Statistics and Data Science, Minor

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@stat.cmu.edu

The Minor in Statistics and Data Science

Overview

The Minor in Statistics and Data Science trains students in the fundamentals of statistical theory, methods, applications, and judgment, such that they can make data-driven impacts in their major area of study. This minor is especially relevant for students who want to focus on the data-centric aspects of their major area of study, as well as students who want to develop data science skills that give them a competitive advantage in their career.

For this minor, students must complete the following training that is a part of all Statistics and Data Science degrees:

- Mathematical Foundations: Calculus and linear algebra, such that students can describe and analyze statistical models and methods.
- Data Analysis: Beginning, intermediate, and advanced data analysis topics, such that students know foundational and state-of-the-art methods for addressing data-specific questions. This includes regression and related methods, so that students can make accurate predictions and justified inferential conclusions about real data problems.
- Probability and Statistical Theory: Fundamental tools to estimate unknown quantities (e.g., averages, variances), quantify uncertainty (e.g., confidence intervals), and conduct inference (e.g., hypothesis tests), which are needed for most data analyses.

Curriculum

In order to complete a minor in Statistics 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 prerequisites:

		Units
21-090	Precalculus	10
21-111 & 21-112	Differential Calculus and Integral Calculus	10-20
or 21-120	Differential and Integral Calculus	
21-256	Multivariate Analysis	9-11
or 21-259	Calculus in Three Dimensions	
or 21-266	Vector Calculus using Matrix Algebra	
or 21-268	Multidimensional Calculus	
21-240	Matrix Algebra with Applications	10-11
or 21-241	Matrices and Linear Transformations	
or 21-242	Matrix Theory	

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 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 Analysis

Choose one (1) 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.

Intermediate Data Analysis

The Intermediate Data Analysis requirement must be completed prior to 36-401, if not, an additional Advanced Analysis and Methodology course is required. Choose one (1) 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

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

Advanced Data Analysis and Methodology

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

Notes:

- 36-401 and 36-402 must be completed in residence Carnegie Mellon University. Transfer credit will not be accepted.
- In order to meet the prerequisite requirements for several 36-4xx courses a grade of at least a C is required in 36-401.
- Special Topics rotate and new ones are regularly added.

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

Advanced Data Analysis and Methodology

Take the following course:

36-401	Modern Regression	9

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

36-311	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
36-402	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

Notes:

- 36-401 and 36-402 must be taken here at Carnegie Mellon University. Transfer credit will not be accepted.
- In order to meet the prerequisite requirements for several 36-4xx courses a grade of at least a C is required in 36-401.
- Special Topics rotate and new ones 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:

36-235	Probability and Statistical Inference I *	9
36-236	Probability and Statistical Inference II **	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.

Comments:

(i) In order to be in good standing and to continue with the minor, 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 Minor advisor regarding double counting.

Sample Programs for the Minor

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

The first schedule uses 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		