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

B.S. in Statistics (Statistics and 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

The Bachelor of Science in Statistics and Data Science (Neuroscience Track) trains students in statistical theory, methods, applications, and judgment, such that they can make real-world impacts. Furthermore, students gain foundational training in psychology and neuroscience, so that students can apply statistical techniques to complex scenarios that arise from data on the brain and neurological activity. This degree is especially relevant for students interested in quantitative psychology and behavioral science, as well as pre-med students and those who want to pursue related graduate programs.

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 and 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 four courses:		
21-256	Multivariate Analysis	9
21-259	Calculus in Three Dimensions	10
21-266	Vector Calculus using Matrix Algebra	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

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.
- 21-266 and 21-268 are intended only for students with a very strong mathematical background.
- 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. 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
36-290	Introduction to Statistical Research Methodology	9
36-303	Sampling, Survey and Society	9
36-309	Experimental Design for Behavioral & Social Sciences	9
36-315	Statistical Graphics and Visualization	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:

36-311	Statistical Analysis of Networks	9
36-313	Statistics of Inequality and Discrimination	9
36-318	Introduction to Causal Inference	9
36-319	Statistics and Machine Learning for the Physical Sciences	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:

36-311	Statistical Analysis of Networks	9
36-313	Statistics of Inequality and Discrimination	9
36-318	Introduction to Causal Inference	9
36-319	Statistics and Machine Learning for the Physical Sciences	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:

36-235	Probability and Statistical Inference I *	9
36-236	Probability and Statistical Inference II **	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) in place of 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.

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 to 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	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.

Note: 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

Notes:

- 36-40136-40136-40136-40136-40136-40136-40136-40136-40136-401and 36-40236-40236-40236-40236-40236-40236-40236-40236-40236-402must be taken in residence at Carnegie Mellon University. Transfer credit will not be accepted.
- In order to meet the prerequisite requirements, a grade of at least a C is required in 36-40136-40136-40136-40136-40136-40136-40136-40136-40136-401

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 Neuropsychology 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**

*Note: 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.

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 and Data Science 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).

Sample Programs

The following sample programs illustrate two ways (of many) to satisfy the requirements for the B.S. in Statistics (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	Intermediate Data Analysis course	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	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
			-----