FW8051 Statistical Modeling of Ecological Data using R and WinBugs/JAGS[1]

Professors Dr. John Fieberg, jfieberg@umn.edu
Office Hours (B52 Skok): Tuesday/Friday 9:00-10:00am or by appointment

Location Monday and Wednesday 1:55-3:50 in 302 Kaufert Laboratory.

Course Description: Ecological data pose many challenges to statistical inference. Most data come from observational studies rather than designed experiments; observational units are frequently followed over time, resulting in multiple, non-independent measurements; response data are often binary (e.g., presence-absence data) or non-negative integers (e.g.., counts), and therefore, the data do not fit the standard assumptions of linear regression (Normality, constant variance. This course will familiarize students with modern statistical methods that address these complexities, with an emphasis on Bayesian implementations of commonly used regression models. We will begin with a review of linear regression, emphasizing the role of design matrices in model construction. We will learn how to create explanatory variables (e.g., when fitting models with categorical predictors) and also generate model-based estimates of predicted values along with associated measures of uncertainty. With this foundation in place, we will then consider extensions to non-Normal data (e.g., generalized linear models) and correlated data (mixed models, generalized estimating equations). Exercises will utilize program R and general Bayesian modeling software JAGs and will make extensive use of both real and simulated data sets.

Prerequisites: A graduate-level statistics class (e.g., ST 5021), and an understanding of key statistical concepts, including hypothesis tests, the Normal distribution, and linear regression. In addition, you should have a working knowledge of the R programming language (e.g., be able to read in data, work with common object types [lists, matrices, data frames], install and load packages, access help functions, and construct simple plots). Students without any R experience should seek to obtain a basic level of understanding prior to the course through self-study. One possibility is the online course located here: http://tryr.codeschool.com.

Learning Objectives: The overarching goal of the course is to train students to effectively analyze the data they collect as part of their research. By the end of this course, you should be able to:

- Construct models that address specific biological objectives.
- Understand the role of random variables and common statistical distributions in formulating modern statistical regression models.
- Identify key model assumptions, utilize diagnostic tools to assess validity of these assumptions, and conduct sensitivity analyses to evaluate model robustness to assumption violations.
- Gain an appreciation for challenges associated with selecting among competing models and performing multi-model inference.

[1] Disclaimer: This syllabus is tentative and the instructor reserves the right to make changes to the syllabus as deemed necessary. Students will be notified in a timely manner of any syllabus changes via email or Moodle. Please remember to check your email and the course page on Moodle often.

- Critique statistical methods used in the applied literature, identify strengths and weaknesses of different modeling approaches, and select appropriate analyses in your work.
- Conduct research using a workflow that maximizes 'reproducibility' of your work.

To achieve the above learning objectives, you will be expected to develop new statistical modeling and computing skills (see **Skills Objectives**).

Skills Objectives: By the end of this course, you should be able to:

- Construct covariates that allow fitting of models with categorical predictors and that allow for non-linear relationships between explanatory and response variables.
- Fit and evaluate a variety of regression models in both Frequentist (using R) and Bayesian frameworks (using JAGs).
- Use simulation methods to test your understanding and evaluate power and robustness of different regression models.
- Obtain model-based estimates of predicted responses along with confidence and prediction intervals for a variety of commonly used regression models.
- Model non-Normal data using generalized linear models.
- Fit models to correlated data using mixed models and generalized estimating equations; estimate robust standard errors by performing a cluster-level bootstrap (resampling independent observational units).

Textbooks: Several sections of the course will draw upon material in the book by Marc Kéry, below:

Kéry, M. 2010. Introduction to WinBugs for Ecologists. Academic Press, Elsevier, New York.

In addition, we will frequently use Zuur et al.'s book for illustrative examples and applications:

Zuur, A. F. et al. 2009. Mixed Effects Models and Extension in Ecology with R. Springer, New York.

Lastly, these texts will be supplemented with readings from online e-books and journal articles.

Some Important notes on the books:

- You can download an electronic version of Zuur et al. or purchase a softcover version for $25 through the UMN library via this link: http://link.springer.com.ezp2.lib.umn.edu/book/10.1007/978-0-387-87458-6
- You can find an electronic single-use versions of Kery's book here: http://www.sciencedirect.com.ezp3.lib.umn.edu/science/book/9780123786050

Software:

We will make extensive use of **R** and **JAGS** during the course. R is a modern statistical computing package supported by a large network of scientists worldwide. JAGS is a popular software platform that makes Bayesian modeling 'easy' (or at least more accessible to a wider audience). Although the learning curve associated with these programs can be steep, invest the

time to become comfortable now and you will see huge dividends in the future. Importantly, these programs are free for Microsoft, Apple, and Linux platforms so you can take the skills you learn anywhere you go.

You should download current versions of these programs to your personal computer. In addition, you will be required to use RStudio along with the knitr package (and related ezknitr package) to produce reproducible html documents that summarize your work. You can download necessary software here:

1. R: http://streaming.stat.iastate.edu/CRAN/
2. JAGS (works with Windows, Macintosh, and Linux operating systems): http://www-ice.iarc.fr/~martyn/software/jags/
3. RStudio: http://www.rstudio.com/

Lastly, to prepare for class you should install the following packages in R: car, nlme, lme4, knitr, ezknitr, R2jags, rjags, lattice, latticeExtra, rms, ggplot2, runjags, mcmcplots.

We will frequently work through examples in class, so bring your laptop every day!

Course Notes and Website: Readings, as well as all lectures and homework assignments will be posted on a classroom website within Moodle. For most topics, there will be *before* and *after* versions of lecture handouts; the former will have some material removed to help facilitate classroom discussion.

Grades: Grades will be assigned in a manner consistent with the University's Grading Standards: http://policy.umn.edu/Policies/Education/Education/GRADINGTRANSCRIPTS.html

Component of grade	%
Homework	40
Mid-term course project	25
Individual component (70% of project)	
Group component (30% of project)	
Final exam	25
Class Participation	10

Homework: homework assignments will offer you a chance to practice implementing the statistical methods we cover in class. With each assignment, you will be expected to analyze real (or simulated) data and turn in a short report that includes your R code and associated output as well as answers to any questions posed as part of the assignment. These reports should be produced using functions in the knitr package of program R. It is OK to work together on homework assignments (in fact, it is encouraged), but each individual is responsible for writing his or her own computer code and for producing a written report.

Homework Assessments: homework assignments will be self- and peer-graded using the workshop feature in Moodle. I will provide you with an answer key and a simple rubric for each assignment. This process will allow for timely feedback, help develop your evaluation skills, and most importantly, facilitate learning by: 1) requiring you to reflect on the parts of the assignment

where you may have struggled to understand important concepts; b) exposing you to alternative solutions provided by your peers (and via an instructor-provided answer key). I will review all self-assessments to determine where to focus my efforts in class and will also review and re-grade all assignments where self- and peer-assessments differ.

Late Policy: *implementing the self- and peer-assessments requires that all assignments are turned in on time!* Assignments will often take considerable time – you should plan on starting them early; it will be difficult to complete them in a single sitting. Penalty-free extensions may be granted in rare cases (e.g., ***documented*** illness or emergencies), but in general, I expect you to plan ahead for sanction events (e.g., intercollegiate athletic events, University activities, religious observances, etc.) so that you can turn in assignments on time.

Course Project: the course project will serve as an active learning component, emphasizing critical thinking and problem solving skills. You will be required to build a predictive model using a real data set. In real life, no one will tell you exactly how you should analyze your data. Ecological data are often messy, they may not fit the assumptions of common regression methods, and there will often be no clear "right" way to analyze your data. Faced with this level of ambiguity, you must come up with a method of analysis that you can defend (e.g., to your advisor, to reviewers, to managers who may or may not like the results of the analysis). Thus, it is not enough to know *how* to implement a variety of regression methods. You must be able to understand the strengths and limitations of various approaches and be able to choose among several (potentially imperfect) methods. You must also be able to explain your results in a way that faithfully represents the information in your data.

The project will be completed in 2-steps. Initially you will work independently to produce a potential solution to the problem. You will then confer in small groups, evaluate each other's approaches, and then submit a single group report for final grading.

Final Exam: The final exam will be held during the regularly scheduled exam time on Tuesday, May 9 (8:00-10:00am). The exam will cover core concepts from throughout the course.

Scholastic Dishonesty:
You are expected to do your own academic work. Obvious forms of plagiarism on homework assignments and course projects will result in a 0 for the assignment. In addition, I will be forced to file a formal report to the Office for Student Conduct and Academic Integrity. If you ever have any questions about what might or might not be permissible, ask!

Student Conduct Code:
As a student at the University you are expected adhere to Board of Regents Policy: *Student Conduct Code*. To review the Student Conduct Code, please see:

http://regents.umn.edu/sites/default/files/policies/Student_Conduct_Code.pdf.

Sexual Harassment:
"Sexual harassment" means unwelcome sexual advances, requests for sexual favors, and/or other verbal or physical conduct of a sexual nature. Such conduct has the purpose or effect of

unreasonably interfering with an individual's work or academic performance or creating an intimidating, hostile, or offensive working or academic environment in any University activity or program. Such behavior is not acceptable in the University setting. For additional information, please consult Board of Regents Policy:

http://regents.umn.edu/sites/default/files/policies/SexHarassment.pdf

Equity, Diversity, Equal Opportunity, and Affirmative Action:
The University provides equal access to and opportunity in its programs and facilities, without regard to race, color, creed, religion, national origin, gender, age, marital status, disability, public assistance status, veteran status, sexual orientation, gender identity, or gender expression. For more information, please consult Board of Regents Policy:

http://regents.umn.edu/sites/default/files/policies/Equity_Diversity_EO_AA.pdf.

Disability Accommodations:
The University of Minnesota is committed to providing equitable access to learning opportunities for all students. Disability Services (DS) is the campus office that collaborates with students who have disabilities to provide and/or arrange reasonable accommodations.

If you have, or think you may have, a disability (e.g., mental health, attentional, learning, chronic health, sensory, or physical), please contact DS at 612-626-1333 to arrange a confidential discussion regarding equitable access and reasonable accommodations.

If you are registered with DS and have a current letter requesting reasonable accommodations, please share your letter with me as soon as possible in order to secure accommodations in a timely manner.

For more information, please see the DS website, https://diversity.umn.edu/disability/.

Mental Health and Stress Management:
As a student you may experience a range of issues that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance and may reduce your ability to participate in daily activities. University of Minnesota services are available to assist you. You can learn more about the broad range of confidential mental health services available on campus via the Student Mental Health Website: http://www.mentalhealth.umn.edu.

Academic Freedom and Responsibility:
Academic freedom is a cornerstone of the University. Within the scope and content of the course as defined by the instructor, it includes the freedom to discuss relevant matters in the classroom and conduct relevant research. Along with this freedom comes responsibility. Students are encouraged to develop the capacity for critical judgment and to engage in a sustained and independent search for truth. Students are free to take reasoned exception to the views offered in any course of study and to reserve judgment about matters of opinion, but they are responsible

for learning the content of any course of study for which they are enrolled.[2] When conducting research, pertinent institutional approvals must be obtained and the research must be consistent with University policies.

Reports of concerns about academic freedom are taken seriously, and there are individuals and offices available for help. Contact the instructor, the Department Chair, your adviser, the associate dean of the college, or the Vice Provost for Faculty and Academic Affairs in the Office of the Provost.

[2] *Language adapted from the American Association of University Professors "Joint Statement on Rights and Freedoms of Students".*