FORS 538 Statistical Models for Ecological Data Analysis

Fall 2017: 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: Tues 1-2:30; or by appointment

David Affleck
Office: CHCB 405
Phone: 4186
Email: david.affleck@umontana.edu
Office hours: Fri 1:00-2:30 pm.

Course Description:
This is an ***applied*** course covering advanced statistical modeling techniques using examples from forestry, ecology, and the environmental sciences. We 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, we will explore various parametric and semi-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., multiple 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 we 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

We 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 with using R. Also, at the end of specific sections in the course we 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, we ask that you break into groups of two students. Students are likely to have varying backgrounds and experience with statistics and R. Consequently, we 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, we require that you rotate through partners over the course of the semester.

- Grades (120 total points)

There will be 5 synthesis assignments given out over the course of the semester. Each is worth 20 points. Each group of two students will turn in a single assignment and will receive the same grade for that assignment. There will also be short problems handed out weekly worth 1-2 points each.

- 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 is quickly becoming the standard that is being 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 28	Introduction Introduction to R	Dobrowski/ Affleck	
2	Sept 4	R Syntax and scripting, Univariate data analysis, Simple Linear Regression	Dobrowski	
3	Sept 11	Bootstrapping, resampling, empirical distributions	Dobrowski	1.Univariate data analysis and LM
4	Sept 18	Factors/ Factor response Exploratory analysis	Dobrowski	
5	Sept 25	Classification and regression trees	Dobrowski	2. CART and exploratory data analysis
6	Oct 2	Multiple linear regression & regression diagnostics	Affleck	
7	Oct 9	Development and applications of linear regression models	Affleck	3. Multiple linear regression and diagnostics
8	Oct 16	Generalized Linear Models (GLMs) – motivations and mechanics	Affleck	
9	Oct 23	GLMs – deviance and diagnostics	Affleck	

week	date	Lecture/discussion topic	Instructor	Synthesis Assignment
10	Oct 30	GLMs – logistic models	Affleck	
11	Nov 6	GLM – overdispersion	Affleck	4. GLM
12	Nov 13	Smoothers/ Generalized Additive Models	Dobrowski	
13	Nov 20	Smoothers/ Generalized Additive Models	Dobrowski	
14	Nov 27	Model selection/ validation	Dobrowski	5. Synthesis assignment
15	Dec 4	Advanced topics	Dobrowski/ Affleck	