Bayesian Data Analysis for Natural and Social Sciences
EEB 614 & Anthropology 604

The University of Tennessee – Fall 2025
Tuesdays & Thursdays 8:10 – 9:25 A.M. 427 Hesler Biology

Instructor:	Dr. Benjamin Auerbach
E-mail:	auerbach@utk.edu
	(Dr. Auerbach typically responds within 24 hours. Urgent emails should have "URGENT" in the subject line.)
Office:	520 Hesler Biology
Office Hours:	Tuesdays & Thursdays, 9:30 – 10:30 A.M.

Course at-a-Glance

- Emphasis is on advanced research design, including skills in programming and critical thinking about the application of quantitative analytical approaches.
- You will learn which statistical approaches best fit the questions you ask of data, with a focus on Bayesian analytical methods.
- Assessment is based on formal critiques of published papers, independent research paper, and weekly participation.
- Expect to read approximately 50 pages per week and give yourself plenty of time to work through tutorials in R.

Course Description

This course introduces an assortment of advanced, specialized statistical methods that draw on methods already familiar to you, as well as introducing modeling approaches that are novel to you. An emphasis is placed on Bayesian inference and the estimation of parameters for your data. For example, how likely is the mean in your data given what you know about the population from which you drew a sample? Or how might we predict the frequency of a behavior given a set of predictor variables? To learn these methods, you will be drawing from examples in the biological sciences and behavioral & social sciences.

While the course introduces you to Bayesian methods and logic, none of the statistical methods encountered this semester are exhaustively covered. Some methods will be introduced with the intention of exposure to basic concepts and application, though those of you seeking proficiency should seek out specialty courses or be prepared to undertake rigorous independent study. With the knowledge gained, however, you will have the basic tools with which to venture into these additional studies.

I place a focus on two general areas throughout the course: conceptual thinking and programming. You will be asked to think about parameter space, distributions of data, and how to link these to the outcomes of questions that you ask of data. Thus, we will spend time thinking about the logic of research questions, design, and analysis. Programming in the R environment is emphasized and taught throughout the course.

Ultimately, the focus of this course is on the ability to develop and execute a research project. The ability to develop a research design to assess a central question, followed by identifying the appropriate statistical methods for evaluating hypotheses generated to test that question, is at the core of this project. All students are encouraged to tie these projects into dissertation or thesis research, and, furthermore, should consider producing research that may be turned into a peer-reviewed publication.

Course Objectives

- Have a clear understanding of the variety of statistical methods available within and beyond the general linear model, including their limitations and assumptions.
- Be able to independently determine which statistical methods are most appropriately applied to your data.
- Learn the critical evaluation skills necessary to judge the statistical veracity of your interpretations and those made by others.
- Become more proficient with and expand experience with computer programming, especially within the R environment.

Prerequisite

Students enrolled must have completed EEB 560 or ANTH 504 with at least a B, or have completed comparable courses (e.g., STAT 537 & 538) with a B. EEB 614 / ANTH 604 is a Level B Course in the Intercollegiate Graduate Statistics Program (IGSP).

Technology and Generative Artificial Intelligence Policies

You will need to have R installed on your computer, as well as R Studio. Any computer that is compatible with R 4.0+ (which is downloadable for free from **cran.r-project.org**) is acceptable. R Studio is free and may be downloaded from **rstudio.com**. You will additionally need to install JAGS and Stan on your computer, which are compliable C++ programs. I will explain how to install these in a separate document provided on Canvas.

In this course, it is expected that all submitted work is produced by the students themselves. You may be tempted to use generative AI tools like ChatGPT to develop R code, **but I strongly discourage this**. Why? For one, you will become a better coder and user of R if you learn the coding yourself, especially as in many instances generalized code in R will not be applicable to the specific needs of your research question. If you do not understand how to write and read code in R, furthermore, you will not be able to evaluate if the code generated by an AI tool is accurate if you do not know how to write the code yourself! Do not use generative AI tools on your critiques or your term paper project. Use of a Generative AI tool to complete an assignment constitutes academic dishonesty.

Course Structure

Class will meet twice weekly, on Tuesdays and Thursdays, as a lecture-based seminar on Tuesdays and a practicum on Thursdays. Below is a summary of what to expect for general class structure each week:

Open discussion. The first part of class consists of a brief open discussion of questions and topics from the previous course meeting. This is meant to be a brief, informal discussion before lecture.

Lecture. I will deliver a seminar-style lecture on each Tuesday unless otherwise specified in the course schedule. The goal of these lectures is to introduce the statistical theory for analyzing specific analytical cases (see the Course Schedule). As noted in the course objectives, it essential for you to understand which statistical solutions to choose for your questions and data; lectures will be guided toward helping you develop confidence in determining what analytical approaches to take, and how to interpret the results. Some math and equations will be introduced as appropriate.

Practicum. Each Thursday class session will consist of a practicum in which simulated research questions and real data are introduced for statistical analyses in R. R script files will be provided via Canvas to guide you through the practicum. During some weeks, you will be asked to work in small groups, and others we will work together as a class. You are encouraged to submit any data sets you have to Dr. Auerbach for use during these practica; please do this at least a week before you want it to be explored in class. Open discussion is strongly encouraged.

Course Web Site

All course announcements and materials—including scanned copies of course readings, data sets, and supplemental study materials—are available from UT's Canvas site: online.utk.edu.

Course Readings

The course largely draws from one textbook, which you are encouraged to purchase:

Inchausti, P. (2023). *Statistical Modeling with R: A Dual Frequentist and Bayesian Approach for Life Scientists*. Oxford. (ISBN: 978-0-19-285902-0)

In addition, course readings will be taken from the following sources. These are posted to Canvas:

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, *567*(7748), 305–307. https://doi.org/10.1038/d41586-019-00857-9 (and comments)

Kruschke JK. 2015. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Second edition. New York: Academic Press. (ISBN: 978-0124058880) *A new edition is anticipated in October 2026.

McElreath. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Second edition. Boca Raton, FL: CRC Press.

Stang, A., Poole, C., & Kuss, O. (2010). The ongoing tyranny of statistical significance testing in biomedical research. *European Journal of Epidemiology*, *25*(4), 225–230. https://doi.org/10.1007/s10654-010-9440-x

VanderWeele, T. J. (2010). Re: The ongoing tyranny of statistical significance testing in biomedical research. *European Journal of Epidemiology*, *25*(11), 843–845. https://doi.org/10.1007/s10654-010-9507-8

Assessment

Some advice: While the deadlines below are absolute deadlines, you should aim at submitting critiques, the term paper abstract, and the term paper before those dates. Your professor would be grateful if all the papers were *not* turned in at the last minute!

PARTICIPATION (15% of grade): This is a graduate level course, and so you are fully responsible for your active presence in the classroom. If you are not attending class (barring legitimate reasons, such as those related to health, family, research, or conferences), then your knowledge will suffer. Learning is a collaborative exercise, and at this advanced level, you will get out of the course what you put into it. At minimum, I expect you to participate in the practica by working on analyses and, when taking part in a small group, contributing to the task at hand. We all look forward to sharing thoughtful conversations about the concepts, limitations, and applications of methods.

CRITIQUES (15% of grade each, totaling 30% of grade): An explanation of critique writing and grading is provided in a separate document on Canvas. You are required to independently select a peer-reviewed journal paper and critique the research design and the use of statistics by the author(s). You will write two critiques, the first due on **30 September** and the second on **6 November. Submit your critique and the paper that you are critiquing to me via e-mail (to auerbach@utk.edu).** In each critique, you should briefly summarize the hypotheses and goals of the study, explain the kinds of data collected by the authors, describe the statistical methods used, and argue the efficacy of the approach and analysis. *Remember that constructive criticism involves weighing both the merits and shortcomings of an argument, including the validity of interpretations based on the evidence provided by the analyses performed.* I provide an example of a good critique for you on Canvas.

TERM PAPER (45% of grade): In lieu of a final examination, you should spend the semester developing a simple research question that may be tested using your own data, or data available from many online repositories (e.g., Dryad, MorphoSource, or figshare, among others). You should write up to a full-length journal manuscript (25-page maximum, 12-point double-spaced) paper outlining your hypothesis, justifying the types of data you choose to utilize in testing the hypothesis and the statistical methods you decide are appropriate. The paper should then provide the results of the statistical analyses and a short interpretation of those results. It is not expected that you perform an extensive amount of background reading for this paper, though a demonstration of relevance to other studies previously performed is necessary. **Detailed guidelines are available on Canvas.**

You must have a research topic and data selected no later than 14 October, and preferably before. Please turn in a proposal up to 300 words outlining your question and planned statistical methods by that date to me via e-mail by 5:00 P.M. on that date. Term papers must be submitted to me via e-mail by 12:00 P.M. on 5 December. You must also include a copy of the R script you used and all statistical output generated in constructing your analysis.

Do not turn in drafts of papers previously submitted for publication, or copies of your master's thesis; this paper should reflect relatively new work. MORE IMPORTANTLY, YOUR PAPER MUST REPRESENT YOUR INDEPENDENT WORK. DO NOT COLLABORATE WITH YOUR CLASSMATES OR OTHERS ON YOUR TERM PAPER. (However, of course, I am always available via e-mail to discuss your paper and address questions or problems you may be encountering.)

TERM PAPER PRESENTATION (10% of grade): In addition to the term paper itself, you will be required to present your term paper results in a professional meeting format talk in class. These presentations should include a PowerPoint or similar slide presentation and should be rehearsed. **Each talk should be 10 minutes**. Presentations consist of:

- Establishing the question that you are asking (1 slide)
- Providing context for the question (1 slide)
- Summary of source of data (1 slide)
- An overview of methods used to analyze the question (1 slide)
- Presentation of results and conclusions (3-4 slides as necessary)

You will be evaluated on the completeness of your presentation, the appropriate presentation of information on slides, slide legibility, and professional presentation. **Presentations will take place on December 1st (at a time TBD) and December 2nd in class.**

Course Schedule

See the table on page 6. All the deadlines for paper submissions in this course are hard, final target dates. Submission of assignments before those deadlines is encouraged!

IMPORTANT DATES:

30 September – 1st Critique Due
6 November – 2nd Critique Due
14 October – Term paper proposal due
1 & 2 December – Term paper presentations
Term papers are due on December 5th

Additional resources

There are numerous resources available to help you learn more about Bayesian methods and their applications. Students seeking more information about the history and application of statistical methods may want to seek these books:

Chivers, T. (2024). *Everything is Predictable: How Bayesian Statistics Explain Our World*. Atria Publishers. **This is a popular press attempt at exploring how Bayesian reasoning underlies modern thought and biases.**

Gelman, A., Hill, J., & Vehtari, A. (2021). *Regression and Other Stories*. Cambridge University Press. **This is a great look at the practical use of data in analysis.**

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis.* Third edition. CRC Press. **This is a good technical resource for all types of Bayesian analysis.**

McGrayne, S. B. (2012). *The theory that would not die: How Bayes' rule cracked the Enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy*. Yale University Press. **This is a good overview of the history of Bayes Theorem.**

For questions about R coding, a good place to start are helpfiles and package guides provided through CRAN. For specific use questions, you may seek resources on the internet including Stack Overflow, though be aware that sites like this can be of mixed utility (and people posting are often snarky). A good place to look is the *Use R!* series published through Springer. These are available through the library, though be aware some may be dated given updates to R and to packages occur faster than the rate at which books are published.

Students with Learning Needs

If you require accommodation because of special needs in learning, please contact the Office of Disability Services through their website: https://sds.utk.edu. You may also call them at 865-974-6087. Please also contact me immediately via e-mail after you register with the Office of Disability Services. Arrangements will be made to adjust the course to fit your needs.

Make-Up Policy

If you become sick (with Covid, flu, or any other cause), with notice, you will be accommodated. Understandably, if you are sick, I do not expect you to attend lectures, even virtually, though you will need to make up any work missed. Legitimate athletic, religious, legal or medical reasons all qualify for eligibility to make up assignments or request extensions on course deadlines. If you must miss a lecture or cannot turn in any materials required over the semester, you must contact Dr. Auerbach ***before*** the lecture or deadline.

COURSE SCHEDULE: Bayesian Data Analysis for Natural and Social Sciences

DATE	TOPIC	PRACTICUM	READINGS
19 & 21 August	Introduction to the course History of statistics Frequentist vs. Bayesian	R markdown and scripts	*Read by Thursday:* Kruschke, chapter 3 Inchausti, chapter 2
26 & 28 August	Problematizing NHST Probability & Bayesian modeling	Random sampling and data management in R	Stang et al. 2010 VanderWeele 2010 Amrhein et al. 2019 Kruschke, chapter 4
2 & 4 September	Estimating parameters	Modeling parameters in R	Inchausti, chapter 3
9 & 11 September	Power & effect size in statistics	An introduction to MCMC & JAGS	Kruschke, chapter 13 (Tuesday) Kruschke, chapters 7 & 8 (Thursday)
16 & 18 September	General linear models: simple regression	Bayesian model fitting and validation	Inchausti, chapter 4.1 – 4.7
23 & 25 September	General linear models: multiple / polynomial regression	Bayesian model fitting & standardization	Inchausti, chapter 4.8 – 4.12
30 Sept. & 2 October	General linear models: categorical variables	Bayesian *t*-test and ANOVA	Inchausti, chapter 5
NO CLASS ON 7 OCTOBER – FALL BREAK			
9 October	Model selection	*No practicum*	Inchausti, chapter 7
14 & 16 October	General linear models: interactions Intro to generalized linear models	Bayesian ANCOVA	Inchausti, chapters 6 and 8
21 & 23 October	Experimental design in biological sciences Tackling maximum entropy	Maximum entropy and GLM	Inchausti, chapter 14 McElreath, chapter 10
28 & 30 October	Binary responses in GLM	Modeling GLM with binary outcomes	Inchausti, chapter 9
NO CLASS ON 4 NOVEMBER – WORK ON TERM PAPERS AND GO VOTE!			
6 November	Count responses in GLM	*No practicum*	Inchausti, chapter 10
11 & 13 November	Continuous responses in GLM	Modeling GLM with continuous outcomes	Inchausti, chapter 12
18 & 20 November	Introduction to hierarchical models	Hierarchical models in R	Inchausti, chapter 13 (skim 15)
25 November	Discussion: implementing Bayesian methods in your research	*No practicum*	--
1 & 2 December	Final project presentations – REMINDER, TERM PAPERS ARE DUE ON 5 DECEMBER		