ZOOL 631: Biometry

Spring 2021

Instructor: Jake Ferguson

Email: jakeferg@hawaii.edu

office: 956-5472

ZOOM
link: https://hawaii.zoom.us/j/91247374057
(Passcode: Data)

Class hours: MWF: 9:30 - 10:20, W: 11:30 - 1:20

Office hours: W: 10:30-11:20, R: 10:00-11:00

Past lectures and notes:
https://drive.google.com/drive/folders/1S-xP6mtSieI2foVmGOJBO88gHdiaCYwf?usp=sharing

Rstudio cloud sharing link

Course Description

Biological data pose many challenges to statistical inference. Most data in ecological and evolutionary studies come from observational studies rather than designed experiments; observational units are frequently sampled repeatedly over time or space, 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 the processes that generate many kinds of biological data, and the foundations of the likelihood inference framework.

The emphasis of this course will be on building biological models, estimating parameters, and coherently interpreting those estimates. We will discuss the underlying mechanics of estimation using likelihood and measuring and interpreting uncertainty in point estimates. We will also spend time on comparing the relative evidence in competing statistical models and on standard linear regression tools.

Learning Objectives

The overarching goal of the course is to train you to effectively analyze the data collected in your research. By the end of this course, you should be able to:

- Understand the role of random variables and common statistical distributions in formulating modern statistical regression models.
- Describe how certain biological processes generate data following common statistical distributions.
- Demonstrate “model literacy” – be able to describe a variety of statistical models and their assumptions using equations and text and match parameters in these equations to estimates in computer output.

- 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.

Skills Objectives

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

- Describe how probabilistic models can be used to describe ecological and evolutionary data.
- Fit and evaluate a variety of regression models using R
- Construct covariates that allow fitting of models with categorical and continuous predictors that allow for non-linear relationships between explanatory and response variables.
- Use simulation methods to perform tests when standard statistical methods fail.
- Obtain model-based estimates of predicted responses along with confidence and prediction intervals for a variety of commonly used regression models.

Textbook

Several sections of the course will draw upon material in the book by Ben Bolker:

> Bolker, Benjamin M. *Ecological models and data in R*. Princeton University Press, 2008.
>
> Free version of an older draft at: https://ms.mcmaster.ca/~bolker/emdbook/book.pdf

Software

We will make extensive use of **R** during the course. R is a modern statistical computing package supported by a large network of scientists worldwide. Although the learning curve associated with both 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.

A nice resource for getting up to speed quickly in R is at datacamp (https://www.datacamp.com/courses/free-introduction-to-r). It does not require installing any software to go through the tutorials.

We will use Rstudio cloud in the course (https://rstudio.cloud/). You can login in using your UH account and create a free account.

Website

Readings, as well as homework assignments will be posted on a classroom website within Laulima. Lectures will be on the whiteboard, you will be expected to take your own notes on the material.

Grades will be assigned in a manner consistent with the University's Grading Standards: http://www.manoa.hawaii.edu/graduate/content/grades-regulations

Component of grade	%
Homework	60

Labs	25
Course project	15

Homework assignments will be due Fridays at 5, submitted on laulima. These assignments will offer you a chance to practice implementing the statistical methods we cover in class. With later assignments, you will be expected to analyze real (or simulated) data and turn in a short report that includes any worked homework problems as well as any relevant R code and associated output. We will eventually have homework reports that will be produced using functions in the knitr package of program R. ***It is encouraged to work together on homework assignments, but each individual is responsible for writing his or her own work including any computer code and written reports.*** To submit your work, you must have the ability to scan paper or write on a tablet device, save as a PDF, and submit on Laulima. If scanning paper, either a scanner or a smartphone app such as Apple Notes, Google Drive, or similar app should be used.

Late Policy. ***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 so that you can turn in assignments on time. You will lose **10%** of your grade for each day that your assignment is late.

Course Project. This will serve as an active learning component, emphasizing critical thinking and problem-solving skills. You will be required to build a 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.

Scholastic Dishonesty. You are expected to do your own homework, though I encourage you to work in groups and discuss the homework problems. 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!

Disability Accommodations. The University of Hawai`i is committed to providing equitable access to learning opportunities for all students. The Kokua program 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 Kokua at (808) 956-7511 to arrange a confidential discussion regarding equitable access and reasonable accommodations.

If you are registered with Kokua 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.