Prerequisites: STAT UN2103. Students without programming experience in R might find STAT UN2102 very helpful.
This course will be taught as a machine learning class. 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; students without programming experience in R might find STAT UN2102 helpful.

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.

<table>
<thead>
<tr>
<th>Course</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GU4203</td>
<td>PROBABILITY THEORY</td>
</tr>
<tr>
<td>GU4204</td>
<td>Statistical Inference</td>
</tr>
<tr>
<td>GU4205</td>
<td>Linear Regression Models</td>
</tr>
<tr>
<td>GU4206</td>
<td>Statistical Computing and Introduction to</td>
</tr>
<tr>
<td></td>
<td>Data Science</td>
</tr>
<tr>
<td>GU4207</td>
<td>Elementary Stochastic Processes</td>
</tr>
</tbody>
</table>
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 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.

Fall 2020: STAT GU4221

<table>
<thead>
<tr>
<th>Course Number</th>
<th>Section/Call Number</th>
<th>Times/Location</th>
<th>Instructor</th>
<th>Points</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT 4221</td>
<td>001/12556</td>
<td>T Th 1:10pm - 2:25pm</td>
<td>Flavio</td>
<td>3</td>
<td>5/55</td>
</tr>
<tr>
<td></td>
<td>Online Only</td>
<td></td>
<td>Bartmann</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

STAT GU4224 BAYESIAN STATISTICS. 3.00 points.
Prerequisites: STAT GU4204 or the equivalent.
Prerequisites: STAT GU4204 or the equivalent. Bayesian data analysis: building, fitting, evaluating and improving probability models. Prior information, hierarchical models and combining information. Linear and nonlinear models. Simulation of fake data and evaluation of methods. Computing using R and Stan.

Fall 2020: STAT GU4224

<table>
<thead>
<tr>
<th>Course Number</th>
<th>Section/Call Number</th>
<th>Times/Location</th>
<th>Instructor</th>
<th>Points</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT 4224</td>
<td>001/12490</td>
<td>M W 6:10pm - 7:25pm</td>
<td>Neath</td>
<td>3.00</td>
<td>36/35</td>
</tr>
<tr>
<td></td>
<td>Online Only</td>
<td></td>
<td>Barmann</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

STAT GU4231 Survival Analysis. 0 points.
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.

Fall 2020: STAT GU4231

<table>
<thead>
<tr>
<th>Course Number</th>
<th>Section/Call Number</th>
<th>Times/Location</th>
<th>Instructor</th>
<th>Points</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT 4231</td>
<td>001/12495</td>
<td>T Th 6:10pm - 7:25pm</td>
<td>Hamroun</td>
<td>3</td>
<td>12/35</td>
</tr>
<tr>
<td></td>
<td>Online Only</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.

Fall 2020: STAT GU4233

<table>
<thead>
<tr>
<th>Course Number</th>
<th>Section/Call Number</th>
<th>Times/Location</th>
<th>Instructor</th>
<th>Points</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT 4233</td>
<td>001/12496</td>
<td>Sa 10:10am - 12:40pm</td>
<td>Haoran</td>
<td>3</td>
<td>17/35</td>
</tr>
<tr>
<td></td>
<td>Online Only</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

STAT GU4234 Sample Surveys. 3 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.

Fall 2020: STAT GU4234

<table>
<thead>
<tr>
<th>Course Number</th>
<th>Section/Call Number</th>
<th>Times/Location</th>
<th>Instructor</th>
<th>Points</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT 4234</td>
<td>001/12492</td>
<td>M W 1:10pm - 2:25pm</td>
<td>Rongning</td>
<td>3</td>
<td>12/25</td>
</tr>
<tr>
<td></td>
<td>Online Only</td>
<td></td>
<td>Wu</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

STAT GU4261 Statistical Methods in Finance. 3 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.

Fall 2020: STAT GU4261

<table>
<thead>
<tr>
<th>Course Number</th>
<th>Section/Call Number</th>
<th>Times/Location</th>
<th>Instructor</th>
<th>Points</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT 4261</td>
<td>001/12494</td>
<td>F 8:40am - 11:25am</td>
<td>El Barmouh</td>
<td>3</td>
<td>25/25</td>
</tr>
<tr>
<td></td>
<td>Online Only</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

STAT GU4263 Statistical Inference and Time Series Modelling. 3 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.

Fall 2020: STAT GU4263

<table>
<thead>
<tr>
<th>Course Number</th>
<th>Section/Call Number</th>
<th>Times/Location</th>
<th>Instructor</th>
<th>Points</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT 4263</td>
<td>001/12495</td>
<td>T Th 6:10pm - 7:25pm</td>
<td>Haoran</td>
<td>3</td>
<td>12/35</td>
</tr>
<tr>
<td></td>
<td>Online Only</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAT 4263</td>
<td>002/12496</td>
<td>Sa 10:10am - 12:40pm</td>
<td>Haoran</td>
<td>3</td>
<td>17/35</td>
</tr>
<tr>
<td></td>
<td>Online Only</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
STAT GU4291 Advanced Data Analysis. 3 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.

Fall 2020: STAT GU4291

<table>
<thead>
<tr>
<th>Course Number</th>
<th>Section/Call Number</th>
<th>Times/Location</th>
<th>Instructor</th>
<th>Points</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT 4291</td>
<td>001/12515</td>
<td>F 6:10pm - 8:40pm Online Only</td>
<td>Demissie Alemayehu</td>
<td>3</td>
<td>25/25</td>
</tr>
</tbody>
</table>

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

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
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 STOCHASTIC PROCESSES-APPLIC
STAT GU4265 Stochastic Methods in Finance