

## *Course Syllabus*

# **Analysis of Environmental Data**

*Course number:* ECO602 and ECO634

*Instructor:* Dr. Kevin McGarigal (mcgarigalk@eco.umass.edu; 577-0655)

## **Description**

This course provides students with an understanding of basic statistical concepts critical to the proper use and understanding of statistics in environmental conservation and prepares students for subsequent ECo courses in statistical modeling. The lecture (required for all ECo Master's level graduate students) covers foundational concepts in statistical modeling; emphasis is on conceptual underpinnings of statistics not methodology, with a focus on defining statistical models and the major inference paradigms in use today.

The lab (optional for ECo graduate students) introduces the statistical computing language R and provides hands-on experience using R to screen and adjust data, examine deterministic functions and probability distributions, conduct classic one- and two-sample tests, utilize bootstrapping and Monte Carlo randomization procedures, and conduct stochastic simulations for ecological modeling. Specifically, lab focuses on the following:

1. *R programming language and statistical computing environment.*—Learning the R language and statistical computing environment, which serves as the computing platform for all ECo statistics courses; emphasis is on learning fundamental R skills that will allow students to grow and expand their expertise in subsequent courses or on their own.
2. *Basic statistical methods.*—Using R to conduct basic statistic procedures, including data screening and adjustments, classic one and two sample tests, resampling methods and stochastic simulations.

## **Objectives**

The overall goal of this course is to provide students with a gentle introduction to the world of statistics in ecology and conservation. The specific objectives are for students to:

- Gain broad understanding of the conceptual underpinnings of statistics in ecology and conservation, the major issues and pitfalls associated with study design, and the key distinctions among statistical methods or classes of methods commonly used in ecology and conservation;
- Work and learn in an interdisciplinary environment; specifically, to develop problem-solving skills in an interdisciplinary team environment;
- Engage in active, student-directed learning in preparation for professional life; and
- Refine their written and oral communication skills.

## **Guiding Philosophy**

The course objectives listed above largely dictate the teaching and learning method used in this course. Specifically, this course is strongly focused on project-based, student-directed learning. Consequently, learning new material from assigned readings and regurgitating it on exams is de-emphasized. Rather, the emphasis is on sharing knowledge and expertise (both pre-existing and newly acquired) with other students (in a small group environment) to address contemporary issues in ecological modeling. The time and energy devoted to this

course is largely in the form of group analysis of problem sets and discussion of those results in reference to the concepts learned from lecture and assigned readings, not studying for exams.

### **Course Format**

The course is logically divided into two components: lecture (required) and lab (optional). The lecture consists of two 75-minute lectures each week covering the material provided, including assigned readings, plus an optional 2-hour meeting each week to review lecture materials and go over the take-home problem sets. The (optional) lab consists of a single 3-hour lab section each week in which students work in small groups to complete exercises. The purpose of the lab section is to provide students with hands-on experience analyzing real data sets using the R software. In addition to the regularly scheduled lecture and lab meetings, special meetings (to be arranged) may be required for presentation and discussion of lecture and lab projects (see below). These special sessions will be scheduled based on student availability; however, the instructor reserves the right to designate a time slot for these meetings if a common time cannot be agreed upon by all students.

### **Student Responsibilities**

Students are responsible for fully understanding all of the information presented in this syllabus. If there are any questions regarding this information, it is the student's responsibility to bring it to the instructor's attention before the first graded assignment. In addition, students are responsible for attending class (including lectures, labs and special meetings), actively participating in class and group discussions, and completing all assignments. Students are expected to contribute equally to all group projects. In addition, students are required to meet outside of class with group members to complete lecture/lab projects. Students are expected to meet with their groups outside of class for at least 3 hours per week. Failure to do any of these things may result in failure of the course (see Grading System below). Any assigned reading material is to be read before the appropriate class/lab session. Students are responsible for asking questions anytime they need clarification (remember, there is no such thing as a bad question, only a poorly timed one). Finally, and most importantly, students are responsible for sharing their views, perspectives, opinions, and experiences with the class. Each student brings to the class a unique world view that has been shaped by their personal experiences and observations. By sharing this world view, each of us will develop a broader and more enlightened world view and, as a result, develop a better understanding of how to apply statistics in ecology and conservation.

### **Attendance Policies**

Students are expected to attend all lectures, labs and special meetings for group presentations, as appropriate. In addition, students are encouraged to attend all discussion sections and complete all take-home assignments. Although attendance is not recorded directly, students are graded on lecture attendance indirectly as described below (see Grading System) and are graded on their participation in all lecture/lab projects by their peers (see Grading System).

### **Accommodation Policies**

The University of Massachusetts Amherst is committed to providing an equal educational opportunity for all students. If you have a documented physical, psychological, or learning disability on file with Disability Services (DS), Learning Disabilities Support Services (LDSS), or Psychological Disabilities Services (PDS), you may be eligible for reasonable academic accommodations to help you succeed in this course. If you have a documented disability that requires an accommodation, please notify the instructor within the first two weeks of the semester so that appropriate arrangements can be made.

### **Academic Dishonesty Policy**

It is expected that all graduate students will abide by the Graduate Student Honor Code and the Academic Honesty Policy (available at the Graduate Dean's Office, the Academic Honesty Office (Ombud's Office) or online at [http://www.umass.edu/gradschool/handbook/univ\\_policies\\_regulations\\_a.htm](http://www.umass.edu/gradschool/handbook/univ_policies_regulations_a.htm)). All students have the right of appeal through the Academic Honesty Board. Any form of academic dishonesty (e.g., plagiarism, cheating, fabrication) will result in a score of zero on that grading item. A second violation will result in an 'F' for the course and potential expulsion from the graduate program. To further eliminate any potential for cheating, grading is in part by peers, as described below (see Grading System).

### **How to get Help**

The instructor and Teaching Assistant (TA) are available to help students in all aspects of the course. In the true spirit of interdisciplinary teamwork, it is recommended that students seek information and assistance first from their teammates on a particular assignment or lab project. If this is unsatisfactory, students can contact the instructor or TA during the regularly scheduled office hours. If students are unable to meet during that time period, they can arrange to meet the instructor or TA individually by appointment.

Name: Dr. Kevin McGarigal  
Office Hours: Tues/Thurs 10:45-12:00 noon  
Office #: Holdsworth 304  
Phone: 577-0655  
Email: [mccgarigalk@eco.umass.edu](mailto:mccgarigalk@eco.umass.edu)

TA: TBD

### **Grading System**

This course utilizes *project-based* learning and *student-directed* learning as the primary means of instruction and student evaluation. Consequently, grading is based largely on student participation and performance in student-directed projects. Grading is based on several items; each item is described in detail below.

Lecture Component:

Grading Items	Points	Percentage of course grade
Daily quizzes	26	26%
Take home problems	24	24%
Critical review	20	20%
Final take home	30	30%

1. Daily Quizzes.—Class attendance and participation is gauged by daily, in-class quizzes. These are short one- or two-question quizzes completed in class and cover material from the preceding class. Each quiz is worth 2 points. The cumulative score on quizzes is adjusted to a 26 point maximum for this grading item.
2. Take home problems.—In this series of 9 take home problems, students work in groups to answer questions pertaining to each major lecture topic. There is a take home problem set for each major topic covered in lecture that serves to reinforce the lecture material and provide real-world examples of environmental data analysis. Each problem set is worth 3 points. Students are expected to complete 8 of 9 problem sets (i.e., one problem set can be missed) for a maximum of 24 points. However, points earned in excess of 24 (max of 27), which is feasible by successfully completing all 9 problem sets, will be counted as “extra credit”.
3. Critical review of paper.—In this take home project, students work in groups to read, discuss and critically evaluate a scientific paper with respect to the foundational concepts discussed in this course. Students gain practical understanding of the design and analysis of a real data set and experience critically evaluating the consistency and appropriateness of the statistical approach adopted. The critical review paper is worth a maximum 20 points.
4. Final take home exam.—In this final exam, students work individually to answer a series of questions pertaining to the material covered in the entire course. The questions require synthesis of the course material and are similar in nature to take-home problems. The final exam is worth a maximum of 30 points.

#### Lab Component:

Grading Items	Points	Percentage of course grade
Project 1	15	15%
Project 2	15	15%
Project 3	15	15%
Project 4	15	15%
Project 5	40	40%

1. Project 1.—*Deterministic functions*.—In this project, groups learn how to examine a range

- of deterministic (or mathematical) functions in R by way of examples and learn the skills needed to develop and/or examine other functions. For the assignment, groups examine a real data set and find at least three alternative mathematical functions to describe the apparent relationship between the independent and dependent variable.
2. Project 2.–*Probability distributions.* In this project, groups learn how to examine a range of common probability distributions in R by way of examples and learn the skills needed to examine other distributions. For the assignment, groups examine a real data set and determine whether any of the existing probability distributions are appropriate for describing the error component of a parametric statistical model designed to describe the apparent relationship between the independent and dependent variable.
  3. Project 3.– *Classical tests.*–In this project, groups learn how to conduct a wide range of classical statistical tests for single-sample and two-sample problems in R. For the assignment, groups conduct a battery of tests on several real data sets.
  4. Project 4.–*Resampling methods.*–In this project, groups learn how to resample data in R for various purposes, focusing on the bootstrap and randomization tests. For the assignment, groups use bootstrapping to compute rarefaction curves for a breeding bird data set and construct a randomization test for the effect of understory treatment on tree seedling density.
  5. Project 5.–*Stochastic simulation.*–In this project, groups learn techniques and ideas related to simulating ecological patterns in R. The main goals are to learn how to generate patterns in order to sharpen intuition and test estimation tools, learn how to estimate statistical power by simulation, and with an example illustrate the flexibility of R for simulating ecological patterns and processes. For the assignment, groups construct a power analysis for a linear trend for a hypothetical observational study involving the date of flowering in a spring annual.

Based on the cumulative score earned on all grading items, the overall course grade (for lecture and lab separately) is determined as follows:

Cumulative Score	Grade
94-100	A
90-93	A-
84-89	B
80-83	B-
74-79	C
<70	F

### General Guidelines for Assignments

With the exception of the daily quizzes and final take home exam, students are self-organized into small groups to complete the graded assignments. Each group prepares a single report.

Every member is responsible for contributing equally to the content of the report. The report must be well organized and presented in a clear and concise manner.

Group assignments are graded according to the following procedure. First, the entire group receives a score (out of a maximum of 100%) based on information content (75%) and presentation (25%) by the course instructor and teaching assistant. MOST IMPORTANTLY, reports must reflect a single voice; i.e., the report must not consist of several separate efforts combined at the end for purposes of the presentation. Moreover, the concepts discussed in lecture must be integrated into the discussion to the extent possible. These are the major challenges and largely determine the grade received for the assignment. Second, each member of the group is evaluated by their peers within the group. Upon completion of the assignment, each member grades the contribution of each individual to the group effort and assigns a score (out of a maximum of 100%) to each individual. The final score for each individual is the average of the instructor and peer scores.

## Readings

The readings consist largely of the lecture notes prepared by the instructor and the recommended course textbook (Bolker 2008). The textbook is strongly recommended for those students taking the lab and those intending to pursue additional training in statistics (as are the additional reference books).

### Recommended textbook:

- Bolker BM. 2008. Ecological Models and Data in R. Princeton University Press, Princeton, NJ. 396 pp.

### Additional reference books:

- Crawley MJ. 2007. The R Book. John Wiley & Sons, Ltd., West Sussex, England. 942 pp.
- Hilborn R, and M Mangel. 1997. The Ecological Detective: Confronting Models with Data. Princeton University Press, Princeton, NJ. 315 pp.
- Quinn G, and M Keough. 2002. Experimental design and data analysis for biologists. Cambridge University Press. 537 pp.

## Lecture Schedule

The following lecture schedule remains subject to change and last minute modifications. In addition, individual listed topics may require multiple class sessions to cover.

No.	Topic	Readings
1	Role of statistics in environmental research	Chapter 1
2	Environmental data	Notes
3	Exploratory analysis	Chapter 2
4	Deterministic functions	Chapter 3

5	Probability distributions	Chapter 4
6	Inference frameworks (frequentist, likelihood and Bayesian)	Notes
7	Nonparametric least squares inference	Notes
8	Maximum likelihood inference	Chapter 6
9	Bayesian inference	Chapter 6
10	Stochastic simulation	Chapter 5
11	Synthesis and summary	Notes

---

## Lab Schedule

The following lab schedule remains subject to change and last minute modifications.

---

No.	Topic
1-2	Introduction to R: <ul style="list-style-type: none"> <li>• What is R?</li> <li>• Installation of R</li> <li>• A Few Important Syntax Conventions in R</li> <li>• Variables and Types</li> <li>• Data Structures</li> <li>• R Operators</li> <li>• Creating Subsets of a Matrix or Data Frame</li> <li>• Row or Column Operations on a Matrix or Data Frame</li> <li>• Functions in R</li> <li>• Getting Data Into and Out of R</li> <li>• Plotting in R</li> <li>• Getting Help in R</li> <li>• Libraries and Packages</li> <li>• Tutorials for Learning R</li> </ul>
3-4	Data screening and adjustments: <ul style="list-style-type: none"> <li>• Summary statistics (e.g., means, standard deviations, quantiles)</li> <li>• Missing data (e.g., single variable and multi-variable imputation)</li> <li>• Frequency of occurrence and abundance plots</li> <li>• Dropping variables</li> <li>• Single variable distributions</li> <li>• Relationships between pairs of variables</li> <li>• Outliers</li> <li>• Data transformations</li> <li>• Data standardizations</li> <li>• Dissimilarity matrices</li> </ul>

- 4-6      Deterministic functions:
- What is a deterministic (mathematical) function
  - Linear function – Oregon birds example
  - Logistic function – Oregon birds example
  - Ricker function – Stripped bass example
  - Bestiary of deterministic functions
- 7-8      Probability distributions:
- What is a probability distribution?
  - Plotting distributions
  - Bestiary of probability distributions
- 9-10     Classical tests:
- Single sample
  - Two samples
- 11-12    Resampling procedures:
- Bootstrap
  - Randomization tests
- 13-14    Stochastic simulation:
- Simulating static ecological processes
  - Simulating dynamic processes – population matrix model
-