**FORS 538 Statistical Models for Ecological Data Analysis**

Fall 2020: Remote online instruction. Lab periods scheduled for Tuesday, Thursday 11:00-12:30; Old Journalism (Stone Hall) 106

**Instructors**:
Solomon Dobrowski
Office: CHCB 438
Phone: 6068
Email: solomon.dobrowski@umontana.edu
Office hours: by appointment

**Course Description**:
This is an ***applied*** course covering advanced statistical modeling techniques using examples from forestry, ecology, and the environmental sciences. I will cover data management, visualization, and scripting with *R*, an open source data analysis and statistics platform which is rapidly becoming the standard in many scientific disciplines. After reviewing the linear regression model and associated diagnostics, I will explore various parametric and non-parametric modeling strategies that allow for non-linear response functions and/or non-Gaussian response distributions. Estimation and inference in the context of generalized linear models, generalized additive models, and classification and regression trees will be discussed using examples from the scientific literature. This course will lay the foundation for subsequent graduate-level analytic coursework that is offered.

**Prerequisites**:
Students must have previous coursework in statistics to the level of STAT 451 and 452 (formerly MATH 444 and 445) or equivalent (i.e., linear regression and ANOVA).

**Course Material**:
Course materials will principally be comprised of primary research articles, online primers, and tutorials. All course materials will be distributed in class, are available online, or will be made available via the course Moodle website.

There is no required textbook although I recommend the following texts, both for gaining traction with R and for further reference on advanced techniques:

Faraway – Linear Models with R and Extending the Linear Model with R
Venables and Ripley - Modern Applied Statistics with S
Murrell - R Graphics

I will make some of these texts available on reserve in the library.

**Objectives**:
1) Gain familiarity with data analysis and visualization techniques using R
2) Explore the utility and limitations of standard linear regression models
3) Apply and interpret generalized linear models and generalized additive models
4) Implement classification and regression trees for exploratory data analysis

5) Examine and apply different model diagnostics to help choose an appropriate modeling strategy for a given analysis

**Procedures**:

Lecture/Lab/Discussion

The course will be a mixture of lectures, interactive labs, and discussion of primary literature. Grading will be based on written assignments.

- Course Assignments

At the end of lectures you will be assigned short problems to develop practical skills using R. Also, at the end of specific sections in the course I will hand out synthesis assignments designed to assess your knowledge of the statistical material presented. Synthesis assignments will be required to have a scientific format with an introduction, methods, results, and discussion. For most of these assignments, I ask that you break into groups of two students. Students are likely to have varying backgrounds and experience with statistics and R. Consequently, I want to emphasize the importance of working collaboratively on these assignments. If you have a strong background in stats, partner with someone with less experience. If you have experience scripting in R, partner with someone who has less. This will facilitate the learning experience for the course as a whole. Also, I require that you rotate through partners over the course of the semester.

- Grades (100 total points)

There will be 4 major assignments given out over the course of the semester. The first three will be group projects worth 20 points each. Each group of two students will turn in a single assignment and will receive the same grade for that assignment. The final assignment is worth 40 points and will be a synthesis project that allows students to use their own data or to work with a sample dataset given out by the instructor.

Software

R – a language and environment for statistical computing and visualization. For those of you with previous coding experience, scripting in R should be nothing new. To those of you with no previous experience in command-line environments, this will be your first introduction to scripting. R is open-source software so it is free. It has become the standard used in the biological sciences for data analysis and visualization. If you plan to continue in this line of work knowledge of R may be the single most important practical skill you take away from this course. There are copious amounts of documentation available for R. In addition to the handouts we provide, we suggest the intro manual located at:

R Intro Manual

**Data**:

A number of datasets will be made available to you for use in labs and for use in the assignments. Some of these datasets are public domain, and some are proprietary or have value added and represent the work of many individuals. You do not have open access permission to these data beyond work conducted in this course. In other words, *you cannot use the data for any other purpose without*

*permission from the instructor or persons or agencies that steward the respective datasets.* We will provide as much information as possible as to the provenance and availability of datasets.

**Schedule**:

| week | date | Lecture/discussion topic | Instructor | Synthesis Assignment |
|---|---|---|---|---|
| 1 | Aug 17 | Introduction<br>Introduction to R | Dobrowski | |
| 2 | Aug 24 | R Syntax and scripting,<br>Univariate data analysis,<br>Simple Linear Regression | Dobrowski | |
| 3 | Aug 31 | Bootstrapping, resampling, empirical distributions | Dobrowski | 1.Univariate data analysis and LM |
| 4 | Sept 7 | Factors/ Factor response<br>Exploratory analysis | Dobrowski | |
| 5 | Sept 14 | Classification and regression trees | Dobrowski | 2. CART and exploratory data analysis |
| 6 | Sept 21 | Multiple linear regression & regression diagnostics | Dobrowski | |
| 7 | Sept 28 | Development and applications of linear regression models | Dobrowski | |
| 8 | Oct 5 | Generalized Linear Models (GLMs) – motivations and mechanics | Dobrowski | |
| 9 | Oct 12 | GLMs – deviance and diagnostics | Dobrowski | |
| 10 | Oct 19 | GLMs – logistic models | Dobrowski | 3. GLM |

| week | date | Lecture/discussion topic | Instructor | Synthesis Assignment |
|---|---|---|---|---|
| 11 | Oct 26 | GLM – overdispersion | Dobrowski | |
| 12 | Nov 2 | Smoothers/ Generalized Additive Models | Dobrowski | 4. Synthesis assignment |
| 13 | Nov 9 | Smoothers/ Generalized Additive Models | TBD | |
| 14 | Nov 16 | Model selection/ validation | TBD | |
| 15 | Nov 23 | Final presentations | Dobrowski | |

***COVID related information***

I am structuring this course such that students can choose to engage in this course in a strictly online fashion. For students who would benefit from in person interaction, I will provide opportunities to do so during our scheduled meeting times. Due to the infectious nature of COVID-19, I require that students wear appropriate face masks during all in-person meetings.

***Students with Disabilities***

- The University of Montana assures equal access to instruction through collaboration between students with disabilities, instructors, and Disability Services for Students (DSS). If you think you may have a disability adversely affecting your academic performance, and you have not already registered with DSS, please contact DSS in Lommason 154 or 406 243 2243. I will work with you and DSS to provide an appropriate accommodation.

***Student Conduct Code***

- All students must practice academic honesty. Academic misconduct is subject to an academic penalty by the course instructor and/or a disciplinary sanction by the University. All students need to be familiar with the Student Conduct Code.