Department of Statistics and Data Science

Rebecca Nugent, Department Head
Peter Freeman, Director of Undergraduate Studies
Samantha Nielsen, Senior Undergraduate Academic Advisor
Glenn Clune, Undergraduate Academic Advisor
Amanda Mitchell, Undergraduate Academic Advisor
Email: statadvising@andrew.cmu.edu
Location: Baker Hall 132
www.stat.cmu.edu/ (http://www.stat.cmu.edu/)

Overview

Uncertainty is inescapable: randomness, measurement error, deception, and incomplete or missing information complicate 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 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 later graduate study in Statistics or to complement your interests in almost any field, including Psychology, Physics, Biology, History, Business, Information Systems, and Computer Science.
• The Minor in Statistics is useful complement to a (primary) major in another Department or College. Almost every field of inquiry must grapple with statistical problems, and the tools of statistical theory and data analysis you will develop in the Statistics minor (or Additional Major) will give you a critical edge.
• 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 data, designs for accurate measurement and comparison, and methods for checking the validity of theoretical assumptions. Statistical Theory informs Data Analysis and vice versa. The Department of Statistics and Data Science curriculum follows both of these threads and helps the student develop the complementary skills required.

Throughout the sections of this catalog, we describe the requirements for the Major in Statistics and the different categories within our basic curriculum, followed by the requirements for the Major in Economics and Statistics, the Major in Statistics and Machine Learning, and the Minor in Statistics.

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, Undergraduate Academic Advisor
Location: Baker Hall 132
statadvising@andrew.cmu.edu (statadvising@stat.cmu.edu)

Students in the Bachelor of Science 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 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) 29–39 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.

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

Sequence 1

<table>
<thead>
<tr>
<th>Course</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-111  Differential Calculus</td>
<td>10</td>
</tr>
<tr>
<td>21-112  Integral Calculus</td>
<td>10</td>
</tr>
<tr>
<td>and one of the following</td>
<td></td>
</tr>
<tr>
<td>21-256  Multivariate Analysis</td>
<td>9</td>
</tr>
<tr>
<td>21-259  Calculus in Three Dimensions</td>
<td>9</td>
</tr>
<tr>
<td>21-268  Multidimensional Calculus</td>
<td>10</td>
</tr>
</tbody>
</table>

Sequence 2

<table>
<thead>
<tr>
<th>Course</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-120  Differential and Integral Calculus</td>
<td>10</td>
</tr>
<tr>
<td>and one of the following</td>
<td></td>
</tr>
<tr>
<td>21-256  Multivariate Analysis</td>
<td>9</td>
</tr>
<tr>
<td>21-259  Calculus in Three Dimensions</td>
<td>9</td>
</tr>
<tr>
<td>21-268  Multidimensional Calculus</td>
<td>10</td>
</tr>
</tbody>
</table>

Notes:
- Passing the MSC 21-120 assessment test is an acceptable alternative to completing 21-120.

Linear Algebra**:
Complete one of the following three courses:

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

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 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:

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

Intermediate*
Choose one of the following courses:

<table>
<thead>
<tr>
<th>Course</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>36-202  Methods for Statistics &amp; Data Science **</td>
<td>9</td>
</tr>
<tr>
<td>36/70-208  Regression Analysis</td>
<td>9</td>
</tr>
<tr>
<td>36-309  Experimental Design for Behavioral &amp; Social Sciences</td>
<td>9</td>
</tr>
<tr>
<td>36-290  Introduction to Statistical Research Methodology</td>
<td>9</td>
</tr>
</tbody>
</table>

** Must take prior to 36-401, if not, an additional Advanced Data Analysis Elective is required

Advanced Data Analysis Elective
Choose one of the following courses:

<table>
<thead>
<tr>
<th>Course</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>36-303  Sampling, Survey and Society</td>
<td>9</td>
</tr>
<tr>
<td>36-311  Statistical Analysis of Networks</td>
<td>9</td>
</tr>
<tr>
<td>36-313  Statistics of Inequality and Discrimination</td>
<td>9</td>
</tr>
<tr>
<td>36-315  Statistical Graphics and Visualization</td>
<td>9</td>
</tr>
<tr>
<td>36-461  Special Topics: Statistical Methods in Epidemiology</td>
<td>9</td>
</tr>
<tr>
<td>36-462  Special Topics: Methods of Statistical Learning</td>
<td>9</td>
</tr>
<tr>
<td>36-463  Special Topics: Multilevel and Hierarchical Models</td>
<td>9</td>
</tr>
<tr>
<td>36-464  Special Topics: Applied Multivariate Methods</td>
<td>9</td>
</tr>
<tr>
<td>36-465  Special Topics: Conceptual Foundations of Statistical Learning</td>
<td>9</td>
</tr>
<tr>
<td>36-466  Special Topics: Statistical Methods in Finance</td>
<td>9</td>
</tr>
<tr>
<td>36-467  Special Topics: Data over Space &amp; Time</td>
<td>9</td>
</tr>
<tr>
<td>36-468  Special Topics: Text Analysis</td>
<td>9</td>
</tr>
<tr>
<td>36-469  Special Topics: Statistical Genomics and High Dimensional Inference</td>
<td>9</td>
</tr>
<tr>
<td>36-490  Undergraduate Research</td>
<td>9</td>
</tr>
<tr>
<td>36-493  Sports Analytics Capstone</td>
<td>9</td>
</tr>
<tr>
<td>36-497  Corporate Capstone Project</td>
<td>9</td>
</tr>
</tbody>
</table>

Students can also take a second 36-46x (see section #5).

and take the following two courses:

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

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

Advanced Data Analysis Electives
Choose two of the following courses:

<table>
<thead>
<tr>
<th>Course</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>36-303  Sampling, Survey and Society</td>
<td>9</td>
</tr>
<tr>
<td>36-311  Statistical Analysis of Networks</td>
<td>9</td>
</tr>
<tr>
<td>36-313  Statistics of Inequality and Discrimination</td>
<td>9</td>
</tr>
<tr>
<td>36-315  Statistical Graphics and Visualization</td>
<td>9</td>
</tr>
<tr>
<td>36-461  Special Topics: Statistical Methods in Epidemiology</td>
<td>9</td>
</tr>
<tr>
<td>36-462  Special Topics: Methods of Statistical Learning</td>
<td>9</td>
</tr>
<tr>
<td>36-463  Special Topics: Multilevel and Hierarchical Models</td>
<td>9</td>
</tr>
<tr>
<td>36-464  Special Topics: Applied Multivariate Methods</td>
<td>9</td>
</tr>
<tr>
<td>36-465  Special Topics: Conceptual Foundations of Statistical Learning</td>
<td>9</td>
</tr>
<tr>
<td>36-466  Special Topics: Statistical Methods in Finance</td>
<td>9</td>
</tr>
<tr>
<td>36-467  Special Topics: Data over Space &amp; Time</td>
<td>9</td>
</tr>
</tbody>
</table>
36-468 Special Topics: Text Analysis  9
36-469 Special Topics: Statistical Genomics and High Dimensional Inference  9
36-490 Undergraduate Research  9
36-493 Sports Analytics Capstone  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 requirement take the following two courses:

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

**It is possible to substitute 36-218, 36-219, or 21-325 for 36-225. (36-225 is the standard introduction to probability, 36-219 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 to meet the prerequisite requirements, a grade of at least a C is required in 36-225 (or equivalent), 36-226 (or equivalent), and 36-401.

4. Statistical Computing:  9 units

36-350 Statistical Computing  9

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.). 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. Note that all Special Topics are not offered every semester, and new Special Topics are regularly added.

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 the 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.

15-110 Principles of Computing  10
15-112 Fundamentals of Programming and Computer Science  12
15-121 Introduction to Data Structures  10
15-122 Principles of Imperative Computation  10
10-301 Introduction to Machine Learning  12
10-315 Introduction to Machine Learning (SCS Majors)  12
15-388 Practical Data Science  9

21-127 Concepts of Mathematics  10
21-260 Differential Equations  9
21-292 Operations Research I  9
21-301 Combinatorics  9
21-355 Principles of Real Analysis I  9
80-220 Philosophy of Science  9
80-221 Philosophy of Social Science  9
80-310 Formal Logic  9
85-310 Research Methods in Cognitive Psychology  9
85-320 Research Methods in Developmental Psychology  9
85-340 Research Methods in Social Psychology  9
88-223 Decision Analysis  12
88-302 Behavioral Decision Making  9

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. Concentration Area

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 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 require 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 has no other additional major or minor. The requirement does not apply for students who pursue an additional major in statistics.

Recommendations

Students in the 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 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.

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 will need to 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 applications. 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 21-300, 36-303, 36-490, 36-493, 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, 36-493 Sports Analytics Capstone, or 36-497 Corporate Capstone Project. Note that 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://coursescatalog.web.cmu.edu/dietrichcollegeofhumanitiesandsocialsciences/#collegeservicesandprograms) to gain research experience.

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.

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 Methods for Statistics &amp; Data Science</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Junior</th>
<th>Senior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>Spring</td>
</tr>
<tr>
<td>36-401 Modern Regression</td>
<td>36-402 Advanced Methods for Data Analysis</td>
</tr>
<tr>
<td>-----</td>
<td>-----</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>21-120 Differential and Integral Calculus</td>
<td>21-256 Multivariate Analysis</td>
</tr>
<tr>
<td>36-200 Reasoning with Data</td>
<td>-----</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Junior</th>
<th>Senior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>Spring</td>
</tr>
<tr>
<td>36-401 Modern Regression</td>
<td>36-402 Advanced Methods for Data Analysis</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
</tr>
</tbody>
</table>

### B.S. in Statistics (Mathematical Sciences Track)

Glenn Clune, Undergraduate Academic Advisor
Location: Baker Hall 132
statadvising@andrew.cmu.edu (statadvising@stat.cmu.edu)

Students in the Bachelor of Science 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 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 (Mathematical Sciences Track) are detailed below and are organized by categories #1-#7.

### Curriculum

#### 1. Mathematical Foundations (Prerequisites)

29-39 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.

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

**Sequence 1**

<table>
<thead>
<tr>
<th>21-111</th>
<th>Differential Calculus</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-112</td>
<td>Integral Calculus</td>
<td>10</td>
</tr>
</tbody>
</table>

and one of the following

<table>
<thead>
<tr>
<th>21-256</th>
<th>Multivariate Analysis</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-259</td>
<td>Calculus in Three Dimensions</td>
<td>9</td>
</tr>
<tr>
<td>21-268</td>
<td>Multidimensional Calculus</td>
<td>10</td>
</tr>
</tbody>
</table>

**Sequence 2**

| 21-120 | Differential and Integral Calculus | 10 |

and one of the following

<table>
<thead>
<tr>
<th>21-256</th>
<th>Multivariate Analysis</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-259</td>
<td>Calculus in Three Dimensions</td>
<td>9</td>
</tr>
<tr>
<td>21-268</td>
<td>Multidimensional Calculus</td>
<td>10</td>
</tr>
</tbody>
</table>

**Notes:**

- Passing the MSC 21-120 assessment test is an acceptable alternative to completing 21-120.

**Linear Algebra**: Complete one of the following three courses:

<table>
<thead>
<tr>
<th>21-240</th>
<th>Matrix Algebra with Applications</th>
<th>10</th>
</tr>
</thead>
<tbody>
<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>
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 satisfy 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

**Intermediate**

Choose one of the following courses:

- 36-202 Methods for Statistics & Data Science ** 9
- 36/70-208 Regression Analysis 9
- 36-309 Experimental Design for Behavioral & Social Sciences 9
- 36-290 Introduction to Statistical Research Methodology 9

* Or extra data analysis course in Statistics 9

** Must take prior to 36-401, if not, an additional Advanced Data Analysis Elective is required

Advanced Data Analysis Elective

Choose one of the following courses:

- 36-303 Sampling, Survey and Society 9
- 36-311 Statistical Analysis of Networks 9
- 36-313 Statistics of Inequality and Discrimination 9
- 36-315 Statistical Graphics and Visualization 9
- 36-461 Special Topics: Statistical Methods in Epidemiology 9
- 36-462 Special Topics: Methods of Statistical Learning 9
- 36-463 Special Topics: Multilevel and Hierarchical Models 9
- 36-464 Special Topics: Applied Multivariate Methods 9
- 36-465 Special Topics: Conceptual Foundations of Statistical Learning 9
- 36-466 Special Topics: Applied Multivariate Methods 9
- 36-467 Special Topics: Statistical Methods in Finance 9
- 36-468 Special Topics: Data over Space & Time 9
- 36-469 Special Topics: Statistical Genomics and High Dimensional Inference 9

Students can also take a second 36-46x (see section #5), and take the following two courses:

- 36-401 Modern Regression 9
- 36-402 Advanced Methods for Data Analysis 9

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
- 36-313 Statistics of Inequality and Discrimination 9
- 36-315 Statistical Graphics and Visualization 9
- 36-461 Special Topics: Statistical Methods in Epidemiology 9
- 36-462 Special Topics: Methods of Statistical Learning 9
- 36-463 Special Topics: Multilevel and Hierarchical Models 9
- 36-464 Special Topics: Applied Multivariate Methods 9
- 36-465 Special Topics: Conceptual Foundations of Statistical Learning 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
- 36-469 Special Topics: Statistical Genomics and High Dimensional Inference 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
- 36-226 Introduction to Statistical Inference 9
- 36-326 Mathematical Statistics (Honors) 9

** It is possible to substitute 36-218, 36-219, or 21-325 for 36-225. (36-225 is the standard introduction to probability, 36-219 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 to meet the prerequisite requirements, a grade of at least a C is required in 36-225 (or equivalent), 36-226 (or equivalent), and 36-401.

4. Statistical Computing: 9 units

- 36-350 Statistical Computing 9

---
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.). 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. Note that all Special Topics are not offered every semester, and new Special Topics are regularly added.

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 the 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.

- 15-110 Principles of Computing  
- 15-112 Fundamentals of Programming and Computer Science  
- 15-121 Introduction to Data Structures  
- 15-122 Principles of Imperative Computation  
- 10-301 Introduction to Machine Learning  
- 10-315 Introduction to Machine Learning (SCS Majors)  
- 15-388 Practical Data Science  
- 21-127 Concepts of Mathematics  
- 21-260 Differential Equations  
- 21-292 Operations Research I  
- 21-301 Combinatorics  
- 21-355 Principles of Real Analysis I  
- 80-220 Philosophy of Science  
- 80-221 Philosophy of Social Science  
- 80-310 Formal Logic  
- 85-310 Research Methods in Cognitive Psychology  
- 85-320 Research Methods in Developmental Psychology  
- 88-223 Decision Analysis  
- 88-302 Behavioral Decision Making

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

Mathematical Statistics Track  
46–52 units

- 21-127 Concepts of Mathematics  
- 21-355 Principles of Real Analysis I  
- 36-410 Introduction to Probability Modeling

And two of the following:

- 21-228 Discrete Mathematics  
- 21-257 Models and Methods for Optimization  
- 21-292 Operations Research I  
- 21-301 Combinatorics  
- 21-344 Numerical Linear Algebra  
- 21-356 Principles of Real Analysis II  
- 36-700 Probability and Mathematical Statistics

Recommendations

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

Additional Major in Statistics (Mathematical Science Track)

Students who elect Statistics (Mathematical Science Track) as an additional major must fulfill all Statistics (Mathematical Science Track) degree requirements. 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 (Mathematical Science Track).

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 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, 36-493, 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, 36-493 Sports Analytics Capstone, 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/dietrich collegeofhumanitiesand socialsciences/ collegeservic esandprograms) for additional research experience.

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 intermediate data analysis course). This schedule has more emphasis on statistical theory and probability.
B. S. in Statistics (Statistics and Neuroscience Track)

Glenn Clune, Undergraduate Academic Advisor
Location: Baker Hall 132
statadvising@andrew.cmu.edu (statadvising@stat.cmu.edu)

Students in the Bachelor of Science 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 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 (Neuroscience Track) are detailed below and are organized by categories #1-#7.

Curriculum

1. Mathematical Foundations (Prerequisites)  29-39 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.

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
- 21-112  Integral Calculus
and one of the following
- 21-256  Multivariate Analysis

**Sequence 2**
- 21-259  Calculus in Three Dimensions
- 21-268  Multidimensional Calculus

Notes:
- **Beginning**
  Choose one of the following courses:
  - 36-200  Reasoning with Data
  - 36-207  Engineering Statistics and Quality Control
  - 36-247  Statistics for Lab Sciences
  - 36-270  Probability and Statistics for Business Applications
  - 36-220  Engineering Statistics and Quality Control
  - 36-247  Statistics for Lab Sciences

Note: Students who enter the program with AP or SAT scores of 4 or 5 are recommended to complete one of the following courses: 36-200 Reasoning with Data, 36-247 Statistics for Lab Sciences.

Intermediate**:
Choose one of the following courses:
- 36-202  Methods for Statistics & Data Science
- 36-207  Engineering Statistics and Quality Control
- 36-220  Engineering Statistics and Quality Control
- 36-247  Statistics for Lab Sciences
- 36-290  Introduction to Statistical Research Methodology

*Or extra data analysis course in Statistics

Department of Statistics and Data Science  7
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:
  - **36-226** Introduction to Statistical Inference 9
  - **36-326** Mathematical Statistics (Honors) 9

**It is possible to substitute 36-218, 36-219 or 21-325 for 36-225. (36-225 is the standard introduction to probability; 36-219 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 to meet the prerequisite requirements, a grade of at least a C is required in 36-225 (or equivalent), 36-226 (or equivalent) and 36-401.

4. Statistical Computing: 9 units

- **36-350** Statistical Computing 9

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.). 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. Note that all Special Topics are not offered every semester, and new Special Topics are regularly added.

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 the 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.

- **15-110** Principles of Computing 10
- **15-112** Fundamentals of Programming and Computer Science 12
- **15-121** Introduction to Data Structures 10
- **15-122** Principles of Imperative Computation 10
- **10-301** Introduction to Machine Learning 12
- **10-315** Introduction to Machine Learning (SCS Majors) 12
- **15-388** Practical Data Science 9
- **21-127** Concepts of Mathematics 10
- **21-260** Differential Equations 9
- **21-292** Operations Research I 9
- **21-301** Combinatorics 9
- **21-355** Principles of Real Analysis I 9
- **80-220** Philosophy of Science 9
- **80-221** Philosophy of Social Science 9
- **80-310** Formal Logic 9
- **85-310** Research Methods in Cognitive Psychology 9
- **85-320** Research Methods in Developmental Psychology 9
- **85-340** Research Methods in Social Psychology 9
- **88-223** Decision Analysis 12
- **88-302** Behavioral Decision Making 9

**Statistics and Neuroscience Track** 45–54 units

- **85-211** Cognitive Psychology 9
- **85-219** Biological Foundations of Behavior 9

And three electives (at least one from Methodology and Analysis and at least one within the Neuroscience Background listed below):
Methodology and Analysis
36-700 Probability and Mathematical Statistics 12
10-301 Introduction to Machine Learning 12
18-290 Signals and Systems 12
85-314 Cognitive Neuroscience Research Methods 9
42/86-631 Neural Data Analysis 12

Neuroscience Background
03-362 Cellular Neuroscience 9
03-363 Systems Neuroscience 9
15-386 Neural Computation 9
85-414 Cognitive Neuropsychology 9
85-419 Introduction to Parallel Distributed Processing 9

Total Number of Units for the Major: 146-185*
Total Number of Units for the Degree: 360

* Note: This number can vary depending on the calculus sequence and on the concentration area. Students take in addition this number includes the 36 units of the "Concentration Area" category which may not be required (see category 7 above for details).

Recommendations
Students in the 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 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 Biostatistics (or related programs) after completing their undergraduate degree are strongly encouraged to pursue the Mathematical Statistics Track.

Additional Major in Statistics (Neuroscience Track)
Students who elect Statistics (Neuroscience Track) as an additional major must fulfill all Statistics (Neuroscience Track) degree requirements. 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 may 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 if many of the other major's requirements overlap with the requirements for a Major in Statistics (Neuroscience Track).

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 & 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 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 applications. 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, 36-493, 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, 36-493 Sports Analytics Capstone, or 36-497 Corporate Capstone Project. Note that 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), for additional research experience.

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 intermediate data analysis course). This schedule has more emphasis on statistical theory and probability.

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 Methods for Data Analysis</td>
</tr>
<tr>
<td>36-401 Modern Regression</td>
<td>85-211 Cognitive Psychology</td>
</tr>
</tbody>
</table>

Junior | Senior |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>Spring</td>
</tr>
<tr>
<td>36-401 Modern Regression</td>
<td>36-402 Advanced Methods for Data Analysis</td>
</tr>
<tr>
<td>Neuroscience Track Elective</td>
<td>Neuroscience Track Elective</td>
</tr>
<tr>
<td>-----</td>
<td>-----</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>36-200 Reasoning with Data</td>
<td>36-202 Methods for Data Analysis</td>
</tr>
<tr>
<td>21-111 Differential Calculus</td>
<td>21-112 Integral Calculus</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
</tr>
</tbody>
</table>

Junior | Senior |
<table>
<thead>
<tr>
<th></th>
<th></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>85-219 Biological Foundations of Behavior</td>
<td>Neuroscience Track Elective</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
</tr>
</tbody>
</table>
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 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
   21-268 Multidimensional Calculus 10

   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 Methods for Statistics & Data Science ** 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, if not, an additional Advanced Statistics Elective is required.

**Advanced Statistics Elective**

Choose **two** of the following courses:

36-303 Sampling, Survey and Society 9
36-311 Statistical Analysis of Networks 9
36-313 Statistics of Inequality and Discrimination 9
36-315 Statistical Graphics and Visualization 9
36-461 Special Topics: Statistical Methods in Epidemiology 9
36-462 Special Topics: Methods of Statistical Learning 9
36-463 Special Topics: Multilevel and Hierarchical Models 9
36-464 Special Topics: Applied Multivariate Methods 9
36-465 Special Topics: Conceptual Foundations of Statistical Learning 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
36-469 Special Topics: Statistical Genomics and High Dimensional Inference 9
36-490 Undergraduate Research 9
36-493 Sports Analytics Capstone 9
36-497 Corporate Capstone Project 9

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

**Advanced Statistics Electives**

Choose **three** of the following courses:

36-303 Sampling, Survey and Society 9
36-311 Statistical Analysis of Networks 9
36-313 Statistics of Inequality and Discrimination 9
36-315 Statistical Graphics and Visualization 9
36-461 Special Topics: Statistical Methods in Epidemiology 9
36-462 Special Topics: Methods of Statistical Learning 9
36-463 Special Topics: Multilevel and Hierarchical Models 9
36-464 Special Topics: Applied Multivariate Methods 9
36-465 Special Topics: Conceptual Foundations of Statistical Learning 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
36-469 Special Topics: Statistical Genomics and High Dimensional Inference 9
36-490 Undergraduate Research 9
36-493 Sports Analytics Capstone 9
36-497 Corporate Capstone Project 9

**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
Students seeking transfer credit for those requirements from Economics and Statistics.

Other major's requirements overlap with the requirements for a Major in Economics and Statistics. This has a complex set of requirements and prerequisites or when many of the requirements for the 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 (73-xxx) and three Statistics (36-xxx)] that do not count for their primary major. If students do not have at least three ECON and three STA classes, they will need to 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 Methods for Data Analysis</td>
</tr>
<tr>
<td>36-200 Reasoning with Data</td>
<td>21-256 Multivariate Analysis</td>
</tr>
<tr>
<td>73-102 Principles of Microeconomics</td>
<td>73-103 Principles of Macroeconomics</td>
</tr>
<tr>
<td>73-060 Economics: BaseCamp *not required</td>
<td>73-265 Economics and Data Science</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Junior</th>
<th>Senior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>Spring</td>
</tr>
<tr>
<td>36-401 Modern Regression</td>
<td>73-270 Professional Communication for Economists</td>
</tr>
<tr>
<td>73-374 Econometrics II</td>
<td>73-274 Econometrics I</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 are advised to 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, Senior Academic Advisor
Amanda Mitchell, Academic Advisor

Location: Baker Hall 132
statadvising@andrew.cmu.edu (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)

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
21-268 Multidimensional Calculus 10

**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
21-268 Multidimensional Calculus 10

**Notes:**

- Passing the Mathematical Sciences 21-120 assessment test is an acceptable alternative to completing 21-120
- Integration and Approximation
  21-122 Integration and Approximation 10
- 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

* 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:

21-127 Concepts of Mathematics 10

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 satisfies 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 (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.

**Intermediate**

Choose one of the following courses:

36-202 Methods for Statistics & Data Science ** 9
36/70-208 Regression Analysis 9
36-309 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 or will need to take an additional Advanced Data Analysis Elective

**Advanced Data Analysis Electives**

Choose two of the following courses:

36-303 Sampling, Survey and Society 9
36-311 Statistical Analysis of Networks 9
36-313 Statistics of Inequality and Discrimination 9
36-315 Statistical Graphics and Visualization 9
36-461 Special Topics: Statistical Methods in Epidemiology 9
36-462 Special Topics: Methods of Statistical Learning 9
36-463 Special Topics: Multilevel and Hierarchical Models 9
36-464 Special Topics: Applied Multivariate Methods 9
36-465 Special Topics: Conceptual Foundations of Statistical Learning 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
36-469 Special Topics: Statistical Genomics and High Dimensional Inference 9
36-490 Undergraduate Research 9
36-493 Sports Analytics Capstone 9
36-497 Corporate Capstone Project 9

All Special Topics are not offered every semester. They are on a rotation and new Special Topics are regularly added.

and take the following two courses:

36-401 Modern Regression 9
36-402 Advanced Methods for Data Analysis 9

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

**Advanced Data Analysis Electives**

Choose three of the following courses:

36-303 Sampling, Survey and Society 9
36-311 Statistical Analysis of Networks 9
36-313 Statistics of Inequality and Discrimination 9
36-315 Statistical Graphics and Visualization 9
36-461 Special Topics: Statistical Methods in Epidemiology 9
36-462 Special Topics: Methods of Statistical Learning 9
36-463 Special Topics: Multilevel and Hierarchical Models 9
36-464 Special Topics: Applied Multivariate Methods 9
36-465 Special Topics: Conceptual Foundations of Statistical Learning 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
36-469 Special Topics: Statistical Genomics and High Dimensional Inference 9
36-490 Undergraduate Research 9
36-493 Sports Analytics Capstone 9
36-497 Corporate Capstone Project 9

All Special Topics are not offered every semester. They are on a rotation and new Special Topics are regularly added.

and take the following two courses:

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

- 36-401 Modern Regression
- 36-402 Advanced Methods for Data Analysis

### 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
- 36-226 Introduction to Statistical Inference
  
  *It is possible to substitute 36-218, 36-219or 21-325 for 36-225. 36-225 is the standard introduction to probability, 36-219 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 meet the prerequisite requirements, a grade of at least a C is required in 36-225 (or equivalent), 36-226 (or equivalent) and 36-401.

### 4. Statistical Computing 9 units

36-350 Statistical Computing

### 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 inards 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. 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 major's requirements overlap with the requirements for a Major in Statistics and Machine Learning.

and take one of the following Machine Learning Advanced Electives:

- 05-434 Machine Learning in Practice
- 10-403 Deep Reinforcement Learning & Control
- 10-703 Deep Reinforcement Learning & Control
- 10-405 Machine Learning with Large Datasets (Undergraduate)
- 10-605 Machine Learning with Large Datasets
- 10-417 Intermediate Deep Learning

- 10-418 Machine Learning for Structured Data
- 10-707 Advanced Deep Learning
- 11-344 Machine Learning in Practice
- 11-411 Natural Language Processing
- 11-441 Machine Learning for Text Mining
- 11-485 Introduction to Deep Learning
- 11-661 Language and Statistics
- 11-761 Language and Statistics
- 15-281 Artificial Intelligence: Representation and Problem Solving
- 15-386 Neural Computation
- 15-387 Computational Perception
- 16-311 Introduction to Robotics
- 16-385 Computer Vision
- 16-720 Computer Vision

*PhD level ML course as approved by Statistics advisor

** Independent research with an ML faculty member as approved by Statistics Advisor

***This is not an exhaustive list. Please contact your Academic Advisor if there is a course you are considering taking that is not on this list.

<table>
<thead>
<tr>
<th>Total number of units for the</th>
<th>176–198 units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of units for the degree</td>
<td>360 units</td>
</tr>
</tbody>
</table>

### 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 will need to 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.

- 10-418 Machine Learning for Structured Data
- 10-707 Advanced Deep Learning
- 11-344 Machine Learning in Practice
- 11-411 Natural Language Processing
- 11-441 Machine Learning for Text Mining
- 11-485 Introduction to Deep Learning
- 11-661 Language and Statistics
- 11-761 Language and Statistics
- 15-281 Artificial Intelligence: Representation and Problem Solving
- 15-386 Neural Computation
- 15-387 Computational Perception
- 16-311 Introduction to Robotics
- 16-385 Computer Vision
- 16-720 Computer Vision

*PhD level ML course as approved by Statistics advisor

** Independent research with an ML faculty member as approved by Statistics Advisor

***This is not an exhaustive list. Please contact your Academic Advisor if there is a course you are considering taking that is not on this list.

<table>
<thead>
<tr>
<th>Total number of units for the</th>
<th>176–198 units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of units for the degree</td>
<td>360 units</td>
</tr>
</tbody>
</table>

### 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 will need to 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 waiver of 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></td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td>Spring</td>
</tr>
<tr>
<td>36-200 Reasoning with Data</td>
<td>36-202 Methods for Statistics &amp; Data Science</td>
</tr>
<tr>
<td>21-120 Differential and Integral Calculus</td>
<td>21-256 Multivariate Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Junior</th>
<th>Senior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>Spring</td>
</tr>
<tr>
<td>36-401 Modern Regression</td>
<td>36-402 Advanced Methods for Data Analysis</td>
</tr>
<tr>
<td></td>
<td>15-351 Algorithms and Advanced Data Structures</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.

### Schedule 2

<table>
<thead>
<tr>
<th>Freshman</th>
<th>Sophomore</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td>Spring</td>
</tr>
<tr>
<td>36-200 Reasoning with Data</td>
<td>21-127 Concepts of Mathematics</td>
</tr>
<tr>
<td>21-256 Multivariate Analysis</td>
<td></td>
</tr>
<tr>
<td>15-112 Fundamentals of Programming and Computer Science</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Junior</th>
<th>Senior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>Spring</td>
</tr>
<tr>
<td>36-401 Modern Regression</td>
<td>15-351 Algorithms and Advanced Data Structures</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

**The linear algebra requirement needs to be complete before taking 36-401 Modern Regression or 36-46X Special Topics.

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

### 2. Data Analysis

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 and Methodology 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 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

**Intermediate Data Analysis**

Choose one of the following courses:

- 36-202 Methods for Statistics & Data Science ** 9
- 36-70-208 Regression Analysis 9
- 36-290 Introduction to Statistical Research Methodology 9
- 36-309 Experimental Design for Behavioral & Social Sciences 9

*The Beginning and Intermediate Data Analysis sequence (i.e. 36-200 and 36-202, or equivalents as listed above) can be replaced with an additional Advanced Analysis and Methodology course, shown below in Sequence 2.*

**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: Methods of Statistical Learning 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
- 36-469 Special Topics: Statistical Genomics and High Dimensional Inference 9
- 36-490 Undergraduate Research 9
- 36-493 Sports Analytics Capstone 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

**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)

**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. Make sure to consult your Statistics Minor advisor regarding double counting.

**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 36-202 to satisfy 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>Fall</td>
<td>Fall</td>
</tr>
<tr>
<td>36-225 Introduction to Probability Theory</td>
<td>36-226 Introduction to Statistical Inference</td>
</tr>
<tr>
<td>36-226 Introduction to Statistical Inference</td>
<td>36-202 Methods for Statistics &amp; Data Science</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fall</th>
<th>Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-112 Differential Calculus</td>
<td>21-112 Integral Calculus</td>
</tr>
<tr>
<td>36-200 Reasoning with Data</td>
<td>21-256 Multivariate Analysis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Junior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
</tr>
<tr>
<td>36-225 Introduction to Probability Theory</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Senior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
</tr>
<tr>
<td>36-401 Modern Regression</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 COULDECHOVA, Assistant Professor of Statistics and Public Policy – Ph.D., Stanford University; Carnegie Mellon, 2014–

REBECCA DOERGE, Dean of Mellon College of Science, Professor of Statistics – Ph.D., North Carolina State University; Carnegie Mellon, 2016–

PETER FREEMAN, Associate Teaching Professor; Director of Undergraduate Studies – Ph.D., University of Chicago; Carnegie Mellon, 2004–

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

CHRISTOPHER R. GENOVESE, 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–

AMELIA HAVILAND, Professor of Statistics and Public Policy – Ph.D., Carnegie Mellon University; Carnegie Mellon, 2003–

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

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

ROBERT K. KASS, Maurice Falk Professor of Statistics & Computational Neuroscience – Ph.D., University of Chicago; Carnegie Mellon, 1981–

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

ARUN KUCHIBHOTLA, Assistant Professor – Ph.D., University of Pennsylvania; Carnegie Mellon, 2020–

MIKAEL KUUSELA, Assistant Professor – Ph.D, Ecole Polytechnique Federale de Lausanne; Carnegie Mellon, 2018–

ANN LEE, Associate Professor, Co-Director of PhD program – 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–

ROBIN MEJIA, Assistant Research Professor – PhD, UC Berkeley; Carnegie Mellon, 2018–


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, Department Head, Stephen E. and Joyce Fienberg Professor of Statistics & Data Science – Ph.D., University of Washington; Carnegie Mellon, 2006–

AADITYA RAMDAS, Assistant Professor – PhD, Carnegie Mellon; Carnegie Mellon, 2018–

ALEX REINHART, Assistant Teaching Faculty – Ph.D., Carnegie Mellon University; Carnegie Mellon, 2018–

ALESSANDRO RINALDO, Associate Dean for Research, Professor – Ph.D., Carnegie Mellon; Carnegie Mellon, 2005–

KATHRYN ROEDER, UPMC Professor of Statistics and Life Sciences – Ph.D., Pennsylvania State University; Carnegie Mellon, 1994–

CHAD M. SCHAFER, Professor – Ph.D., University of California, Berkeley; Carnegie Mellon, 2004–

TEDDY SEIDENFELD, Herbert A. Simon Professor of Philosophy and Statistics – Ph.D., Columbia University; Carnegie Mellon, 1985–

COSMA SHALIZI, Associate Professor – Ph.D., University of Wisconsin, Madison; Carnegie Mellon, 2005–

RYAN TIBSHIRANI, Professor – Ph.D., Stanford University; Carnegie Mellon, 2011–

VALERIE VENTURA, Professor, Co-Director of PhD program – Ph.D., University of Oxford; Carnegie Mellon, 1997–

ISABELLA VERDINELLI, Professor in Residence – Ph.D., Carnegie Mellon University; Carnegie Mellon, 1991–

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

Emeriti Faculty

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

WILLIAM F. EDDY, John C. Warner Professor of Statistics – Ph.D, Yale University; Carnegie Mellon, 1976–

JOSEPH K. 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 Wisconsin; Carnegie Mellon, 1979–

DALENE STANGL, Teaching Professor – Ph.D., Carnegie Mellon University; Carnegie Mellon, 2017–

Adjunct Faculty

OLGA CHILINA, Lecturer – MS, University of Toronto; Carnegie Mellon, 2016–

Special Faculty

MOHAMED FARAG, Lecturer – D. Engineering, George Washington University; Carnegie Mellon, 2020–


JAMIE MCGOVERN, Director of the Master’s in the Statistical Practice Program; Carnegie Mellon, 2020–

GORDON WEINBERG, Senior Lecturer – M.A. Mathematics, University of Pittsburgh; Carnegie Mellon, 2004–
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–