ANTHROPOLOGY 504:
ANTHROPOLOGICAL STATISTICS I

THE UNIVERSITY OF TENNESSEE – KNOXVILLE
FALL 2016

Instructor: **Dr. Benjamin M. Auerbach**
Contact information:
Office: 229 South Stadium Hall
Office hours: Walk-in hours: Thursdays, 3:45 – 5:00 P.M.
By appointment (sign up via direct e-mail to Dr. Auerbach)
E-mail: auerbach@utk.edu

Time: Thursdays, 5:45 P.M. – 8:35 P.M.
Location: 427 Hesler Biology Building

Course Description

Three fundamentals are universally essential in any good research: 1) you know how to determine the kinds of data necessary to address your research questions (i.e., research design); you know the appropriate methods with which to analyze those data; and, most importantly, you have the skills to critically interpret the results of the analyses. All three of these, especially the latter two, fall within the realm of statistics.

This is the first in a two-part course sequence on research design and analysis tailored to address anthropological research. In this course, we will develop the foundations for statistical analysis, with special emphasis placed on the fundamentals above. We will discuss methods within the context of specific research circumstances. For example, how do you compare two samples? The number of possible approaches will surprise you, but not all are equally valid!

Our focus will be on frequentist methods, especially the parametric general linear model and non-parametric rank statistics. However, I will note Bayesian methods whenever they are appropriate. The distinction between these schools of statistical analysis will be explained early in the course. A central focus in the course will be on how to interpret results, and an understanding of statistical significance (including the problems of null hypothesis statistical testing, or NHST).

The course uses the R programming environment. Do not worry if you have never had any programming experience; you will gain knowledge and basic competency over the course of the semester. Likewise, some basic mathematical knowledge will be emphasized, though no math beyond linear (matrix) algebra will be used in the course.

Course Objectives

- Learn research design for your specific research questions, including identification the kinds of data that you will need to collect to address those questions.
- Develop competency for independent research and analysis.
- Gain the critical evaluative skills necessary to discern the statistical veracity of your conclusions based on the analytical methods that you choose to use.
- Become familiar with and begin to develop fluency with computer programming, especially within the R environment.

Prerequisite

There are no course prerequisites. However, undergraduates who enroll in the course should be aware that this course is taught exclusively at a graduate level, and expects a workload commensurate with that level of advanced training. Anthropological knowledge and a background in linear algebra will help significantly, but are not required.

Course Structure

Class will meet once a week as a lecture-based seminar with a practicum at the end of all classes. You are encouraged to bring a laptop computer to use for taking notes as well as the practicum portion of the class. Any computer that is compatible with R 3.2+ (which is downloadable for free from http://cran.r-project.org) is acceptable. Students using Windows-based computers should also look to acquire R Studio, which provides a GUI for R. Here 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 formal lecture.

Formal lecture. Dr. Auerbach will deliver a formal lecture for the first half of class. The goal of these lectures is to introduce the formal statistical theory for analyzing specific analytical cases (see the Course Schedule). Frequentist approaches will be emphasized, but Bayesian alternatives will be introduced whenever possible. As noted in the course objectives, it essential for you to understand which statistical solutions to choose for your questions and data; formal lectures will be guided toward helping you develop confidence in determining what analytical approaches to take, and how to interpret the results. Some basic math and equations will be introduced as appropriate.

Practicum. The last portion of each class session will consist of a practicum, in which simulated research questions and real data are introduced for statistical analyses. During some weeks, you will be asked to work in small groups, and others we will work together as a class. Analyses will be performed using R in the classroom. 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 during this portion of class meetings. Dr. Auerbach may show how to perform analyses in other programs (e.g., SPSS or STATA) outside of class meeting times by student request, though you are strongly encouraged to use R.

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 within one semester, 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 along your first steps, Dr. Auerbach will be providing tutorials, and you should look at acquiring a copy of *Discovering Statistics Using R*. 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. For example, a couple individuals recently initiated an internet-based resource called Bayesian First Aid. So, search engines may be your best friend when all else fails.

Required and Suggested Texts

Anthropological Statistics I uses Zar's *Biostatistical Analysis* as a principal resource, though be aware that it is one of many sources that will be usefully accessed throughout the course. You are strongly encouraged to also acquire a copy of Andy Field and colleagues' *Discovering Statistics Using R*. Other resources (see below) are provided during the course via the course web site.

Zar, Jerrold H. 2009. *Biostatistical Analysis*. Fifth edition. New York: Prentice Hall.

Field, Andy, Jeremy Miles, Zoë Field. 2012. *Discovering Statistics Using R*. London: SAGE Publications Ltd.

Other resources used in this course

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

Bolstad WM. 2007. *Introduction to Bayesian Statistics*. Second edition. New York: Wiley-Interscience. (ISBN: 978-0-470-14115-1)

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

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

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

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 Blackboard site: **bblearn.utk.edu.** Make visiting the site a regular practice!

Assessment and Term Paper

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 (25% 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 in class (barring legitimate reasons, such as those related to health, family, research, or conferences), then your knowledge will suffer—as will your grade. With up to 15 students enrolled, I do not expect you to engage in the open discussion every week. However, I do 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.

CRITIQUES (15% of grade each): 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 **3 October and 7 November by 6:00 PM**. Submit both your critique and the paper you are critiquing. 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.*

An explanation of critique writing and grading is provided in a separate document on Blackboard. An example of a good critique may also be found on Blackboard.

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 provided by Dr. Auerbach. You should write a brief (15 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.

You must have a research topic and data selected no later than October 17th. Please turn in an abstract (up to 300 words) outlining your question and planned statistical methods by that date to Dr. Auerbach via e-mail. Term papers must be submitted to Dr. Auerbach by 5:00 P.M. on December 5th. 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 Masters 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, Dr. Auerbach is always available via e-mail and in his office to discuss your paper and address questions or problems you may be encountering.)

Students with Special 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 Dr. Auerbach 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.

Course Schedule

Again, 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

18 August – First course meeting
3 October – Critique #1 hard deadline
17 October – Term paper abstract hard deadline
7 November – Critique #2 hard deadline
9 December – Term paper hard deadline

COURSE SCHEDULE: ANTHROPOLOGICAL STATISTICS I (ANTH 504) – FALL 2016

Optional readings are in burgundy text. Readings from sources other than Zar will be made available on Blackboard.

DATE	TOPIC	PRACTICUM	READINGS
18 August	Introduction Philosophy of science & logic Null hypothesis significance testing	A brief introduction to R	Warner, Chapter 3 Zar, Chapter 1 Field et al., Chapters 1 & 3
25 August	Statistical models & probability Data Parameters and distributions	Data entry in R	Zar, Chapters 1 – 6 Field et al., Chapters 2 – 3
8 September	Graphics & statistical reporting Assumptions in statistics Statistical power	Dataframes and Quartz in R	Warner, Chapter 3 Field et al., Chapters 4 & 5 *Examples provided on Blackboard*
15 September	Bayes' Theorem Odds ratios	Descriptive statistics in R	Konigsberg and Frankenberg, 2013 Buck et al., Chapters 1 & 2 Bolstad, Chapter 4
22 September	One-sample inference	One-sample *t*-tests in R	Zar, Chapter 7 Field et al., Chapter 9
29 September	Two-sample inference	Two-sample *t*-tests in R Bayes Factor for competing models	Zar, Chapters 8 & 9 Albert, Chapter 8 Field et al., Chapter 9
3 October	First Critique Due By 6:00 P.M. (e-mail to Dr. Auerbach)		
13 October	Nonparametric methods Chi-square and Wilcoxin statistics	Non-parametric approaches in R	Zar, Chapters 7-9 & 22 Field et al., Chapter 15
20 October	Data transformations Resampling & iterative methods	MCMC in R	Albert, Chapter 6
27 October	Correlation & causation Introduction to regression	Correlation in R	Zar, Chapters 17 & 19 Field et al., Chapter 6
3 November	The general linear model (GLM) Regression Is GLM appropriate for anthropology?	Regression in R	Zar, Chapters 17 & 18 Field et al., Chapter 7

DATE	TOPIC	PRACTICUM	READINGS
7 November	Second Critique Due By 6:00 P.M. (e-mail to Dr. Auerbach)		
10 November	Analysis of variance (ANOVA)	Regression in R (continued)	Zar, Chapter 10
17 November	ANOVA (continued) Multiple regression	ANOVA in R	Zar, Chapters 20 & 12 Field et al., Chapters 7 & 10
21 November	Introducing MANOVA Kruskal-Wallis test	MANOVA in R	Zar, Chapters 10 & 16 Warner, Chapter 19 Field et al., Chapter 16
1 December	Discriminant analysis as a post hoc Post hoc analyses	Post hoc analyses in R	Field et al., Chapter 16 Warner, Chapter 18
5 December	**TERM PAPERS ARE DUE BY 5:00 P.M.!**		