ANTHROPOLOGY 604: ANTHROPOLOGICAL STATISTICS II

THE UNIVERSITY OF TENNESSEE – SPRING 2021

Instructor: Dr. Benjamin M. Auerbach
Office: 416 Strong Hall (Dr. Auerbach will not be available at his office)
Office Hours: Dr. Auerbach is available for individual meetings via Zoom, requested via e-mail.
E-mail: auerbach@utk.edu
(Dr. Auerbach typically responds within 24 hours of receiving an e-mail. More urgent matters should be marked as URGENT in the e-mail subject line.)
Course website: Canvas page (utk.instructure.com)

COURSE MEETING DETAILS

This is an advanced graduate lecture course in multivariate statistics and Bayesian approaches. **All course meetings will take place exclusively online via Zoom video conferencing.** Our course meetings will occur synchronously throughout the semester. Participation and attendance on Zoom is mandatory barring legitimate reasons for absence and notification to me before class. If you experience health or family-related concerns over the course of the semester, please know that I will work with you to minimize any impact anent course content and assignments. See the attendance policy at the end of this syllabus for more details. Please note that all course meetings will be recorded for those who cannot attend. Details about course meeting structure may be found on the next page.

Meeting Times: Tuesdays and Thursdays, 11:30 A.M. – 12:45 P.M.
Zoom meeting address: **https://tennessee.zoom.us/j/92178104012**
The password for the Zoom session will be posted to Canvas.

COURSE AT-A-GLANCE

- Emphasis is on advanced research design, including skills in programming and critical thinking about the application of quantitative analytical approaches.
- Statistical concepts covered include multivariate models, data reduction methods, clustering methods, spatial & distance statistics, path analysis & structural equation modeling, Bayesian approaches to linear models, and generalized linear mixed models.
- Assessment is based on formal critiques of published papers, independent research paper, and weekly participation.
- Expect to read approximately 100 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. Many are discussed within the context of anthropological inquiry, though examples are drawn from throughout the scientific literature to illustrate the proper use (as well as misuses) of these methods. Techniques include frequentist methods for data reduction and association, as well as methods associated with maximum likelihood and Bayesian methods. Some of these methods are still in nascent stages of adoption within anthropology, but it is important for you to be aware of them given their potential utility in addressing a variety of scenarios.

We will approach this course as a survey of research design and the appropriation of quantitative methods to those analyses. That is, 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 in ANTH 604, 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: mathematical conceptual thinking and programming. Knowledge of advanced mathematics is not required, but conversancy with linear algebra and the concepts of basic calculus are useful. For this reason, we take a couple weeks to discuss these fundamental concepts and their relationship with familiar statistics. Programming, namely in the R environment, is emphasized and taught throughout the course.

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

COURSE STRUCTURE

Class will meet twice weekly, on Tuesdays and Thursdays, as a lecture-based seminar on Tuesdays and a practicum on Thursdays. 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. 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 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 on Zoom, 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.

WHAT IS R? WHY AREN'T WE USING SAS / NCSS / SPSS / ETC.?

R is a freely downloadable computer environment that allows for great flexibility in programming, especially for statistical purposes. While other statistical software packages are in widespread use (such as SAS, NCSS, SPSS, STATA, and JMP), and allow for some user flexibility in modifying existing analytical packages, these programs are deficient in that they have idiosyncratic, regimented data handling structures, and generally allow a limited scope of statistical analyses. For most of your research needs, you will likely find that any of these other software packages will be more than adequate. Yet, if you want to use a less common statistical method, a new statistical approach, or most Bayesian approaches, these other software packages often do not offer solutions. R offers a full spectrum of statistical analyses, from Student's *t*-tests to Bayesian generalized linear mixed models, and everything in between.

You are not expected to become an R Guru™, but you should expect to develop basic competency in the language. Anyone who has worked extensively with R will tell you that most of the learning for the software is through individual trial and error, so be persistent and don't give up easily! To help you develop more advanced R programming abilities, Dr. Auerbach will be providing tutorials, and you should look at acquiring a copy of *Biostatistical Design and Analysis Using R* and *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Moreover, there are abundant help resources available:

- R itself has built-in query and help functions. The authors of packages in the software write many of these, so the help may be of varying quality, but it's the first place to look.
- There are excellent internet resources. A personal favorite is Stack Overflow, a Q&A site for programmers. The people behind CRAN maintain a journal and regularly post manuals and questions.
- New resources are always appearing to make R more useable and user-friendly. For example, a handful of individuals initiated an internet-based resource called Bayesian First Aid in 2014 that they maintain on Github (https://github.com/rasmusab/bayesian_first_aid). So, search engines may be your best friend when all else fails.

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.

REQUIRED AND SUGGESTED TEXTS

All course readings will be provided on Canvas as scanned copies. There is no single textbook that covers the variety of statistics addressed in ANTH 604 (though we will rely a lot on Kruschke's 2015 textbook). A bibliography of all texts listed in the course schedule is provided below. You may look into acquiring a copy of the books listed below, including those by Logan, by Strang, and by Warner. Full digital copies of Kruschke's and McElreath's books are provided on Canvas, though you if can find hard copies of these affordably I recommend acquiring them.

Albert J. 2009. *Bayesian Computation with R*. Second edition. New York: Springer. (ISBN: 978-0-387-92297-3)

Amrhein V, Greenland S, and McShane B. 2019. Retire statistical significance. *Nature* 567:305-307. (and comments)

Buck CE, Cavanagh WG, and Litton CD. 1996. *Bayesian Approach to Interpreting Archaeological Data*. New York: John Wiley & Sons. (ISBN: 0-471-96197-3)

Claude J. 2008. *Morphometrics with R.* New York: Springer. (ISBN: 978-0-387-77789-4)

Hadfield J. 2018. MCMCglmm Course Notes. Online Publication: https://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf.

Hair JF Jr., Black WC, Babin BJ, and Anderson RE. 2010. *Multivariate Data Analysis*. Seventh edition. New York: Prentice Hall. (ISBN: 978-0-13-813263-7)

Konigsberg LW, and Frankenberg SR. 2013. Bayes in biological anthropology. *American Journal of Physical Anthropology* 57:153-184.

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

Lee S-Y, and Song X-Y. Bayesian structural equation model. *WIREs Comput Stat* 6:276-287.

Logan M. 2010. *Biostatistical Design and Analysis Using R*. New York: Wiley-Blackwell. (ISBN: 978-8126550708)

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

Otárola-Castillo E, and Torquato MG. 2018. Bayesian statistics in archaeology. *Annual Reviews in Anthropology* 47:435-453.

Pugesek BH, Tomer A, and von Eye A (eds.). 2003. *Structural Equation Modeling*. Cambridge: Cambridge University Press. (ISBN: 0521781337)

Song X-Y, and Lee S-Y. 2012. A tutorial on the Bayesian approach for analyzing structural equation models. *J Math Psychol* 56:135-148.

Smith RJ. 2018. The continuing misuse of null hypothesis significance testing in biological anthropology. *American Journal of Physical Anthropology*. DOI: 10.1002/ajpa.23399

Strang G. 2016. *Introduction to Linear Algebra*. Fifth edition. Wellesley, MA: Wellesley Cambridge Press. (ISBN: 978-0-9802327-7-6)

Warner RM. 2012. *Applied Statistics: From Bivariate through Multivariate Techniques*. Second edition. Los Angeles: SAGE Publications. (ISBN: 978-1412991346)

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 of the papers were *not* turned in at the last minute!

PARTICIPATION (10% of grade): This is a graduate level course, and so you are fully responsible for your active presence in the classroom, even as we meet virtually. 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). Two of these should be produced and submitted via e-mail on **26 February** and **26 March** by 5:00 PM. 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, data available from many online repositories (e.g., Dryad, MorphoSource, or figshare, among others), or data that I can provide. 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 March 4th. Please turn in an abstract (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 May 1st. You must also include a copy of all statistical output used 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 (15% 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 will be given a 15-minute presentation period, followed by group Q&A. You will be evaluated on the completeness of your presentation (setting up the question, providing background, explaining methods, and presenting results & conclusions), the appropriate presentation of information on slides, slide

legibility, and professional presentation. **Presentations will take place on April 30th from 12:00 to 2:00 and from 2:30 to 4:00 P.M.**

COURSE SCHEDULE

See the table on the next page. We will take a break during the week of April 5th, which you should use to work on your projects and take a break from Zoom.

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

Important Dates

21 January – First course meeting
26 February – Critique #1 hard deadline
4 March – Term paper abstract hard deadline
26 March – Critique #2 hard deadline
30 April – Term paper presentations
1 May – Term paper hard deadline

STUDENTS WITH TECHNOLOGICAL NEEDS

If you do not have access to a stable internet connection, or lack the computing resources necessary to access the materials for this course, you may contact the Office of Information Technology (http://oit.utk.edu) to request a personal hotspot or a loaner laptop for use during the semester. Due to high demand, please contact them well in advance of the semester if possible.

STUDENTS WITH LEARNING NEEDS

If you require accommodation because of special needs in learning, please contact the Office of Disability Services at 2227 Dunford Hall (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 the novel coronavirus, 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: ANTHROPOLOGICAL STATISTICS II (ANTH 604) – SPRING 2021

DATE	TOPIC	PRACTICUM	READINGS
21 January	Introduction to the course Fundamental statistical concepts Power & effect size in statistics	An introduction to R (self-guided)	Logan, chapters 1-6 Kruschke, chapters 3 & 13
26 & 28 January	A primer in linear algebra concepts	Matrices in R	Healy 1986 (Strang, chs. 1-3, 5, 6)
2 & 4 February	Linear algebra in statistics A matrix algebra view of MANOVA	MANOVA in R	Strang, chapter 12 Warner, chapters 19 & 18
9 & 11 February	Data reduction methods Clustering methods & distances	Factor analysis & Hclust in R	Warner, chapter 20 Hair et al. 2010
16 & 18 February	Morphometric data analysis	Morphometrics in R	Claude, chapters 1-4 (skim 5 & 6)
23 & 25 February	Problematizing NHST Probability & Bayesian modeling	Introducing Bayesian packages in R (e.g., Bayesian First Aid)	Amrhein et al. 2019 (and responses) Smith 2018 Kruschke, chs. 4-6, & 11-12
2 & 4 March	The Markov Chain Monte Carlo Sampling parameter space	An introduction to JAGS & MCMCglmm	Kruschke, chapters 7 & 8 McElreath, chapter 9 Hadfield 2018
9 & 11 March	Hierarchical models	Hierarchical modeling in R	Kruschke, chapters 9 & 10 McElreath, chapter 13
16 & 18 March	Generalized linear models Thinking about model fits	GLM in R	McElreath, chapter 10 Kruschke, chapter 15 *Additional reading*: Albert, chs. 6-8
23 & 25 March	Bayesian linear models	Bayesian linear models in R	Kruschke, chapters 16 & 17
30 March & 1 April	Bayesian ANOVA	Bayesian ANOVA in R	Kruschke, chapter 19
NO CLASS WEEK OF 6 APRIL – Work on your projects!			
13 & 15 April	Reporting Bayesian results Bayesian applications to archaeology	Change point analysis and other archaeological applications	Kruschke, 25.1 *Readings from* Buck et al. 1996
20 & 22 April	Multiple regression and structural equation modeling	Multiple regression in R	*Readings from* Pugesek et al. 2003 Kruschke, chapter 18 Song & Lee 2012; Lee & Song 2014
27 April	Discussion: Implementing Bayesian approaches – when and why	--	Konigsberg & Frankenberg, 2013 Otárola-Castillo & Torquato, 2018

30 April – Term Paper Presentations via Zoom
1 May – Term Papers Due