**SYLLABUS**

**Multivariate statistics for environmental science - EAS 501.077**

*Revised 15 November 2021 (check course web page for updates)*

**Academic semester**: Fall

**Academic year**: 2021-2022 (expected to repeat annually each fall)

**Class schedule**: **Mondays and Wednesdays: 1:00pm – 2:30pm**
***Lectures will be held in-person only* in Dana 1024**

**Prerequisites:** EAS 538 (or equivalent, per instructor approval)

**Note regarding COVID:** As they have throughout the past year-and-a-half, policies around academic and public health are subject to change as the pandemic evolves. This course will follow all policies issued by the University, which are documented on the Campus Blueprint's FAQ. These policies may change over the course of the term, so please review the Campus Blueprint's FAQ for the most up to date information.

**Masking:** Effective Aug. 11, a face covering is required for anyone in a campus building or on campus transit, regardless of vaccination status. The change applies across the entire University, including in the classroom. You can read a copy of the policy here.

**Instructor**: Andrew Gronewold, Ph.D., P.E., Associate Professor
School for Environment and Sustainability
Department of Earth and Environmental Sciences
Department of Civil and Environmental Engineering
E-mail: drewgron@umich.edu
Phone: 734-764-6286
Office hours: [In-person, Dana 2008; Zoom available upon request]
Mondays, 2:30pm – 3:30pm
Thursdays,
8:00am – 9:00am (9/30, 10/14, 10/28, 11/11, 12/9)
8:00am – 10:00am (9/23, 10/7, 10/21, 11/4, 11/18, 12/2)

**GSI:** Tiffany Dias
School for Environment and Sustainability
E-mail: tadias@umich.edu
Office hours: Starting Sept 8, https://umich.zoom.us/j/98138819349
Passcode: 053578
Wednesday, 3:00pm – 4:00pm (or upon request)

**SPECIAL STATEMENT RELATED TO RECORDINGS**

*Course lectures may be audio/video recorded and made available to other students in this course. As part of your participation in this course, you may be recorded. If you do not wish to be recorded, please contact me (drewgron@umich.edu) to discuss alternative arrangements. Students may not record or distribute any class activity without written permission from the instructor, except as necessary as part of*

**SYLLABUS**

**Multivariate statistics for environmental science - EAS 501.077**

*Revised 15 November 2021 (check course web page for updates)*

*approved accommodations for students with disabilities. Any approved recordings may only be used for the student's own private use.*

**Class web-pages:** The class has two web-pages:

- The UM Canvas course page
- Our Google sites course web page

## I. Overview

Multivariate analysis provides the foundational tools for understanding and analyzing data sets that require simultaneous examination of multiple variables. This course aims to enable students with the ability to describe, explore, and find order in data, and to extract underlying structure and patterns. While the course is intended to provide students with a background suitable for a variety of applications, the emphasis will be on the environmental sciences. Further emphasis will be placed on a working knowledge of conventional statistical methods, use of software packages (primarily R), and problem solving.

Students will be expected to have a reasonable level of competence in statistics (i.e. will have taken a basic statistics course). Basic knowledge of R or similar software is helpful, but not required. The course will culminate with submittal of an individual semester-long project designed to apply fundamentals from the course to a data set selected by the student.

PRIMARY TEXT(S): The course will draw primarily on material from the following text(s), available in electronic format through the UM library system:

- Everitt, B.S., and Hothorn, T (2011). *An Introduction to Applied Multivariate Analysis with R*. Springer.
- Zelterman, Daniel (2015). *Applied Multivariate Statistics with R.* Springer.

ADDITIONAL TEXT(S): Some course material (such as examples in R) may be drawn from additional texts. These texts (identified as Supplemental) are listed on the course web-site (see link above).

METHOD OF EVALUATION: Students will be evaluated based on a combination of a semester-long course project, homework assignments, and quizzes:

| ITEM | PERCENTAGE OF FINAL GRADE |
|---|---|
| Course project (intermediate and final products) | 60% |
| Homework assignments (due at beginning of designated class) | 40% |

**SYLLABUS**

**Multivariate statistics for environmental science - EAS 501.077**

*Revised 15 November 2021 (check course web page for updates)*

NOTES ON INDIVIDUAL COURSE PROJECTS: The final project will be in the form of a written report describing a student's analysis of a "real-world" data set selected by the student that employs at least two (2) multivariate analysis (MVA) methods; these methods may, but do not necessarily have to, come from the course. Details of the project expectations are included in a separate document, available through the course web-page. In general, projects should include a clear and testable hypothesis based on either scientific theory or historical empirical evidence, a description of the requisite data and the tools needed to analyze that data, and a summary of findings (including an appendix with details of the analysis including plots, tables, and computer output). Students are encouraged to align their project scope and structure with requirements for a graduate thesis, or for submittal to a peer-reviewed publication.

## II. Learning objectives

The objective of this course is to provide a solid foundation for applied statistical analysis with an emphasis on traditional components of multivariate methods and use of data sets from the environmental sciences.

## III. Learning outcomes

Students will meet for class twice per week throughout the semester for a lecture-style presentation. Progress in the course will be expected through outside-of-class readings, homework assignments, and a semester-long individual applied project. By the end of this course, students are expected to demonstrate proficiency in:

- The tools and concepts of multivariate statistics
- Application of multivariate statistics to real environmental data sets
- R software for performing multivariate analysis

**SYLLABUS**

**Multivariate statistics for environmental science - EAS 501.077**

*Revised 15 November 2021 (check course web page for updates)*

## IV. Course schedule (subject to change)

*NOTE: Syllabus and schedule may change periodically; check the course web site for updates*

| Week | Day(s) | Topic(s) | Due dates (for homework and semester project) | Recommended reading [sections in brackets are more advanced and generally beyond the scope of our course] | |
|---|---|---|---|---|---|
| | | | | Everitt & Hothorn (2011) | Zelterman (2015) |
| | ***CLASS WILL BE REMOTE ONLY FOR WEEKS 1 AND 2*** | | | | |
| 1 | M – Aug 30<br>W – Sep 1 | Course overview<br>Introduction to MVA | | **Preface; 1.1 - 1.3** | **Chapter 1 (all)** |
| 2 | ***M – Sep 6***<br>W – Sep 8 | ***Labor day (Sep 7) – no class***<br>R: coding and data visualization<br>Data management practices | | **1.4; 2.1 - 2.4;<br>2.6 - 2.9** | **2.1 - 2.4, 2.8<br>[2.5 – 2.7]<br>Chapter 3 (all)** |
| 3 | M – Sep 13<br>W – Sep 15 | Covariance, correlation, distance<br>Linear algebra: the basics | **HW 1 due (Mon, Sep 13)[1]** | **1.5** | **[Chapter 4]** |
| 4 | M – Sep 20<br>W – Sep 22 | Univariate normal distribution<br>Multivariate normal distribution | | **[1.6]** | **[Chapters 5-7]** |
| 5 | M – Sep 27<br>W – Sep 29 | Principle components analysis (PCA) | | **3.1-3.2, 3.4, 3.6, 3.10-3.12<br>[3.3, 3.5, 3.7, 3.8, 3.9]** | **[8.1 – 8.4]** |
| 6 | M – Oct 4<br><br>W – Oct 6 | Factor analysis<br><br>Cluster analysis | **PROJECT PHASE 1 DUE[1] (Wed, Oct 6)**<br>(data, hypothesis, research questions) | **5.1 - 5.3, 5.9 - 5.11<br>[5.4 – 5.8]<br>6.1 – 6.4, 6.6-6.7; [6.5]** | **[8.5]<br><br>11.1-11.2 [11.3]** |
| 7 | M – Oct 11<br>W – Oct 13 | Correspondence analysis<br>More methods for categorical data | **HW 2 due (Wed, Oct 13)[1]** | *No specific reading from E&H or Zelterman; recommend materials on STHDA web-site and other on-line sources* | |
| 8 | ***M – Oct 18***<br>W – Oct 20 | ***Fall study break – no class***<br>More methods: (MCA, MFA, FAMD) | | | |
| 9 | M – Oct 25<br>W – Oct 27 | Multiple and multivariate regression | | [3.10] | [Chapter 9] |
| 10 | M – Nov 1<br>W – Nov 3 | More on logistic regression<br>Multidimensional scaling | ~~HW 3 due (Mon, Nov 1)[2]~~<br>**HW 3 due (Wed, Nov 3)[1]** | Chapter 4 | - |
| 11 | M – Nov 8<br>W – Nov 10 | Structural equation models | **PROJECT PHASE 2 DUE[1] (Mon, Nov 8)**<br>(analysis & results) | Chapter 7 | - |
| 12 | M – Nov 15<br>W – Nov 17 | Case study 1<br>Case study 2 (T. Dias guest lecture) | | - | - |
| 13 | M – Nov 22<br>***W – Nov 24*** | Case study 3<br>***No Class – Thanksgiving break*** | **HW 4 due (Mon, Nov 22)[1]** | | |
| 14 | *M – Nov 29*<br>*W – Dec 1* | Case study 4<br>Selecting and interpreting methods | **PROJECT PHASE 3 DUE[1] (Wed, Dec 1)**<br>(full report in draft form) | - | - |
| 15 | *M – Dec 6*<br>*W – Dec 8* | Working review sessions:<br>(R coding, MVA topics, project work) | | | |
| | | | | | |

**SYLLABUS**

**Multivariate statistics for environmental science - EAS 501.077**

*Revised 15 November 2021 (check course web page for updates)*

| | | | | | |
|---|---|---|---|---|---|
| | Dec 11 – 17 | FINAL EXAM PERIOD<br>(No final exam for this class) | **FINAL PROJECTS DUE[1]**<br>**(Wed, Dec 15)** | | |

NOTES:

1) Homework assignments and project updates (phases 1 – 3 and final report) should be submitted via Canvas by 1:00pm on the due date.
2) Some syllabus structure, lecture notes, and assignments were borrowed from Dr. Andreas Hamann; University of Alberta (Class RenR 690)