ZOL/PLB 851

STATISTICAL METHODS IN ECOLOGY AND EVOLUTION

(FORMERLY QUANTITATIVE METHODS IN ECOLOGY AND EVOLUTION)

Fall 2014

Instructor: Dr. Ian Dworkin

Office: 328 Giltner Hall
e-mail: idworkin@msu.edu

Office hours: Tuesdays 12:30-1:30
Other times by appointment

Class Info: Tuesday, Thursday 10:20am –12:10pm
115 Berkeley Hall

Optional class sessions:
In addition to the regularly schedule classes, I may have optional R computer sessions to help people get started with R. Times and dates TBA. I may also set up "master classes" later in the semester for people who want to explore particular topics in more detail. If possible bring your laptops to these sessions.

The Course

Course Description
Interpretation and analysis of data in ecology and evolutionary biology. Statistical computer software. (3 credits)

Recommended Background
STT465 (Statistics for Biologists II) **&** STT814 or an equivalent course (CSE845).
I assume that you are familiar with regression, ANOVA, ANCOVA & have a basic grasp of the utility of probability distributions and are at least vaguely familiar with some of them (normal, binomial, t-distributions, χ^2, F distributions). While this course will not deal with many formal mathematical proofs, I do assume a certain level of comfort with algebraic concepts and a dim recollection of derivatives (we won't be doing any calculus, but we may look at a few derivates with respect to maximum likelihood). Some basic linear algebra is also useful.

In particular the material that you need to be somewhat familiar with (even if it is rusty):

A) Basics of probability (in particular some recollection of conditional probability)
B) Random variables, what are they ? how do we use them in statistics?
C) Basic understanding of some probability distributions (normal/Gaussian, binomial , poisson, t, chi-square, F distribution). Also some understanding of PDFs and CDFs in relation to the above.
D) Familiarity with ANOVA's, how to set up a basic model, and some recollection of sum of squares, mean squares and F ratios
E) Regression analysis and interpretation and maybe some dim idea of multiple regression
F) Analysis of covariance

G) Maybe a dim recollection of how ANOVA's, ANCOVA and regressions are really all part of the family of general linear models (GLM).
H) While it is no way required if you remember even basic matrix algebra this would not hurt (although we won't be using it a lot). Ditto for derivatives and integrals.

If you do not have the background as suggested above, I just want to make sure that you are aware well ahead of time so that you are not shocked in the course. I am happy to suggest some remedial reading, but be prepared for a very challenging semester.

Course Justification
Professional biologists interested in ecological and evolutionary questions inevitably deal with variable data. In fact variability is a fundamental concept in both ecology and evolution. The ability to manage, analyze and interpret data collected either through planned experiments or through opportunistic or monitoring programs represents a fundamental skill for all ecologists and evolutionary biologists. This course will further develop your skills and understanding of quantitative methods for the analysis, interpretation and presentation of ecological and evolutionary data. In particular we will focus on model building, and will spend a fair bit of time considering alternative approaches (Parametric, Likelihood, Bayesian and Resampling) for parameter (point and interval) estimation for statistical models, assessing uncertainty in these estimates and using these to make biological and statistical inferences. It is likely that through your career you will not only read about new techniques that you are unfamiliar with, but will often be confronted with data that requires you to learn new approaches to data analysis. I consider it one of the ***key goals of the course*** to help you develop skills to aid in this process.

Goals of this course:

1. Discuss the philosophical and historical context within which we collect, measure, analyze and interpret data.
2. Overview contemporary techniques for the analysis of ecological and evolutionary data, examining different approaches for both estimation and inference.
3. Provide a thorough introduction to model based approaches to analysis that allow for an easier approach to translating statistical inference to biological inference.
4. Introduce R, a powerful and useful programming environment for data analysis and presentation. Develop skills at writing R functions, with the goal of being able to perform advanced, computationally intensive analyses.
5. Provide hands-on experience analyzing data relevant to each student's particular field of study.
6. Provide you with sufficient and appropriate background to be able to teach yourself new methods for data analysis as needed.

It is worth pointing out that no single Statistics course can teach you everything. People do whole PhD's in each topic we cover, so this course introduces you to a lot of powerful approaches to help answer your biological questions using statistical inference. However, I can consider this course more about opening the door to these subjects, to aid you as you develop your skills; not as the "final" statistics knowledge you will ever need.

What we will NOT cover in this class: Given the limited time we have together I have decided to exclusively focus on univariate response statistical models. That is, models where there is a single "response" variable whose variation we are trying to explain, with one or more "explanatory" variables. While many multivariate techniques (such as MANOVA and other multivariate extensions of GLM, principal components analysis, factor analysis, path analysis and multi-dimensional scaling) are all extremely important methods, we simply will not have the time to cover them in this course. I will happily recommend books that are useful for multivariate approaches. I apologize for this (and I wish I could discuss multivariate methods, most of my own work utilizes them). We will also not discuss many of the statistical/machine learning approaches that have become popular recently, especially for high-dimensional problems with substantial non-linearity. This is for two important reasons. These methods require additional expertise in computational methods, and some slightly more complicated mathematical techniques, and also most of these techniques (support vector machines, neural networks, bagging and boosting for example) are extremely useful for discovery, classification and prediction, but far less useful for hypothesis testing and to explain scientific patterns. The other extremely important topic that we will not formally cover (in lecture) is experimental design, which I am **ASSUMING** that students are fairly comfortable with. Although we will discuss experimental design for some particular examples. The rationale for this is that STT814 covers a wide variety of experimental designs, and in addition, many ecologists also need to learn how to deal with observational data, not just planned experiments.

R as a Statistical Platform for the Course

In addition to providing general information on quantitative approaches, I will also provide specific instruction on the use of one particular platform for data management, analysis and graphical presentation. There are many commercial statistical software packages available (e.g. SAS, SPSS, STATISTICA, STATA, S-PLUS) and each has its own pros and cons. The statistical programming environment that we will be using is called R. This statistical computing and graphics package is widely used in statistical methodological research and its use in ecology and evolutionary biology is increasing. One of the advantages of using R is that it is freely available online (www.r-project.org/), runs on a variety of operating systems (Windows, Mac OS X, Linux), which has encouraged the development of many sophisticated statistical packages and an active and generous R-help message board. You will need to download the most recent version of R onto a personal computer, or use one of the microcomputing facilities on campus to complete your assignments. The most recent version is currently version 3.1.1 (*Sock it to me*). We will use this version for the entire semester, so download it and please do not update during the semester. When discussing specific R commands in R during lecture I will provide electronic copies of the code so that you can follow along more easily and use the material as needed. ***Please install this version of R onto whatever computer you plan to use for your analyses.*** You may also consider using R Studio (http://www.rstudio.com/). This has some benefits for both version control (using git and github) a powerful integrated development environment (IDE) and makes it easy to generate reports in many formats (markdown, html and PDFs being the primary ones). The only downside is I do not use it R studio at all, and some of its features are a bit foreign to me, so I will not be able to help you with those particular aspects. i.e. I will plead ignorance on anything that is clearly an R Studio issues.

Coding in R: Given how large the class size now is, and the fact that each student will be largely working with their own dataset, I will be imposing a strict requirement for an R style guide, which I will share on the first day of class. I expect that **ALL R scripts** associated with assignments from each student will be written following the style guide. This will make it far easier for me to be able to read and evaluate your R code.
https://www.msu.edu/~idworkin/ZOL851_style_guide.html

Version Control: We will spend some time this semester discussing the issues of version control and reproducible research (and how to combine markdown or LaTeX with R code into single documents). We will be using the git version control system (http://git-scm.com/). There are a number of ways (from command line, which is how I use it to very simple GUIs). We will likely be using the free service github, so please get yourself an account there (https://github.com/).

Your Dataset

It is often easiest to grasp new concepts through their application to familiar problems. In addition to discussing lecture examples from ecological and evolutionary literature, you will also analyze a dataset of your own. As a result, one of your first tasks will be to search out a dataset that you can analyze for your final paper (see below). This can be data that you collected yourself or a dataset that you have acquired from a supervisor or colleague that is related to your thesis (in a general or specific way). It is to your advantage to uncover a dataset relevant to your unique interests but if you are unable to do so, I will provide you with one.

Some required features for your data set:
At least 50 observations/samples
The response variable should be continuous
At least one categorical predictor/explanatory variable
At least one continuous predictor/explanatory variable
At least one other predictor (either categorical or continuous)
(optional) a dichotomous response variable (for assignment 5/6)

Grading

This course will be graded on a point system as outlined below. Grades may be scaled up if results for the class are low, but they will not be scaled down.

4.0	90-100 points	2.0	70-74 points
3.5	85-89 points	1.5	65-69 points
3.0	80-84 points	1.0	60-64 points
2.5	75-79 points	0	< 60 points

Grading for the course will be based on:

Item	**Points**	**Due Date**
Assignment 1	5	TBD
Assignment 2	5	TBD
Assignment 3	10	TBD
Assignment 4	10	TBD
Assignment 5	10	TBD
Assignment 6	10	TBD
Final Paper	40	TBD (Exam week)
Class Participation	10	
Total	100	

My Policy on Due Dates

Assignments will be due at the start of class on the date in which they are due. I appreciate that graduate students have many responsibilities in addition to course work. Due dates for assignments can be adjusted individually with prior approval (at least two weeks prior to due date). Grades for late assignments will be deducted at a rate of 10% per day. **The Final paper cannot be handed in late** (since grades are due 3 days later).

Assignments

Assignments are designed to ensure that you have practical working knowledge of the statistical approaches discussed in lecture and to provide you with experience presenting the results of statistical analyses. There will be 6 assignments during this course and each will cover a set of concepts discussed in lecture. An example dataset (if necessary) and relevant biological background or a list of required readings as well as specific instructions will be provided separately for each assignment. Most assignments will consist primarily of annotated output from R, but you will also be expected to present the results of analyses in text, figures or tables and to apply concepts and techniques from the assignments to your own dataset. Assignments are to be completed independently, unless otherwise advised. ***You are expected to work on your assignments and final paper independently unless otherwise advised. No help from other members of your lab, your PI or a statistical consultant on this. If this happens, this will result in a failed assignment***.

Final Paper

At the end of the semester you will submit a final paper of your own, which will resemble a published paper, including an Introduction, methods section, results, discussion and any relevant figures and tables, but will also include an appendix with annotated output from R. The statistical techniques used in your study will come from class, but the specific biological questions and approaches will be entirely determined by you. Your goal should be for this final paper to form a solid portion of a thesis chapter or manuscript for publication.

Participation grade
My goal is to provide a stimulating and interactive learning environment but I cannot do this entirely on my own. It is my belief that students, and graduate students in particular, provide an important component of the learning environment for others in the class. As a result, your participation grade will reflect your contribution to the overall learning environment in the classroom. Clearly you cannot contribute positively to such an environment if you are absent, and overt disruptions of others will be penalized. Full points will be given only to students who communicate insightful questions, perspectives and comments, and encourage similar contributions from others in the class. In addition part of your participation mark will depend on completing the class surveys at the beginning and end of the semester. Completion of the class surveys is also considered part of the participation grade.

Course Evaluation: Michigan State University takes seriously the opinion of students in the evaluation of the effectiveness of instruction, and has implemented the SIRS (Student Instructional Rating System) process to gather student feedback. This course utilizes the "online SIRS" system, and you will receive an e-mail sometime during the last two weeks of class asking you to fill out the SIRS web form at your convenience. As a further reminder for you to fill out the SIRS evaluation form, the final grade for this course will not be accessible on STUINFO during the week following the submission of grades unless you respond to the SIRS web site. You have the option on the online SIRS form to decline to participate in the evaluation of the course – we hope, however, that you will be willing to give us your frank and constructive feedback so that we may instruct students even better in the future.

What you should do if you are going to miss one or more classes: Given the size of the class, my suggestions are as follows;

1) Email me and let me know when you are missing class.
2) Talk with a classmate to arrange to get notes, readings assignments, and if at all possible **an audio recording of the lecture**. I talk **A LOT**, and far too quickly (but feel free to ask me to slow down). My power point presentations are really just a sketch, and I fill in a lot of the details verbally or at the chalkboard. Therefore I highly recommend you get a chance to listen to the lecture while going through the lecture notes and R scripts.

Class Outline

Date	Topic	Date	Topic
August 28th	Outline of course and objectives. Introductions. Intro to R? Bolker: Chapter 1 Chamberlin, Platt, Stephens	Sept. 2	Approaches to science and statistics: Intro to R_1 Dalgaard: Chapters 1 & 2 (R) Bolker: Chapter 2 (R) QuinnDunham, Loehle, Crawley: Chapter 2 (R)
Sept. 4	Measurement Theory: Intro to R_2. Crawley: Chapter 2 (R) Crawley: Chapter 8, pages 279-315. Houle ?, Guthery, Shrader-Frechette, StephensEtAl, Lukacs	Sept. 9	Exploratory data analysis in R. Effect Sizes Bolker: Chapter 2 GelmanHill: Appendix B. DworkinEtal2005 (optional, we will be using this data set a lot in class)
Sept. 11	Exploratory data analysis in R; Presentation of Data and Results using R. Dalgaard: Chapter 10 (working with data in R). ZuurA, Wainer Crawley: Chapters 3,4	Sept. 16	Introduction to GLM: ANOVAs and regression as one big happy family. Dalgaard: Chapters 6,7,11,12 GelmanHill: Chapters 3,4, Appendix A R_intro_guide chapter 11 (a lot of useful advice for R syntax for using lm()).
Sept. 18	General Linear Models – ANCOVA, assumptions and diagnostics. Interactions, SS Types. Crawley: Chapter 9 Chapter 10:388-417 Chapter 11:449-486 Chapter 12: 489-509 Packard_Boardman, mcArdle, Garcioa-Berthou, Jackson	Sept. 23	A Biologists field guide to probability. Dalgaard: Chapter 3,4. Bolker: Chapter 4. GelmanHill: Chapter 2 Grinstead (this is a probability book, which is mostly for your reference and pleasure reading).
Sept. 25	The Utility of Probability distributions	Sept. 30	Simulations for power analyses and inference. Bolker: Chapter 5 GelmanHill: Chapters 7-8 Hoenig, Thomas, Steidl.
Oct. 2	Simulations for power analyses and inference. Dalgaard: Chapter 9 Bolker: Chapter 5	Oct. 7	Resampling methods for estimation and inference. Moore (this is a very introductory and gentle introduction if needed). Crawley: 284, 287, 418-421 Fox_appendix, Crowley
Oct. 9	Maximum Likelihood Estimation - $L(\theta \mid y) \propto P(y \mid \theta)$ Bolker: Chapter 6	Oct. 14	Maximum Likelihood Estimation & inference Bolker: Chapter 6

	Blume 2002		
Oct. 16	Maximum Likelihood Estimation & inference Bolker Chapter 7:222-232 TBA	Oct. 21	Bayesian philosophy & estimation I – $p(\theta\|y)$
Oct. 23	Bayesian estimation & MCMC- an introduction Bolker: Chapter 7 (the remaining bits)	Oct. 28	Estimating Natural Selection Mitchell-Olds; Conner 89, Conner96
Oct. 30	Model Selection I $P(\mathbf{y}\|\boldsymbol{\theta}, \text{model})$ $\boldsymbol{L}(\boldsymbol{\theta}\|\mathbf{y}, \text{model})$ Burnham2001;Whittingham2006; HobbsHilborn2006.	Nov. 4	Model Selection II , introduction to Generalized linear models (GLiM) Bolker: Chapter 9. JohnsonOmland, Anderson2000
Nov. 6	GLiM 2 – logistic regression Dalgaard: Chapter 13 Bolker: Chapter 9. GelmanHill: Chapter 5	Nov. 11	GLiM 3 – poisson, quasi-poisson, negative binomial. Dalgaard: Chapter 15 GelmanHill: Chapter 6 Zeleis; Potts; Richards; Hoef
Nov. 13	Make your own statistics: Process models Bolker: Chapter 3,8	Nov. 18	TBD
Nov. 20	Make your own statistics: Process models: Population Growth models; Population genetic models.	Nov. 25th	Mixed/Random Effects Models Bolker: Chapter 10 GelmanHill: Chapters 11,12 Cressie2009; Wilson2010; Bolker 2008
Dec. 2	Mixed/Random Effects Models GelmanHill: Chapter 13,14	Dec. 4	Mixed/Random Effects Models GelmanHill: Chapter 18. Crawley: Chapter 19. Hadfield CourseNotes MCMCglmm.

Notes on readings: background readings (stuff you should already know but may have forgotten). Required text readings. Required PDF readings. Advanced readings (not required).

Readings

Primary Literature

Just as you need to keep yourself abreast of the developments in your particular field of study (from theory to important empirical studies) it is equally important to keep informed with developments in the statistical methods that inform your area of expertise. Readings from the primary literature are designed to supplement the material presented in lecture and will increase your general understanding and facilitate discussion. All readings will be made available through ANGEL in the appropriate folder for that topic.

Highly Recommended or required Texts – Most of the topics we cover with be associated with these texts. It is worth your while to have these books. However, in general we will not be working directly out of the texts, and different ones may suit you. Please note that several of the books listed below are available as digital copies via the MSU library, links to these E-books can be found in the ANGEL resources section.

Bolker, B. 2008. *Ecological Models and Data in R*. Princeton University Press.

Required: I think that this book is destined to change the skills of EEBB'ers for the next generation. It presents a fantastic overview of process modeling using Likelihood and Bayesian approaches using examples from R. From the absolute basics to very advanced models. If you put the effort into going through this book you will be well rewarded in all of your future endeavors. We will cover more than half of the chapters in the book (1,2,3,4,5,6, 9 & some of ten) before we begin to focus on various aspects of linear models (using Gelman and Hill) for the course.

Outdated drafts of chapters may still be available at the authors website, along with resources related to the book.http://www.zoo.ufl.edu/bolker/emdbook/index.html

The book also has a wiki set up for errata and clarifications
http://emdbolker.wikidot.com/

Gelman, A. & Hill, J. 2007. *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge.

Required: This book presents an alternative approach to modeling complex data using a GLMM framework, and is quite clear. It teaches concepts using an example based approach, and provides many hints, and "rules of thumbs" to help develop your statistical intuition. It is a highly recommended for situations where you are dealing with complexities like repeated measures, time series or error variation that is not constant across treatments. It uses a combination of parametric, simulation and Bayesian approaches. The examples are largely not from the biological literature however. One additional and important warning, the multi-level model approach used throughout this book is quite different in implementation from hierarchical models that we will work with in class.

Dalgaard, P. 2008. *Introductory Statistics with R*. 2nd ed. Springer. While you should already be familiar with all of the statistical material covered in this book, it may be useful in helping develop you basic skills with using R for performing basic manipulations with your data set, as well as simple plotting and statistical procedures. It will also be useful as review of some basic statistical concepts, for which you may need to be reminded. It is quite clear, and its examples are easy to follow. The Maindonald & Braun book (see below) is also quite useful and the choice may be a matter of personal preference. **This text is available online VIA the MSU library**.

Links to the book website
http://staff.pubhealth.ku.dk/~pd/ISwR.htm

Crawley, M.J. 2007. THE R BOOK. Wiley.
This is a behemoth of a book, but it provides fantastic coverage of R from introduction to intermediate (and some advanced methods) with respect to programming, plotting, data management and statistics. MSU also has online access to the book, so there is no reason to not use it. **This text is available online VIA the MSU library**.

Recommended texts **(in alphabetical order, not importance)**: These are a variety of texts that may not only serve you well for this class, but for your graduate work and beyond. I am not suggesting that you purchase all of these, books, but this list may come in handy as you develop your skills, and want to further explore other areas of interest.

Adler, J. 2009. R In a Nutshell. O'Rielly.
This book is a great introduction to PROGRAMMING in R, including a lot of cools tips and tricks to make life easier. In particular basic programming, data management and plotting are covered well. What is not covered so well is statistical analysis (which would serve as a good place to look up functions, but not for understanding them). Still a good overall.

Bolstad, W.A. 2007. Introduction to Bayesian Statistics. 2nd edition. WILEY.
This provides an introduction to the basic ideas and methods of the Bayesian approach, and is probably the clearest books in this regard. While it does not deal with models more complex than regression, it does an excellent job of building an intuition for what a prior distribution is, and how it can affect the "results" (i.e. the posterior). It does not have much on MCMC (and that is mostly in an appendix).

Braun, W.J. & Murdoch, DJ. 2008. *A First course in Statistical Programming with R*. Cambridge.
This slim book covers a lot of the concepts with regards to how to efficiently write your code in R. If you plan to use R to do more than run regression models and make very simple graphics, I recommend it. It will be especially useful for people who plan to use R for power analysis (or any other) simulations, and numerical optimization (for maximum likelihood for instance).

Burnham, K.P. & D.R. Anderson. 2002. ***Model selection and multi-model inference: A practical information theoretic approach***. **Springer**.
A thorough introduction to model selection using information theoretic approaches (AIC & BIC). This text is available online VIA the MSU library.

Faraway, J. J. 2005. *Linear Models with R*. Chapman & Hall/CRC, Boca Raton, FL.

Faraway, J. J. 2006. *Extending the Linear Model with R: generalized linear, mixed effects and nonparametric regression models*. Chapman & Hall/CRC, Boca Raton, FL. These two texts used to be required for the course, but while they are quite good, I felt they were too pricey. These two texts cover the majority of the parametric statistical approaches that we will discuss in this course including general (ANOVA, regression & ANCOVA) and generalized (logistic, poisson and so many other) linear models. It does not delve a great deal into statistical theory, but instead focuses on practical aspects of performing analyses and model evaluation. While they are a bit pricey as a pair, I highly recommend them, as they will be a valuable resource for a long time to come.

These sites link to the authors information about the book (including errata etc..)
http://www.maths.bath.ac.uk/~jjf23/LMR/

http://www.maths.bath.ac.uk/~jjf23/ELM/

Good, P. 2006. *Resampling methods: A Practical Guide to Data Analysis*. 3rd ed. Birkhauser.
This books serves as a decent introduction to both bootstrapping and permutation/randomization resampling methods (which we will be using in class a fair bit). It is a gentle introduction to the subject with example code for R (and numerous other platforms), but it lacks any great depth. As a book it will get you started with resampling, but not to get you to the high powered tools (which we will not really cover in the course anyways).

Gotelli, N.J. & Ellison, A.M. 2004. A Primer of Ecological Statistics. Sinauer.
This slim book represents a good introduction for many of the basic concepts in inferences, probability, experimental design and statistics. It has a basic introduction to a number of different approaches for estimation and inference, and in particular its short guide to experimental design is quite good. It is not really a "hands-on" approach, so you will not learn how to actually run the analyses or examine diagnostics of models. However, if you have never taken any statistics before (and you are crazy enough to be taking this course given that), a book like this would be an essential introduction.

Hilborn,R. & Mangel, M.1997. *The Ecological Detective: Confronting models with Data*. Princeton University Press.

This is a classic text for thinking about how to analyze data for ecological and evolutionary data. While the majority of the nitty-gritty is covered well by Ben Bolker's book, I still highly recommend reading this book, especially the first few chapters.

Maindonald, J. & Braun, J. 2006. *Data Analysis and Graphics Using R*. 2nd ed. Cambdridge.
I own the first edition of this book and it helped me a lot when I was just getting started using R. It will not present a great deal of new statistical material for you, but it will show you how to do all of the basic statistical techniques in R. The book by Dalgaard covers many of the same subjects, and it may be a personal choice (or what you have access to).

McCarthy, M.A. 2007. *Bayesian Methods for Ecology*. Cambridge.
This book provides an extremely gentle introduction to Bayesian thinking and methodology using the Bayesian equivalents of otherwise common statistical methods (ANOVA and regression for instance). It is quite clear, and I recommend it for people who might struggle with some of the ideas of Bayesian methods. One word of caution, it is extremely one-sided with respect to the value of Bayesian methods, and treats all other methods as if they should be left in the dust-bin. I do no subscribe to this philosophy at all (as you will learn in this course). It exclusively uses the BUGS environment (a language similar to R, but different in a few key ways) for Bayesian analysis.

Pinheiro, J.C. & D.M. Bates. 2000. *Mixed Effect Models in S/S-Plus*. Springer.
This is one of the classic texts describing all of the details (including all of the computational and algebraic gore) required to really understand mixed effect models. It also describes in detail how to use the original package for mixed modeling in R (nlme), which has been largely supplanted by lme4 (also by Douglas Bates). This text is available online VIA the MSU library.

Quinn, G.P. Keough, M.J. 2002. Experimental Design and data analysis for biologists.
This covers much of the same material as Gotelli and Ellison, and is a good introductory book, and reference book. It has some useful material on multivariate statistics as well. No real practical examples in R (or SAS).

Seefeld, K. Linder, E. 2007. Statistics Using R with Biological Examples. A free online book (PDF is available on the ANGEL site for the course). This book teaches statistics starting at a reasonably introductory level within a strongly Bayesian and computational framework. In particular the chapters on probability are excellent!
http://cran.r-project.org/doc/contrib/Seefeld_StatsRBio.pdf

Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S*. 4th edition. Springer, New York, NY.

One of the classic texts used for both R & S (R is derived from S, and is extremely similar). This is not a "how to do/learn statistics" text, but how to use R/S to perform statistics. Still it is extremely invaluable as a resource. This text is available online VIA the MSU library.

R Resources

To download R (and the starting point for virtually all things R)
http://cran.r-project.org/

The manuals for using & programming in R . "An Introduction to R" & "R Data Import/Export" are the important ones for this class. The rest are largely useful if you really get into programming.
http://cran.r-project.org/manuals.html

The contributed documents section on the CRAN site is **REALLY USEFUL**. It has lots of books and tutorials of exceptional quality. Some of my favorites are up on the ANGEL site for the course.
http://cran.r-project.org/other-docs.html

The "Task views" section describes some of the packages/libraries in R that are useful for particular tasks (examples include "multivariate", "Bayesian", "Ecological and environmental data" & "spatial data" to name a few).
http://cran.r-project.org/web/views/

Tips for using R
http://pj.freefaculty.org/R/Rtips.html

kick-starting R
http://cran.r-project.org/doc/contrib/Lemon-kickstart/

Course notes for "Statistical Programming in R/S"
http://socserv.mcmaster.ca/jfox/Courses/R-course/index.html

Programming in R web site with examples
http://www.faculty.ucr.edu/~tgirke/Documents/R_BioCond/R_Programming.html

Examples of high end graphics and plots in R with source codes
http://addictedtor.free.fr/graphiques/

The R wiki –lots of useful example code bits. In particular this has useful information on Regression & mixed models in R, and examples from the Ecological Detective for R.
http://wiki.r-project.org

Other (statistics) web sites that may be useful to you.

Bruce Walsh's website including his course notes for biostats, and Quantitative Genetics.
http://nitro.biosci.arizona.edu/

Electronic Statistics textbook resource – **EXCELLENT RESOURCE!!!!**
http://www.statsoft.com/textbook/stathome.html

Statistical Analysis in Ecology and Evolution – A course with very similar goals to our own. Lecture notes, R code and assignments available on the site.
http://www.unc.edu/courses/2006spring/ecol/145/001/index.html