The Major in Statistics and Data Science

Statistics is the science and art of prediction and explanation. The mathematical foundation of statistics lies in the theory of probability, which is applied to problems of making inferences and decisions under uncertainty. Practical statistical analysis also uses a variety of computational techniques, methods of visualizing and exploring data, methods of seeking and establishing structure and trends in data, and a mode of questioning and reasoning that quantifies uncertainty.

Data Science expands on Statistics to encompass the entire lifecycle of data, from its specification, gathering and cleaning, through its management and analysis, to its use in making decisions and setting policy. It is a natural outgrowth of Statistics that incorporates advances in Machine Learning, Data Mining and High-Performance Computing along with domain expertise in the Social Sciences, Natural Sciences, Engineering, Management, Medicine and Digital Humanities.

Students majoring in Statistics and Data Science take courses in both mathematical and practical foundations. They are also encouraged to take courses in areas of application.

The B.A. in Statistics and Data Science is designed to acquaint students with the fundamental techniques in the field. The B.S. should prepare students to participate in research efforts or pursue graduate school in Data Science.

Courses for nonmajors: S&DS 100 and 101 through 109 only assume knowledge of high-school mathematics. Students who complete one of these courses should consider taking S&DS 230 (Introductory Data Analysis).

Prerequisite: Multivariable calculus is required and should be taken before or during the sophomore year. This requirement may be satisfied by MATH 120, ENAS 151, Math 230, or the equivalent.

Getting Started: Students who wish to major in S&DS are encouraged to take S&DS 200. But, students may also enter the major by taking a 100-level course followed by STAT 230. Students should complete the Calculus prerequisite and Linear Algebra requirement as early as possible, as they provide mathematical background that is required in many courses.

Requirements of the B.A.: The B.A. requires 11 courses, including a course in Linear Algebra (MATH 222 or 225), two courses on Core Probability and Statistics, two courses that provide Computational Skills, two courses on Methods, and the Senior Requirement. This leaves three more courses to be chosen from the categories described below.

Requirements of the B.S.: The B.S. requires 14 courses, including Linear Algebra (MATH 222 or 225), another course on Mathematical Theory, S&DS 242 (Theory of Statistics), another course on Core Probability and Statistics, two courses that provide Computational Skills, two courses

on Methods, at least two more courses from categories A through E below, and the Senior Requirement. This leaves three more courses to be chosen from categories A through G below.

The Senior Requirement: Students can complete the senior requirement either by taking a capstone course, which now is only S&DS 425 (presently STAT 325, Case Studies), or by completing an individual research project. Research projects may be advised by a member of the department of Statistics and Data Science or by a faculty member in an application area. Students may do their research project either individually (by registering for 491 or 492), or in the Senior Seminar (490). *Students must complete a research project to be eligible for Distinction in the Major.*

A. Core Probability and Statistics: These are essential courses in probability and statistics. Every major should take at least two of these courses, and should probably take more. Students completing the BS must take STAT 242.

- STAT 238 (Probability and Statistics)
- STAT 241 (Probability Theory)
- STAT 242 (Theory of Statistics)
- STAT 251 (Stochastic Processes)
- STAT 312 (Linear Models)

B. Computational Skills: Every student in Data Science should be able to compute with data. While the main purpose of some of these courses is not computing, students who have taken at least two of these courses should be capable of digesting and processing data. While there are other courses that require a lot of programming, these ones are essential.

- S&DS 200 or STAT 230 (Introductory Data Analysis)
- CPSC 100 or 112, or ENAS 130. Substitution of CPSC 201 or 223 permitted.
- STAT 325 (Statistical Case Studies)
- STAT 262 (Computational Tools for Data Science)

C. Methods of Data Science: These courses teach fundamental methods for dealing with data. They range from the practical to the theoretical. Every major must take at least two of these courses.

- STAT 361 (Data Analysis)
- STAT 363 (Multivariate Statistics for Social Sciences)
- STAT 365 (Applied Data Mining and Machine Learning)
- STAT 369 (Theory of Data Mining and Machine Learning)
- CPSC 477 (Natural Language Processing)
- AMTH 437 (Optimization Techniques)

D. Mathematical Foundations and Theory

All students in the major must know linear algebra. If they have learned linear algebra through other courses (such as MATH 230/231), they may substitute another course from this category. Students pursuing the B.S. must take at least two courses from this list. Students who wish to pursue graduate school should take many.

- MATH 222/225 (Linear Algebra).
- MATH 244 (Discrete Mathematics)
- MATH 260 (Basic Analysis in Function Spaces)
- MATH 300 (Topics in Analysis) or MATH 301 (Introduction to Analysis)
- STAT 330 (Advanced Probability)
- STAT 364 (Information Theory)
- CPSC 365 (Design and Analysis of Algorithms)
- STAT 410 (Statistical Inference, currently numbered 610)
- CPSC 469 (Randomized Algorithms)

E. Efficient Computation and Big Data: These courses are for students who want to do serious programming or implement large-scale analyses. None are required for the major. Students who wish to work in the software industry should take at least one of these.

- CPSC 223 (Data Structures and Programming Techniques)
- CPSC 323 (Introduction to Systems Programming and Computer Organization)
- CPSC 437 (Introduction to Databases)
- CPSC 424 (Parallel Programming Techniques)

F. Data Science in Context: Students are encouraged to take courses that involve the study of data in application areas. These courses will teach students how these data are obtained, how reliable they are, how they are used, and the types of inferences that can be made from them. These course selections should be approved by the DUS. Examples of such courses include

- PSYC 235 - Research Methods in Psychology
- ANTH 376 - Observing and Measuring Behavior
- EVST 362 - Observing Earth from Space

G. Methods in Application Areas: These are methods courses in areas of applications. They help expose students to the cultures of fields that explore data. These course selections should be approved by the DUS. Examples of such courses include:

- MB&B 452: Biological Data Science, Mining and Modeling
- FES 753: Regression Modeling
- STAT 674 / FES 781: Applied Spatial Statistics
- EENG 310: Signals and Systems
- EENG 445: Biomedical Image Processing and Analysis
- ECON 136: Econometrics
- CPSC 475: Computational Vision & Biological Perception

Current Statistics BA	Proposed S&DS BA
Prerequisites for Major: 2 credits Requirements for Major: 10 credits (2 electives) Total: 12 credits	Prerequisites for Major: 1 credits Requirements for Major: 11 credits (3 electives) Total: 12 credits
Prerequisites: Multivariable Calculus: Math 120 or ENAS 151 Linear Algebra: Math 222 or 225	**Prerequisite:** Multivariable Calculus: Math 120 or ENAS 151
Theory and Applications of Probability (2 courses): STAT 241 and 251	**A. Probability and Statistics (at least 2):** Stat 238, 241, 242, 251, 312
Computing (1 course): CPSC 112 ENAS 130	**B. Computational Skills (at least 2):** CPSC 100 or 112 S&DS 200 or 230 STAT 262 STAT 325
Data Analysis (2 courses): STAT 230 STAT 361, 363	**C. Methods of Data Science (at least 2):** STAT 361, 363, 365, 369 CPSC 477 (Natural Language Processing) AMTH 437 (Optimization Techniques)
	D. Mathematical Foundations and Theory: **Math 222/225 (required)** MATH 244 (Discrete Mathematics) MATH 260 (Basic Analysis in Function Spaces) MATH 301 (Introduction to Analysis) STAT 330 (Advanced Probability) STAT 364 (Information Theory) CPSC 365 (Design and Analysis of Algorithms) STAT 610 (Statistical Inference) CPSC 469 (Randomized Algorithms)
Statistical Inference (2 courses): STAT 242, 312 Are in category **A** of S&DS	**E. Efficient Computation and Big Data:** CPSC 223, 323, 424, 437
	F. Data Science in Context: Examples: PSYC 235, ANTH 376, EVST 362
	G. Methods in Application Areas (examples): MB&B 452, FES 753, EENG 310, EENG 475
Senior Requirement (1 credit): Senior Research Project (STAT 490)	**Senior Requirement (1 credit):** Senior Research Project (S&DS 490 or 491) Case Studies Course (S&DS 425)

Current Statistics BS	Proposed S&DS BS
Prerequisites for Major: 2 credits Requirements for Major: 12 credits (3 electives) Total: 14 credits	Prerequisites for Major: 1 credits Requirements for Major: 14 credits 5 electives: 3 from any category; 2 more from A-E Total: 15 credits
Prerequisites: Multivariable Calculus: Math 120 or ENAS 151 Linear Algebra: Math 222 or 225	**Prerequisite:** Multivariable Calculus: Math 120 or ENAS 151
Theory and Applications of Probability (2 courses): STAT 241 and 251	**A. Probability and Statistics (at least 2):** **STAT 242 (required)** STAT 238, 241, 251, 312
Computing (1 course): CPSC 112 ENAS 130	**B. Computational Skills (at least 2):** CPSC 100 or 112 S&DS 200 or 230 STAT 262 STAT 325
Data Analysis (2 courses): STAT 230 STAT 361, 363	**C. Methods of Data Science (at least 2):** STAT 361, 363, 365, 369 CPSC 477 (Natural Language Processing) AMTH 437 (Optimization Techniques)
Mathematical Analysis (1 course): MATH 260 MATH 300 MATH 301	**D. Mathematical Foundations and Theory: (at least 2):** **Math 222/225 (required)** MATH 244 (Discrete Mathematics) MATH 260 (Basic Analysis in Function Spaces) MATH 301 (Introduction to Analysis) STAT 330 (Advanced Probability) STAT 364 (Information Theory) CPSC 365 (Design and Analysis of Algorithms) STAT 610 (Statistical Inference) CPSC 469 (Randomized Algorithms)
Statistical Inference (2 courses): STAT 242, 312 Are in category **A** of S&DS	**E. Efficient Computation and Big Data:** CPSC 223, 323, 424, 437
	F. Data Science in Context: Examples: PSYC 235, ANTH 376, EVST 362
	G. Methods in Application Areas (examples): MB&B 452, FES 753, EENG 310, EENG 475
Senior Requirement (1 credit): Senior Research Project (STAT 490)	**Senior Requirement (1 credit):** Senior Research Project (S&DS 490 or 491) Case Studies Course (S&DS 425)

Current Applied Math BA	Proposed S&DS BA
Prerequisites for Major: 3 credits Requirements for Major: 11 credits Total: 14 credits	Prerequisites for Major: 1 credits Requirements for Major: 11 credits (3 electives) Total: 12 credits
Prerequisites: Multivariable Calculus: Math 120 or ENAS 151 Linear Algebra: Math 222 or 225 Intro Programming: CPSC 110, 112 or ENAS 130	**Prerequisite:** Multivariable Calculus: Math 120 or ENAS 151
Probability and Statistics (1 course): Stat 238 or 241	**A. Probability and Statistics (at least 2):** Stat 238, 241, 242, 251, 312
Differential Equations (1 course): Math 246 or ENAS 194	**B. Computational Skills (at least 2):** CPSC 100 or 112 S&DS 200 or 230 STAT 262 STAT 325
Data Analysis (1 course): STAT 230 or 361	**C. Methods of Data Science (at least 2):** STAT 361, 363, 365, 369 CPSC 477 (Natural Language Processing) AMTH 437 (Optimization Techniques)
Discrete Mathematics (1 course): MATH 244 or CPSC 202	**D. Mathematical Foundations and Theory:** **Math 222/225 (required)** MATH 244 (Discrete Mathematics) MATH 260 (Basic Analysis in Function Spaces) MATH 301 (Introduction to Analysis) STAT 330 (Advanced Probability) STAT 364 (Information Theory) CPSC 365 (Design and Analysis of Algorithms) STAT 610 (Statistical Inference) CPSC 469 (Randomized Algorithms)
Breadth Requirement (3 courses): Chosen from Optimization, Probability and Statistics, Differential Equations and Analysis, Algorithms and Numerical Methods, Mathematical Economics, Electrical Engineering, Data Mining & Machine Learning, Physical Sciences, Engineering.	
	E. Efficient Computation and Big Data: CPSC 223, 323, 424, 437
Concentration Requirement (3 courses): Three mathematical courses in an area of application, subject to approval by the DUS.	**F. Data Science in Context:** Examples: PSYC 235, ANTH 376, EVST 362
	G. Methods in Application Areas (examples): MB&B 452, FES 753, EENG 310, EENG 475
Senior Requirement (1 credit): Senior Research Project (AMTH 490 or 491)	**Senior Requirement (1 credit):** Senior Research Project (S&DS 490 or 491) Case Studies Course (S&DS 425)

Current Applied Math BS	Proposed S&DS BS
Prerequisites for Major: 3 credits Requirements for Major: 14 credits Total: 17 credits The courses below, plus 2 advanced courses	Prerequisites for Major: 1 credits Requirements for Major: 14 credits 5 electives: 3 from any category; 2 more from A-E Total: 15 credits
Prerequisites: Multivariable Calculus: Math 120 or ENAS 151 Linear Algebra: Math 222 or 225 Intro Programming: CPSC 110, 112 or ENAS 130	**Prerequisite:** Multivariable Calculus: Math 120 or ENAS 151
Probability and Statistics (1 course): Stat 238 or 241	**A. Probability and Statistics (at least 2):** **STAT 242 (required)** STAT 238, 241, 251, 312
Differential Equations (1 course): Math 246 or ENAS 194	**B. Computational Skills (at least 2):** CPSC 100 or 112 S&DS 200 or 230 STAT 262 STAT 325
Data Analysis (1 course): STAT 230 or 361	**C. Methods of Data Science (at least 2):** STAT 361, 363, 365, 369 CPSC 477 (Natural Language Processing) AMTH 437 (Optimization Techniques)
Discrete Mathematics (1 course): MATH 244 or CPSC 202	**D. Mathematical Foundations and Theory: (at least 2):** **Math 222/225 (required)** MATH 244 (Discrete Mathematics) MATH 260 (Basic Analysis in Function Spaces) MATH 301 (Introduction to Analysis) STAT 330 (Advanced Probability) STAT 364 (Information Theory) CPSC 365 (Design and Analysis of Algorithms) STAT 610 (Statistical Inference) CPSC 469 (Randomized Algorithms)
Mathematical Analysis (1 course): MATH 300 or MATH 301	
Breadth Requirement (3 courses): Chosen from Optimization, Probability and Statistics, Differential Equations and Analysis, Algorithms and Numerical Methods, Mathematical Economics, Electrical Engineering, Data Mining & Machine Learning, Physical Sciences, Engineering.	
	E. Efficient Computation and Big Data: CPSC 223, 323, 424, 437
Concentration Requirement (3 courses): Three mathematical courses in an area of application, subject to approval by the DUS.	**F. Data Science in Context:** Examples: PSYC 235, ANTH 376, EVST 362
	G. Methods in Application Areas (examples): MB&B 452, FES 753, EENG 310, EENG 475
Senior Requirement (1 credit): Senior Research Project (AMTH 490 or 491)	**Senior Requirement (1 credit):** Senior Research Project (S&DS 490 or 491) Case Studies Course (S&DS 425)