COURSE INFORMATION

Biostatistics – EEB411
10:20 – 11:10 am. MWF. Strong Hall 103
University of Tennessee - Knoxville
SPRING 2023

INSTRUCTOR:

James Fordyce jfordyce@utk.edu
Office: 540 Hesler Hall
Office hours: Monday 12-1pm and by appointment. Office hours will be done via Zoom at: https://tennessee.zoom.us/j/93424451927
Correspondence: I will typically respond to email within 24 hours during the week and 48 hours on the weekend. Please do not hesitate to reach out to either me or a TA for clarification or questions related to the material.

TA:

Anchal Padukone apadukon@vols.utk.edu
Office: 430 Hesler Hall
Office hours: Thursday 12-1pm via open Zoom and by appointment (Meeting link for the whole semester: https://tennessee.zoom.us/j/783051043 ; password: 370011)
Correspondence: I usually respond to email within 24 hours during the week (I am typically most responsive in the afternoons), and within 48 hours on the weekend. Please reach out if you have questions!

TEXT: These are optional. I mention them as a useful resource. There are also many free resources on the web. Wiki is actually rather good on the stats front. A simple google search for stats or R related questions will often direct you to stackoverflow.com – which is a (usually) really helpful forum.

Crawley, M.J. 2012. The R Book. Wiley (this is free through the UTK library site)
Dalgaard, P. 2002. Introductory Statistics with R. (also free through the UTK library site)
Quinn, G.P. & M.J. Keough 2009. Experimental Design and Data Analysis for Biologists

Other good ones:
Legendre, P & L. Lengendre. 2012. Numerical Ecology.
Kruschke, J. K. 2015. Doing Bayesian Data Analysis.
McElreath, R. 2020. Statistical Rethinking: A Bayesian course with examples in R and Stan.

COURSE OBJECTIVES: The use of statistics is ubiquitous in biology. An understanding of statistics is not only important for the design of experiments and analysis of one's own data, but also for the ability to critically read the literature (including literature that you might be asked to review). It is important for scientists to understand *what* exactly various statistical approaches are doing. Are they testing a null hypothesis (P value alert!)? Are they on a parameter interpreting voyage to Bayesian Land (Come for the stats, stay for the rides!)? So many options. So many flavors. They are all models. Simply models. “Models” can sound intimidating,

but they're really not if you approach them with a competent level of understanding. Actually, you all have been doing statistical models since grade school – Ever calculate the average of something? Yup, that's modeling the data. The goal of this course is to introduce basic statistical approaches in a way that we might understand what statistical hypothesis, or modelling objective, is being addressed and how we might interpret this biologically. No course (certainly not this one) can cover the whole body of statistical approaches used by researchers. One important objective of this course is for you to feel comfortable with the language of statistics so that you can use books or consultants efficiently and effectively. Another objective of the course is to encourage (strongly) carefully thinking about the design of a study – from idea, to question, to hypothesis, to experimental design, to analysis, to (finally) interpretation of analysis. In a perfect world (and we all want a perfect world) all these steps should be done before the first datum is collected. One should never collect data and then ask, "How should I analyze this?". This often doesn't end well (or at least as "clean and pretty" as one would hope).

Statistics is a big field. It's really easy to quickly feel overwhelmed. A little dose of math, a dab of Greek letters, some pedantic language, and (my personal favorite) statements like "It can be easily seen that…" is a real turn-off for many students (and scientists!). This is something with which I can truly empathize. I'm going to explicitly try NOT to do that. If at any point you feel as though something was opaque, or I try to link two concepts together and you are not seeing the connection, PLEASE LET ME KNOW. My approach to this material is to develop concepts using baby steps. By no means do I use "baby steps" as condescension. What I mean is, let's set a path where we slowly begin from the simplest concept and build upon that, one small step at a time. Once the very basics are understood, you'll appreciate that, for example, there's really no difference between "regression" and "factorial anova"; or, that mixed models are simply a way for us to account for variation that is not central to our question and to keep track of how many independent observations we really have. So, baby steps will work in our favor – let's make sure we "get" this stuff. Metaphorically speaking: Let's learn how a major scale is put together before we try Bach's *Neverending Canon*. Let's learn that salt, sugar, and spices can interact in interesting ways before we jump into a beef wellington. Nearly all the best musicians and chefs started with baby steps (*twinkle twinkle little star* and scrambled eggs). Let's do that with statistics.

This course should culminate with a solid understanding of construction and interpretation of linear models. It doesn't sound like much, but that covers the vast majority experimental statistical approaches you're likely to encounter.

There are always new approaches and more to learn. I'm sure there are a few people in the world that are true stats masters and understand everything in the world of statistics. Except for those 2 or 3 savants on the planet, we're all continuing to learn. My personal goal for this course is to set you up with the tools for understanding the basic approaches, and to (with a little effort) teach yourselves alternative approaches that might meet your analytical, research, and scientific needs.

Stats are actually a lot of fun. I promise.

LECTURES: Lectures will be in person. Lectures will also be provided via YouTube vignettes. The YouTube vignettes were put together when we were all online – so, they will mostly match (perhaps in even more detail) precisely the content of the in-person lectures. They will be accompanied by detailed notes and computer code (posted to the Files folder on the Canvas site). Links to the video lectures can be found at the end of this syllabus.

ASSIGNMENT SUBMISSION: Weekly HW exercises and take-home R exams will largely consist of annotated computer code. **Email us your completed R scripts at <biostatisticseeb411@gmail.com> by noon the day they are due.** Note that this email address is primarily for receiving submissions and not checked as frequently as the instructors' emails. If you have questions about the assignments, please contact the instructors directly. Find detailed assignment submission instructions at this link (https://www.youtube.com/watch?v=Xk5MOhC71zU).

The online conceptual portions of the Midterm and Final will be administered as timed Quizzes via Canvas. You will need to use Chrome as your browser and install the Chrome plugin Proctorio to take these quizzes. You can find download instructions in the "Secure Exam Proctor (Proctorio)" tab in the course navigation pane on our Canvas course site.
We will make each quiz available over a window of a couple of days; you will need an uninterrupted chunk of time during this window to take it.

We will use the Announcements feature on Canvas to communicate when assignments and quizzes have been posted, and to post important announcements or updates. Please check this page regularly or check that you are notified of new Announcements.

ASSESSMENT:

Weekly Exercises	60%	-	11 total, lowest score will be dropped (actually, not dropped. Consider it “extra credit”)
Mid-term	20%	-	22-24 March (It’s part take-home R coding, part online conceptual)
Final	20%	-	12 May (It’s part take-home R coding, part online conceptual)

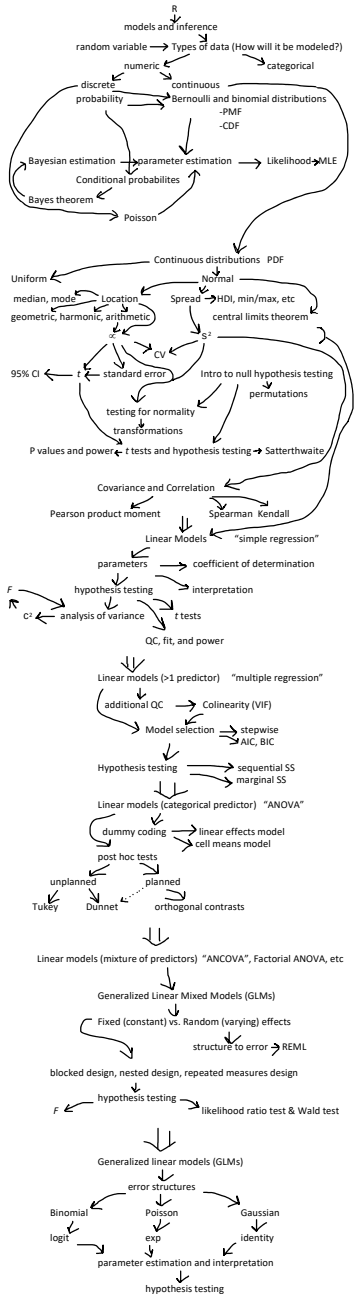
Grading will be based on the “standard” >89 is an A, <90 and >79 is a B, <80 and >69 is a C, <70 and >59 is a D, else is an F.

SOFTWARE:
We will be using the statistical programming language, R.
R is free at http://www.r-project.org/. R has become a standard tool in ecology and evolutionary biology (look through a recent issue of Evolution, Ecology, Science or Nature and you'll notice R is commonly used). R also has great graphics abilities. We will like R. We will like it very much.

TENTATIVE WEEKLY SCHEDULE (Homework will usually be assigned on Friday of week):

Week of	Topic	HW (assigned)
23-Jan	Introduction	X
30-Jan	Discrete distributions – Binomial & Poisson	X
6-Feb	Conditional probability, Bayes theorem	X
13-Feb	Continuous distributions, normal distribution, hypothesis testing, moments, transformations, *t* distribution, comparing groups, power, p-values	X
20-Feb	Covariance and correlation	X
27-Feb	Linear Models, analysis of variance, X^2 and *F* distribution	X
6-Mar	Linear models (multiple regression) and model selection	X
13-Mar	Spring Break	
20-Mar	Midterm	
27-Mar	Linear models, dummy coding (anova designs)	X
3-April	Linear models, dummy coding (anova, ancova, factorial designs)	
10-April	Linear models, dummy coding (anova, ancova, factorial designs)	X
17-April	Mixed Models	X
24-April	Generalized Linear Models	X
1-May	Bonus fun stuff	

A COURSE FLOW CHART: I am fully aware of how ridiculous this figure might look. But bear with me. First, past students have explicitly asked me for a flow chart explaining, at least pedagogically, what my strategy is for teaching this material. Second, notice how things become much more linear once we get about halfway through. The first half is mainly explaining *how* the tools work, whereas the second half is applying those tools to models (which are essentially bigger tools). This is why it is important to let us know if you're starting to feel lost. New/bigger tools require old tools. Third, it does look at least a little bit cool, right?



UNIVERSITY POLICIES

Academic Integrity
From the University of Tennessee Honor Statement, "An essential feature of the University of Tennessee, Knoxville is a commitment to maintaining an atmosphere of intellectual integrity and academic honesty. '*As a student of the university, I pledge that I will neither knowingly give nor receive any inappropriate assistance in academic work, thus affirming my own personal commitment to honor and integrity*.'"

University Civility Statement
"Civility is genuine respect and regard for others: politeness, consideration, tact, good manners, graciousness, cordiality, affability, amiability and courteousness. Civility enhances academic freedom and integrity, and is a prerequisite to the free exchange of ideas and knowledge in the learning community. Our community consists of students, faculty, staff, alumni, and campus visitors. Community members affect each other's well-being and have a shared interest in creating and sustaining an environment where all community members and their points of view are valued and respected. Affirming the value of each member of the university community, the campus asks that all its members adhere to the principles of civility and community adopted by the campus." See the UT Principles of Civility and Community.

Disability Services
The University of Tennessee, Knoxville, is committed to providing an inclusive learning environment for all students. If you anticipate or experience a barrier in this course due to a chronic health condition, a learning, hearing, neurological, mental health, vision, physical, or other kind of disability, or a temporary injury, you are encouraged to contact Student Disability Services (SDS) at 865-974-6087 or sds@utk.edu. An SDS Coordinator will meet with you to develop a plan to ensure you have equitable access to this course. If you are already registered with SDS, please contact your instructor to discuss implementing accommodations included in your course access letter.

Your Role in Improving Teaching and Learning Through Course Assessment
At UT, it is our collective responsibility to improve the state of teaching and learning. During the semester, you may be requested to assess aspects of this course either during class or at the completion of the class. You are encouraged to respond to these various forms of assessment as a means of continuing to improve the quality of the UT learning experience.

Key Campus Resources for Students
- Center for Career Development (Career counseling and resources; HIRE-A-VOL job search system)
- Course Catalogs (Listing of academic programs, courses, and policies)
- Hilltopics (Campus and academic policies, procedures and standards of conduct)
- OIT HelpDesk **(865)** 974-9900
- Schedule of Classes/Timetable
- Student Health Center (visit the site for a list of services)
- Student Success Center (Academic support resources)
- Undergraduate Academic Advising (Advising resources, course requirements, and major guides)

- University Libraries (Access to library resources, databases, course reserves, and services)

LINKS TO YOUTUBE VIGNETTES

Here are the YouTube links to the lectures. These are a legacy of COVID times, so the exact content (examples, order, dad jokes) might vary from what is said in lecture- but the content should be consistent. Note that they are subject to change (if I edit them and re-upload, they'll have a different link) – I will send an announcement to the class if this happens.

Intro to R
https://www.youtube.com/watch?v=9rsgpT7nrdE&t=2385s

Discrete distributions – Binomial & Poisson
https://www.youtube.com/watch?v=RMVia55tXWs&t=16s
https://www.youtube.com/watch?v=is1tM-eCqXo

Conditional probability, Bayes theorem, Continuous distributions, normal distribution, hypothesis testing, moments, transformations
https://www.youtube.com/watch?v=seTURBr8KX8&t=2779s
https://www.youtube.com/watch?v=4_Lhwk3M-oo&t=2s
https://www.youtube.com/watch?v=pNKpl7zHFag&t=41s
https://www.youtube.com/watch?v=bGMJXAbzIrM
https://www.youtube.com/watch?v=sNLwZxjtTw4&t=3s

t distribution, comparing groups, power, p-values
https://www.youtube.com/watch?v=dZobBCdRfoQ
https://www.youtube.com/watch?v=lBLBRlMz4Co

Covariance and correlation
https://www.youtube.com/watch?v=1b5_eMp90Lk

Linear Models, analysis of variance, X^2 and *F* distribution
https://www.youtube.com/watch?v=mxuTVC0TnyQ&t=19s
https://www.youtube.com/watch?v=XIOvgRytmxI&t=14s
https://www.youtube.com/watch?v=FLdVRI3Y12Q

Linear models (multiple regression) and model selection
https://www.youtube.com/watch?v=sgq1sg79BtQ
https://www.youtube.com/watch?v=wlLOV9BobU0

Linear models, dummy coding (anova designs)
https://www.youtube.com/watch?v=sJkxeR6akMs
Linear models, dummy coding (anova, ancova, factorial designs)
https://www.youtube.com/watch?v=d7QZutlQcEc

Mixed models
https://www.youtube.com/watch?v=aQm1JZqo5GQ&t=198s

Generalized linear models
https://www.youtube.com/watch?v=8djp0iv_09Q