Department of Statistics and Data Science

Christopher R. Genovese, Department Head
Rebecca Nugent, Director of Undergraduate Studies
Samantha Nielsen, Lead Academic Advisor
Glenn Clune, Academic Advisor
Email: statadvising@stat.cmu.edu
Location: Baker Hall 132
www.stat.cmu.edu/

Overview

Uncertainty is inescapable: randomness, measurement error, deception, and incomplete or missing information complicate all our lives. Statistics is the science and art of making predictions and decisions in the face of uncertainty. Statistical issues are central to big questions in public policy, law, medicine, industry, computing, technology, finance, and science. Indeed, the tools of Statistics apply to problems in almost every area of human activity where data are collected.

Statisticians must master diverse skills in computing, mathematics, decision making, forecasting, interpretation of complicated data, and design of meaningful comparisons. Moreover, statisticians must learn to collaborate effectively with people in other fields and, in the process, to understand the substance of these other fields. For all these reasons, Statistics students are highly sought-after in the marketplace.

Recent Statistics majors at Carnegie Mellon have taken jobs at leading companies in many fields, including the National Economic Research Association, Boeing, Morgan Stanley, Deloitte, Rosetta Marketing Group, Nielsen, Proctor and Gamble, Accenture, and Goldman Sachs. Other students have taken research positions at the National Security Agency, the U.S. Census Bureau, and the Science and Technology Policy Institute or worked for Teach for America. Many of our students have also gone on to graduate study at some of the top programs in the country including Carnegie Mellon, the Wharton School at the University of Pennsylvania, Johns Hopkins, University of Michigan, Stanford University, Harvard University, Duke University, Emory University, Yale University, Columbia University, and Georgia Tech.

The Department and Faculty

The Department of Statistics and Data Science at Carnegie Mellon University is world-renowned for its contributions to statistical theory and practice. Research in the department runs the gamut from pure mathematics to the hottest frontiers of science. Current research projects are helping make fundamental advances in neuroscience, cosmology, public policy, finance, and genetics.

The faculty members are recognized around the world for their expertise and have garnered many prestigious awards and honors. (For example, three members of the faculty have been awarded the COPSS medal, the highest honor given by professional statistical societies.) At the same time, the faculty is firmly dedicated to undergraduate education. The entire faculty, junior and senior, teach courses at all levels. The faculty are accessible and are committed to involving undergraduates in research.

The Department augments all these strengths with a friendly, energetic working environment and exceptional computing resources. Talented graduate students join the department from around the world, and add a unique dimension to the department's intellectual life. Faculty, graduate students, and undergraduates interact regularly.

How to Take Part

There are many ways to get involved in Statistics at Carnegie Mellon:

• The Bachelor of Science in Statistics in the Dietrich College of Humanities and Social Sciences (DC) is a broad-based, flexible program that helps you master both the theory and practice of Statistics. The program can be tailored to prepare you for a graduate program in Statistics or to complement your interests in almost any field, including Psychology, Physics, Biology, History, Business, Information Systems, and Computer Science.

• The Bachelor of Science in Economics and Statistics provides an interdisciplinary course of study aimed at students with a strong interest in the empirical analysis of economic data. Jointly administered by the Department of Statistics and Data Science and the Undergraduate Economics Program, the major's curriculum provides students with a solid foundation in the theories and methods of both fields. (See Dietrich College Interdepartmental Majors as well later in this section)

• The Bachelor of Science in Statistics and Machine Learning is a program housed in the Department of Statistics and Data Science and is jointly administered with the Department of Machine Learning. In this major students take courses focused on skills in computing, mathematics, statistical theory, and the interpretation and display of complex data. The program is geared toward students interested in statistical computation, data science, and "big data" problems.

• The Statistics Concentration and the Operations Research and Statistics Concentration in the Mathematical Sciences Major (see Department of Mathematical Sciences) are administered by the Department of Mathematical Sciences with input from the Department of Statistics and Data Science.

• There are several ongoing exciting research projects in the Department of Statistics and Data Science, and the department enthusiastically seeks to involve undergraduates in this work. Both majors and non-majors are welcome.

• Non-majors are eligible to take most of our courses, and indeed, they are required to do so by many programs on campus. Such courses offer one way to learn more about the Department of Statistics and Data Science and the field in general.

Curriculum

Statistics consists of two intertwined threads of inquiry: Statistical Theory and Data Analysis. The former uses probability theory to build and analyze mathematical models of data in order to devise methods for making effective predictions and decisions in the face of uncertainty. The latter involves techniques for extracting insights from complicated designs, data, and incomplete or missing information complicate all our lives. Statistics is central to big questions in public policy, law, medicine, industry, computing, technology, finance, and science. Indeed, the tools of Statistics apply to problems in almost every area of human activity where data are collected.

Note: We recommend that you use the information provided below as a general guideline, and then schedule a meeting with a Statistics Undergraduate Advisor (statadvising@stat.cmu.edu) to discuss the requirements in more detail, and build a program that is tailored to your strengths and interests.

B.S. in Statistics

Glenn Clune, Academic Advisor
Peter Freeman, Faculty Advisor
Location: Baker Hall 132
statadvising@stat.cmu.edu

Students in the Bachelor of Science program develop and master a wide array of skills in computing, mathematics, statistics, mathematical theory, and the interpretation and display of complex data. In addition, Statistics majors gain experience in applying statistical tools to real problems in other fields and learn the nuances of interdisciplinary collaboration. The requirements for the Major in Statistics are detailed below and are organized by categories #1-#7.

Curriculum

1. Mathematical Foundations (Prerequisites)

Mathematics is the language in which statistical models are described and analyzed, so some experience with basic calculus and linear algebra is an important component for anyone pursuing a program of study in Statistics.
Calculus*: Complete one of the following three sequences of mathematics courses at Carnegie Mellon, each of which provides sufficient preparation in calculus:

**Sequence 1**
- 21-111 Differential Calculus 10
- 21-112 Integral Calculus 10
and one of the following:
- 21-256 Multivariate Analysis 9
- 21-259 Calculus in Three Dimensions 9

**Sequence 2**
- 21-120 Differential and Integral Calculus 10
and one of the following:
- 21-256 Multivariate Analysis 9
- 21-259 Calculus in Three Dimensions 9

Notes:
- * It is recommended that students complete the calculus requirement during their freshman year.
- ** The linear algebra requirement needs to be completed before taking 36–401 Modern Regression.

Linear Algebra**: Complete one of the following three courses:
- 21-240 Matrix Algebra with Applications 10
- 21-241 Matrices and Linear Transformations 10
- 21-242 Matrix Theory 10

2. Data Analysis: 36-45 units

Data analysis is the art and science of extracting insight from data. The art lies in knowing which displays or techniques will reveal the most interesting features of a complicated data set. The science lies in understanding the various techniques and the assumptions on which they rely. Both aspects require practice to master.

The Beginning Data Analysis courses give a hands-on introduction to the art and science of data analysis. The courses cover similar topics but differ slightly in the examples they emphasize. 36–200 draws examples from many fields and satisfies the DC College Core Requirement in Statistical Reasoning. This course is therefore recommended for students in the College. (Note: A score of 4 or 5 on the Advanced Placement (AP) Exam in Statistics may be used to waive this requirement). Other courses emphasize examples in business (36–207), engineering and architecture (36–220), and the laboratory sciences (36–247).

The Intermediate Data Analysis courses build on the principles and methods covered in the introductory course, and more fully explore specific types of data analysis methods in more depth.

The Advanced Data Analysis courses draw on students’ previous experience with data analysis and understanding of statistical theory to develop advanced, more sophisticated methods. These core courses involve extensive analysis of real data with emphasis on developing the oral and writing skills needed for communicating results.

Sequence 1 (For students beginning their freshman or sophomore year)

**Beginning***
Choose one of the following courses:
- 36-200 Reasoning with Data 9
- 36/70-207 Probability and Statistics for Business Applications 9
- 36-220 Engineering Statistics and Quality Control 9
- 36-247 Statistics for Lab Sciences 9

Note: Students who enter the program with 36-225 or 36-226 should discuss options with an advisor. Any 36-300 or 36-400 level course in Data Analysis that does not satisfy any other requirement for a Statistics Major and Minor may be counted as a Statistical Elective.

**Intermediate***
Choose one of the following courses:
- 36-202 Statistics & Data Science Methods ** 9
- 36/70-208 Regression Analysis 9
- 36-209 Experimental Design for Behavioral & Social Sciences 9
- 36-290 Introduction to Statistical Research Methodology 9

* Or extra data analysis course in Statistics
** Must take prior to 36-401

Advanced
Choose one of the following courses:
- 36-303 Sampling, Survey and Society 9
- 36-315 Statistical Graphics and Visualization 9
- 36-311 Statistical Analysis of Networks 9
- 36-461 Special Topics: Statistical Methods in Epidemiology 9
- 36-462 Special Topics: Data Mining 9
- 36-463 Special Topics: Multilevel and Hierarchical Models 9
- 36-464 Special Topics: Applied Multivariate Methods 9
- 36-466 Special Topics: Statistical Methods in Finance 9
- 36-467 Special Topics: Data over Space & Time 9
- 36-468 Special Topics: Text Analysis 9
- 36-490 Undergraduate Research 9
- 36-497 Corporate Capstone Project 9

and take the following two courses:
- 36-401 Modern Regression 9
- 36-402 Advanced Methods for Data Analysis 9

Students can also take a second 36-46x (see section #5). Sequence 2 (For students beginning later in their college career)

Advanced
Choose two of the following courses:
- 36-303 Sampling, Survey and Society 9
- 36-311 Statistical Analysis of Networks 9
- 36-315 Statistical Graphics and Visualization 9
- 36-461 Special Topics: Statistical Methods in Epidemiology 9
- 36-462 Special Topics: Data Mining 9
- 36-463 Special Topics: Multilevel and Hierarchical Models 9
- 36-464 Special Topics: Applied Multivariate Methods 9
- 36-466 Special Topics: Statistical Methods in Finance 9
- 36-467 Special Topics: Data over Space & Time 9
- 36-468 Special Topics: Text Analysis 9
- 36-490 Undergraduate Research 9
- 36-497 Corporate Capstone Project 9

** All Special Topics are not offered every semester, and new Special Topics are regularly added. See section 5 for details.

and take the following two courses:
- 36-401 Modern Regression 9
- 36-402 Advanced Methods for Data Analysis 9

3. Probability Theory and Statistical Theory: 18 units

The theory of probability gives a mathematical description of the randomness inherent in our observations. It is the language in which statistical models are stated, so an understanding of probability is essential for the study of statistical theory. Statistical theory provides a mathematical framework for making inferences about unknown quantities from data. The theory reduces statistical problems to their essential ingredients to help devise and evaluate inferential procedures. It provides a powerful and wide-ranging set of tools for dealing with uncertainty.

To satisfy the theory requirement take the following two courses:
- 36-225 Introduction to Probability Theory ** 9

and one of the following two courses:
<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>36-226</td>
<td>Introduction to Statistical Inference</td>
<td>9</td>
</tr>
<tr>
<td>36-326</td>
<td>Mathematical Statistics (Honors)</td>
<td>9</td>
</tr>
</tbody>
</table>

*It is possible to substitute 36-217, 36-218, or 21-325 for 36-225. (36-225 is the standard introduction to probability, 36-217 is tailored for engineers and computer scientists, 36-218 is a more mathematically rigorous class for Computer Science students and more mathematically advanced Statistics students (Statistics students need advisor approval to enroll), and 21-325 is a rigorous probability theory course offered by the Department of Mathematics.)*

**Comment:**
(i) In order meet the prerequisite requirements for the major, a grade of C or better is required in 36-225, 36-226 and 36-401. In particular, a grade of C or higher is required in order to be able to continue in the major.

### 4. Statistical Computing: 9 units
36-350 Statistical Computing *

*In rare circumstances, a higher level Statistical Computing course, approved by your Statistics advisor, may be used as a substitute.

### 5. Special Topics 9 units
The Department of Statistics and Data Science offers advanced courses that focus on specific statistical applications or advanced statistical methods. These courses are numbered 36-46x (36-461, 36-462, etc.). Two of these courses will be offered every year, one per semester. Past topics included Statistical Learning, Data Mining, Statistics, and the Law, Bayesian Statistics, Nonparametric Statistics, Statistical Genetics, Multilevel and Hierarchical Models, and Statistical Methods in Epidemiology. The objective of the course is to expose students to important topics in statistics and/or interesting applications which are not part of the standard undergraduate curriculum.

To satisfy the Special Topics requirement choose one of the 36-46x courses (which are 9 units).

Note: All 36-46x courses require 36-401 as a prerequisite or corequisite.

### 6. Statistical Elective: 9-10 units
Students are required to take one elective which can be within or outside the Department of Statistics and Data Science. Courses within Statistics can be any 300 or 400 level course (that is not used to satisfy any other requirement for the statistics major).

The following is a partial list of courses outside Statistics that qualify as electives as they provide intellectual infrastructure that will advance the student’s understanding of statistics and its applications. Other courses may qualify as well, consult with the Statistics Undergraduate Advisor.

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-110</td>
<td>Principles of Computing</td>
<td>10</td>
</tr>
<tr>
<td>15-112</td>
<td>Fundamentals of Programming and Computer Science</td>
<td>12</td>
</tr>
<tr>
<td>15-121</td>
<td>Introduction to Data Structures</td>
<td>10</td>
</tr>
<tr>
<td>15-122</td>
<td>Principles of Imperative Computation</td>
<td>10</td>
</tr>
<tr>
<td>10-301</td>
<td>Introduction to Machine Learning</td>
<td>12</td>
</tr>
<tr>
<td>10-315</td>
<td>Introduction to Machine Learning (Undergrad)</td>
<td>12</td>
</tr>
<tr>
<td>15-388</td>
<td>Practical Data Science</td>
<td>9</td>
</tr>
<tr>
<td>21-260</td>
<td>Differential Equations</td>
<td>9</td>
</tr>
<tr>
<td>21-292</td>
<td>Operations Research I</td>
<td>9</td>
</tr>
<tr>
<td>21-301</td>
<td>Combinatorics</td>
<td>9</td>
</tr>
<tr>
<td>21-355</td>
<td>Principles of Real Analysis I</td>
<td>9</td>
</tr>
<tr>
<td>80-220</td>
<td>Philosophy of Science</td>
<td>9</td>
</tr>
<tr>
<td>80-221</td>
<td>Philosophy of Social Science</td>
<td>9</td>
</tr>
<tr>
<td>80-310</td>
<td>Formal Logic</td>
<td>9</td>
</tr>
<tr>
<td>85-310</td>
<td>Research Methods in Cognitive Psychology</td>
<td>9</td>
</tr>
<tr>
<td>85-320</td>
<td>Research Methods in Developmental Psychology</td>
<td>9</td>
</tr>
<tr>
<td>85-340</td>
<td>Research Methods in Social Psychology</td>
<td>9</td>
</tr>
<tr>
<td>88-223</td>
<td>Decision Analysis</td>
<td>12</td>
</tr>
<tr>
<td>88-302</td>
<td>Behavioral Decision Making</td>
<td>9</td>
</tr>
</tbody>
</table>

Note: Additional prerequisites are required for some of these courses. Students should carefully check the course descriptions to determine if additional prerequisites are necessary.

### 7. Tracks*:

**Self-Defined Concentration Area (with advisor’s approval):** 36 units

The power of Statistics, and much of the fun, is that it can be applied to answer such a wide variety of questions in so many different fields. A critical part of statistical practice is understanding the questions being asked so that appropriate methods of analysis can be used. Hence, a critical part of statistical training is to gain experience applying the abstract tools to real problems.

The Concentration Area is a set of four related courses outside of Statistics that prepares the student to deal with statistical aspects of problems that arise in another field. These courses are usually drawn from a single discipline of interest to the student and must be approved by the Statistics Undergraduate Advisor. While these courses are not in Statistics, the concentration area must complement the overall Statistics degree.

For example, students intending to pursue careers in the health or biomedical sciences could take further courses in Biology or Chemistry, or students intending to pursue graduate work in Statistics could take further courses in advanced Mathematics.

The concentration area can be fulfilled with a minor or additional major, but not all minors and additional majors fulfill this requirement. Please make sure to consult the Undergraduate Statistics Advisor prior to pursuing courses for the concentration area. Once the concentration area is approved, any changes made to the previously agreed upon coursework requires re-approval by the Undergraduate Advisor.

**Concentration Approval Process**
- Submit the below materials to the Undergraduate Statistics Advisor
- List of possible coursework to fulfill the concentration*
- 150-200 word essay describing how the proposed courses complement the Statistics degree.

* These courses can be amended later, but must be re-approved by the Statistics Undergraduate Advisor, if amended.

**Note:** The concentration/track requirement is only for students whose *primary* major is statistics and have no other additional major or minor. The requirement does not apply for students who pursue an additional major in statistics.

#### Mathematics Statistics Track 46-52 units

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-127</td>
<td>Concepts of Mathematics</td>
<td>10</td>
</tr>
<tr>
<td>21-355</td>
<td>Principles of Real Analysis I</td>
<td>9</td>
</tr>
<tr>
<td>36-410</td>
<td>Introduction to Probability Modeling</td>
<td>9</td>
</tr>
</tbody>
</table>

And two of the following:

- 36-700 Probability and Mathematical Statistics
- or 36-705 Intermediate Statistics
- 21-228 Discrete Mathematics
- 21-257 Models and Methods for Optimization
- 21-292 Operations Research I
- 21-301 Combinatorics
- 21-356 Principles of Real Analysis II

#### Statistics and Neuroscience Track 45-54 units

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>85-211</td>
<td>Cognitive Psychology</td>
<td>9</td>
</tr>
<tr>
<td>85-219</td>
<td>Biological Foundations of Behavior</td>
<td>9</td>
</tr>
</tbody>
</table>

And three electives (at least one from Methodology and Analysis and at least one from Neuroscientific Background):

#### Methodology and Analysis

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>36-700</td>
<td>Probability and Mathematical Statistics</td>
<td>12</td>
</tr>
<tr>
<td>or 36-705</td>
<td>Intermediate Statistics</td>
<td>12</td>
</tr>
<tr>
<td>10-301</td>
<td>Introduction to Machine Learning</td>
<td>12</td>
</tr>
<tr>
<td>18-290</td>
<td>Signals and Systems</td>
<td>12</td>
</tr>
<tr>
<td>85-314</td>
<td>Cognitive Neuroscience Research Methods</td>
<td>9</td>
</tr>
<tr>
<td>42-86-631</td>
<td>Neural Data Analysis</td>
<td>9</td>
</tr>
</tbody>
</table>

#### Neuroscience Background

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>03-362</td>
<td>Cellular Neuroscience</td>
<td>9</td>
</tr>
<tr>
<td>03-363</td>
<td>Systems Neuroscience</td>
<td>9</td>
</tr>
<tr>
<td>15-386</td>
<td>Neural Computation</td>
<td>9</td>
</tr>
<tr>
<td>85-414</td>
<td>Cognitive Neuropsychology</td>
<td>9</td>
</tr>
<tr>
<td>85-419</td>
<td>Introduction to Parallel Distributed Processing</td>
<td>9</td>
</tr>
</tbody>
</table>
Recommendations for Prospective PhD Students

Students interested in pursuing a PhD in Statistics or Biostatistics (or related programs) after completing their undergraduate degree are strongly recommended to pursue the Mathematical Statistics Track.

Additional Major in Statistics

Students who elect Statistics as a second or third major must fulfill all Statistics degree requirements except for the Concentration Area requirement. Majors in many other programs would naturally complement a Statistics Major, including Tepper’s undergraduate business program, Social and Decision Sciences, Policy and Management, and Psychology.

With respect to double-counting courses, it is departmental policy that students must have at least five statistics courses that do not count for their primary major. If students do not have at least five, they typically take additional advanced data analysis electives.

Students are advised to begin planning their curriculum (with appropriate advisors) as soon as possible. This is particularly true if the other major has a complex set of requirements and prerequisites or when many of the other major’s requirements overlap with the requirements for a Major in Statistics.

Many departments require Statistics courses as part of their Major or Minor programs. Students seeking transfer credit for those requirements from substitute courses (at Carnegie Mellon or elsewhere) should seek permission from their advisor in the department setting the requirement. The final authority in such decisions rests there. The Department Statistics does not provide approval or permission for substitution or waiver of another department’s requirements.

If a waiver or substitution is made in the home department, it is not automatically approved in the Department of Statistics and Data Science. In many of these cases, the student will need to take additional courses to satisfy the Statistics major requirements. Students should discuss this with a Statistics advisor when deciding whether to add an additional major in Statistics.

Research

One goal of the Statistics program is to give students experience with statistical research. The department gives students research experience through various courses focused on real world experiences and application. There is a variety of research projects in the department as well, and students who would like to pursue working on a project with faculty will need to contact that faculty directly to discuss that possibility.

Before graduation, students are encouraged to participate in a research project under faculty supervision. Students mostly do this through projects in specific courses, such as 36-290, 36-303, 36-490, and/or 36-497. Students can also pursue an independent study, or a summer research position.

Qualified students are also encouraged to participate in an advanced research project through 36-490 Undergraduate Research or 36-497 Corporate Capstone Project. Note that both of these courses require an application. Students who maintain a quality point average of 3.25 overall may also apply to participate in the Dietrich College Senior Honors Program (http://coursecatalog.web.cmu.edu/dietrichcollegeofhumanitiesandsocialsciences/#collegeservicesandprograms).

Sample Programs

The following sample programs illustrate three (of many) ways to satisfy the requirements of the Statistics Major. However, keep in mind that the program is flexible enough to support many other possible schedules and to emphasize a wide variety of interests.

The first schedule uses calculus sequence 1.

The second schedule is an example of the case when a student enters the program through 36-225 and 36-226 (and therefore skips the beginning data analysis sequence). This schedule has more emphasis on statistical theory and probability.

The third schedule is an example of the Mathematical Statistics track.

In these schedules, C.A. refers to Concentration Area courses.

Schedule 1

<table>
<thead>
<tr>
<th>Freshman</th>
<th>Sophomore</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>Spring</td>
</tr>
<tr>
<td>36-200 Reasoning with Data</td>
<td>36-202 Statistics &amp; Data Science Methods</td>
</tr>
<tr>
<td>21-111 Differential Calculus</td>
<td>21-112 Integral Calculus</td>
</tr>
</tbody>
</table>

Schedule 2

<table>
<thead>
<tr>
<th>Freshman</th>
<th>Sophomore</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>Spring</td>
</tr>
<tr>
<td>36-225 Introduction to Probability Theory</td>
<td>36-226 Introduction to Statistical Inference</td>
</tr>
<tr>
<td>Stat Elective</td>
<td>36-350 Statistical Computing</td>
</tr>
<tr>
<td>36-200 Reasoning with Data</td>
<td></td>
</tr>
</tbody>
</table>

Schedule 3 - Mathematics Track Only

<table>
<thead>
<tr>
<th>Freshman</th>
<th>Sophomore</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>Spring</td>
</tr>
<tr>
<td>21-120 Differential and Integral Calculus</td>
<td>21-256 Multivariate Analysis</td>
</tr>
</tbody>
</table>

B.S. in Economics and Statistics

Samantha Nielsen, Statistics & Data Science Lead Academic Advisor
Kathleen Conway, Economics Senior Academic Advisor
Rebecca Nugent and Edward Kennedy, Faculty Advisors
Carol Goldberg, Executive Director, Undergraduate Economics Program

Statistics & Data Science Location: Baker Hall 132
statadvising@stat.cmu.edu

* Note: This number can vary depending on the calculus sequence and on the concentration area a student takes. In addition this number includes the 36 units of the “Concentration Area” category which may not be required (see category 7 above for details).
Economics Location: Tepper 2400
econprog@andrew.cmu.edu

The B.S. in Economics and Statistics is jointly advised by the Department of Statistics and Data Science and the Undergraduate Economics Program.

The Major in Economics and Statistics provides an interdisciplinary course of study aimed at students with a strong interest in the empirical analysis of economic data. With joint curriculum from the Department of Statistics and Data Science and the Undergraduate Economics Program, the major provides students with a solid foundation in the theories and methods of both fields. Students in this major are trained to advance the understanding of economic issues through the analysis, synthesis and reporting of data using the advanced empirical research methods of statistics and econometrics. Graduates are well positioned for admission to competitive graduate programs, including those in statistics, economics and management, as well as for employment in positions requiring strong analytic and conceptual skills - especially those in economics, finance, education, and public policy.

All economics courses counting towards an economics degree must be completed with a grade of "C" or higher.

The requirements for the B.S. in Economics and Statistics are the following:

I. Prerequisites 38-39 units

1. Mathematical Foundations 38-39 units
   Calculus
   21-120 Differential and Integral Calculus 10
   and one of the following:
   21-256 Multivariate Analysis 9
   21-259 Calculus in Three Dimensions 9
   Note: Passing the MSC 21-120 assessment test is an acceptable alternative to completing 21-120.
   Note: Taking/having credit for both 21-111 and 21-112 is equivalent to 21-120. The Mathematical Foundations total is then 48-49 units. The Economics and Statistics major would then total 201-211 units.

Linear Algebra
   One of the following three courses:
   21-240 Matrix Algebra with Applications 10
   21-241 Matrices and Linear Transformations 10
   21-242 Matrix Theory 10
   Note: 21-241 and 21-242 are intended only for students with a very strong mathematical background.

II. Foundations 18-36 units

2. Economics Foundations 18 units
   73-102 Principles of Microeconomics 9
   73-103 Principles of Macroeconomics 9

3. Statistical Foundations 9-18 units
   Sequence 1 (For students beginning their freshman or sophomore year)
   Beginning*
   Choose one of the following courses:
   36-200 Reasoning with Data 9
   36/70-207 Probability and Statistics for Business Applications 9
   36-220 Engineering Statistics and Quality Control 9
   36-247 Statistics for Lab Sciences 9
   Note: Students who enter the program with 36-225 or 36-226 should discuss options with an advisor. Any 36-300 or 36-400 level course in Data Analysis that does not satisfy any other requirement for the Economics and Statistics Major may be counted as a Statistical Elective.
   Intermediate*
   Choose one of the following courses:
   36-202 Statistics & Data Science Methods ** 9
   36-208 Regression Analysis 9
   36-290 Introduction to Statistical Research Methodology 9

36-309 Experimental Design for Behavioral & Social Sciences 9
* Or extra data analysis course in Statistics
** Must take prior to 36-401 Modern Regression.

Advanced
Choose two of the following courses:
36-303 Sampling, Survey and Society 9
36-311 Statistical Analysis of Networks 9
36-315 Statistical Graphics and Visualization 9
36-461 Special Topics: Statistical Methods in Epidemiology 9
36-462 Special Topics: Data Mining 9
36-463 Special Topics: Multilevel and Hierarchical Models 9
36-464 Special Topics: Applied Multivariate Methods 9
36-466 Special Topics: Statistical Methods in Finance 9
36-467 Special Topics: Data over Space & Time 9
36-468 Special Topics: Text Analysis 9
36-490 Undergraduate Research 9
36-497 Corporate Capstone Project 9

Sequence 2 (For students beginning later in their college career)

Advanced
Choose three of the following courses:
36-303 Sampling, Survey and Society 9
36-311 Statistical Analysis of Networks 9
36-315 Statistical Graphics and Visualization 9
36-461 Special Topics: Statistical Methods in Epidemiology 9
36-462 Special Topics: Data Mining 9
36-463 Special Topics: Multilevel and Hierarchical Models 9
36-464 Special Topics: Applied Multivariate Methods 9
36-466 Special Topics: Statistical Methods in Finance 9
36-467 Special Topics: Data over Space & Time 9
36-468 Special Topics: Text Analysis 9
36-490 Undergraduate Research 9
36-497 Corporate Capstone Project 9

**All Special Topics are not offered every semester, and new Special Topics are regularly added. See section 5 for details.

III. Disciplinary Core 126 units

1. Economics Core 45 units
   73-230 Intermediate Microeconomics 9
   73-240 Intermediate Macroeconomics 9
   73-270 Professional Communication for Economists 9
   73-265 Economics and Data Science 9
   73-274 Econometrics I 9
   73-374 Econometrics II 9

2. Statistics Core 36 units
   36-225 Introduction to Probability Theory ** 9
   and one of the following two courses:
   36-226 Introduction to Statistical Inference * 9
   36-326 Mathematical Statistics (Honors) * 9
   and both of the following two courses:
   36-401 Modern Regression * 9
   36-402 Advanced Methods for Data Analysis 9
   *In order meet the prerequisite requirements for the major, a grade of C or better is required in 36-225 (or equivalents), 36-226 or 36-326 and 36-401.
   #It is possible to substitute 36-217, 36-218, or 21-325 for 36-225 36-225
   36-22536-225. (36-225 36-225 36-22536-225 is the standard introduction to probability, 36-217 is tailored for engineers and computer scientists, 36-218 is a more mathematically rigorous class for Computer Science students and more mathematically advanced Statistics students (Statistics students need advisor approval to enroll), and 21-325 21-325 21-32521-325

Department of Statistics and Data Science 5
Department of Statistics and Data Science

is a rigorous Probability Theory course offered by the Department of Mathematics.

3. Computing
36-350 Statistical Computing 9 units

*In rare circumstances, a higher level Statistical Computing course, approved by your Statistics advisor, may be used as a substitute.

4. Advanced Electives
Students must take two advanced Economics elective courses (numbered 73-300 through 73-495, excluding 73-374) and two (or three - depending on previous coursework, see Section 3) advanced Statistics elective courses (numbered 36-303, 36-311, 36-315, 36-46x, 36-490, or 36-497).

Students pursuing a degree in Economics and Statistics also have the option of earning a concentration area (https://www.cmu.edu/tepper/programs/undergraduate-economics/curriculum/concentrations) by completing a set of interconnected electives. While a concentration area is not required for this degree, it is an additional option that allows students to explore a group of aligned topics and/or develop a specialized and advanced skill set appropriate for a desired career path. The electives required for this degree may count towards your concentration area. To fulfill a concentration, students must take four courses from the designated set of electives. Please make sure to consult an advisor when choosing these courses.

| Total number of units for the major | 191-201 units |
| Total number of units for the degree | 360 units |

Professional Development

Students are strongly encouraged to take advantage of professional development opportunities and/or coursework. One option is 73-210 Economics Colloquium I, a fall-only course that provides information about careers in Economics, job search strategies, and research opportunities. The Department of Statistics and Data Science also offers a series of workshops pertaining to resume preparation, graduate school applications, careers in the field, among other topics. Students should also take advantage of the Career and Professional Development Center.

Additional Major in Economics and Statistics

Students who elect Economics and Statistics as a second or third major must fulfill all Economics and Statistics degree requirements. Majors in many other programs would naturally complement an Economics and Statistics Major, including Tepper’s undergraduate business program, Social and Decision Sciences, Policy and Management, and Psychology.

With respect to double-counting courses, it is departmental policy that students must have at least six courses (three Economics and three Statistics) that do not count for their primary major. If students do not have at least six, they typically take additional advanced data analysis or economics electives, depending on where the double counting issue is.

Students are advised to begin planning their curriculum (with appropriate advisors) as soon as possible. This is particularly true if the other major has a complex set of requirements and prerequisites or when many of the other major’s requirements overlap with the requirements for a Major in Economics and Statistics.

Many departments require Statistics courses as part of their Major or Minor programs. Students seeking transfer credit for those requirements from substitute courses (at Carnegie Mellon or elsewhere) should seek permission from their advisor in the department setting the requirement. The final authority in such decisions rests there. The Department of Statistics and Data Science does not provide approval or permission for substitution or waiver of another department’s requirements.

If a waiver or substitution is made in the home department, it is not automatically approved in the Department of Statistics and Data Science. In many of these cases, the student will need to take additional courses to satisfy the Economics and Statistics major requirements. Students should discuss this with a Statistics advisor when deciding whether to add an additional major in Economics and Statistics.

Sample Program

The following sample program illustrates one way to satisfy the requirements of the Economics and Statistics Major. Keep in mind that the program is flexible and can support other possible schedules (see footnotes below the schedule).

<table>
<thead>
<tr>
<th>Freshman</th>
<th>Sophomore</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>Spring</td>
</tr>
<tr>
<td>21-120 Differential and Integral Calculus</td>
<td>36-202 Statistics &amp; Data Science Methods</td>
</tr>
<tr>
<td>36-200 Reasoning with Data</td>
<td>21-256 Multivariate Analysis</td>
</tr>
<tr>
<td>73-103 Principles of Microeconomics</td>
<td>73-103 Principles of Macroeconomics</td>
</tr>
<tr>
<td>73-060 Economics: BaseCamp <em>not required</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td>73-265 Economics and Data Science</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*In each semester, ----- represents other courses (not related to the major) which are needed in order to complete the 360 units that the degree requires.

Prospective PhD students might add 21-127 fall of sophomore year, replace 21-240 with 21-241, add 21-260 in spring of junior year and 21-355 in fall of senior year.

B.S. in Statistics and Machine Learning

Samantha Nielsen, Academic Advisor
Ryan Tibshirani and Ann Lee, Faculty Advisors
Location: Baker Hall 132
statadvising@stat.cmu.edu

Students in the Statistics and Machine Learning program develop and master a wide array of skills in computing, mathematics, statistical theory, and the interpretation and display of complex data. In addition, Statistics and Machine Learning majors gain experience in applying statistical tools to real problems in other fields and learn the nuances of interdisciplinary collaboration. This program is geared towards students interested in statistical computation, data science, or “Big Data” problems. The requirements for the Major in Statistics and Machine Learning are detailed below and are organized by categories.

Curriculum

1. Mathematical Foundations (Prerequisites) 49-59 units

Mathematics is the language in which statistical models are described and analyzed, so some experience with basic calculus and linear algebra is an important component for anyone pursuing a program of study in Statistics and Machine Learning.

Calculus*:

Complete one of the following sequences of mathematics courses at Carnegie Mellon, each of which provides sufficient preparation in calculus:

**Sequence 1**

| 21-111 | Differential Calculus | 10 |
| 21-112 | Integral Calculus | 10 |

and one of the following:

| 21-256 | Multivariate Analysis | 9 |
| 21-259 | Calculus in Three Dimensions | 9 |
### Sequence 2

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-120</td>
<td>Differential and Integral Calculus</td>
<td>10</td>
</tr>
<tr>
<td>21-256</td>
<td>Multivariate Analysis</td>
<td>9</td>
</tr>
<tr>
<td>21-259</td>
<td>Calculus in Three Dimensions</td>
<td>9</td>
</tr>
</tbody>
</table>

**Notes:**
- Passing the Mathematical Sciences 21-120 assessment test is an acceptable alternative to completing 21-120

### Integration and Approximation

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-122</td>
<td>Integration and Approximation</td>
<td>10</td>
</tr>
</tbody>
</table>

### Linear Algebra**:

Complete one of the following three courses:

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-240</td>
<td>Matrix Algebra with Applications</td>
<td>10</td>
</tr>
<tr>
<td>21-241</td>
<td>Matrices and Linear Transformations</td>
<td>10</td>
</tr>
<tr>
<td>21-242</td>
<td>Matrix Theory</td>
<td>10</td>
</tr>
</tbody>
</table>

* It is recommended that students complete the calculus requirement during their freshman year.

**The linear algebra requirement needs to be completed before taking 36-401 Modern Regression.

21-241 and 21-242 are intended only for students with a very strong mathematical background.

### Mathematical Theory:

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-127</td>
<td>Concepts of Mathematics</td>
<td>10</td>
</tr>
</tbody>
</table>

## 2. Data Analysis 45-54 units

Data analysis is the art and science of extracting insight from data. The art lies in knowing which displays or techniques will reveal the most interesting features of a complicated data set. The science lies in understanding the various techniques and the assumptions on which they rely. Both aspects require practice to master.

The Beginning Data Analysis courses give a hands-on introduction to the art and science of data analysis. The courses cover similar topics but differ slightly in the examples they emphasize. 36-200 draws examples from many fields and satisfy the Dietrich College Core Requirement in Statistical Reasoning. One of these courses is therefore recommended for students in the College. (Note: A score of 4 or 5 on the Advanced Placement (AP) Exam in Statistics may be used to waive this requirement). Other courses emphasize examples in business (36-207), engineering and architecture (36-220), and the laboratory sciences (36-247).

The Intermediate Data Analysis courses build on the principles and methods covered in the introductory course, and more fully explore specific types of data analysis methods in more depth.

The Advanced Data Analysis courses draw on students' previous experience with data analysis and understanding of statistical theory to develop advanced, more sophisticated methods. These core courses involve extensive analysis of real data with emphasis on developing the oral and writing skills needed for communicating results.

### Sequence 1

**Beginning***

Choose one of the following courses:

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>36-200</td>
<td>Reasoning with Data</td>
<td>9</td>
</tr>
<tr>
<td>36/70-207</td>
<td>Probability and Statistics for Business Applications</td>
<td>9</td>
</tr>
<tr>
<td>36-220</td>
<td>Engineering Statistics and Quality Control</td>
<td>9</td>
</tr>
<tr>
<td>36-247</td>
<td>Statistics for Lab Sciences</td>
<td>9</td>
</tr>
</tbody>
</table>

Note: Students who enter the program with 36-225 or 36-226 should discuss options with an advisor. Any 36-300 or 36-400 level course in Data Analysis that does not satisfy any other requirement for a Statistics Major and Minor may be counted as a Statistical Elective.

**Intermediate***

Choose one of the following courses:

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>36-202</td>
<td>Statistics &amp; Data Science Methods</td>
<td>9</td>
</tr>
<tr>
<td>36/70-208</td>
<td>Regression Analysis</td>
<td>9</td>
</tr>
</tbody>
</table>

### 3. Probability Theory and Statistical Theory 18 units

The theory of probability gives a mathematical description of the randomness inherent in our observations. It is the language in which statistical models are stated, so an understanding of probability is essential for the study of statistical theory. Statistical theory provides a mathematical framework for making inferences about unknown quantities from data. The theory reduces statistical problems to their essential ingredients to help devise and evaluate inferential procedures. It provides a powerful and wide-ranging set of tools for dealing with uncertainty.

To satisfy the theory requirement take the following two courses**:

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>36-225</td>
<td>Introduction to Probability Theory</td>
<td>9</td>
</tr>
<tr>
<td>36-326</td>
<td>Introduction to Statistical Inference</td>
<td>9</td>
</tr>
</tbody>
</table>

**It is possible to substitute 36-217, 36-218, or 36-325 for 36-225. (36-225 is the standard introduction to probability, 36-217 is tailored for engineers and computer scientists, 36-218 is a more mathematically rigorous class for Computer Science students and more mathematically advanced Statistics students.**

---

**Advanced**

Choose two of the following courses:

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>36-303</td>
<td>Sampling, Survey and Society</td>
<td>9</td>
</tr>
<tr>
<td>36-311</td>
<td>Statistical Analysis of Networks</td>
<td>9</td>
</tr>
<tr>
<td>36-315</td>
<td>Statistical Graphics and Visualization</td>
<td>9</td>
</tr>
<tr>
<td>36-461</td>
<td>Special Topics: Statistical Methods in Epidemiology</td>
<td>9</td>
</tr>
<tr>
<td>36-462</td>
<td>Special Topics: Data Mining</td>
<td>9</td>
</tr>
<tr>
<td>36-463</td>
<td>Special Topics: Multilevel and Hierarchical Models</td>
<td>9</td>
</tr>
<tr>
<td>36-464</td>
<td>Special Topics: Applied Multivariate Methods</td>
<td>9</td>
</tr>
<tr>
<td>36-466</td>
<td>Special Topics: Statistical Methods in Finance</td>
<td>9</td>
</tr>
<tr>
<td>36-467</td>
<td>Special Topics: Data over Space &amp; Time</td>
<td>9</td>
</tr>
<tr>
<td>36-468</td>
<td>Special Topics: Text Analysis</td>
<td>9</td>
</tr>
<tr>
<td>36-490</td>
<td>Undergraduate Research</td>
<td>9</td>
</tr>
<tr>
<td>36-497</td>
<td>Corporate Capstone Project</td>
<td>9</td>
</tr>
</tbody>
</table>

Special Topics rotate and new ones are regularly added.

and take the following two courses:

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>36-401</td>
<td>Modern Regression</td>
<td>9</td>
</tr>
<tr>
<td>36-402</td>
<td>Advanced Methods for Data Analysis</td>
<td>9</td>
</tr>
</tbody>
</table>

---

**Sequence 2 Advanced**

Choose three of the following courses:

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>36-303</td>
<td>Sampling, Survey and Society</td>
<td>9</td>
</tr>
<tr>
<td>36-311</td>
<td>Statistical Analysis of Networks</td>
<td>9</td>
</tr>
<tr>
<td>36-315</td>
<td>Statistical Graphics and Visualization</td>
<td>9</td>
</tr>
<tr>
<td>36-461</td>
<td>Special Topics: Statistical Methods in Epidemiology</td>
<td>9</td>
</tr>
<tr>
<td>36-462</td>
<td>Special Topics: Data Mining</td>
<td>9</td>
</tr>
<tr>
<td>36-463</td>
<td>Special Topics: Multilevel and Hierarchical Models</td>
<td>9</td>
</tr>
<tr>
<td>36-464</td>
<td>Special Topics: Applied Multivariate Methods</td>
<td>9</td>
</tr>
<tr>
<td>36-466</td>
<td>Special Topics: Statistical Methods in Finance</td>
<td>9</td>
</tr>
<tr>
<td>36-467</td>
<td>Special Topics: Data over Space &amp; Time</td>
<td>9</td>
</tr>
<tr>
<td>36-468</td>
<td>Special Topics: Text Analysis</td>
<td>9</td>
</tr>
<tr>
<td>36-490</td>
<td>Undergraduate Research</td>
<td>9</td>
</tr>
<tr>
<td>36-497</td>
<td>Corporate Capstone Project</td>
<td>9</td>
</tr>
</tbody>
</table>

All Special Topics are not offered every semester, and new special topics are regularly added.

and take the following two courses:

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>36-401</td>
<td>Modern Regression</td>
<td>9</td>
</tr>
<tr>
<td>36-402</td>
<td>Advanced Methods for Data Analysis</td>
<td>9</td>
</tr>
</tbody>
</table>

---

**Notes:**

- **Must take prior to 36-401**
- **It is possible to substitute 36-217, 36-218, or 36-325 for 36-225.**
students (Statistics students need advisor approval to enroll), and 21-325 is a rigorous Probability Theory course offered by the Department of Mathematics.) 36-326 Mathematical Statistics (Honors) can be substituted for 36-226 Introduction to Statistical Inference and is considered an honors course.

Comments:
(i) In order to meet the prerequisite requirements, a grade of at least a C is required in 36-225, 36-226 and 36-401.

4. Statistical Computing 9 units
36-350 Statistical Computing * 9

*In rare circumstances, a higher level Statistical Computing course, approved by your Statistics advisor, may be used as a substitute.

5. Machine Learning/Computer Science 46-48 units

Statistical modeling in practice nearly always requires computation in one way or another. Computational algorithms are sometimes treated as "black-boxes", whose innards the statistician need not pay attention to. But this attitude is becoming less and less prevalent, and today there is much to be gained from a strong working knowledge of computational tools. Understanding the strengths and weaknesses of various methods allows the data analyst to select the right tool for the job; understanding how they can be adapted to work in new settings greatly extends the realm of problems that he/she can solve. While all Majors in Statistics are given solid grounding in computation, extensive computational training is really what sets the Major in Statistics and Machine Learning apart.

15-112 Fundamentals of Programming and Computer Science 12
15-122 Principles of Imperative Computation 10
15-351 Algorithms and Advanced Data Structures 12
10-301 Introduction to Machine Learning 12
or 10-315 Introduction to Machine Learning (Undergrad) 12

and take one of the following Machine Learning Advanced Electives:

10-405 Machine Learning with Large Datasets (Undergraduate) 12
10-605 Machine Learning with Large Datasets 12
10-703 Deep Reinforcement Learning & Control 12
10-707 Topics in Deep Learning 12
11-411 Natural Language Processing 12
11-441 Machine Learning for Text Mining 9
11-661 Language and Statistics 12
15-381 Artificial Intelligence: Representation and Problem Solving 9
15-386 Neural Computation 9
15-387 Computational Perception 9
16-311 Introduction to Robotics 12
16-385 Computer Vision 12
16-720 Computer Vision 12
11-761 Language and Statistics 12

*PhD level ML course as approved by Statistics advisor

** Independent research with an ML faculty member

Total number of units for the major 176–198 units
Total number of units for the degree 360 units

Recommendations

Students in the Dietrich College of Humanities and Social Sciences who wish to major or minor in Statistics are advised to complete both the calculus requirement (one Mathematical Foundations calculus sequence) and the Beginning Data Analysis course 36-200 Reasoning with Data by the end of their freshman year. The linear algebra requirement is a prerequisite for the course 36-401. It is therefore essential to complete this requirement during your junior year at the latest!

Recommendations for Prospective PhD Students

Students interested in pursuing a PhD in Statistics or Machine Learning (or related programs) after completing their undergraduate degree are strongly recommended to take additional Mathematics courses. They should see a faculty advisor as soon as possible. Students should consider 36-326 Mathematical Statistics (Honors) as an alternative to 36-226. Although 21-240 Matrix Algebra with Applications is recommended for Statistics majors, students interested in PhD programs should consider taking 21-241 Matrices and Linear Transformations or 21-242 Matrix Theory instead. Additional courses to consider are 21-228 Discrete Mathematics, 21-260 Differential Equations, 21-341 Linear Algebra, 21-355 Principles of Real Analysis I, and 21-356 Principles of Real Analysis II.

Additional experience in programming and computational modeling is also recommended. Students should consider taking more than one course from the list of Machine Learning electives provided under the Computing section.

Additional Major in Statistics and Machine Learning

Students who elect Statistics and Machine Learning as a second or third major must fulfill all degree requirements.

With respect to double-counting courses, it is departmental policy that students must have at least six courses (three Computer Science/Machine Learning and three Statistics) that do not count for their primary major. If students do not have at least six, they typically take additional advanced data analysis or ML electives, depending on where the double counting issue is.

Students are advised to begin planning their curriculum (with appropriate advisors) as soon as possible. This is particularly true if the other major has a complex set of requirements and prerequisites or when many of the other major’s requirements overlap with the requirements for a Major in Statistics and Machine Learning.

Many departments require Statistics courses as part of their Major or Minor programs. Students seeking transfer credit for those requirements from substitute courses (at Carnegie Mellon or elsewhere) should seek permission from their advisor in the department setting the requirement. The final authority in such decisions rests there. The Department of Statistics and Data Science does not provide approval or permission for substitution or waiver of another department’s requirements.

If a waiver or substitution is made in the home department, it is not automatically approved in the Department of Statistics and Data Science. In many of these cases, the student will need to take additional courses to satisfy the Statistics and Machine Learning major requirements. Students should discuss this with a Statistics advisor when deciding whether to add an additional major in Statistics and Machine Learning.

Sample Programs

The following sample program illustrates one way to satisfy the requirements of the Statistics and Machine Learning program. Keep in mind that the program is flexible and can support other possible schedules (see footnotes below the schedule). Sample program 1 is for students who have not satisfied the basic calculus requirements. Sample program 2 is for students who have satisfied the basic calculus requirements and choose option 2 for their data analysis courses (see section #2)

Schedule 1

<table>
<thead>
<tr>
<th>Freshman</th>
<th>Sophomore</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>Spring</td>
</tr>
<tr>
<td>36-200 Reasoning with Data</td>
<td>36-202 Statistics &amp; Data Science Methods</td>
</tr>
<tr>
<td>21-120 Differential and Integral Calculus</td>
<td>21-256 Multivariate Analysis</td>
</tr>
</tbody>
</table>
In order to get a minor in Statistics a student must satisfy all of the following requirements:

1. Mathematical Foundations (Prerequisites)  
   **29–39 units**

   **Calculus**:  
   Complete one of the following two sequences of mathematics courses at Carnegie Mellon, each of which provides sufficient preparation in calculus:

   **Sequence 1**
   - 21-111 Differential Calculus  
   - 21-112 Integral Calculus

   and one of the following:
   - 21-256 Multivariate Analysis  
   - 21-259 Calculus in Three Dimensions

   **Sequence 2**
   - 21-120 Differential and Integral Calculus

   and one of the following:

   **Linear Algebra**
   - 21-240 Matrix Algebra with Applications  
   - 21-241 Matrices and Linear Transformations  
   - 21-242 Matrix Theory

2. Data Analysis  
   **36 units**

   Data analysis is the art and science of extracting insight from data. The art lies in knowing which displays or techniques will reveal the most interesting features of a complicated data set. The science lies in understanding the various techniques and the assumptions on which they rely. Both aspects require practice to master.

   The Beginning Data Analysis courses give a hands-on introduction to the art and science of data analysis. The courses cover similar topics but differ slightly in the examples they emphasize. 36-200 draws examples from many fields and satisfy the DC College Core Requirement in Statistical Reasoning. One of these courses is therefore recommended for students in the College. (Note: A score of 4 or 5 on the Advanced Placement (AP) Exam in Statistics may be used to waive this requirement). Other courses emphasize examples in business (36-207), engineering and architecture (36-220), and the laboratory sciences (36-247).

   The Intermediate Data Analysis courses build on the principles and methods covered in the introductory course, and more fully explore specific types of data analysis methods in more depth.

   The Advanced Data Analysis courses draw on students’ previous experience with data analysis and understanding of statistical theory to develop advanced, more sophisticated methods. These core courses involve extensive analysis of real data with emphasis on developing the oral and writing skills needed for communicating results.

   **Sequence 1** (For students beginning their freshman or sophomore year)

   **Beginning Data Analysis**  
   Choose one of the following courses:
   - 36-200 Reasoning with Data  
   - 36/70-207 Probability and Statistics for Business Applications  
   - 36-220 Engineering Statistics and Quality Control  
   - 36-223 Statistical graphics  
   - 36-237 Experimental Design for Behavioral & Social Sciences

   **Intermediate Data Analysis**  
   Choose one of the following courses:
   - 36-202 Statistics & Data Science Methods  
   - 36-208 Regression Analysis  
   - 36-290 Introduction to Statistical Research Methodology  
   - 36-309 Experimental Design for Behavioral & Social Sciences

   **Or extra data analysis course in Statistics**

   **Note**: Students who enter the program with 36-225 or 36-226 should discuss options with an advisor. Any 36-300 or 36-400 level course in Data Analysis that does not satisfy any other requirement for a Statistics Major and Minor may be counted as a Statistical Elective.

   **Intermediate Data Analysis**
   Choose one of the following courses:
   - 36-202 Statistics & Data Science Methods  
   - 36-208 Regression Analysis  
   - 36-290 Introduction to Statistical Research Methodology  
   - 36-309 Experimental Design for Behavioral & Social Sciences

   **Or extra data analysis course in Statistics**

   **Must take prior to 36-401**
Advanced Data Analysis and Methodology

Take the following course:

36-401 Modern Regression 9

and one of the following courses:

36-402 Advanced Methods for Data Analysis 9
36-410 Introduction to Probability Modeling 9
36-461 Special Topics: Statistical Methods in Epidemiology 9
36-462 Special Topics: Data Mining 9
36-463 Special Topics: Multilevel and Hierarchical Models 9
36-464 Special Topics: Applied Multivariate Methods 9
36-465 Special Topics: An Introduction to Bayesian Inference 9
36-466 Special Topics: Statistical Methods in Finance 9
36-467 Special Topics: Data over Space & Time 9
36-468 Special Topics: Text Analysis 9
36-490 Undergraduate Research 9
36-497 Corporate Capstone Project 9

Special Topics rotate and new ones are regularly added.

Sequence 2 (For students beginning later in their college career)

Advanced Data Analysis and Methodology

Take the following course:

36-401 Modern Regression 9

and take two of the following courses (one of which must be 400-level):

36-303 Sampling, Survey and Society 9
36-311 Statistical Analysis of Networks 9
36-315 Statistical Graphics and Visualization 9
36-402 Advanced Methods for Data Analysis 9
36-410 Introduction to Probability Modeling 9
36-461 Special Topics: Statistical Methods in Epidemiology 9
36-462 Special Topics: Data Mining 9
36-463 Special Topics: Multilevel and Hierarchical Models 9
36-464 Special Topics: Applied Multivariate Methods 9
36-465 Special Topics: An Introduction to Bayesian Inference 9
36-466 Special Topics: Statistical Methods in Finance 9
36-467 Special Topics: Data over Space & Time 9
36-468 Special Topics: Text Analysis 9
36-490 Undergraduate Research 9
36-497 Corporate Capstone Project 9

Special Topics rotate and new ones are regularly added.

3. Probability Theory and Statistical Theory  18 units

To satisfy the theory requirement take the following two courses:

36-225 Introduction to Probability Theory 9
36-226 Introduction to Statistical Inference 9

or 36-326 Mathematical Statistics (Honors) 9

**It is possible to substitute 36-217, 36-218 or 21-325 for 36-225.** (36-225 is the standard introduction to probability, 36-217 is tailored for engineers and computer scientists, 36-218 is a more mathematically rigorous class for Computer Science students and more mathematically advanced Statistics students (Statistics students need advisor approval to enroll), and 21-325 is a rigorous Probability Theory course offered by the Department of Mathematics.) 36-326 Mathematical Statistics (Honors) can be substituted for 36-226 Introduction to Statistical Inference and is considered an honors course.

Comments:

(i) In order to be a Major or a Minor in good standing, a grade of at least a C is required in 36-225, 36-226 and 401. In particular, a grade of C or higher is required in order to be able to continue in the major.

Total number of units required for the minor 83 Units

Double Counting

With respect to double-counting courses, it is departmental policy that students must have at least three statistics courses (36-xxx) that do not count for their primary major. If students do not have at least three, they need to take additional advanced electives.

Sample Programs for the Minor

The following two sample programs illustrates two (of many) ways to satisfy the requirements of the Statistics Minor. Keep in mind that the program is flexible and can support many other possible schedules.

The first schedule uses calculus sequence 1, and 36-202t satisfies the intermediate data analysis requirement. The second schedule is an example of the case when a student enters the Minor through 36-225 and 36-226 (and therefore skips the beginning data analysis course). The schedule uses calculus sequence 2, and an advanced data analysis elective (to replace the beginning data analysis course).

Schedule 1

<table>
<thead>
<tr>
<th>Freshman</th>
<th>Sophomore</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fall</strong></td>
<td><strong>Spring</strong></td>
</tr>
<tr>
<td>36-211 Differential Calculus</td>
<td>36-212 Integral Calculus</td>
</tr>
<tr>
<td>36-200 Reasoning with Data</td>
<td>36-256 Multivariate Analysis</td>
</tr>
</tbody>
</table>

Schedule 2

<table>
<thead>
<tr>
<th>Freshman</th>
<th>Sophomore</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fall</strong></td>
<td><strong>Spring</strong></td>
</tr>
<tr>
<td>36-225 Introduction to Probability Theory</td>
<td>36-226 Introduction to Statistical Inference</td>
</tr>
</tbody>
</table>

Substitutions and Waivers

Many departments require Statistics courses as part of their Major or Minor programs. Students seeking transfer credit for those requirements from substitute courses (at Carnegie Mellon or elsewhere) should seek permission from their advisor in the department setting the requirement. The final authority in such decisions rests there. The Department of Statistics and Data Science does not provide approval or permission for substitution or waiver of another department's requirements.

However, the Statistics Director of Undergraduate Studies will provide advice and information to the student’s advisor about the viability of a proposed substitution. Students should make available as much information as possible concerning proposed substitutions. Students seeking waivers may be asked to demonstrate mastery of the material.

If a waiver or substitution is made in the home department, it is not automatically approved in the Department of Statistics and Data Science. In many of these cases, the student will need to take additional courses to satisfy the Statistics major requirements. Students should discuss this with a Statistics advisor when deciding whether to add an additional major in Statistics.

Statistics Majors and Minors seeking substitutions or waivers should speak to the Academic Advisor in Statistics.

Faculty

ZACHARY BRANSON, Assistant Teaching Professor – M.S. in Statistics, Harvard University; Carnegie Mellon, 2019–
DAVID CHOI, Assistant Professor of Statistics and Information Systems – Ph.D., Stanford University; Carnegie Mellon, 2004–

ALEXANDRA CHOULDECHOVA, Assistant Professor of Statistics and Public Policy – Ph.D., Stanford University; Carnegie Mellon, 2014–

PETER FREEMAN, Assistant Teaching Faculty – Ph.D., University of Chicago; Carnegie Mellon, 2004–

MAX G’SELL, Assistant Professor – Ph.D., Stanford University; Carnegie Mellon, 2014–

CHRISTOPHER R. GENOVESE, Department Head and Professor of Statistics – Ph.D., University of California, Berkeley; Carnegie Mellon, 1994–

JOEL B. GREENHOUSE, Professor of Statistics – Ph.D., University of Michigan; Carnegie Mellon, 1982–


JIASHUN JIN, Professor of Statistics – Ph.D., Stanford University; Carnegie Mellon, 2007–

BRIAN JUNKER, Associate Dean and Professor of Statistics – Ph.D., University of Illinois; Carnegie Mellon, 1990–

ROBERT E. KASS, Professor of Statistics – Ph.D., University of Chicago; Carnegie Mellon, 1981–

EDWARD KENNEDY, Assistant Professor – Ph.D., University of Pennsylvania; Carnegie Mellon, 2016–

ANN LEE, Associate Professor - Ph.D., Brown University; Carnegie Mellon, 2005–

JOHN P. LEHOCZKY, Thomas Lord Professor of Statistics – Ph.D., Stanford University; Carnegie Mellon, 1969–

JING LEI, Associate Professor - Ph.D., University of California, Berkeley; Carnegie Mellon, 2011–

ANJALI MAZUMDER, Assistant Research Professor


MATEY NEYKOV, Assistant Professor – Ph.D., Harvard University; Carnegie Mellon, 2017–

NYNKE NIEZINK, Assistant Professor – Ph.D., University of Groningen; Carnegie Mellon, 2017–

REBECCA NUGENT, Assistant Department Head, Teaching Professor - Ph.D., University of Washington; Carnegie Mellon, 2006–

ALEX REINHART, Assistant Teaching Faculty – Ph.D., Carnegie Mellon University; Carnegie Mellon, 2011–

ALESSANDRO RINALDO, Professor – Ph.D., Carnegie Mellon University; Carnegie Mellon, 2005–

LARRY WASSERMAN, Professor of Statistics – Ph.D., University of Toronto; Carnegie Mellon, 1988–

YUTING WEI, Assistant Professor – Ph.D, University of California; Carnegie Mellon, 2019–

Emeriti Faculty

George T. Duncan, Professor of Statistics and Public Policy – Ph.D., University of Minnesota; Carnegie Mellon, 1974–

William F. Eddy, John C. Warmer Professor of Statistics – Ph.D, Yale University; Carnegie Mellon, 1976–

Joseph B. Kadane, Leonard J. Savage Professor of Statistics and Social Sciences – Ph.D., Stanford University; Carnegie Mellon, 1969–

Mark J. Schervish, Professor of Statistics – Ph.D., University of Illinois; Carnegie Mellon, 1979–

Dale Lan, Teaching Professor – Ph.D., Carnegie Mellon University; Carnegie Mellon, 2001–

Adjunct Faculty

Olga Chilina, Lecturer – MS, University of Toronto; Carnegie Mellon, 2016–

April Galyardt – Ph.D., Carnegie Mellon University; Carnegie Mellon, 2017–

Christopher Peter Makris, Adjunct Lecturer – MSP, Carnegie Mellon University; Carnegie Mellon, 2018–

Ross O’Connell – Ph.D., University of Michigan; Carnegie Mellon, 2016–

Gordon Weinberg, Senior Lecturer – M.A. Mathematics, University of Pittsburgh; Carnegie Mellon, 2004–

Special Faculty

Robin Mejia

Affiliated Faculty

Anthony Brockwell – Ph.D., Melbourne University; Carnegie Mellon, 1999–

Bernie Devlin – Ph.D., Pennsylvania State University; Carnegie Mellon, 1994–

Sam Ventura – Ph.D., Carnegie Mellon University; Carnegie Mellon, 2015–