

Department of Statistics and Data Science

Rebecca Nugent, Department Head

Peter Freeman, Director of Undergraduate Studies

Zach Branson, Assistant Director of the Undergraduate Program

Samantha Nielsen, Associate Director of Academic Programs

Amanda Mitchell, Lead Senior Academic Advisor

Glenn Clune, Academic Program Manager

Sylvie Aubin, Academic Program Manager

Peter Long, Academic Advisor

Email: statadvising@andrew.cmu.edu
Location: Baker Hall 129

www.stat.cmu.edu/ (<http://www.stat.cmu.edu/>)

Overview

Uncertainty is inescapable: randomness, measurement error, deception, and incomplete or missing information all 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 have diverse skills in computing, mathematics, decision making, designing experiments, forecasting, and interpreting and communicating analysis results. Moreover, effective statisticians actively collaborate with people in other fields and, in the process, learn about other fields. Statistics & Data Science students who master core concepts and collaboration are highly sought after in the marketplace.

Recent statistics majors at Carnegie Mellon have taken jobs at leading companies in many fields, including the National Economic Research Association, Boeing, Morgan Stanley, Deloitte, Rosetta Marketing Group, Nielsen, Proctor & Gamble, Accenture, and Goldman Sachs. Others 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 also go on to graduate study at some of the top programs in the country including Carnegie Mellon, Harvard, MIT, Yale, NYU, Penn, Johns Hopkins, Duke, Michigan, Chicago, Northwestern, Washington, Stanford, and California.

The Department and Faculty

The Department of Statistics & Data Science at Carnegie Mellon University is world-renowned for its contributions to statistical theory and practice. Research in the department spans 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 and Data Science 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 (or Additional Major) in Statistics and Data Science is a 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 & 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 & 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 & Data Science.
- 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 & Data Science and the field in general.

Curriculum

Statistics and Data Science 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 & Data Science curriculum follows both of these threads and helps students develop required skills.

Throughout the sections of this catalog, we describe the requirements for the Major in Statistics and Data Science (the core major as well as the Mathematics and Neuroscience tracks), followed by the requirements for the Major in Economics and Statistics, the Major in Statistics and Machine Learning, and the Minor in Statistics and Data Science.

Note: We recommend that you use the information provided below as a general guideline, and then schedule a meeting with a Statistics and Data Science 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 and Data Science

Peter Freeman, *Undergraduate Program Director*

Location: Baker Hall 229
pfreeman@andrew.cmu.edu

Zach Branson, *Assistant Director of the Undergraduate Program*
Location: Baker Hall 232
zbranson@andrew.cmu.edu

Amanda Mitchell, *Lead Senior Academic Advisor*
Glenn Clune, *Academic Program Manager*
Sylvie Aubin, *Academic Program Manager*

Peter Long, *Academic Advisor*
 Location: Baker Hall 129
 statadvising@andrew.cmu.edu (statadvising@stat.cmu.edu)

Students in the Bachelor of Science in Statistics and Data 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 and Data Science majors gain experience in applying statistical tools to real problems in other fields and learn the nuances of interdisciplinary collaboration. The requirements for the B.S. in Statistics and Data Science are detailed below and are organized by categories #1-7.

Curriculum

1. Mathematical Foundations (Prerequisites) 39-52 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 & Data Science.

Complete the following:

21-090	Precalculus	10
Complete one of the following options:		
21-111	Differential Calculus	10
21-112	Integral Calculus	10
OR		
21-120	Differential and Integral Calculus	10
And one of the following three courses:		
21-256	Multivariate Analysis	9
21-259	Calculus in Three Dimensions	10
21-268	Multidimensional Calculus	11
And one of the following three courses:		
21-240	Matrix Algebra with Applications	10
21-241	Matrices and Linear Transformations	11
21-242	Matrix Theory	11

NOTE: Passing the Mathematical Sciences assessment tests available during First-Year Orientation is an acceptable alternative to completing 21-090 and/or 21-120.

- 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. draws examples from many fields and satisfies the Dietrich College Core Requirement in Statistical Reasoning. This course is therefore required for students in the college. (**Note:** a score of 5 on the Advanced Placement (AP) Exam in Statistics may be used to waive this requirement). 36-220 emphasizes examples in engineering.

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.

Beginning Data Analysis

Choose *one* of the following courses:

36-200	Reasoning with Data *	9
36-220	Engineering Statistics and Quality Control	9

* A score of 5 on the Advanced Placement (AP) Exam in Statistics may be used to waive this requirement.
NOTE: Students who enter the program with credit for probability and statistical inference should discuss options with an advisor.

Sequence 1

Intermediate Data Analysis

Choose *one* of the following courses:

36-202	Methods for Statistics & Data Science *	9
36-309	Experimental Design for Behavioral & Social Sciences	9
36-290	Introduction to Statistical Research Methodology	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
36-318	Introduction to Causal Inference	9
36-390	Study Abroad Experience in Statistics and Data Science	9
36-396	Tartan Athletics Analytics	9
36-490	Undergraduate Research	9
36-493	Sports Analytics Capstone	9
36-497	Corporate Capstone Project	9

Students can also take a second Special Topics (36-46x or 36-47x) course to fulfill an advanced data analysis elective requirement (see section #5).

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

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
36-318	Introduction to Causal Inference	9
36-390	Study Abroad Experience in Statistics and Data Science	9
36-396	Tartan Athletics Analytics	9
36-490	Undergraduate Research	9
36-493	Sports Analytics Capstone	9
36-497	Corporate Capstone Project	9

Students can also take a second Special Topics (36-46x or 36-47x) course to fulfill an advanced data analysis elective requirement (see section #5).

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, complete the following:

Take one of the following courses:

36-235	Probability and Statistical Inference I *	9
36-225	Introduction to Probability Theory	9

And one of the following three courses:

36-236	Probability and Statistical Inference II **	9
36-226	Introduction to Statistical Inference	9
36-326	Mathematical Statistics (Honors)	9

* It is possible to substitute 36-218, 36-219, 36-225, 15-259 or 21-325 for 36-235. 36-235 is the standard (and recommended) introduction to probability, 36-219 is tailored for engineers and computer scientists, 36-218 and 15-259 are more mathematically rigorous classes for Computer Science students and more mathematically advanced (students need advisor approval to enroll), and 21-325 is a rigorous probability theory course offered by the Department of Mathematics.)

** It is possible to substitute 36-226 or 36-326 (honors course) for 36-236. 36-236 is the standard (and recommended) introduction to statistical inference.

NOTE: Students who enter the program with credit for probability and statistical inference should discuss options with an advisor.

Please note that students who complete 36-235 are expected to take 36-236 to complete their theory requirements. Students who choose to take 36-225 instead will be required to take 36-226 afterward. They will not be eligible to take 36-236.

Comment:

(i) In order to meet the prerequisite requirements, a grade of at least a C is required in 36-235 (or equivalent) and 36-236 (or equivalent).

4. Statistical Computing 19-21 units

Fundamental to the practice of statistics and data science is the ability to effectively code data processing and analysis tasks. Within the domain of statistics, the use of the programming language R is ubiquitous, and thus we expose students to it throughout the curriculum (and in depth in Statistical Computing). Within the larger domain of data science, the use of the programming language Python is also ubiquitous, and thus we require all majors to gain, at a minimum, basic competency in the language by taking either Principles of Computing, or Fundamentals of Programming and Computer Science. We would advise those students who are considering receiving course credit for one of these two courses given their score on the AP Computer Science A exam to actually take one (or both) of them at Carnegie Mellon instead, as within data science as a whole Python is far more widely used than Java.

Take one of the following courses:

15-110	Principles of Computing	10
15-112	Fundamentals of Programming and Computer Science	12
02-120	Undergraduate Programming for Scientists	12

Complete the following course:

36-350	Statistical Computing	9
--------	-----------------------	---

5. Special Topics 9 units

The Department of Statistics & 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.) or 36-47x (36-470, 36-471, 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. **Please note that all Special Topics are not offered every semester, and new Special Topics are regularly added.**

To satisfy the Special Topics requirement complete one of the following:

36-460	Special Topics: Sports Analytics	9
36-461	Special Topics: Statistical Methods in Epidemiology	9
36-462	Special Topics: Statistical Machine Learning	9
36-463	Special Topics: Multilevel and Hierarchical Models	9
36-464	Special Topics: Psychometrics: A Statistical Modeling Approach	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
36-470	Special Topics: Statistical Methods in Health Sciences	9

36-471	Special Topics: Time Series	9
36-472	Special Topics: Computational Statistical Methods in Life Sciences	9
36-473	Special Topics: Statistical Principles of Generative AI	9

NOTE: 36-46x and 36-47x courses require 36-401 as a prerequisite or corequisite.

6. Modern Regression and Advanced Methodology 18 UNITS

Central to the practice of statistics is the implementation and interpretation of statistical models. The purpose of statistical models is to represent data-generating processes, such that predictions and inferential conclusions can be made about real-world phenomena. Proper modeling involves not just coding, but also thinking critically about data, research goals, and the validity of the models themselves, given their intrinsic assumptions. The courses 36-401 and 36-402 focus on the theory of statistical models (especially linear models and their extensions), how they are applied in real data analyses, and how to interpret and present these analyses in written reports.

To satisfy these requirements, complete the following:

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

NOTE: In order to meet the prerequisite requirements, a grade of at least a C is required in 36-401.

7. 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 & Data Science 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 your Statistics Undergraduate Director. While these courses are not in Statistics & Data Science, the concentration area must complement the overall 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 careers in industry could look for appropriate business courses.

The concentration area can be fulfilled with a minor or additional major, but not all minors and additional majors fulfill this requirement. Due to other major options we already offer, we will not consider concentrations related to Economics, Machine Learning, Mathematics, or Neuroscience.

Concentration approval process: Please make sure to consult your Statistics & Data Science undergraduate advisor *prior* to pursuing courses for the concentration area. Students will submit a form provided by their advisor to have their concentration reviewed. Any changes or amendments to the concentration must be approved by the advisor. Once the concentration area is approved, any changes made to the previously agreed upon coursework require re-approval by an advisor.

* These courses can be amended later but must be re-approved by your 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.

Total number of units for the major	175-193* Units
Total number of units for the degree	360 Units

* This number can vary depending on the courses chosen for the concentration area that a student takes. Speak with an academic advisor for more details.

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 by the end of their freshman year.

The linear algebra requirement is a prerequisite for the course 36-401. It is therefore essential that students complete this requirement by their junior years at the latest.

Recommendations for Prospective Ph.D. Students

Students interested in pursuing a Ph.D. in Statistics or Biostatistics (or related programs) after completing their undergraduate degree are strongly recommended to pursue the **B.S. in Statistics and Data Science (Mathematical Sciences Track)** or to take additional Mathematics courses. 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, 21-341 Linear Algebra, 21-355 Principles of Real Analysis I, and 21-356 Principles of Real Analysis II. We also recommend that students interested in pursuing a Ph.D. gain some research experience during their undergraduate degree, as discussed further in the Research section below. Internships that involve meaningful real data analysis are also looked upon favorably in PhD programs.

Additional Major in Statistics and Data Science

Students who elect the B.S. in Statistics and Data Science as an additional major must fulfill all degree requirements except for the Concentration Area requirement. Majors in many other programs would naturally complement a statistics and data science 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 the B.S. in Statistics and Data Science.

Substitutions and Waivers

Many departments require Statistics & Data Science 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 & Data Science. In many of these cases, the student will need to take additional courses to satisfy major requirements. Students should discuss this with a Statistics advisor when deciding whether to add an additional major in Statistics.

Research

The Statistics & Data Science program encourages students to gain research experience. Opportunities within the department include Summer Undergraduate Research Apprenticeships (SURA), run in association with the university's Office of Undergraduate Research and Scholar Development, and the departmental capstone courses 36-490 Undergraduate Research, 36-493 Sports Analytics Capstone or 36-497 Corporate Capstone Project. (Note that these courses require an application.) Additionally, students can pursue independent study. For those students who maintain a quality point average of 3.25 overall or above, there is also the Dietrich College Senior Honors Program (<http://coursecatalog.web.cmu.edu/schools-colleges/dietrichcollegeofhumanitiesandsocialsciences/#collegeservicesandprograms>).

The faculty in the Statistics & Data Science department largely work within the domains of statistical theory and methodological development, areas that require advanced mathematical training. Thus we encourage students to search broadly for research opportunities: faculty, post-doctoral researchers, and graduate students in many departments throughout the university have data to analyze and would welcome the help of undergraduate statistics students.

Sample Programs

The following sample programs illustrate two ways (of many) to satisfy the requirements for the B.S. in Statistics and Data Science. 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 second schedule is an example of the case when a student enters the program through 36-235 and 36-236.

Schedule 1

First-Year		Second-Year	
Fall	Spring	Fall	Spring
36-200 Reasoning with Data	36-202 Methods for Statistics & Data Science	36-235 Probability and Statistical Inference I	36-236 Probability and Statistical Inference II
21-111 Differential Calculus	21-112 Integral Calculus	21-256 Multivariate Analysis	36-350 Statistical Computing
----	One of the following two courses:	----	21-240 Matrix Algebra with Applications
----	15-110 Principles of Computing	----	----
----	15-112 Fundamentals of Programming and Computer Science	----	----

Third-Year		Fourth-Year	
Fall	Spring	Fall	Spring
36-401 Modern Regression	36-402 Advanced Methods for Data Analysis	Course toward concentration	Course toward concentration
36-3xx or 36-4xx Advanced Data Analysis Elective	36-46x Special Topics course	----	----
Course toward concentration	Course toward concentration	----	----
----	----	----	----

Schedule 2

First-Year		Second-Year	
Fall	Spring	Fall	Spring
21-090 Precalculus	21-120 Differential and Integral Calculus	36-235 Probability and Statistical Inference I	36-236 Probability and Statistical Inference II
36-200 Reasoning with Data	One of the following two courses:	21-256 Multivariate Analysis	21-240 Matrix Algebra with Applications
----	15-110 Principles of Computing	----	36-350 Statistical Computing
----	15-112 Fundamentals of Programming and Computer Science	----	----
----	----	----	----

Third-Year		Fourth-Year	
Fall	Spring	Fall	Spring
36-401 Modern Regression	36-402 Advanced Methods for Data Analysis	36-46x Special Topics	Course toward concentration
36-3xx or 36-4xx Advanced Data Analysis Elective	Course toward concentration	Course toward concentration	36-3xx or 36-4xx Advanced Data Analysis Elective
Course toward concentration	----	----	----
----	----	----	----

B.S. in Statistics and Data Science (Mathematical Sciences Track)

Peter Freeman, *Undergraduate Program Director*
 Location: Baker Hall 229
pfreeman@andrew.cmu.edu

Zach Branson, *Assistant Director of the Undergraduate Program*
 Location: Baker Hall 232
 zbranson@andrew.cmu.edu

Amanda Mitchell, *Lead Senior Academic Advisor*
 Glenn Clune, *Academic Program Manager*
 Sylvie Aubin, *Academic Program Manager*
 Peter Long, *Academic Advisor*
 Location: Baker Hall 129
 statadvising@andrew.cmu.edu (statadvising@stat.cmu.edu)

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

Curriculum

1. Mathematical Foundations (Prerequisites) 49-62 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 & Data Science.

Complete the following:

21-090	Precalculus	10
--------	-------------	----

Complete one of the following course options:

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

OR

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

Complete the following course:

21-122	Integration and Approximation	10
--------	-------------------------------	----

And one of the following three courses:

21-256	Multivariate Analysis	9
21-259	Calculus in Three Dimensions	10
21-268	Multidimensional Calculus	11

And one of the following three courses:

21-240	Matrix Algebra with Applications	10
21-241	Matrices and Linear Transformations	11
21-242	Matrix Theory	11

NOTE: Passing the Mathematical Sciences assessment tests available during First-Year Orientation is an acceptable alternative to completing 21-090 and/or 21-120.

- 21-122 is a required prerequisite for 21-355 Principles of Real Analysis I, a requirement for the Mathematical Sciences Track major concentration.
- 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 Dietrich College Core Requirement in Statistical Reasoning. This course is therefore required for students in the college. (Note: a score of 5 on the Advanced Placement (AP) Exam in Statistics may be used to waive this requirement). 36-220 emphasizes examples in engineering.

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.

Beginning Data Analysis

Choose *one* of the following courses:

36-200	Reasoning with Data *	9
36-220	Engineering Statistics and Quality Control	9

* A score of 5 on the Advanced Placement (AP) Exam in Statistics may be used to waive this requirement.

NOTE: Students who enter the program with credit for probability and statistical inference should discuss options with an advisor.

Sequence 1

Intermediate Data Analysis

Choose *one* of the following courses:

36-202	Methods for Statistics & Data Science *	9
36-309	Experimental Design for Behavioral & Social Sciences	9
36-290	Introduction to Statistical Research Methodology	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
36-318	Introduction to Causal Inference	9
36-396	Tartan Athletics Analytics	9
36-490	Undergraduate Research	9
36-493	Sports Analytics Capstone	9
36-497	Corporate Capstone Project	9

Students can also take a second Special Topics (36-46x or 36-47x) course to fulfill an advanced data analysis elective requirement (see section #5).

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

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
36-318	Introduction to Causal Inference	9
36-396	Tartan Athletics Analytics	9
36-490	Undergraduate Research	9
36-493	Sports Analytics Capstone	9
36-497	Corporate Capstone Project	9

Students can also take a second Special Topics (36-46x or 36-47x) course to fulfill an advanced data analysis elective requirement (see section #5).

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, complete the following:

Take one of the following courses:

36-235	Probability and Statistical Inference I *	9
21-325	Probability **	9
36-225	Introduction to Probability Theory	9

And one of the following three courses:

36-226	Introduction to Statistical Inference	9
36-236	Probability and Statistical Inference II ***	9
36-326	Mathematical Statistics (Honors)	9

* It is possible to substitute 36-218, 36-219, 36-225, 15-259, or 21-325 for 36-235. 36-235 is the standard (and recommended) introduction to probability, 36-219 is tailored for engineers and computer scientists, 36-218 and 15-259 are more mathematically rigorous classes for Computer Science students and more mathematically advanced (students need prior approval to enroll), and 21-325 is a rigorous probability theory course offered by the Department of Mathematics).

**Students in this major may want to consider taking 21-325 Probability over 36-235 as it is a more theoretical course and may better suit those interested in pursuing a PhD in the future.

***It is possible to substitute 36-226 or 36-326 (honors course) for 36-236. 36-236 is the standard (and recommended) introduction to statistical inference.

NOTE: Students who enter the program with credit for probability and statistical inference should discuss options with an advisor.

Please note that students who complete 36-235 are expected to take 36-236 to complete their theory requirements. Students who choose to take 36-225 will be required to take 36-226 afterward. They will not be eligible to take 36-236.

Comment:

(i) In order to meet the prerequisite requirements, a grade of at least a C is required in 36-235 (or equivalent) and 36-236 (or equivalent).

4. Statistical Computing 19-21 units

Fundamental to the practice of statistics and data science is the ability to effectively code data processing and analysis tasks. Within the domain of statistics, the use of the programming language R is ubiquitous, and thus we expose students to it throughout the curriculum (and in depth in Statistical Computing). Within the larger domain of data science, the use of the programming language Python is also ubiquitous, and thus we require all majors to gain, at a minimum, basic competency in the language by taking either Principles of Computing, or Fundamentals of Programming and Computer Science. We would advise those students who are considering receiving course credit for one of these two courses given their score on the AP Computer Science A exam to actually take one (or both) of them at Carnegie Mellon instead, as within data science as a whole Python is far more widely used than Java.

Take one of the following courses:

15-110	Principles of Computing	10
15-112	Fundamentals of Programming and Computer Science	12
02-120	Undergraduate Programming for Scientists	12

Complete the following course:

36-350	Statistical Computing	9
--------	-----------------------	---

5. Special Topics 9 units

The Department of Statistics & 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.) or 36-47x (36-470, 36-471, 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. **Please note that all Special Topics are not offered every semester, and new Special Topics are regularly added.**

To satisfy the Special Topics requirement complete *one* of the following:

36-460	Special Topics: Sports Analytics	9
36-461	Special Topics: Statistical Methods in Epidemiology	9
36-462	Special Topics: Statistical Machine Learning	9
36-463	Special Topics: Multilevel and Hierarchical Models	9

36-464	Special Topics: Psychometrics: A Statistical Modeling Approach	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
36-470	Special Topics: Statistical Methods in Health Sciences	9
36-471	Special Topics: Time Series	9
36-472	Special Topics: Computational Statistical Methods in Life Sciences	9
36-473	Special Topics: Statistical Principles of Generative AI	9

NOTE: 36-46x and 36-47x courses require 36-401 as a prerequisite or corequisite.

6. Modern Regression and Advanced Methodology 18 UNITS

Central to the practice of statistics is the implementation and interpretation of statistical models. The purpose of statistical models is to represent data-generating processes, such that predictions and inferential conclusions can be made about real-world phenomena. Proper modeling involves not just coding, but also thinking critically about data, research goals, and the validity of the models themselves, given their intrinsic assumptions. The courses 36-401 and 36-402 focus on the theory of statistical models (especially linear models and their extensions), how they are applied in real data analyses, and how to interpret and present these analyses in written reports.

To satisfy these requirements, complete the following:

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

NOTE: In order to meet the prerequisite requirements, a grade of at least a C is required in 36-401.

7. Mathematical Statistics Track 46-52 UNITS

21-127	Concepts of Mathematics *	12
21-355	Principles of Real Analysis I	9
36-410	Introduction to Probability Modeling	9

* Students with little to no previous experience in theoretical mathematics and/or theoretical proofs are encouraged to consider taking 21-108 Introduction to Mathematical Concepts either prior to or with 21-127 for additional support.

NOTE: 21-127 is a prerequisite for 21-355 and must be completed before students can register for the course.

And two of the following:

21-228	Discrete Mathematics	9
21-301	Combinatorics	9
21-344	Numerical Linear Algebra	9
21-356	Principles of Real Analysis II	9
21-373	Algebraic Structures	9
36-700	Probability and Mathematical Statistics	12

Total number of units for the major **187-219 Units***
Total number of units for the degree **360 Units**

* This number can vary depending on the courses chosen for the concentration area that a student takes. Speak with an academic advisor for more details.

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 by the end of their freshman year.

The linear algebra requirement is a prerequisite for the course 36-401. It is therefore essential that students complete this requirement by their junior years at the latest.

Recommendations for Prospective Ph.D. Students

Students interested in pursuing a Ph.D. in Statistics or Biostatistics (or related programs) after completing their undergraduate degree are strongly recommended to pursue the **B.S. in Statistics and Data Science (Mathematical Sciences Track)** or to take additional Mathematics courses. 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, 21-341 Linear Algebra, 21-355 Principles of Real Analysis I, and 21-356 Principles of Real Analysis II. We also recommend that students interested in pursuing a Ph.D. gain some research experience during their undergraduate degree, as discussed further in the Research section below. Internships that involve meaningful real data analysis are also looked upon favorably in PhD programs.

Additional Major in Statistics and Data Science (Mathematical Sciences Track)

Students who elect the B.S. in Statistics and Data Science (Mathematical Sciences Track) as an additional major must fulfill all Statistics and Data Science (Mathematical Sciences Track) degree requirements. With respect to double-counting courses, it is departmental policy that students must have at least six courses [three Statistics courses (36-xxx) and three Mathematical Sciences Track electives] that do not count for their primary major. If students do not have at least six, they typically take additional advanced data analysis and/or math 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 B.S. in Statistics and Data Science (Mathematical Sciences Track).

Substitutions and Waivers

Many departments require Statistics & Data Science 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 & Data Science. In many of these cases, the student will need to take additional courses to satisfy major requirements. Students should discuss this with a Statistics advisor when deciding whether to add an additional major in Statistics.

Research

The Statistics & Data Science program encourages students to gain research experience. Opportunities within the department include Summer Undergraduate Research Apprenticeships (SURA), run in association with the university's Office of Undergraduate Research and Scholar Development, and the departmental capstone courses 36-490 Undergraduate Research or 36-497 Corporate Capstone Project. (Note that these courses require an application.) Additionally, students can pursue independent study. For those students who maintain a quality point average of 3.25 overall or above, there is also the Dietrich College Senior Honors Program (<http://coursecatalog.web.cmu.edu/schools-colleges/dietrichcollegeofhumanitiesandsocialsciences/#collegeservicesandprograms>).

The faculty in the Statistics & Data Science department largely work within the domains of statistical theory and methodological development, areas that require advanced mathematical training. Thus we encourage students to search broadly for research opportunities: faculty, post-doctoral researchers, and graduate students in many departments throughout

the university have data to analyze and would welcome the help of undergraduate statistics students.

Sample Programs

The following sample programs illustrate two ways (of many) to satisfy the requirements for the B.S. in Statistics and Data Science (Mathematical Sciences Track). 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 second schedule is an example of the case when a student enters the program through 36-235 and 36-236.

SCHEDULE 1

First-Year		Second-Year	
Fall	Spring	Fall	Spring
36-200 Reasoning with Data	36-202 Methods for Statistics & Data Science	21-122 Integration and Approximation	36-236 Probability and Statistical Inference II
21-111 Differential Calculus	21-256 Multivariate Analysis	21-127 Concepts of Mathematics	36-350 Statistical Computing
-----	21-112 Integral Calculus	36-235 Probability and Statistical Inference I	21-240 Matrix Algebra with Applications
-----	-----	One of the two following courses:	-----
		15-110 Principles of Computing	
		15-112 Fundamentals of Programming and Computer Science	

Third-Year		Fourth-Year	
Fall	Spring	Fall	Spring
36-401 Modern Regression	36-402 Advanced Methods for Data Analysis	36-46x Special Topics	36-410 Introduction to Probability Modeling
Math Track Elective	36-3xx or 36-4xx Advanced Data Analysis Elective	21-355 Principles of Real Analysis I	Math Track Elective
-----	-----	-----	-----
-----	-----	-----	-----

Schedule 2

First-Year		Second-Year	
Fall	Spring	Fall	Spring
21-090 Precalculus	21-120 Differential and Integral Calculus	36-235 Probability and Statistical Inference I	36-236 Probability and Statistical Inference II
36-200 Reasoning with Data	21-241 Matrices and Linear Transformations	21-256 Multivariate Analysis	21-127 Concepts of Mathematics
Take one of the following courses:		21-122 Integration and Approximation	-----
15-110 Principles of Computing		-----	
15-112 Fundamentals of Programming and Computer Science		-----	

Third-Year		Fourth-Year	
Fall	Spring	Fall	Spring
36-350 Statistical Computing	36-402 Advanced Methods for Data Analysis	Math Track Elective	36-410 Introduction to Probability Modeling
36-401 Modern Regression		36-3xx or 36-4xx Advanced Data Analysis Elective	Math Track Elective
21-355 Principles of Real Analysis I	36-3xx or 36-4xx Advanced Data Analysis Elective	-----	-----
-----	-----	-----	-----
-----	-----	-----	-----

B.S. in Statistics and Data Science (Neuroscience Track)

Peter Freeman, *Undergraduate Program Director*
 Location: Baker Hall 229

pfreeman@andrew.cmu.edu

Zach Branson, *Assistant Director of the Undergraduate Program*
Location: Baker Hall 232
zbranson@andrew.cmu.edu

Amanda Mitchell, *Lead Senior Academic Advisor*
Glenn Clune, *Academic Program Manager*
Sylvie Aubin, *Academic Program Manager*
Peter Long, *Academic Advisor*
Location: Baker Hall 129
statadvising@andrew.cmu.edu (statadvising@stat.cmu.edu)

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

Curriculum

1. Mathematical Foundations (Prerequisites) 39-52 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 & Data Science.

Complete the following:

21-090	Precalculus	10
Complete one of the following options:		
21-111	Differential Calculus	10
21-112	Integral Calculus	10
OR		
21-120	Differential and Integral Calculus	10
And one of the following three courses:		
21-256	Multivariate Analysis	9
21-259	Calculus in Three Dimensions	10
21-268	Multidimensional Calculus	11
Add one of the following three courses:		
21-240	Matrix Algebra with Applications	10
21-241	Matrices and Linear Transformations	11
21-242	Matrix Theory	11

NOTE: Passing the Mathematical Sciences assessment tests available during First-Year Orientation is an acceptable alternative to completing 21-090 and/or 21-120.

- 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 Dietrich College Core Requirement in Statistical Reasoning. This course is therefore required for students in the college. (Note: a score of 5 on the Advanced Placement (AP) Exam in Statistics may be used to waive this requirement). 36-220 emphasizes examples in engineering and architecture.

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.

Beginning Data Analysis

Choose *one* of the following courses:

36-200	Reasoning with Data *	9
36-220	Engineering Statistics and Quality Control	9

* A score of 5 on the Advanced Placement (AP) Exam in Statistics may be used to waive this requirement.

NOTE: Students who enter the program with credit for probability and statistical inference should discuss options with an advisor.

Sequence 1

Intermediate Data Analysis

Choose *one* of the following courses:

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

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

Advanced Data Analysis Electives

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
36-318	Introduction to Causal Inference	9
36-396	Tartan Athletics Analytics	9
36-490	Undergraduate Research	9
36-493	Sports Analytics Capstone	9
36-497	Corporate Capstone Project	9

Students can also take a second Special Topics (36-46x or 36-47x) course to fulfill an advanced data analysis elective requirement (see section #5).

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

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
36-318	Introduction to Causal Inference	9
36-396	Tartan Athletics Analytics	9
36-490	Undergraduate Research	9
36-493	Sports Analytics Capstone	9
36-497	Corporate Capstone Project	9

Students can also take a second Special Topics (36-46x or 36-47x) course to fulfill an advanced data analysis elective requirement (see section #5).

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, complete the following:

Take one of the following courses:

36-235	Probability and Statistical Inference I	9
36-225	Introduction to Probability Theory	9
and one of the following three courses:		
36-226	Introduction to Statistical Inference	9
36-236	Probability and Statistical Inference II **	9
36-326	Mathematical Statistics (Honors)	9

* It is possible to substitute 36-218, 36-219, 36-225, 15-259, or 21-325 for 36-235. 36-235 is the standard (and recommended) introduction to probability, 36-219 is tailored for engineers and computer scientists, 36-218 and 15-259 are more mathematically rigorous classes for Computer Science students and more mathematically advanced (students need advisor approval to enroll), and 21-325 is a rigorous probability theory course offered by the Department of Mathematics.

** It is possible to substitute 36-226 or 36-326 (honors course) in place of 36-236. 36-236 is the standard (and recommended) introduction to statistical inference.

NOTE: Students who enter the program with credit for probability and statistical inference should discuss options with an advisor.

Please note that students who complete 36-235 are expected to take 36-236 to complete their theory requirements. Students who choose to take 36-225 instead will be required to take 36-226 afterward. They will not be eligible to take 36-236.

Comment:

(i) In order to meet the prerequisite requirements, a grade of at least a C is required in 36-235 (or equivalent) and 36-236 (or equivalent).

4. Statistical Computing 19-21 units

Fundamental to the practice of statistics and data science is the ability to effectively code data processing and analysis tasks. Within the domain of statistics, the use of the programming language R is ubiquitous, and thus we expose students to it throughout the curriculum (and in depth in Statistical Computing). Within the larger domain of data science, the use of the programming language Python is also ubiquitous, and thus we require all majors to gain, at a minimum, basic competency in the language by taking either Principles of Computing, or Fundamentals of Programming and Computer Science. We would advise those students who are considering receiving course credit for one of these two courses given their score on the AP Computer Science A exam to actually take one (or both) of them at Carnegie Mellon instead, as within data science as a whole Python is far more widely used than Java.

Take one of the following courses:

15-110	Principles of Computing	10
15-112	Fundamentals of Programming and Computer Science	12
02-120	Undergraduate Programming for Scientists	12
Complete the following course:		
36-350	Statistical Computing	9

5. Special Topics 9 units

The Department of Statistics & 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.) or 36-47x (36-470, 36-471, 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. **Please note that all Special Topics are not offered every semester, and new Special Topics are regularly added.**

To satisfy the Special Topics requirement complete one of the following:

36-460	Special Topics: Sports Analytics	9
36-461	Special Topics: Statistical Methods in Epidemiology	9
36-462	Special Topics: Statistical Machine Learning	9
36-463	Special Topics: Multilevel and Hierarchical Models	9
36-464	Special Topics: Psychometrics: A Statistical Modeling Approach	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
36-470	Special Topics: Statistical Methods in Health Sciences	9
36-471	Special Topics: Time Series	9
36-472	Special Topics: Computational Statistical Methods in Life Sciences	9
36-473	Special Topics: Statistical Principles of Generative AI	9

6. Modern Regression and Advanced Methodology 18 UNITS

Central to the practice of statistics is the implementation and interpretation of statistical models. The purpose of statistical models is to represent data-generating processes, such that predictions and inferential conclusions can be made about real-world phenomena. Proper modeling involves not just coding, but also thinking critically about data, research goals, and the validity of the models themselves, given their intrinsic assumptions. The courses 36-401 and 36-402 focus on the theory of statistical models (especially linear models and their extensions), how they are applied in real data analyses, and how to interpret and present these analyses in written reports.

To satisfy these requirements, complete the following:

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

NOTE: In order to meet the prerequisite requirements, a grade of at least a C is required in 36-401.

7. Statistics and Neuroscience Track 45-54 UNITS

85-110	Cognitive Psychology	9
85-170	Foundations of Brain and 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

10-301	Introduction to Machine Learning	12
18-290	Signals and Systems	12
42-630	Introduction to Neural Engineering	12
42-632	Neural Signal Processing	12
36-700	Probability and Mathematical Statistics	12
42/86-631	Neural Data Analysis	12
85-310	Research Methods in Cognitive Psychology	9
85-370	Cognitive Neuroscience Research Methods	9

Neuroscience Background

03-362	Cellular Neuroscience	9
03-363	Systems Neuroscience	9
15-386	Neural Computation	9
85-408	Visual Cognition	9
85-413	Perception	9
85-419	Introduction to Parallel Distributed Processing	9
85-472	Cognitive Neuropsychology	9

Total Number of Units for the Major: 175-211* Units

Total Number of Units for the Degree: 360 Units

* This number can vary depending on the courses chosen for the concentration area that a student takes. Speak with an academic advisor for more details.

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 by the end of their freshman year.

The linear algebra requirement is a prerequisite for the course 36-401. It is therefore essential that students complete this requirement by their junior years at the latest.

Recommendations for Prospective Ph.D. Students

Students interested in pursuing a Ph.D. in Statistics or Biostatistics (or related programs) after completing their undergraduate degree are strongly recommended to pursue the **B.S. in Statistics (Mathematical Sciences Track)** or to take additional Mathematics courses. 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, 21-341 Linear Algebra, 21-355 Principles of Real Analysis I, and 21-356 Principles of Real Analysis II. We also recommend that students interested in pursuing a Ph.D. gain some research experience during their undergraduate degree, as discussed further in the Research section below. Internships that involve meaningful real data analysis are also looked upon favorably in PhD programs.

Additional Major in Statistics and Data Science (Neuroscience Track)

Students who elect the B.S. in Statistics and Data Science (Neuroscience Track) as an additional major must fulfill all Statistics and Data Science (Neuroscience Track) degree requirements. With respect to double-counting courses, it is departmental policy that students must have at least six courses [three Statistics courses (36-xxx) and three Neuroscience Track electives] that do not count for their primary major. If students do not have at least six, they typically take additional advanced data analysis and/or neuroscience 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 the B.S. in Statistics and Data Science (Neuroscience Track).

Substitutions and Waivers

Many departments require Statistics & Data Science 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 & Data Science. In many of these cases, the student will need to take additional courses to satisfy major requirements. Students should discuss this with a Statistics advisor when deciding whether to add an additional major in Statistics.

Research

The Statistics & Data Science program encourages students to gain research experience. Opportunities within the department include Summer Undergraduate Research Apprenticeships (SURA), run in association with the university's Office of Undergraduate Research and Scholar Development, and the departmental capstone courses 36-490 Undergraduate Research, 36-493 Sports Analytics Capstone, or 36-497 Corporate Capstone Project. (Note that these courses require an application.) Additionally, students can pursue independent study. For those students who maintain a quality point average of 3.25 overall or above, there is also the Dietrich College Senior Honors Program (<http://coursecatalog.web.cmu.edu/schools-colleges/dietrichcollegeofhumanitiesandsocialsciences/#collegeservicesandprograms>).

The faculty in the Statistics & Data Science department largely work within the domains of statistical theory and methodological development, areas that require advanced mathematical training. Thus we encourage students to search broadly for research opportunities: faculty, post-doctoral researchers, and graduate students in many departments throughout the university have data to analyze and would welcome the help of undergraduate statistics students.

Sample Programs

The following sample programs illustrate two ways (of many) to satisfy the requirements for the B.S. in Statistics and Data Science (Neuroscience

Track). 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 second schedule is an example of the case when a student enters the program through 36-235 and 36-236.

schedule 1

First-Year		Second-Year	
Fall	Spring	Fall	Spring
21-111 Differential Calculus	21-112 Integral Calculus	21-256 Multivariate Analysis	36-236 Probability and Statistical Inference II
36-200 Reasoning with Data	36-202 Methods for Statistics & Data Science	36-235 Probability and Statistical Inference I	36-350 Statistical Computing
85-110 Cognitive Psychology	And one of the following two courses:	85-170 Foundations of Brain and Behavior	21-240 Matrix Algebra with Applications
-----	15-110 Principles of Computing	-----	-----
-----	15-112 Fundamentals of Programming and Computer Science	-----	-----

Third-Year		Fourth-Year	
Fall	Spring	Fall	Spring
36-401 Modern Regression	36-402 Advanced Methods for Data Analysis	36-46x or 47x Special Topics	36-3xx or 36-4xx Advanced Data Analysis Elective
Neuroscience Track Elective	Neuroscience Track Elective	Neuroscience Track Elective	-----
-----	-----	-----	-----
-----	-----	-----	-----

Schedule 2

First-Year		Second-Year	
Fall	Spring	Fall	Spring
21-090 Precalculus	21-120 Differential and Integral Calculus	21-256 Multivariate Analysis	21-240 Matrix Algebra with Applications
36-200 Reasoning with Data	85-110 Cognitive Psychology	85-170 Foundations of Brain and Behavior	36-3xx or 36-4xx Advanced Data Analysis Elective
-----	Take one of the following two courses:	-----	-----
-----	15-110 Principles of Computing	-----	-----
-----	15-112 Fundamentals of Programming and Computer Science	-----	-----

Third-Year		Fourth-Year	
Fall	Spring	Fall	Spring
36-235 Probability and Statistical Inference I	36-236 Probability and Statistical Inference II	36-401 Modern Regression	36-402 Advanced Methods for Data Analysis
-----	36-350 Statistical Computing	Neuroscience Track Elective	36-46x or 47x - Special Topics
-----	Neuroscience Track Elective	36-3xx or 36-4xx Advanced Data Analysis Elective	Neuroscience Track Elective
-----	-----	-----	-----

B.S. in Economics and Statistics

Peter Freeman, *Undergraduate Program Director*

Location: Baker Hall 229
pfreeman@andrew.cmu.edu

Zach Branson, *Assistant Director of the Undergraduate Program*

Location: Baker Hall 232
zbranson@andrew.cmu.edu

Amanda Mitchell, *Lead Senior Academic Advisor*

Sylvie Aubin, *Academic Program Manager*

Location: Baker Hall 129
statadvising@andrew.cmu.edu (statadvising@stat.cmu.edu)

The Major in Economics and Statistics provides an interdisciplinary course of study aimed at students with a strong interest in the empirical analysis of economic data. With a joint curriculum from the Department of Statistics and Data Science and the Undergraduate Economics Program, the major provides students with a solid foundation in the theories and

methods of both fields. Students in this major are trained to advance the understanding of economic issues through the analysis, synthesis and reporting of data using the advanced empirical research methods of statistics and econometrics. Graduates are well positioned for admission to competitive graduate programs, including those in statistics, economics and management, as well as for employment in positions requiring strong analytical and conceptual skills - especially those in economics, finance, education, and public policy.

Curriculum

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

1. MATHEMATICAL FOUNDATIONS (PREREQUISITES) 39-52 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 Economics and Statistics.

Complete the following:

21-090	Precalculus	10
Complete one of the following options:		
21-111	Differential Calculus	10
21-112	Integral Calculus	10
OR		
21-120	Differential and Integral Calculus	10
And one of the following three courses:		
21-256	Multivariate Analysis	9
21-259	Calculus in Three Dimensions	10
21-268	Multidimensional Calculus	11
And one of the following three courses:		
21-240	Matrix Algebra with Applications	10
21-241	Matrices and Linear Transformations	11
21-242	Matrix Theory	11

NOTES Passing the Mathematical Sciences assessment tests available during First-Year Orientation is an acceptable alternative to completing 21-090 and/or 21-120.

- 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. ECONOMICS FOUNDATIONS 18 UNITS

Take one of the following courses:

73-102	Principles of Microeconomics *	9
73-104	Principles of Microeconomics Accelerated **	9
Take the following course:		
73-103	Principles of Macroeconomics	9

* Students who place out of 73-102 based on the economics placement exam will receive a pre-req waiver for 73-102 and are waived from taking 73-102.

** This course requires students to complete a 4 or 5 on the AP Microeconomics exam or qualifying score on the IB/Cambridge Exams. 73-104 will substitute for any 73-102 prerequisite requirement in other courses. 73-104 is a more rigorous introduction to microeconomics, is taught at a faster pace than 73-102, and dives a bit deeper into key topics. It is designed for students who have prior knowledge to fundamental economic concepts through AP/IB/Cambridge coursework. Enrollment in 73-104 requires special permission. Students who wish to take this course should add themselves to the 73-104 waitlist once registration opens. The Tepper School will verify the advancement placement scores and will enroll students in 73-104.

3. STATISTICAL FOUNDATIONS 27 UNITS

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 Dietrich College Core Requirement in Statistical Reasoning. This course is therefore required for students in the college. (**Note:** a score of 5 on the Advanced Placement (AP) Exam in Statistics may be used to waive this requirement). 36-220 emphasizes examples in engineering.

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.

Beginning Data Analysis

Choose *one* of the following courses:

36-200	Reasoning with Data *	9
36-220	Engineering Statistics and Quality Control	9

* A score of 5 on the Advanced Placement (AP) Exam in Statistics may be used to waive this requirement.

NOTES Students who enter the program with credit for probability and statistical inference should discuss options with an advisor.

Advanced Data Analysis 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
36-318	Introduction to Causal Inference	9
36-460	Special Topics: Sports Analytics	9
36-461	Special Topics: Statistical Methods in Epidemiology	9
36-462	Special Topics: Statistical Machine Learning	9
36-463	Special Topics: Multilevel and Hierarchical Models	9
36-464	Special Topics: Psychometrics: A Statistical Modeling Approach	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
36-470	Special Topics: Statistical Methods in Health Sciences	9
36-471	Special Topics: Time Series	9
36-490	Undergraduate Research	9
36-493	Sports Analytics Capstone	9
36-497	Corporate Capstone Project	9

NOTES Special Topics are not offered every semester, and new Special Topics are regularly added.

4. ECONOMICS CORE 27 UNITS

73-230	Intermediate Microeconomics	9
73-240	Intermediate Macroeconomics	9
70-340	Business Communications	9

5. ECONOMICS QUANTITATIVE ANALYSIS REQUIREMENTS 27 UNITS

Complete the following:

73-265	Economics and Data Science	9
73-274	Econometrics I	9

Take one of the following courses:

73-374	Econometrics II	9
--------	-----------------	---

73-423	Forecasting for Economics and Business	9
70-467	Machine Learning for Business Analytics	9

6. 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, complete the following:

Take one of the following courses:

36-235	Probability and Statistical Inference I *	9
36-225	Introduction to Probability Theory	9

Take one of the following courses:

36-236	Probability and Statistical Inference II **	9
36-226	Introduction to Statistical Inference	9
36-326	Mathematical Statistics (Honors)	9

* It is possible to substitute 36-218, 36-219, 36-225, 15-259, or 21-325 for 36-235. 36-235 is the standard introduction to probability, 36-219 is tailored for engineers and computer scientists, 36-218 and 15-259 are more mathematically rigorous classes 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.

** It is possible to substitute 36-226 or 36-326 for 36-236. 36-236 is the standard introduction to statistical inference.

Please note that students who complete 36-235 are expected to take 36-236 to fulfill their theory requirements. Students who choose to take 36-225 instead will be required to take 36-226 afterward, they will not be eligible to take 36-236.

Comment:

(i) In order to meet the prerequisite requirements for the major, a grade of C or better is required in 36-235 (or equivalents), 36-236 or 36-326 and 36-401.

7. MODERN REGRESSION AND ADVANCED METHODOLOGY 18 UNITS

Central to the practice of statistics is the implementation and interpretation of statistical models. The purpose of statistical models is to represent data-generating processes, such that predictions and inferential conclusions can be made about real-world phenomena. Proper modeling involves not just coding, but also thinking critically about data, research goals, and the validity of the models themselves, given their intrinsic assumptions. The courses 36-401 and 36-402 focus on the theory of statistical models (especially linear models and their extensions), how they are applied in real data analyses, and how to interpret and present these analyses in written reports.

Complete the following courses:

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

NOTE: In order to meet the prerequisite requirements, a grade of at least a C is required in 36-401.

8. STATISTICAL COMPUTING 19-21 UNITS

Fundamental to the practice of statistics and data science is the ability to effectively code data processing and analysis tasks. Within the domain of statistics, the use of the programming language R is ubiquitous, and thus we expose students to it throughout the curriculum (and in depth in Statistical Computing). Within the larger domain of data science, the use of the programming language Python is also ubiquitous, and thus we require all majors to gain, at a minimum, basic competency in the language by taking either Principles of Computing, or Fundamentals of Programming and Computer Science. We would advise those students who are considering receiving course credit for one of these two courses given their score on the AP Computer Science A exam to actually take one (or both) of them at Carnegie Mellon instead, as within data science as a whole Python is far more widely used than Java.

Take one of the following two courses:

15-110	Principles of Computing	10
15-112	Fundamentals of Programming and Computer Science	12
02-120	Undergraduate Programming for Scientists	12

Complete the following course:

36-350	Statistical Computing	9
--------	-----------------------	---

9. ECONOMICS ELECTIVES 18 units

Students must take two advanced Economics elective courses, numbered 73-3xx and higher (excluding 73-497 Senior Project and other colloquium and related courses).

73-374 Econometrics II or 73-423 Forecasting for Economics and Business can count toward the Advanced Economics Elective requirement if they are not being used to fulfill the Advanced Quantitative Analysis requirement.

Total number of units for the major	202-218 Units
Total number of units for the degree	360 Units

Recommendations for Prospective Ph.D. Students

Students interested in pursuing a Ph.D. in Statistics or Biostatistics (or related programs) after completing their undergraduate degree are strongly recommended to pursue the **B.S. in Statistics (Mathematical Sciences Track)** or to take additional Mathematics courses. 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, 21-341 Linear Algebra, 21-355 Principles of Real Analysis I, and 21-356 Principles of Real Analysis II. Students interested in pursuing a PhD in Economics are strongly encouraged to conduct research in Economics. We also recommend that students interested in pursuing a Ph.D. gain some research experience during their undergraduate degree, as discussed further in the Research section below. Internships that involve meaningful real data analysis are also looked upon favorably in PhD programs.

Additional Major in Economics and Statistics

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

With respect to double-counting courses, it is departmental policy that students must have at least six courses [three Economics (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.

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. 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.

Note on additional major and minor pairings

All students pursuing the B.S. in Economics and Statistics as a primary major are prohibited from also pursuing a minor or additional major in

Economics due to the significant overlap in their respective curricula. Please see your academic advisor if you have further questions.

Research

The Statistics & Data Science program encourages students to gain research experience. Opportunities within the department include Summer Undergraduate Research Apprenticeships (SURA), run in association with the university's Office of Undergraduate Research and Scholar Development, and the departmental capstone courses 36-490 Undergraduate Research, 36-493 Sports Analytics Capstone, or 36-497 Corporate Capstone Project. (Note that these courses require an application.) Additionally, students can pursue independent study. For those students who maintain a quality point average of 3.25 overall or above, there is also the Dietrich College Senior Honors Program (<http://coursecatalog.web.cmu.edu/schools-colleges/dietrichcollegeofhumanitiesandsocialsciences/#collegeservicesandprograms>).

The faculty in the Statistics & Data Science department largely work within the domains of statistical theory and methodological development, areas that require advanced mathematical training. Thus we encourage students to search broadly for research opportunities: faculty, post-doctoral researchers, and graduate students in many departments throughout the university have data to analyze and would welcome the help of undergraduate statistics students.

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).

First-Year		Second-Year	
Fall	Spring	Fall	Spring
21-090 Precalculus	21-120 Differential and Integral Calculus	21-256 Multivariate Analysis	36-236 Probability and Statistical Inference II
36-200 Reasoning with Data	73-103 Principles of Macroeconomics	36-235 Probability and Statistical Inference I	21-240 Matrix Algebra with Applications
73-102 Principles of Microeconomics	70-340 Business Communications	73-230 Intermediate Microeconomics	73-240 Intermediate Macroeconomics
First-Year Writing	15-110 Principles of Computing	73-265 Economics and Data Science	73-274 Econometrics I
	-----	-----	-----
	-----	-----	-----

Third-Year		Fourth-Year	
Fall	Spring	Fall	Spring
36-350 Statistical Computing	36-402 Advanced Methods for Data Analysis	36-3xx or 36-4xx Advanced Data Analysis Elective	36-3xx or 36-4xx Advanced Data Analysis Elective
36-401 Modern Regression	-----	Economics Elective	Economics Elective
Advanced Quantitative Analysis Course	-----	-----	-----
-----	-----	-----	-----
-----	-----	-----	-----

NOTE: 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

Peter Freeman, *Undergraduate Program Director*

Location: Baker Hall 229
pfreeman@andrew.cmu.edu

Zach Branson, *Assistant Director of the Undergraduate Program*
Location: Baker Hall 232
zbranson@andrew.cmu.edu

Amanda Mitchell, *Lead Senior Academic Advisor*
Glenn Clune, *Academic Program Manager*

Sylvie Aubin, *Academic Program Manager*

Peter Long, *Academic Advisor*
Location: Baker Hall 129
statadvising@andrew.cmu.edu (statadvising@stat.cmu.edu)

Students in the Bachelor of Science in 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 B.S. in Statistics and Machine Learning are detailed below and are organized by categories.

Curriculum

1. Mathematical Foundations (Prerequisites) 51-64 units

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

Complete the following:

21-090	Precalculus	10
Complete one of the following options:		
21-111	Differential Calculus	10
21-112	Integral Calculus	10

OR

21-120	Differential and Integral Calculus	10
And one of the following three courses:		
21-256	Multivariate Analysis	9
21-259	Calculus in Three Dimensions	10
21-268	Multidimensional Calculus	11

And one of the following three courses:

21-240	Matrix Algebra with Applications	10
21-241	Matrices and Linear Transformations	11
21-242	Matrix Theory	11

Mathematical Theory

21-127	Concepts of Mathematics *	12
--------	---------------------------	----

Notes:

- Passing the Mathematical Sciences assessment tests available during First-Year Orientation is an acceptable alternative to completing 21-090 and/or 21-120.
- 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.

*Students with little to no previous experience in theoretical mathematics and/or theoretical proofs are encouraged to consider taking 21-108 Introduction to Mathematical Concepts either prior to or with 21-127 for additional support.

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. This course is therefore required for students in the college. (Note: a score of 5 on the Advanced Placement [AP] Exam in Statistics may be used to waive this requirement). 36-220 emphasizes examples in engineering and architecture.

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.

Beginning Data Analysis

Choose *one* of the following courses:

36-200	Reasoning with Data *	9
36-220	Engineering Statistics and Quality Control	9

*A score of 5 on the Advanced Placement (AP) Exam in Statistics may be used to waive this requirement. 36-220 emphasizes examples in engineering and Architecture.

Note: Students who enter the program with credit for probability and statistical inference should discuss options with an advisor.

Sequence 1

Intermediate Data Analysis

Choose *one* of the following courses:

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

*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
36-318	Introduction to Causal Inference	9
36-396	Tartan Athletics Analytics	9
36-460	Special Topics: Sports Analytics	9
36-461	Special Topics: Statistical Methods in Epidemiology	9
36-462	Special Topics: Statistical Machine Learning	9
36-463	Special Topics: Multilevel and Hierarchical Models	9
36-464	Special Topics: Psychometrics: A Statistical Modeling Approach	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
36-470	Special Topics: Statistical Methods in Health Sciences	9
36-471	Special Topics: Time Series	9
36-472	Special Topics: Computational Statistical Methods in Life Sciences	9
36-473	Special Topics: Statistical Principles of Generative AI	9
36-490	Undergraduate Research	9
36-493	Sports Analytics Capstone	9
36-497	Corporate Capstone Project	9

All Special Topics (36-46x or 36-47x) courses are not offered every semester. They are on a rotation and new Special Topics are regularly added.

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
36-318	Introduction to Causal Inference	9
36-396	Tartan Athletics Analytics	9
36-460	Special Topics: Sports Analytics	9
36-461	Special Topics: Statistical Methods in Epidemiology	9
36-462	Special Topics: Statistical Machine Learning	9
36-463	Special Topics: Multilevel and Hierarchical Models	9
36-464	Special Topics: Psychometrics: A Statistical Modeling Approach	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
36-470	Special Topics: Statistical Methods in Health Sciences	9
36-471	Special Topics: Time Series	9
36-472	Special Topics: Computational Statistical Methods in Life Sciences	9
36-473	Special Topics: Statistical Principles of Generative AI	9
36-490	Undergraduate Research	9
36-493	Sports Analytics Capstone	9
36-497	Corporate Capstone Project	9

All Special Topics (36-46x or 36-47x) courses are not offered every semester. They are on a rotation and new Special Topics are regularly added.

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, complete the following:

Take one of the following courses:

36-235	Probability and Statistical Inference I *	9
36-225	Introduction to Probability Theory	9

And one of the three following courses:

36-226	Introduction to Statistical Inference	9
36-236	Probability and Statistical Inference II **	9
36-326	Mathematical Statistics (Honors)	9

Note: Students who enter the program with credit for probability and statistical inference should discuss options with an advisor.

*It is possible to substitute 36-218, 36-219, 36-225, 15-259, or 21-325 for 36-235. 36-235 is the standard (and recommended) introduction to probability, 36-219 is tailored for engineers and computer scientists, 36-218 and 15-259 are more mathematically rigorous classes for Computer Science students and more mathematically advanced (students need advisor approval to enroll), and 21-325 is a rigorous probability theory course offered by the Department of Mathematics.)

**It is possible to substitute 36-226 or 36-326 (honors course) for 36-236. 36-236 is the standard (and recommended) introduction to statistical inference.

Please note that students who complete 36-235 are expected to take 36-236 to complete their theory requirements. Students who choose to take 36-225 instead will be required to take 36-226 afterward. They will not be eligible to take 36-236.

Comments:

(i) In order to meet the prerequisite requirements, a grade of at least a C is required in 36-235 (or equivalent) and 36-236 (or equivalent).

4. Statistical Computing 9 units

Fundamental to the practice of statistics and data science is the ability to effectively code data processing and analysis tasks. Within the domain of statistics, the use of the programming language R is ubiquitous, and thus we expose students to it throughout the curriculum (and in depth in Statistical Computing).

36-350	Statistical Computing	9
--------	-----------------------	---

5. Modern Regression and Advanced Methodology 18 UNITS

Central to the practice of statistics is the implementation and interpretation of statistical models. The purpose of statistical models is to represent data-generating processes, such that predictions and inferential conclusions can be made about real-world phenomena. Proper modeling involves not just coding, but also thinking critically about data, research goals, and the validity of the models themselves, given their intrinsic assumptions. The courses 36-401 and 36-402 focus on the theory of statistical models (especially linear models and their extensions), how they are applied in real data analyses, and how to interpret and present these analyses in written reports.

To satisfy these requirements, complete the following:

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

Note: In order to meet the prerequisite requirements, a grade of at least a C is required in 36-401.

6. Machine Learning/Computer Science 57-60 units

Statistical modeling in practice nearly always requires computation in one way or another. Computational algorithms are sometimes treated as “black boxes,” whose innards the statistician need not pay attention to. But this attitude is becoming less and less prevalent, and today there is much to be gained from a strong working knowledge of computational tools. Understanding the strengths and weaknesses of various methods allows the data analyst to select the right tool for the job; understanding how they can be adapted to work in new settings greatly extends the realm of problems that he/she can solve. While all majors in Statistics & Data Science are given solid grounding in computation, extensive computational training is really what sets the B.S. in Statistics and Machine Learning program apart. Note that we would advise those students who are considering receiving course credit for Fundamentals of Programming and Computer Science given their score on the AP Computer Science A exam to actually take the course at Carnegie Mellon instead, as within data science as a whole Python is far more widely used than Java.

15-112	Fundamentals of Programming and Computer Science	12
or 02-120	Undergraduate Programming for Scientists	
15-122	Principles of Imperative Computation	12
15-351	Algorithms and Advanced Data Structures	12
or 15-451	Algorithm Design and Analysis	
10-301	Introduction to Machine Learning	12

and take one of the following Machine Learning Advanced Electives:

05-434	Machine Learning in Practice	12
10-403	Deep Reinforcement Learning & Control	12
10-703	Deep Reinforcement Learning & Control	12
10-405	Machine Learning with Large Datasets (Undergraduate)	12
10-605	Machine Learning with Large Datasets	12
10-417	Intermediate Deep Learning	12
10-418	Machine Learning for Structured Data	12
10-707	Advanced Deep Learning	12
11-344	Machine Learning in Practice	12
11-411	Natural Language Processing	12
11-441	Machine Learning with Graphs	9

11-485	Introduction to Deep Learning	9
11-661	Language and Statistics	12
11-761	Language and Statistics	12
15-386	Neural Computation	9
15-387	Computational Perception	9
16-311	Introduction to Robotics	12
16-385	Computer Vision	12
16-720	Computer Vision	12

*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.

Total number of units for the major	180-205 Units
Total number of units for the degree	360 Units

Recommendations

Students in the Dietrich College of Humanities and Social Sciences who wish to declare a Statistics and Machine Learning major 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 that students complete this requirement by their junior years at the latest.

Recommendations for Prospective Ph.D. Students

Students interested in pursuing a Ph.D. in Statistics or Machine Learning (or related programs) after completing their undergraduate degree are strongly recommended to take additional Mathematics courses. 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, 21-341 Linear Algebra, 21-355 Principles of Real Analysis I, and 21-356 Principles of Real Analysis II. We also recommend that students interested in pursuing a Ph.D. gain some research experience during their undergraduate degree, as discussed further in the Research section below. Internships that involve meaningful real data analysis are also looked upon favorably in PhD programs.

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 an additional 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 the B.S. in Statistics and Machine Learning.

Substitutions and Waivers

Many departments require Statistics & Data Science 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 & Data Science. In many of these cases, the student will need to take additional courses to satisfy major requirements. Students should discuss this with a Statistics advisor when deciding whether to add an additional major in Statistics and Machine Learning.

Research

The Statistics & Data Science program encourages students to gain research experience. Opportunities within the department include Summer Undergraduate Research Apprenticeships (SURA), run in association with the university's Office of Undergraduate Research and Scholar Development, and the departmental capstone courses 36-490 Undergraduate Research, 36-493 Sports Analytics Capstone or 36-497 Corporate Capstone Project. (Note that these courses require an application.) Additionally, students can pursue independent study. For those students who maintain a quality point average of 3.25 overall or above, there is also the Dietrich College Senior Honors Program (<http://coursecatalog.web.cmu.edu/schools-colleges/dietrichcollegeofhumanitiesandsocialsciences/#collegeservicesandprograms>).

The faculty in the Statistics & Data Science department largely work within the domains of statistical theory and methodological development, areas that require advanced mathematical training. Thus we encourage students to search broadly for research opportunities: faculty, post-doctoral researchers, and graduate students in many departments throughout the university have data to analyze and would welcome the help of undergraduate statistics students.

Sample Programs

The following sample programs illustrates two ways (of many) to satisfy the requirements for the B.S. in Statistics and Machine Learning. 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 second schedule is an example of the case when a student enters the program through 36-235 and 36-236.

Schedule 1

First-Year		Second-Year	
Fall	Spring	Fall	Spring
21-111 Differential Calculus	36-202 Methods for Statistics & Data Science	21-256 Multivariate Analysis	36-236 Probability and Statistical Inference II
36-200 Reasoning with Data	21-112 Integral Calculus	36-235 Probability and Statistical Inference I	21-241 Matrices and Linear Transformations
15-112 Fundamentals of Programming and Computer Science	21-127 Concepts of Mathematics	15-122 Principles of Imperative Computation	36-350 Statistical Computing
----	----	----	----

Third-Year		Fourth-Year	
Fall	Spring	Fall	Spring
10-301 Introduction to Machine Learning	36-402 Advanced Methods for Data Analysis	36-3xx or 36-4xx Advanced Data Analysis Elective	36-3xx or 36-4xx Advanced Data Analysis Elective
36-401 Modern Regression	15-351 Algorithms and Advanced Data Structures	Machine Learning Elective	----
----	----	----	----
----	----	----	----

*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

First-Year		Second-Year	
Fall	Spring	Fall	Spring
21-090 Precalculus	21-127 Concepts of Mathematics	21-256 Multivariate Analysis	36-236 Probability and Statistical Inference II
36-200 Reasoning with Data	21-120 Differential and Integral Calculus	36-235 Probability and Statistical Inference I	21-241 Matrices and Linear Transformations
15-112 Fundamentals of Programming and Computer Science	----	15-122 Principles of Imperative Computation	36-3xx or 36-4xx Advanced Data Analysis Elective
----	----	----	----

Third-Year		Fourth-Year	
Fall	Spring	Fall	Spring
36-350 Statistical Computing	36-402 Advanced Methods for Data Analysis	15-351 Algorithms and Advanced Data Structures	Machine Learning Advanced Elective
36-401 Modern Regression	10-301 Introduction to Machine Learning	36-3xx or 36-4xx Advanced Data Analysis Elective	36-3xx or 36-4xx Advanced Data Analysis Elective
36-3xx or 36-4xx Advanced Data Analysis Elective	----	----	----
----	----	----	----

*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.

The Minor in Statistics and Data Science

Peter Freeman, *Undergraduate Program Director*
 Location: Baker Hall 229
 pfreeman@andrew.cmu.edu

Zach Branson, *Assistant Director of the Undergraduate Program*
 Location: Baker Hall 232
 zbranson@andrew.cmu.edu

Amanda Mitchell, *Lead Senior Academic Advisor*
 Location: Baker Hall 129
 statadvising@andrew.cmu.edu (statadvising@stat.cmu.edu)

The Minor in Statistics and Data Science develops skills that complement major study in other disciplines. The program helps the student master the basics of statistical theory and advanced techniques in data analysis. This is a good choice for deepening understanding of statistical ideas and for strengthening research skills.

In order to complete a minor in Statistics and Data Science a student must satisfy all of the following requirements:

1. Mathematical Foundations (Prerequisites) 39-52 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 & Data Science.

Complete the following:

- 21-090 Precalculus 10
- Complete one of the following options:
- 21-111 Differential Calculus 10
- 21-112 Integral Calculus 10
- OR
- 21-120 Differential and Integral Calculus 10
- And one of the following three courses:
- 21-256 Multivariate Analysis 9
- 21-259 Calculus in Three Dimensions 10
- 21-268 Multidimensional Calculus 11
- And one of the following three courses:
- 21-240 Matrix Algebra with Applications 10
- 21-241 Matrices and Linear Transformations 11
- 21-242 Matrix Theory 11

NOTE: Passing the Mathematical Sciences assessment tests available during First-Year Orientation is an acceptable alternative to completing 21-090 and/or 21-120.

- 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 or 36-47X Special Topics.
- 21-241 and 21-242 are intended only for students with a very strong mathematical background.

2. Data Analysis 27-36 UNITS

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

The Beginning Data Analysis courses give a hands-on introduction to the art and science of data analysis. The courses cover similar topics but differ slightly in the examples they emphasize. 36-200 draws examples from many fields and satisfies the Dietrich College Core Requirement in Statistical Reasoning. This course is therefore required for students in the College. (Note: A score of 5 on the Advanced Placement (AP) Exam in Statistics may be used to waive this requirement). 36-220 is another course that can complete the Beginning Data Analysis requirement that emphasizes examples in engineering and architecture.

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 Analysis¹

Choose *one* of the following courses:

36-200	Reasoning with Data *	9
36-220	Engineering Statistics and Quality Control	9

* A score of 5 on the Advanced Placement (AP) Exam in Statistics may be used to waive this requirement.

Intermediate Data Analysis¹

Choose *one* of the following courses:

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

¹ 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.

NOTE: Intermediate Data Analysis requirement must be completed prior to 36-401, if not, an additional Advanced Analysis and Methodology course is required.

Advanced Data Analysis and Methodology

Complete the following:

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
36-460	Special Topics: Sports Analytics	9
36-461	Special Topics: Statistical Methods in Epidemiology	9

36-462	Special Topics: Statistical Machine Learning	9
36-463	Special Topics: Multilevel and Hierarchical Models	9
36-464	Special Topics: Psychometrics: A Statistical Modeling Approach	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
36-470	Special Topics: Statistical Methods in Health Sciences	9
36-471	Special Topics: Time Series	9
36-472	Special Topics: Computational Statistical Methods in Life Sciences	9
36-473	Special Topics: Statistical Principles of Generative AI	9
36-490	Undergraduate Research	9
36-493	Sports Analytics Capstone	9
36-497	Corporate Capstone Project	9

NOTE: Special Topics are not offered every semester, and new Special Topics 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
36-313	Statistics of Inequality and Discrimination	9
36-315	Statistical Graphics and Visualization	9
36-318	Introduction to Causal Inference	9
36-396	Tartan Athletics Analytics	9
36-402	Advanced Methods for Data Analysis	9
36-410	Introduction to Probability Modeling	9
36-460	Special Topics: Sports Analytics	9
36-461	Special Topics: Statistical Methods in Epidemiology	9
36-462	Special Topics: Statistical Machine Learning	9
36-463	Special Topics: Multilevel and Hierarchical Models	9
36-464	Special Topics: Psychometrics: A Statistical Modeling Approach	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
36-470	Special Topics: Statistical Methods in Health Sciences	9
36-471	Special Topics: Time Series	9
36-472	Special Topics: Computational Statistical Methods in Life Sciences	9
36-473	Special Topics: Statistical Principles of Generative AI	9
36-490	Undergraduate Research	9
36-493	Sports Analytics Capstone	9
36-497	Corporate Capstone Project	9

NOTE: Special Topics are not offered every semester, and new Special Topics are regularly added.

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, complete the following:

Take one of the following courses:

36-235	Probability and Statistical Inference I *	9
36-225	Introduction to Probability Theory	9

And one of the following three courses:

36-236	Probability and Statistical Inference II **	9
36-226	Introduction to Statistical Inference	9
36-326	Mathematical Statistics (Honors)	9

* It is possible to substitute 36-218, 36-219, 36-225, 15-259, or 21-325 for 36-235. (36-235 is the standard (and recommended) introduction to probability, 36-219 is tailored for engineers and computer scientists, 36-218 and 15-259 are more mathematically rigorous classes for Computer Science students and more mathematically advanced (students need advisor approval to enroll), and 21-325 is a rigorous Probability Theory course offered by the Department of Mathematics.) 36-326 is not offered every semester/year but can be substituted for 36-226 and is considered an honors course.

** It is possible to substitute 36-226 or 36-326 (honors course) for 36-236. 36-236 is the standard (and recommended) introduction to statistical inference.

Please note that students who complete 36-235 are expected to take 36-236 to fulfill their theory requirements. Students who choose to take 36-225 instead will be required to take 36-226 afterward, they will not be eligible to take 36-236.

Comment:

(i) In order to meet the prerequisite requirements, a grade of at least a C is required in 36-235 (or equivalent) and 36-236 (or equivalent).

Total number of units required for the minor 84-106 UNITS

Double Counting

With respect to double-counting courses, it is departmental policy that students must have at least three statistics courses (36-xxx) that do *not* count for their primary major. If students do not have at least three, they need to take additional advanced electives. Make sure to consult your Statistics and Data Science 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 and Data Science Minor. Keep in mind that the program is flexible and can support many other possible schedules.

The first schedule uses Sequence 1 to satisfy the intermediate data analysis requirement. The second schedule is an example of the case when a student enters the Minor through 36-235 and 36-236 (and therefore skips the beginning data analysis course).

Schedule 1

First-Year		Second-Year	
Fall	Spring	Fall	Spring
21-111 Differential Calculus	21-112 Integral Calculus	21-256 Multivariate Analysis	21-240 Matrix Algebra with Applications
36-200 Reasoning with Data	36-202 Methods for Statistics & Data Science	36-235 Probability and Statistical Inference I	36-236 Probability and Statistical Inference II

Third-Year

Fall	Spring
36-401 Modern Regression	Advanced Analysis and Methodology course (36-4xx)

Schedule 2

First-Year		Second-Year	
Fall	Spring	Fall	Spring
21-090 Precalculus	21-120 Differential and Integral Calculus	21-256 Multivariate Analysis	21-240 Matrix Algebra with Applications

Third-Year		Fourth-Year	
Fall	Spring	Fall	Spring
36-235 Probability and Statistical Inference I	36-236 Probability and Statistical Inference II	36-401 Modern Regression	Advanced Analysis and Methodology course (36-4xx)
	36-3xx or 36-4xx Advanced Data Analysis Elective		

Statistics & Data Science Dietrich Senior Honors Thesis

Eligibility

Eligibility is determined by Dietrich College. Students who are eligible will be notified prior to their senior year.

Dietrich College Requirements (<https://www.cmu.edu/dietrich/students/undergraduate/programs/senior-honors/>):

- Students *must have a major in Dietrich College*, either as a primary or an additional major; or be in the BHA (<https://www.cmu.edu/interdisciplinary/programs/bha.html>) program.
- Cumulative QPA through the end of the junior year of at least 3.25 overall, and 3.50 in the Dietrich College major associated with the proposed project.
- Departmental sponsorship in the form of an agreement by a faculty member to serve as advisor for the 2-semester/18 unit Honors project (graduate students may not serve as advisors; adjunct faculty may do so, but only in collaboration with a regular faculty member), and approval by the department head.

Statistics & Data Science Requirements Overview

The below guidelines apply to any Statistics & Data Science students who are doing an honors thesis that has been *approved through the Statistics & Data Science department* (i.e. our department signs off on the thesis paperwork). If you are a Stat & DS student pursuing a Dietrich senior honors thesis through another department (i.e. a different department than Stat & DS is signing off on it) then these guidelines do not apply to you.

In order to be approved for a thesis with the Stat & DS department the project needs to have a significant statistical component. This will be discussed and confirmed during the proposal approval phase of applying.

Honors Thesis Timeline

Senior Year - Fall Semester

The Dietrich College senior honors thesis is a year-long project. As such, after the fall semester of a student's senior year a progress report will be due to Undergraduate Program Director, Peter Freeman (pfreeman@andrew.cmu.edu), for review.

Progress Paper Requirements:

- Minimum length - 5 pages of text (not including graphs/figures/results)
- This paper should build substantially on the proposal, and lay out what work has been done up to this point, as well as an action plan for the spring semester.
- Must be sent to Undergraduate Program Director, Peter Freeman (pfreeman@andrew.cmu.edu), by the last day of classes for the fall semester (typically the first week of December).

Senior Year - Spring Semester

Final Thesis Requirements:

In alignment with a typical advanced data analysis (ADA) project in the field of Statistics the minimum required length of the final thesis must be

a minimum of 15 written pages, no more than 18 single-spaced pages, 12-point font. *This does *not* include figures.*

- Figures can be embedded within the text (so long as the overall text length requirement is met) but can also be provided as appendices after the main body of the text.
- Reports should be written in IMRaD format (Introduction, Methods, Results, and Discussion), where the "Introduction" can be a Background and Significance section followed by a Data section.
- All theses are due to the Undergraduate Program Director, Peter Freeman (pfreeman@andrew.cmu.edu), and Department Head, Rebecca Nugent (rnugent@andrew.cmu.edu), at the end of the 12th week of class in spring semester (roughly the first week of April).

Substitutions and Waivers

Many departments require Statistics & Data Science 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.

However, the Statistics & Data Science department's Director of Undergraduate Studies can 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 & 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 a departmental academic advisor.

Faculty

SIVARAMAN BALAKRISHNAN, Professor - Ph.D., Carnegie Mellon; Carnegie Mellon, 2015-

ELI BEN-MICHAEL, Assistant Professor, Joint With Heinz College - Ph.D., University of California; Carnegie Mellon, 2022-

ZACHARY BRANSON, Associate Teaching Professor; Assistant Director for the Undergraduate Program - Ph.D., Harvard University; Carnegie Mellon, 2019-

DAVID CHOI, Associate Professor of Statistics and Information Systems - Ph.D., Stanford University; Carnegie Mellon, 2004-

PETER E. FREEMAN, Associate Teaching Professor; Director of the Undergraduate Program - Ph.D., University of Chicago; Carnegie Mellon, 2004-

CHRISTOPHER R. GENOVESE, Professor - Ph.D., University of California; Carnegie Mellon, 1994-

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

AMELIA HAVILAND, E.J. Barone Professor of Health Systems Management - Ph.D., Carnegie Mellon University; Carnegie Mellon, 2003-

JIAHUN JIN, Professor - Ph.D., Stanford University; Carnegie Mellon, 2007-

ROBERT E. 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, Associate Professor - Ph.D., University of Pennsylvania; Carnegie Mellon, 2020-

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

ANN LEE, Professor, Co-Director of PhD program - Ph.D., Brown University; Carnegie Mellon, 2005-

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

GONZALO E. MENA, Assistant Professor - Ph.D., Columbia University; Carnegie Mellon, 2023-

DANIEL NAGIN, Teresa and H. John Heinz III Professor of Public Policy - Ph.D., Carnegie Mellon University; Carnegie Mellon, 1976-

NYNKE NIEZINK, Associate 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-

TAEYONG PARK, Assistant Teaching Professor, CMU-Qatar - Ph.D., Washington University in St. Louis; Carnegie Mellon, 2018-

ANKIT PENNSIA, Assistant Professor - Ph.D., University of Wisconsin, Madison; Carnegie Mellon, 2023-

AADITYA RAMDAS, Associate Professor - Ph.D., Carnegie Mellon; Carnegie Mellon, 2018-

ALEX REINHART, Associate Teaching Professor - Ph.D., Carnegie Mellon University; Carnegie Mellon, 2018-

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

CHAD M. SCHAFFER, 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-

YANDI SHEN, Assistant Professor - Ph.D., University of Washington;

WEIJING TANG, Assistant Professor - Ph.D., University of Michigan; Carnegie Mellon, 2023-

WILL TOWNES, Assistant Professor - Ph.D., Harvard University; Carnegie Mellon, 2022-

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, UPMC University Professor of Statistics - Ph.D., University of Toronto; Carnegie Mellon, 1988-

RON YURKO, Assistant Teaching Professor - Ph.D., Carnegie Mellon; Carnegie Mellon, 2022-

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-

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

JOSEPH B. KADANE, Leonard J. Savage Professor of Statistics and Social Sciences - Ph.D., Stanford University; Carnegie Mellon, 1969-

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

MARK J. SCHERVISH, Professor - Ph.D., University of Illinois; Carnegie Mellon, 1979-

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

Special Faculty

F. SPENCER KOERNER, Lecturer - Ph.D., Carnegie Mellon; Carnegie Mellon, 2022-

JAMIE MCGOVERN, Director: Master's in Applied Data Science Program - B.A., Rice University; Carnegie Mellon, 2020-

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

Affiliated Faculty

ANTHONY BROCKWELL - Ph.D., Melbourne University; Carnegie Mellon, 1999-

PHILIPP BURCKHARDT - Ph.D., Carnegie Mellon; Carnegie Mellon, 2022-

BERNIE DEVLIN - Ph.D., Pennsylvania State University; Carnegie Mellon, 1994-

SAM VENTURA - Ph.D., Carnegie Mellon University; Carnegie Mellon, 2015-