Biometry

ZOOL631

Instructor: Jake Ferguson

Snyder Hall 414D (attached to Edmondson Hall)

Email: jakeferg@hawaii.edu

office: 956-5472

cell: (406) 579-0723

ZOOM MEETING LINK: https://hawaii.zoom.us/j/381207001

In-class hours

MWF: 9:30 – 10:20 ~~in St John 7~~

W: 11:30 – 1:20 ~~in Webster 101~~

Office hours

W: 10-30-11:20

R: 10:00-11:00

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. We will also discuss the frequentist and Bayesian frameworks (time permitting).

The emphasis of this course will be on building and estimating parameters in biological models. 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 modeling tools.

Prequisites

Some calculus will be required.

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:

- 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 generate ecological and evolutionary data.
- Fit and evaluate a variety of regression models in both likelihood (using R) and Bayesian frameworks (using JAGs).
- 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.

Textbooks:

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

A few readings from other books are recommended as well:

> Allen, Linda J. S. An introduction to stochastic processes with application to biology.
>
> Available at: https://uhawaii-manoa.primo.exlibrisgroup.com/permalink/01UHAWAII_MANOA/16bed9j/pq_ebook_centralEBC1575386

We will draw on examples using JAGS from Kéry's book for illustrative examples and applications if we make it to Bayesian inference:

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

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. We may also use **JAGS**, a popular software platform that makes Bayesian modeling 'easy' (or at least more accessible to a wider audience). 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 in the course. Homeworks and labs will be implemented in Rstudio cloud (https://rstudio.cloud/). You can login in using your UH account.

You can also use the desktop version of Rstudio if you wish. If so, you need to download current version of **R** to your personal computer. In addition, you will be required to use **RStudio** along with the *knitr* package (and related *ezknitr* and *rmarkdown* packages) to produce reproducible html documents that summarize your work. You can download necessary software here:

1. R: http://streaming.stat.iastate.edu/CRAN/

2. RStudio: http://www.rstudio.com/

Lastly, to prepare for class you should install the following packages in R: knitr, ezknitr, rmarkdown. I am sure we will add more to this list as the semester progresses.

We may work through examples in class, so I recommend bringing your laptop every day!

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	85
Course project	15

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 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.***

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 for sanction events (e.g., intercollegiate athletic events, University activities, religious observances, etc.) 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 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!

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.

Tentative Lecture and Lab Schedule Spring 2020

Wk	Date	Topic(s)	Reading	Notes
1	13-Jan	Review syllabus, skills quiz		
	15-Jan	Properties of probability	Section 4.1, 4.2	Discussion 1:
	17-Jan	The binomial distribution	Section 4.5.1.1	HW 1 due
2	20-Jan	**MLK day – no class**		
	22-Jan	Characterizing distributions: mean and variance of discrete distribution	Sections 4.4.1 – 4.4.5	Discussion 2:
	24-Jan	Working with the Bernoulli and binomial distributions		HW 2 due
3	27-Jan	Discrete random walks	link. Read up to the section on Brownian motion	
	29-Jan	Discrete random walks: examples		Discussion 3:
	31-Jan	The chain binomial model	link.	HW 3 due

4	3-Feb	The Poisson & geometric distributions		
	5-Feb	Point estimation with likelihood	Section 6.1 through 6.3	Discussion 4:
	7-Feb	Profile likelihood & confidence intervals	Section 6.4	HW 4 due
5	10-Feb	Jensen's inequality & transforming estimates	Section 7.5.2 and A.7	
	12-Feb	Numerical optimization	Sections 7.1, 7.2, 7.4	Discussion 5:
	14-Feb	Numerical optimization II	Sections 7.5, 7.6	HW 5 due
6	17-Feb	**Presidents’s day – no class**		
	19-Feb	Continuous probability: Normal distribution	Section 4.5.2	Discussion 6
	21-Feb	The central limit theorem	TBD	HW 6 due
7	24-Feb	The central limit theorem II: Wald intervals	Section 6.5	
	26-Feb	The log-normal distribution		Discussion 7
	28-Feb	The Chi-square distribution: likelihood ratio tests	6.6.1	HW 7 due
8	2-Mar	Poisson processes I		
	4-Mar	Poisson processes II		Discussion 8
	6-Mar	Linear regression I	Section 9.2	HW 8 due
9	9-Mar	Linear regression II		
	11-Mar	Linear regression III		Discussion 9
	13-Mar	ANOVA	Section 9.2.3 - 9.2.4	HW 9 due
10	16-Mar	**Spring recess – no class**		
	18-Mar	**Spring recess – no class**		
	20-Mar	**Spring recess – no class**		
11	23-Mar	Computational approaches to statistics	link.	
	25-Mar	The model selection problem		Discussion 10
	27-Mar	Information criterion	Section 6.6.2 through 6.7	HW 10 due
12	30-Mar	Model averaging		

	1-Apr	GLM's	Section 9.4	Discussion 11
	3-Apr	GLM's II		HW 11 due
13	6-Apr	Logistic regression		
	8-Apr	Zero-inflation models	Section 4.7.3	Discussion 12
	10-Apr	**Good Friday – no class**		
14	13-Apr	Multivariate normal distribution		
	15-Apr	Multi-level models I	Sections 10.1-10.3	Discussion 13
	17-Apr	Multi-level models II	Sections 10.4-10.5	HW 12 due
15	20-Apr	Multi-level models III		
	22-Apr	Bayesian inference I	TBD	Discussion 14
	24-Apr	Bayesian inference II		HW 13 due
16	27-Apr	Using JAGS		
	29-Apr	Using JAGS II		Discussion 15
	1-May	Multi-level models in JAGS		HW 14 due
17	4-May	TBD		
	5-May	TBD		
	7-May	TBD		