STATISTICS AND DATA SCIENCE

Melissa Schori, Regents Math 307
507-786-3113
schori1@stolaf.edu
wp.stolaf.edu/statistics/

(Mathematics, Statistics, and Computer Science)

With the growing abundance of data gathered in nearly every field, statistics and data science methods have become invaluable for transforming data into useful information. As a subject, statistics and data science lies at the intersection of mathematics, computer science, and statistics and connects naturally to the sciences (natural and social), the humanities, and even the arts.

As a part of the Department of Mathematics, Statistics, and Computer Science (MSCS) students that major in statistics and data science will have access to faculty and resources in these related but distinct disciplines. Faculty in SDS come from a variety of backgrounds, and use their applied research experience to guide students through collaborative projects and research opportunities that prepare students to contribute meaningfully to their chosen field.

Overview of the Major

A statistics and data science (SDS) major at St. Olaf gives students the quantitative, analytical, and communication skills to navigate an increasingly data-rich world. Students pursuing a statistics and data science major at St. Olaf will take courses that blend theoretical and practical concepts, exploring foundational ideas in computational thinking, statistical modeling, and mathematical underpinnings. SDS emphasizes a hands-on approach, and students will gain proficiency in programming languages and statistical software commonly used in the field. In addition, SDS courses encourage students to examine ethical considerations and societal implications of data collection and analysis, a particularly important consideration in the growing world of "big data".

To find out more about the statistics and data science major, visit the Statistics and Data Science program.

Intended Learning Outcomes for the Major

The Statistics and Data Science major is an integrated and interconnected set of courses, reflecting the interdisciplinary nature of the field and welcoming students from all backgrounds and experiences. Students will be grounded in foundational ideas from mathematics, such as probability, linear algebra, optimization, and multivariate thinking, alongside foundational ideas from computer science, such as algorithmic thinking, abstraction, project workflow, and reproducibility. Students will be immersed in collaborative team settings common to professional data scientists. In addition, students will acquire depth in their understanding of both statistics and data science, while being constantly challenged to consider ethical issues in their work and to apply statistics and data science principles to domains of expertise.

Students will demonstrate the ability to:

1. Acquire and curate data of all types.
2. Perform exploratory data analyses through data visualization and numerical summarization.
3. Build, assess, and interpret machine-learning and statistical models.
4. Communicate findings effectively and responsibly to a variety of audiences.
5. Apply analytical thinking to formulate problems, plan data collection/acquisition and engage in the data analysis process to provide insights in an integrated project.
6. Identify and critique multiple perspectives regarding data ethics and privacy.

Requirements

Requirements for the Major

Code	Title	Credits
Required Statistics & Data Science Courses (5 credits)		
SDS 164	Data Science 1	1.00
SDS 172	Statistics 1	1.00
SDS 264	Data Science 2	1.00
SDS 272	Statistics 2	1.00
SDS 341	Algorithms for Decision Making	1.00
Required Mathematics Courses (2 credits)		
MATH 120	Calculus I	1.00
or MATH 119	Introduction to Calculus	
MATH 220	Elementary Linear Algebra	1.00
Required Philosophy Courses (1 credit)		
PHIL 244	Philosophy of Science [2]	1.00
or PHIL 251	Science, Ethics, and Religion	
Electives Courses (2 credits) - Including at least one level III course from Statistics or Data Science courses listed below. [1]		
Statistics Depth Courses (select at least one of the following:)		1.00
Statistics Depth Level II Courses		
SDS 282	Topics in Statistics	
SDS 284	Biostatistics: Design and Analysis	
ECON 384	Econometrics: Cross-Sectional and Panel Data	
ECON 385	Econometrics: Time Series and Forecasting	
SDS 2XX Spatial Data Analysis (prereq SDS 172 and SDS 164)		
Statistics Depth Level III Courses		
SDS 316	Advanced Statistical Modeling	
SDS 322	Statistical Theory	
SDS 382	Advanced Topics in Statistics	
Data Science Depth Courses (1.0 credit)		1.0
Data Science Depth Level II Courses		
BIO 315	Principles of Bioinformatics	
CSCI 379	Foundations of Artificial Intelligence	

SDS 250 Principles of Data Visualization (prereq 164 or 172) [1]	
SDS 280 Topics in Data Science (prerequisites vary) [1]	
Data Science Depth Level III Courses	
SDS 333 High-Dimensional Data Analysis (prereq SDS 164 and MATH 220) [1]	
SDS 380 Topics in Data Science (prerequisites vary) [1]	
Total Credits	**10**

1

Courses that are not yet proposed. All except SDS 280 and SDS 380 have been taught as topics courses previously.

2

The first offering of PHIL 244 Philosophy of Science (with Data Science Module) that will satisfy the SDS major will be in 2025-26.

Courses

SDS 164: *Data Science 1*
Data is the currency of the modern world. At the intersection between statistics and computer science, data science is about gleaning information and making decisions from data. Using data from a variety of contexts and disciplines, students learn to summarize and extract insight from data, create compelling data visualizations, wrangle data, practice literate programming, and explore ethical issues in data science. No prior experience with programming is expected. This course cannot be taken after SDS 264.

SDS 172: *Statistics 1*
A first course in statistical methods, this course addresses study design and its implications as well as exploratory and inferential techniques for analyzing and modeling data. Topics include exploratory graphics, descriptive techniques, randomization tests, statistical designs, hypothesis testing, confidence intervals, and simple/multiple regression. Offered each semester. Enrollment limited for seniors. STAT 110, SDS 172, and ECON 260 all provide an introduction to statistics and students should not take more than one; they all can serve as a prerequisite for further courses. Also counts toward environmental studies major (natural science and social science emphases), kinesiology major, and business and management studies, mathematical biology, and public health studies concentrations.

SDS 264: *Data Science 2*
After mastering the foundations of data science - especially data wrangling and visualization - in SDS 164, data science students are ready to explore impactful applications and deeper fundamentals. Programming fundamentals may include iteration, functions, conditions, data types, SQL, regular expressions, version control, and simulation - all in the context of real data. New applications may include text as data, geospatial mapping, networks, and web scraping to produce dynamic graphics - presented with attention to data ethics and reproducible research. Offered annually.
Prerequisite: SDS 164: Data Science 1.

SDS 272: *Statistics 2*
This course takes a case-study approach to the fitting and assessment of statistical models with application to real data. Specific topics include multiple regression, model diagnostics, logistic regression, experimental design and ANOVA. The approach focuses on problem-solving tools, interpretation, model assumptions underlying analysis methods, and written statistical reports. Offered each semester. Also counts toward environmental studies major (natural science and social science emphases) and business and management studies, mathematical biology, neuroscience, and public health studies concentrations.
Prerequisite: SDS 172, ECON 260 or equivalent preparation (STAT 110 and SDS 164) or (AP Stat and SDS 164), or permission of instructor.

SDS 282: *Topics in Statistics*
Students explore special topics in statistics. Topics vary from year to year. May be repeated if topic is different. Offered periodically.

SDS 284: *Biostatistics: Design and Analysis*
The course investigates issues in health-related settings using a quantitative, research-oriented perspective. Course material focuses on global and public health issues, study design, methods for analyzing health data, and communication of research findings. Design topics include controlled trials, case-control, cohort and other observational studies. Methods include survival analysis and causal inference for observational studies. Communication emphasizes writing up findings and interpreting published research. Also counts toward mathematical biology concentration. Offered alternate years.
Prerequisite: completion of SDS 272 or permission of the instructor.

SDS 294: *Academic Internship*

SDS 298: *Independent Study*

SDS 316: *Advanced Statistical Modeling*
This course extends and generalizes methods introduced in STAT 272 by introducing generalized linear models (GLMs) and correlated data methods. GLMs cover logistic and Poisson regression, and more. Correlated data methods include longitudinal data analysis and multi level models. Applications are drawn from across the disciplines. Offered annually. Also counts toward neuroscience concentration.
Prerequisite: SDS 272.

SDS 322: *Statistical Theory*
This course is an investigation of modern statistical theory along with classical mathematical statistics topics such as properties of estimators, likelihood ratio tests, and distribution theory. Additional topics include Bayesian analysis, bootstrapping, Markov Chain Monte Carlo, and other computationally intensive methods. Offered alternate years. Also counts toward neuroscience concentration.
Prerequisite: SDS 272 and MATH 262.

SDS 341: *Algorithms for Decision Making*
This course introduces students to the subject of machine learning. The primary focus is the development and application of powerful machine learning algorithms applied to complex, real-world data. Topics covered include linear regression, nearest neighbor models, k-means clustering, shrinkage methods, decision trees and forests, boosting, bagging, support vector machines, and hierarchical clustering. Applications are taken from a wide variety of disciplines, including biology, economics, public policy, public health, and sports. Offered on a regular basis. Counts toward computer science and mathematics majors and statistics and data science concentration.
Prerequisite: SDS 164 or SDS 264 or permission of the instructor.

SDS 382: ***Advanced Topics in Statistics***
Students work intensively on a special topic in statistics. Topics may vary from year to year. May be repeated if topics are different. Offered periodically.
Prerequisites: Permission of instructor.

SDS 389: ***Statistics and Data Science Research Methods (0.50)***
Students focus on writing scientific papers, preparing scientific posters, and giving presentations in the context of a specific, year-long, interdisciplinary research project. In addition, this weekly seminar series builds collaborative research skills such as working in teams, performing reviews of math, statistics, and computer science literature, consulting effectively, and communicating proficiently. Exposure to post-graduate opportunities in statistics and data science disciplines is also provided. Open to students accepted into the Center for Interdisciplinary Research.

SDS 394: ***Academic Internship***

SDS 396: ***Directed Undergraduate Research***
This course provides a comprehensive research opportunity, including an introduction to relevant background material, technical instruction, identification of a meaningful project, and data collection. The topic is determined by the faculty member in charge of the course and may relate to their research interests. Offered based on department decision. May be offered as a 1.00 credit course or .50 credit course.
Prerequisite: determined by individual instructor.

SDS 398: ***Independent Research***

Related Courses

CSCI 125: ***Computer Science for Scientists and Mathematicians***
This course teaches introductory programming with a focus on handling data. Emphases include programming concepts and structures, writing computer code to solve quantitative problems, and the use of programming to analyze data. The primary tool is the Python programming language. Students work individually and in teams to apply basic principles and explore real-world datasets with a sustainability theme. Offered each semester. Also counts toward statistics and mathematical biology concentrations; one of CSCI 121, CSCI 125, or CSCI 251 counts toward applied linguistics concentration.
Prerequisite: calculus or permission of the instructor.

ECON 260: ***Introductory Econometrics***
This course emphasizes skills necessary to understand and analyze economic data. Topics include descriptive statistics, probability and random variables, sampling theory, estimation and hypothesis testing, and practical and theoretical understanding of simple and multiple regression analysis. Applications to economic and business problems use real data, realistic applications, and econometric/statistical software. Offered each semester. ECON 260 is required for economics majors who do not take both SDS 272 and either ECON 384 or ECON 385. Credit toward the economics major will not be given for ECON 260 following completion of SDS 272. Also counts toward environmental studies major (social science emphasis) and public health studies concentration.
Prerequisite: MATH 119 or MATH 120 and one of ECON 110 - ECON 121, or permission of instructor.

ECON 384: ***Econometrics: Cross-Sectional and Panel Data***
This course emphasizes theoretical foundations, mathematical structure, and applications of major econometric techniques appropriate for cross-sectional and panel data. Topics to be covered include generalized least squares, dummy variables, non-linear models, instrumental variables techniques, fixed- and random-effects models, and limited dependent variable models. This course is recommended for students interested in analysis of issues in microeconomics and public policy. Offered annually. ECON 384 and ECON 385 may not both be used to satisfy the economic analysis requirements for either the economics or quantitative economics major.
Prerequisite: ECON 262 and one of ECON 260, ECON 263, or SDS 272; or permission of instructor.

ECON 385: ***Econometrics: Time Series and Forecasting***
This course emphasizes the theoretical foundations, mathematical structure, and applications of major econometric techniques appropriate for time-series data. Topics covered include generalized least squares, single-equation time-series models, multi-variable time-series models, forecasting and forecast evaluation, and seasonality. This course is recommended for students interested in analysis of issues in macroeconomics and finance. Offered annually. ECON 384 and ECON 385 may not both be used to satisfy the economic analysis requirements for either the economics or quantitative economics major. Completion of MATH 220 may be helpful but is not required.
Prerequisites: ECON 261 and one of ECON 260 or ECON 263 or SDS 272; or permission of instructor.

MATH 262: ***Probability Theory***
This course introduces the mathematics of randomness. Topics include probabilities on discrete and continuous sample spaces, conditional probability and Bayes' Theorem, random variables, expectation and variance, distributions (including binomial, Poisson, geometric, normal, exponential, and gamma) and the Central Limit Theorem. Students use computers to explore these topics. Offered each semester. Also counts toward business and management studies concentration.
Prerequisite: MATH 126 or MATH 128.

PSYCH 230: ***Research Methods in Psychology***
This course prepares students with tools for understanding how research studies in psychology are conceptualized, designed, and ethically conducted, and how data is analyzed, interpreted, and disseminated. Students apply this understanding in independent and small group research projects. In the process, students develop critical reading, thinking, and scientific writing skills. Students attend lectures plus one two-hour laboratory per week. Offered each semester. Also counts toward environmental studies major and statistics and data science and public health studies concentrations.
Prerequisites: PSYCH 125, and STAT 110 or SDS 172 or ECON 260.

SOAN 371: ***Foundations of Social Science Research: Quantitative Methods***
Students gain the skills necessary to conduct and critically evaluate quantitative research. Students learn the underlying theoretical assumptions and orientations of quantitative research, including research design, sampling techniques, strategies for data collection, and approaches to analysis. Students gain practice in data analysis by conducting are search project and using the Statistical Package for the Social Sciences (SPSS), a standard in sociology. Offered annually in the fall semester. Also counts toward environmental studies major (social science emphasis) and business and management studies and public health studies concentrations.
Prerequisite: STAT 110 or SDS 172; open to junior or senior sociology/anthropology majors only.

STAT 110: ***Principles of Statistics***
This is an introductory course for the liberal arts. Students learn study design principles and develop statistical literacy and reasoning. They learn to describe distributions, assess if known distributions fit their data, estimate population values with confidence intervals, and assess statistical significance with hypothesis tests (e.g., chi-square, *z*-, and *t*-tests, ANOVA, correlation, and regression). Not recommended for students who have completed a term of calculus. STAT 110, SDS 172, and ECON 260 all provide an introduction to statistics, and students should not take more than one; they all can serve as a prerequisite for further courses. Offered each semester. Also counts toward environment studies major (social science emphasis) and kinesiology major.

STAT 270: ***Intermediate Statistics for Social Science Research***
This course focuses on the use of statistics in a social science context. Students investigate three essential questions: How can one reliably measure something? How does one design valid research? How does one analyze research results? Topics include ANOVA designs (for example, one-way and two-way with interaction), data reduction methods, and principles of measurement. Interdisciplinary groups work together on case studies throughout the term. Offered alternate years. Also counts toward public health studies concentration.
Prerequisites: STAT 110 or SDS 172 or ECON 260 or equivalent preparation, or permission of the instructor.

Faculty

Program Director, 2024-2025
Kathryn Ziegler-Graham
Professor of Mathematics, Statistics, and Computer Science
biostatistics

Laura Boehm Vock
Assistant Professor of Mathematics, Statistics, and Computer Science
statistics; spatial data analysis

Jaime I. Davila
Assistant Professor of Mathematics, Statistics, and Computer Science

Francesca Gandini
Assistant Professor of Mathematics, Statistics, and Computer Science

Kimberly (Kim) Mandery
Visiting Instructor of Mathematics, Statistics, and Computer Science

Paul J. Roback (on sabbatical 2024-25)
Kenneth Bjork Distinguished Professor of Mathematics, Statistics, and Computer Science
statistics

Joseph Roith
Associate Professor of Practice in Mathematics, Statistics, and Computer Science
statistics

Martha Zillig
Visiting Assistant Professor of Mathematics, Statistics Computer Science