

STATISTICS

Department of Statistics:

Department website: <https://stat.columbia.edu/>

Office location: 1005 SSW (1255 Amsterdam Avenue);

Office contact: 212-851-2132

Director of Undergraduate Studies:

Ronald Neath, 615 Watson (612 West 115th Street), 212-853-1398;

rcn2112@columbia.edu

Director of Academic Administration:

Dood Kalicharan, 1003 SSW (1255 Amsterdam), 212-853-1398;

dk@stat.columbia.edu

The Study of Statistics

Statistics is the art and science of study design and data analysis.

Probability theory is the mathematical foundation for the study of statistical methods and for the modeling of random phenomena. The Statistics major builds on a foundation in probability and statistical theory to provide practical training in statistical methods, study design, and data analysis. The Statistics major is an appropriate background for graduate study in statistics, social science, epidemiology and public health, genetics, and economics; or for professional work in such areas as drug development, health policy, marketing, opinion polling, insurance, banking and finance, and government.

The Department offers several introductory courses.

- Students interested in learning statistical concepts, with a goal of being educated consumers of statistics, should take STAT UN1001 INTRO TO STATISTICAL REASONING. This course is designed for students who have taken a pre-calculus course, and the focus is on general principles.
- Students seeking an introduction to applied statistics should take STAT UN1101 INTRODUCTION TO STATISTICS. This course is designed for students who wish to learn to conduct statistical analyses, but do not have a background in calculus; the focus is on the implementation of statistical methods, rather than the underlying theory. It is recommended for pre-med students, and students considering the applied track of the statistics minor.
- Students seeking a more mathematically rigorous treatment of the subject should take STAT UN1201 CALC-BASED INTRO TO STATISTICS. This course is designed for students who have taken a semester of college calculus or the equivalent, and the focus is on preparation for further study in probability and statistical theory and methods. It is recommended for students considering the statistics major, or the theoretical track of the minor.
- Students seeking a one-semester calculus-based survey of probability and statistical theory 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 an abridged version of the material covered in the two-semester sequence STAT GU4203 PROBABILITY THEORY and STAT GU4204 STATISTICAL INFERENCE. While some

mathematically mature students may take the 4203–4204 sequence as an introduction to the field, it is generally recommended that students prepare for it by taking STAT UN1201 CALC-BASED INTRO TO STATISTICS.

The Department offers a Major in Statistics, a Minor in Statistics, and interdisciplinary majors with Computer Science, Economics, Mathematics, and Political Science. The major consists of mathematical and computational prerequisites, an introductory course, five core courses in probability and theoretical and applied statistics, plus three electives. The training provided by the undergraduate major is comparable to a master's degree in statistics. The applied track of the minor is suitable for students preparing for academic or professional work in fields where data analysis skills are valued; it can be completed without the mathematical prerequisite required for the major. Students who are more mathematically inclined can opt for the theoretical track, and complete a minor by taking courses from the core sequence of the statistics major.

Student Advising

Statistics Major and Minor Advising:

Ronald Neath, 615 Watson (612 West 115th Street); 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, 615 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, 615 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, 615 Watson;

212-853-1398; rcn2112@columbia.edu

Political Science - Statistic Major Advising:

Political Science: Andrew Gelman, 1016 SSW (1255 Amsterdam);

gelman@stat.columbia.edu

Statistics: Ronald Neath, 615 Watson;

212-853-1398; rcn2112@columbia.edu

Enrolling in Classes

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 minor courses: [STAT UN2103](#) APPLIED LINEAR REG ANALYSIS, [STAT UN2104](#) APPL CATEGORICAL DATA ANALYSIS, [STAT UN3104](#) Applied Bayesian 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-APPLIED](#), 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.

Preparing for Graduate Study

The [BA/MA option](#) allows current Columbia undergraduate students (Columbia College, SEAS, the School of General Studies, and Barnard) the opportunity to complete both the bachelor's degree and the master's degree (BA/MA) in a shorter period of time, thus providing an option that is financially advantageous. The BA/MA in Statistics is open to students from all majors.

Coursework Taken Outside of Columbia

Coursework in fulfillment of a major or minor must be taken at Columbia University unless explicitly noted here and/or expressly permitted by the Director of Undergraduate Studies. Exceptions or substitutions permitted by the Director of Undergraduate Studies should be confirmed in writing by email to the student.

Advanced Placement

Columbia College and the School of General Studies award 3 points of credit for a score of 5 on the AP statistics exam. Students who are required to take STAT UN1101 for their major should check with their major advisor to determine whether this credit provides exemption from their requirement.

Students pursuing a major that requires STAT UN1201 should plan to take that course at Columbia, even if they scored a 5 on the AP statistics exam. AP credit cannot be used to satisfy a requirement for STAT UN1201.

Transfer Courses

When students transfer to Columbia from other institutions, their coursework at their previous institution must first be considered by their school in order to be evaluated for degree credit (e.g., to confirm that the courses will count toward the 124 points of credit that every student is required to complete for the B.A. degree). Only after that degree credit is confirmed, departments may consider whether those courses can also be used to fulfill specific degree requirements toward a major or minor.

No more than two DUS-approved STAT courses toward a Statistics major may be fulfilled with transfer credit.

Not more than one DUS-approved STAT course toward a Statistics joint major or a Statistics minor may be fulfilled with transfer credit.

Study Abroad Courses

Classes taken abroad through Columbia-led programs (i.e., those administered by Columbia's Center for Undergraduate Global Engagement and taught by Columbia instructors) are treated as Columbia courses, equivalent to those taken on the Morningside Heights campus. If they are not explicitly listed by the department as fulfilling requirements in the major or minor, the DUS will need to confirm that they can be used toward requirements in the major/minor.

Classes taken abroad through other institutions and programs are treated as transfer credit to Columbia, and are subject to the same policies as other transfer courses, including limits on the number of approved STAT course that can be applied to the major/minor.

Summer Courses

Summer courses at Columbia are offered through the School of Professional Studies. Courses taken in a Summer Term may be used toward requirements for the Statistics major or minor.

More general policies about Summer coursework can be found in the Academic Regulations section of this Bulletin.

Undergraduate Research

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. Students work under the supervision of Statistics Department faculty mentors. The internship provides a summer housing allowance and a stipend. Applicants should send a brief statement of interest and a copy of their transcript to the Statistics DUS by the end of March to be considered. If summer project descriptions are posted on the Department's website, please indicate your preferred project(s) in your statement 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 coursework when asking for a project.

Department Honors

Students are considered for department honors on the basis of GPA and the comprehensiveness and difficulty of their coursework in Statistics and related disciplines. Generally, no more than 10% of graduating majors receive departmental honors in a given academic year.

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

Guidance for Undergraduate Students in the Department

Program Planning for all Students

Selecting a first course in Statistics:

- Students interested in learning statistical concepts, with a goal of being educated consumers of statistics, should take STAT UN1001 INTRO TO STATISTICAL REASONING. This course is designed for students who have taken a pre-calculus course, and the focus is on general principles.
- Students seeking an introduction to applied statistics should take STAT UN1101 INTRODUCTION TO STATISTICS. This course is designed for students who wish to learn to conduct statistical analyses, but do not have a background in calculus; the focus is on the implementation of statistical methods, rather than the underlying theory. It is recommended for pre-med students, and students considering the applied track of the statistics minor.
- Students seeking a more mathematically rigorous treatment of the subject should take STAT UN1201 CALC-BASED INTRO TO STATISTICS. This course is designed for students who have taken a semester of college calculus or the equivalent, and the focus is on preparation for further study in probability and statistical theory and methods. It is recommended for students considering the statistics major, or the theoretical track of the minor.
- Students seeking a one-semester calculus-based survey of probability and statistical theory 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 an abridged version of the material covered in the two-semester sequence STAT GU4203 PROBABILITY THEORY and STAT GU4204 STATISTICAL INFERENCE. While some mathematically mature students may take the 4203–4204 sequence as an introduction to the field, it is generally recommended that students prepare for it by taking STAT UN1201 CALC-BASED INTRO TO STATISTICS.

Course Numbering Structure

The 1000-level courses (STAT UN1001, STAT UN1101 and STAT UN1201) are introductory courses. Most students will begin their study of statistics with one of these three courses.

The 2000-level courses (STAT UN2102, STAT UN2103 and STAT UN2104) are courses in computational and applied statistics, with STAT UN1101 or STAT UN1201 as a prerequisite. These are important courses in the minor program; students pursuing a statistics major will learn this material by studying the more mathematical treatment given in the 4000-level courses.

The 3000-level courses (STAT UN3104, STAT UN3105 and STAT UN3106) introduce more specialized statistical methods which build on the material introduced in STAT UN2102 and STAT UN2103. While the statistical methods covered in these courses can be quite advanced, the mathematical level remains modest. Again, these courses are part of the minor curriculum, and students completing a statistics major will learn this material elsewhere in the statistics curriculum.

STAT GU4001 is a one-semester calculus-based course in probability and statistics, intended for students who seek a mathematically rigorous

course, but do not intend to major or minor in statistics (for most, this will be a terminal course).

The 4200-level courses are intended for students majoring in statistics and related disciplines.

Courses numbered 4203 through 4207 introduce fundamental material in probability theory, statistical inference, data analysis, and statistical computing; these courses comprise the core of the statistics major.

Courses numbered 4221 through 4234 cover specialized statistical data analysis techniques, and are possible electives for students in the statistics major.

Courses numbered 4241 through 4243 introduce modern tools in machine learning and data science.

Courses numbered between 4261 and 4265 cover statistical and probabilistic theory and methods in modern finance.

Undergraduate Programs of Study

Major in Statistics

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 major requires 14 courses, as follows:

Mathematics Prerequisite (four courses)

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

Computer Science Requirement (one course). Choose one of the following

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

Statistical prerequisite (one course)

STAT UN1201	CALC-BASED INTRO TO STATISTICS
-------------	--------------------------------

Core courses in probability and statistics (five courses):

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

Electives (three courses):

An approved selection of three advanced courses in mathematics, statistics, applied mathematics, industrial engineering and operations research, computer science, or an advanced quantitative course in a social science. At least one elective must be a Statistics Department course numbered between 4221 and 4291

- The mathematics prerequisite can also be satisfied by taking the Honors Mathematics A and B sequence, MATH UN1207 and MATH UN1208.
- 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.

Major in Data Science

In response to the ever increasing importance of “big data” in scientific and policy endeavors, the last few years have seen an explosive growth in theory, methods, and applications of AI and machine learning. The Department of Computer Science and the Department of Statistics jointly offer a Data Science major that emphasizes the interface between the two disciplines.

The major requires 18 courses, as follows. (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.)

Notes:

The mathematics prerequisite can also be satisfied by taking the Honors Mathematics A and B sequence, MATH UN1207 and MATH UN1208.

Mathematical Prerequisites

MATH UN1101	CALCULUS I
MATH UN1102	CALCULUS II
MATH UN1201 or MATH UN1205	CALCULUS III ACCELERATED MULTIVARIABLE CALC
MATH UN2010	LINEAR ALGEBRA

Statistics Required Courses

STAT UN1201	CALC-BASED INTRO TO STATISTICS
STAT GU4203	PROBABILITY THEORY
STAT GU4204	STATISTICAL INFERENCE
STAT GU4205	LINEAR REGRESSION MODELS
STAT GU4241 or COMS W4771	STATISTICAL MACHINE LEARNING MACHINE LEARNING

Statistics Electives

Select two of the following courses:

STAT UN3106	APPLIED MACHINE LEARNING
STAT GU4206	STAT COMP # INTRO DATA SCIENCE
STAT GU4243	APPLIED DATA SCIENCE
STAT GU4224	BAYESIAN STATISTICS
STAT GU4242	Advanced Machine Learning

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	INTRO TO COMP FOR ENG/APP SCI
COMS W1007	

And select one of the following courses:

COMS W3134	Data Structures in Java
COMS W3136	ESSENTIAL DATA STRUCTURES
COMS W3137	HONORS DATA STRUCTURES # ALGOL

Computer Science Required Courses

COMS W3203	DISCRETE MATHEMATICS
CSOR W4231	ANALYSIS OF ALGORITHMS I

Computer Science Electives

Select three of the following courses:

COMS W3261	COMPUTER SCIENCE THEORY
COMS W4236	INTRO-COMPUTATIONAL COMPLEXITY
COMS W4252	INTRO-COMPUTATIONAL LEARN THRY
COMS W4111	INTRODUCTION TO DATABASES

COMS W4130
Any COMS W47xx course EXCEPT W4771

COMS W1005 Introduction to Computer Science and Programming in MATLAB

COMS W1007
ENGI E1006 INTRO TO COMP FOR ENG/APP SCI
STAT UN2102 Applied Statistical Computing

Seminar
ECON GU4918 SEMINAR IN ECONOMETRICS

Major in Economics-Statistics

Please read *Requirements for all Economics Majors, Concentrators, and Interdepartmental Majors* in the [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 18 courses, as follows:

Notes:

- The mathematics prerequisite can also be satisfied by taking the Honors Mathematics A and B sequence, MATH UN1207 and MATH UN1208.

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 CALC and LINEAR ALGEBRA

MATH UN1207 & MATH UN1208
HONORS MATHEMATICS A and HONORS MATHEMATICS B

Statistics

STAT UN1201 CALC-BASED INTRO 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

Major in Mathematics-Statistics

This major program is designed to prepare students 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.

The major requires 14 courses, as follows. (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.):

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
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 STOCHASTIC PROCESSES-APPLICIIONS 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.

- Notes:
- The mathematics prerequisite can also be satisfied by taking the Honors Mathematics A and B sequence, MATH UN1207 and MATH UN1208.
- Students preparing for doctoral study in mathematics or statistics are encouraged to take MATH GU4061 INTRO MODERN ANALYSIS I and MATH GU4062 INTRO MODERN ANALYSIS II.

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.

The political science-statistics major requires a minimum of 6 courses in political science, and 7 or 8 courses in statistics & 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 INTRO TO AMERICAN POLITICS

Comparative Politics:

POLS UN1501 INTRO 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

or 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 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

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

or

Sequence recommend 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 REG ANALYSIS

STAT UN2104 APPL CATEGORICAL DATA ANALYSIS

STAT UN3105 APPLIED STATISTICAL METHODS

STAT UN3106 APPLIED MACHINE LEARNING

Statistics elective:

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

Minor in Statistics

The minor requires five courses, distributed as follows.

Students should select one of the following two tracks.

The requirements for the Applied track of the statistics minor are:

Introduction to statistics (one course): Choose one of the following

STAT UN1101 INTRODUCCION TO STATISTICS

STAT UN1201 CALC-BASED INTRO TO STATISTICS

Applied statistics core (two courses): Take both of the following

STAT UN2102 Applied Statistical Computing

STAT UN2103 APPLIED LINEAR REG ANALYSIS

Statistics electives (two courses): Choose any two of the following

STAT UN2104 APPL CATEGORICAL DATA ANALYSIS

STAT UN3104 Applied Bayesian Analysis

STAT UN3105 APPLIED STATISTICAL METHODS

STAT UN3106 APPLIED MACHINE LEARNING

The requirements for the Theoretical track are:

Introduction to Statistics (one course)

STAT UN1201 CALC-BASED INTRO TO STATISTICS

Probability and statistics core (three courses)

STAT GU4203 PROBABILITY THEORY

STAT GU4204 STATISTICAL INFERENCE

STAT GU4205 LINEAR REGRESSION MODELS

Elective (one course)

One additional STAT course numbered 4206 through 4261

For students who entered Columbia in or before the 2023-24 academic year

Concentrations are available to students who entered Columbia in or before the 2023-2024 academic year. The requirements for the Bachelor of Arts degree, and the role of the concentration in those requirements, can be found in the Academic Requirements section of the Bulletin dated the academic year when the student matriculated at Columbia and the Bulletin dated the academic year when the student was a sophomore and declared programs of study.

Concentrations are not available to students who entered Columbia in or after Fall 2024.

Concentration in Statistics

The concentration requires 6 courses in statistics, as follows.

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.

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 nominally required for the concentration by approved Statistics Department courses.)

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 2025: STAT UN1001

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 1001	001/13986	M W 2:40pm - 3:55pm 602 Hamilton Hall	Victor de la Pena	3.00	80/86
STAT 1001	002/13987	T Th 10:10am - 11:25am 326 Uris Hall	Pratyay Datta	3.00	49/86
STAT 1001	003/13988	T Th 6:10pm - 7:25pm 428 Pupin Laboratories	Anthony Donoghue	3.00	76/86

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

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 2025: STAT UN1101

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 1101	001/13989	T Th 10:10am - 11:25am 501 Northwest Corner	Wayne Lee	3.00	88/160
STAT 1101	002/13991	M W 6:10pm - 7:25pm 614 Schermerhorn Hall	Ha Nguyen	3.00	103/120

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 2025: STAT UN1201

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 1201	001/13992	M W 10:10am - 11:25am 717 Hamilton Hall	Hammou El Barmi	3.00	81/86
STAT 1201	002/13993	M W 10:10am - 11:25am 602 Hamilton Hall	Joyce Robbins	3.00	70/85
STAT 1201	003/13994	T Th 10:10am - 11:25am 717 Hamilton Hall	Joyce Robbins	3.00	74/86
STAT 1201	004/13995	M W 6:10pm - 7:25pm 417 International Affairs Bldg	Banu Baydil	3.00	138/180

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 2025: STAT GU4001

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 4001	001/14003	T Th 6:10pm - 7:25pm 329 Pupin Laboratories	Cristian Pasarica	3.00	77/100
STAT 4001	002/14004	M W 1:10pm - 2:25pm 517 Hamilton Hall	Sumit Mukherjee	3.00	39/86

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 2025: STAT UN2102

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 2102	001/13996	T Th 4:10pm - 5:25pm 428 Pupin Laboratories	Alex Pijyan	3.00	97/120

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 2025: STAT UN2103

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 2103	001/13998	M W 6:10pm - 7:25pm 717 Hamilton Hall	Daniel Rabinowitz	3.00	53/85

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 2025: STAT UN2104

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 2104	001/13999	M W 8:40am - 9:55am 717 Hamilton Hall	Ronald Neath	3.00	45/86

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 2025: STAT UN3106

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 3106	001/14000	T Th 2:40pm - 3:55pm 602 Hamilton Hall	Wayne Lee	3.00	36/86

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 2025: STAT GU4221

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 4221	001/14018	Sa 10:10am - 12:40pm 301 Uris Hall	Franz Rembart	3.00	14/45

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 2025: STAT GU4222

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 4222	001/14019	M W 5:30pm - 6:45pm 903 School Of Social Work	Arian Maleki	3.00	1/25

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 a course in statistical modeling and data analysis, such as STAT GU4205

Spring 2025: STAT GU4224

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 4224	001/14020	T Th 7:40pm - 8:55pm 301 Uris Hall	Dobrin Marchev	3.00	15/30

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

Spring 2025: STAT GU4231

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 4231	001/17402	F 1:10pm - 3:40pm 903 School Of Social Work	Zhiliang Ying	3.00	1/15

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.

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 2025: STAT GU4234

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 4234	001/14021	T Th 2:40pm - 3:55pm 312 Mathematics Building	Rongning Wu	3.00	6/25

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 2025: STAT GU4241

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 4241	001/14024	M W 10:10am - 11:25am 140 Uris 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 2025: STAT GU4261

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 4261	001/14031	Sa 10:10am - 12:40pm 309 Havemeyer Hall	Zhiliang Ying	3.00	18/25

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 2025: STAT GU4291

Course Number	Section/Call Number	Times/Location	Instructor	Points	Enrollment
STAT 4291	001/14034	F 10:10am - 12:40pm 309 Havemeyer 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	STOCHASTIC PROCESSES-APPLICANTS I
STAT GU4265	STOCHASTIC METHODS IN FINANCE
