Research Computing in Evolutionary Anthropology
ANTH 560 and BIOL 519
Spring Term 2014

Instructors: Bret Behiem PhD, Jeffrey Long PhD
Office: Bldg Rm To Be Announced (BB), Anthropology 156 (JL)
Email: babeheim@unm.edu
jlo@unm.edu

Time and place – Tuesday 2:00 – 4:45, Mitchell Hall 216.

Office hours – Bret Beheim, Jeffrey Long by appointment.

Description

Computationally intensive methods now play a central role in many aspects of scientific research. To use such methods requires skills in three distinct domains: computer programming, mathematics, and data analysis. The philosophy of this course is that it is best to learn these skills in a simultaneous integrated fashion. This breaks from a tradition of teaching them in isolation, often in separate classes, and in different departments. This course will introduce students to writing original programs, modeling biological process, and large-scale data analysis. The four following topic areas comprise the body of this course. We will develop each new topic area using the skills and concepts learned in the previous topic areas.

- *Programming and Data Manipulation*: Students will learn to write programs using the package R, which is a language and environment for statistical and graphical analyses. R is a free download that is capable of performing the tasks included in industry standard packages such as SPSS and SAS. R has become the *lingua franca* in the field of informatics, and an interactive researcher community contributes tools and modules to the package.

- *Probability Concepts and Applications:* Students will study basic probability, including Bayes Theorem, random variables, expectation, and biologically useful distributions such as the Beta, Binomial, Exponential, Gamma, Geometric, Normal, Poisson, and Uniform. The students will be introduced to probability models for biological processes. We will explore the Poisson process and Markov chains.

- *Stochastic Simulation:* Students will learn to use (pseudo) random numbers to generate samples from the Uniform distribution. Then, they will learn to generate samples from other specified distributions using methods such as inverse functions, rejection sampling, importance sampling, and Markov chain Monte Carlo.

- *Computational Statistics:* The students will learn to conduct analyses in Likelihood and Bayesian formats. They will learn to use the concept of information for judging the success of statistical models.

Pre-requisites

Graduate standing is required. An undergraduate or graduate course in applied or mathematical statistics will be helpful. The instructors will assume that all students have some knowledge of descriptive statistics and standard tests such as ANOVA and linear regression. All students must have access to a laptop computer for in-class use.

Course Materials

The required text is *Ecological Models and Data in R* by Benjamin Bolker. There are two recommended books for this course – *Introductory Statistics with R* by Peter Dalgaard and *R Graphics* by Paul Murrell. All three books are excellent and relatively inexpensive.

Grades

Students will complete weekly problem sets, take in-class exams, and prepare a data analysis project. Learning computational methods requires immersion and practice.

Problem Sets – There will be weekly problem sets. These problem sets count 40% of your final grade. Students may work together on problem sets, but they must be submitted individually. Portions of each weekly class session are devoted to solving the problems and setting up analyses.

Exams – The in-class exams count 40% of the final grade. These exams will cover knowledge and skills that a student should be able to perform on the fly. Because each new topic area requires using the skills and concepts learned in the previous topic areas, it is essential to demonstrate mastery of the material at regular benchmarks.

Individual Project – The project counts 20% of the final grade. Follow the format of a scientific journal when writing up this project. The maximum number of words is 2,500. The paper should have tables and figures. You are required to attach an appendix to the paper with the scripts for your analyses. You may cite as many references as necessary. The tables, figures, and citations do not enter into your word count. The paper must include original data analysis, but you do not have to collect the data yourself. Simulating data is acceptable. Writing style will contribute to your grade. Perfect spelling and grammar are expected. The paper is due on April 4. We will return the paper to you with a critique on April 18. You have the opportunity to address the critique, revise, and re-submit. The resubmission date is May 2.

Presentation – Students will present their projects to the class in the final two class sessions. The students may present their projects as either a poster, or a slide show. The presentation counts toward the final grade of the project.

Month	Day	Topic	Bolker	Assignment Due	Exams
January	21	Intro to R Environment	Ch1		
	28	Programming in R	Ch 2	Problem Set #1	
February	4	Graphing in R	Ch 3	Problem Set #2	
	11	Basic Probability I	Ch 4		Exam 1
	18	Basic Probability II	Ch 4	Problem Set #3	
	25	Random Numbers & Simulation I	Ch 5	Problem Set #4	
March	4	Simulation II - Rejection	Ch 5	Problem Set #5	
	11	Markov Chains and MCMC			Exam 2
	18	Spring Break			
	25	Maximum Likelihood	Ch 6	Problem Set #6	
April	1	Bayesian Inference	Ch 6	Problem Set #7	
	8	Comparing Models			Exam 3
	15	Mixed Models		Problem Set #8	
	22	MCMC and Gibbs		Problem Set #9	
	29	Project Presentations			
May	6	Project Presentations			