STATISTICS AND DATA SCIENCE

Ellen Haberoth, Regents Math 307
507-786-3113
martinep@stolaf.edu
wp.stolaf.edu/mscs

(Statistics, Mathematics, and Computer Science)

With the growing abundance of data gathered in nearly every field, statistical methods have become invaluable for transforming data into useful information. As a subject, statistics is interdisciplinary, spanning the sciences (natural and social), the humanities, and even the arts. Examples of areas of applications include economics, biology, health, education, actuarial sciences, and law. An increasing number of majors and concentrations require or recommend a statistics course.

Overview of the Concentration

At St. Olaf, students can combine their interests in statistics with any major and acquire a background that leads to graduate study and abundant career opportunities. To find out more about the statistics concentration, visit the Statistics program.

Intended Learning Outcomes for the Concentration

Requirements

The statistics and data science concentration consists of a minimum of four (4) courses, as follows:

<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT 272</td>
<td>Statistical Modeling</td>
<td>1.00</td>
</tr>
<tr>
<td>MSCS 264</td>
<td>Introduction to Data Science</td>
<td>1.00</td>
</tr>
</tbody>
</table>

A prerequisite course prior to STAT 272 can be fulfilled with any of the following:

- AP Statistics
- STAT 110 Principles of Statistics
- STAT 212 Statistics for the Sciences
- ECON 263 Statistics for Economics

Select 1 Level III course from the following: 1.00

- STAT 316 Advanced Statistical Modeling
- STAT 322 Statistical Theory
- MSCS 341 Algorithms for Decision Making
- ECON 385 Econometrics

Select 1 elective from the following: 1.00

- MATH 262 Probability Theory (strongly recommended for mathematics majors)
- PSYCH 230 Research Methods in Psychology
- PSCI 220 Analyzing Politics and Policies
- STAT 270 Intermediate Statistics for Social Science Research
- STAT 282 Topics in Statistics
- STAT 284 Biostatistics: Design and Analysis

CSCI 125 Computer Science for Scientists and Mathematicians

SOAN 371 Foundations of Social Science Research: Quantitative Methods

Any of the Level III courses listed above

Experiential Learning Component (optional, see below)

Total Credits 4

Experiential Learning Component (Optional)

Each concentrator is encouraged to participate in experientially based research or employment that takes statistical methods beyond the traditional classroom. This can occur on- or off-campus. Prior approval by the director of statistics program and a letter after the fact from a supervisor are required to earn credit. Excellent opportunities for experiential learning in statistics are available through STAT 294 Academic Internship, MSCS 390 Mathematics practicum, and MSCS 389 Research Methods (through the Center for Interdisciplinary Research (CIR)). As a CIR fellow, students can work during the academic year or summer with faculty on research from a variety of disciplines.

Note: For students considering graduate school in statistics or a closely related field, the following courses are recommended:

<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATH 126</td>
<td>Calculus II</td>
<td>1.00</td>
</tr>
<tr>
<td>or MATH 128</td>
<td>Honors Calculus II</td>
<td></td>
</tr>
<tr>
<td>MATH 220</td>
<td>Elementary Linear Algebra</td>
<td>1.00</td>
</tr>
<tr>
<td>MATH 226</td>
<td>Multivariable Calculus</td>
<td>1.00</td>
</tr>
<tr>
<td>MATH 230</td>
<td>Differential Equations I</td>
<td>1.00</td>
</tr>
<tr>
<td>MATH 242</td>
<td>Modern Computational Mathematics</td>
<td>1.00</td>
</tr>
<tr>
<td>MATH 244</td>
<td>Real Analysis I</td>
<td>2.00</td>
</tr>
<tr>
<td>&amp; MATH 344</td>
<td>and Real Analysis II</td>
<td></td>
</tr>
<tr>
<td>CSCI 251</td>
<td>Software Design and Implementation</td>
<td>1.25</td>
</tr>
<tr>
<td>&amp; CSCI 252</td>
<td>and Software Design and Implementation Lab (0.25)</td>
<td></td>
</tr>
</tbody>
</table>

STAT 110, STAT 212, and ECON 263 all provide an introduction to statistics, and students should not take more than one; they can all serve as a prerequisite for further courses, although ECON 263 is geared toward majors in economics. STAT 214 is also an introductory course that assumes no background in statistics, but it can also be taken by students coming from STAT 110, ECON 263 or AP Statistics who would like a transition into the statistics concentration.

Experiential Learning Component (Optional)

Each concentrator is encouraged to participate in experientially based research or employment that takes statistical methods beyond the traditional classroom. This can occur on- or off-campus. Prior approval by the director of statistics program and a letter after the fact from a supervisor are required to earn credit. Excellent opportunities for experiential learning in statistics are available through STAT 294 Academic Internship, MSCS 390 Mathematics practicum, and MSCS 389 Research Methods (through the Center for Interdisciplinary Research (CIR)). As a CIR fellow, students can work during the academic year or summer with faculty on research from a variety of disciplines.
during the academic year or summer with faculty on research from a variety of disciplines.

Note: For students considering graduate school in statistics or a closely related field, the following courses are recommended:

<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATH 126 or MATH 128</td>
<td>Calculus II or Honors Calculus II</td>
<td>1.00</td>
</tr>
<tr>
<td>MATH 220</td>
<td>Elementary Linear Algebra</td>
<td>1.00</td>
</tr>
<tr>
<td>MATH 226</td>
<td>Multivariable Calculus</td>
<td>1.00</td>
</tr>
<tr>
<td>MATH 230</td>
<td>Differential Equations I</td>
<td>1.00</td>
</tr>
<tr>
<td>MATH 242</td>
<td>Modern Computational Mathematics</td>
<td>1.00</td>
</tr>
<tr>
<td>MATH 244 &amp; MATH 344</td>
<td>Real Analysis I and Real Analysis II</td>
<td>2.00</td>
</tr>
<tr>
<td>CSCI 251 &amp; CSCI 252</td>
<td>Software Design and Implementation and Software Design and Implementation Lab (0.25)</td>
<td>1.25</td>
</tr>
</tbody>
</table>

STAT 110, STAT 212, and ECON 263 all provide an introduction to statistics, and students should not take more than one; they can all serve as a prerequisite for further courses, although ECON 263 is geared toward majors in economics. Students coming from STAT 110, ECON 263 or AP Statistics who would like to transition into the statistics concentration are encouraged to begin in MSCS 264.

Courses

**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). In addition to lectures, there is a weekly computer lab. Not recommended for students who have completed a term of calculus. STAT 110, STAT 212, and ECON 263 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. Counts toward environment studies major (social science emphasis) and exercise science major.

**STAT 212: Statistics for the Sciences**
A first course in statistical methods for scientists, this course addresses issues for proposing/designing an experiment, as well as exploratory and inferential techniques for analyzing and modeling scientific data. Topics include probability models, exploratory graphics, descriptive techniques, statistical designs, hypothesis testing, confidence intervals, and simple/multiple regression. Offered each semester. Enrollment limited for seniors. STAT 110, STAT 212, and ECON 263 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), exercise science major, and management studies and mathematical biology concentrations.

**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.

**Prerequisites:** STAT 110 or STAT 212 or ECON 263 or equivalent preparation, or permission of the instructor.

**STAT 272: Statistical Modeling**
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, and logistic regression. The approach focuses on problem-solving tools, interpretation, mathematical models underlying analysis methods, and written statistical reports. Offered each semester. Also counts toward environmental studies major (natural science and social science emphases) and management studies, mathematical biology, and neuroscience concentrations.

**Prerequisite:** STAT 110 or STAT 212 or ECON 263 or equivalent preparation, or permission of instructor.

**STAT 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.

**STAT 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 STAT 272 or permission of the instructor.

**STAT 294: Academic Internship**

**STAT 298: Independent Study**

**STAT 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:** STAT 272.

**STAT 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 annually. Also counts toward neuroscience concentration.

**Prerequisite:** STAT 272 and MATH 262.
STAT 394: Academic Internship

STAT 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 his/her 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.

STAT 398: Independent Research

Related Courses

CSCI 125: Computer Science for Scientists and Mathematicians
This course focuses on handling data: visualization, finding patterns, and communicating with data. The primary tools are Python (for transforming data) and R (for visualization and classification). 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 linguistic studies concentration. **Prerequisite:** calculus or consent of the instructor.

ECON 263: Statistics for Economics
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 263 is required for economics majors who do not take both STAT 272 and ECON 385. Credit towards the economics major will not be given for ECON 263 following completion of STAT 272. Also counts toward environmental studies major (social science emphasis). **Prerequisite:** MATH 119 or MATH 120 and one of ECON 110-121, or permission of the instructor.

ECON 385: Econometrics
Ideal for students interested in applying statistical models to economic problems, this course emphasizes theoretical foundations, mathematical structure and applications of major econometric techniques, including ordinary least squares, generalized least squares, dummy variables, non-linear transformations, instrumental variables, simultaneous equation modeling, and time series models. Students in the class complete a sophisticated economic research project of their choice. Offered annually. **Prerequisite:** ECON 261 and either ECON 263 or STAT 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 management studies concentration. **Prerequisite:** MATH 126 or MATH 128.

MSCS 124: A Four-Dimensional Excursion
This elementary introduction to the geometry of four-dimensional Euclidean space begins with an examination of lower-dimensional spaces and their relationships to one another. Generic objects are examined with the goal of understanding how a four-dimensional counterpart might be generated. Students explore what it means to ‘see’ four-dimensional objects and in alternative ways. The term ends with group projects realized as a short research paper, a virtual reality show/object, and a class presentation. Offered periodically during Interim. No prerequisite.

MSCS 150: Statistical and Data Investigations
Students learn basic techniques to analyze, manage, visualize, and model data. Instruction focuses on the analysis of ‘real’ salient datasets in a computer-equipped classroom. In small groups students discuss, analyze, and solve case study-based problems. Class sessions include the Inquiry-Based Learning technique, which engages students in frequent presentations of their solutions to the class. Students use the R statistical software to perform statistical computing and data visualizations. Offered annually.

MSCS 264: Introduction to Data Science
Data is the currency of the modern world, and data science is a field that sits at the intersection between statistics and computer science. At its heart, data science is about gleaning information and making decisions from data; this course provides a solid foundation to the most important data science tools. Students develop a common language for creating visualizations, wrangling with data, programming in a literate manner, producing reproducible research, and communicating results. Offered each semester. Counts toward statistics concentration.

MSCS 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. Familiarity with a programming language such as R or Python is highly recommended. Offered on a regular basis. Counts toward computer science and mathematics majors and statistics and data science concentration. **Prerequisite:** MATH 220, CSCI 251, or STAT 272 or permission of the instructor.

MSCS 389: Math, Statistics, and Computer 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 math, statistics, and computer science disciplines is also provided. Open to students accepted into the Center for Interdisciplinary Research.
**MSCS 390: Mathematics Practicum**
Students work in groups on substantial problems posed by, and of current interest to, area businesses and government agencies. The student groups decide on promising approaches to their problem and carry out the necessary investigations with minimal faculty involvement. Each group reports the results of its investigations with a paper and an hour-long presentation to the sponsoring organization. Offered annually during Interim.

**Prerequisite:** Permission of instructor.

**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, exercise science major, and statistics and data science concentration.

**Prerequisites:** PSYCH 125, and STAT 110 or STAT 212 or ECON 263.

**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 research 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 management studies concentration.

**Prerequisite:** STAT 110 or STAT 212; open to junior or senior sociology/anthropology majors only.

**Faculty**

**Director, 2019-2020**

**Kathryn Ziegler Graham**
Associate Professor of Mathematics, Statistics, and Computer Science
biostatistics

**Anthony Becker**
Professor of Economics
econometrics; forecasting; economic damages; microeconomic theory; public policy

**Sharon J. Lane-Getaz**
Associate Professor of Mathematics, Statistics, and Computer Science
statistics; statistics education

**Julie M. Legler**
Professor of Mathematics, Statistics, and Computer Science
biostatistics

**Jesse Miller**
Visiting Assistant Professor of Mathematics, Statistics, and Computer Science

**Matthew P. Richey**
Professor of Mathematics, Statistics, and Computer Science
applied and computational mathematics

**Paul J. Roback**
Professor of Mathematics, Statistics, and Computer Science
statistics

**Joseph Roith**
Visiting Assistant Professor of Mathematics, Statistics, and Computer Science