ANTHROPOLOGY 604:
ANTHROPOLOGICAL STATISTICS II

THE UNIVERSITY OF TENNESSEE – KNOXVILLE
SPRING 2015

INSTRUCTOR: DR. BENJAMIN M. AUERBACH
CONTACT INFORMATION
OFFICE: 229 SOUTH STADIUM HALL
OFFICE HOURS: WALK-IN HOURS: WEDNESDAYS, 2:30 – 3:30
BY APPOINTMENT (SIGN UP VIA DIRECT E-MAIL TO DR. AUERBACH)
E-MAIL: AUERBACH@UTK.EDU
(DR. AUERBACH DOES NOT READ E-MAILS AFTER 8:00 P.M.)
TIME: TUESDAYS AND THURSDAYS, 5:45 – 8:30 P.M.
LOCATION: 57 HUMANITIES & SOCIAL SCIENCES BUILDING

Course Description

The main goal of ANTH 604 is to introduce analytical methods and demonstrate the practical applications of statistics to anthropological inquiry. This course introduces an assortment of advanced, specialized statistical methods that have utility in various anthropological settings. These techniques include frequentist methods for data reduction and association, as well as methods associated with maximum likelihood and Bayesian methods. Some of these methods (such as spatial analysis, finite mixture and transition analysis) are still in nascent stages of adoption within anthropology, but it is important for students to be aware of them given their potential utility in addressing a variety of scenarios.

Students enrolled in this course should approach it 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 students seeking proficiency should seek out specialty courses or be prepared to undertake rigorous independent study. With the knowledge gained in ANTH 604, however, students will have the basic tools with which to venture into these additional studies.

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 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 3.0+ (which is downloadable for free from http://cran.r-project.org) is acceptable. 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). 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 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, 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 *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

All course readings will be provided on Blackboard as scanned copies. There is no single textbook that covers the variety of statistics addressed in ANTH 604. A bibliography of all texts listed in the course schedule is provided below. Copies of Zar's *Biostatistical Analysis* and Field et al.'s *Discovering Statistics Using R* will be useful references, though most of the statistics discussed in this course are not included in those books.

Students may look into acquiring a copy of the following books, especially if they may be found used or at reduced prices: Strang's *Introduction to Linear Algebra*, Warner's *Applied Statistics*, and Buck et al.'s *Bayesian Approach to Interpreting Archaeological Data*. Full citations of these are below.

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 (10% 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 **24 February and 31 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.*

An explanation of critique writing and grading is provided in a separate document 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 up to a full-length journal manuscript (30 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 March 3rd. 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 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 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.)

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, with five minutes for 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. Additional guidelines will be provided to you as a supplementary document on Blackboard.

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

13 January – First course meeting
24 February – Critique #1 hard deadline
3 March – Term paper abstract hard deadline
31 March – Critique #2 hard deadline
28 April – In-class presentations of term papers
1 May – Term paper hard deadline

COURSE SCHEDULE: ANTHROPOLOGICAL STATISTICS II (ANTH 604) – SPRING 2015

DATE	TOPIC	PRACTICUM	READINGS
13 January	Introduction to the course A recapitulation of basic statistical models	A crash course in R	Field et al., Chapters 3 & 4
20 January	Linear algebra for statistics: basics of matrices, determinants, inversions, transposition, eigenvalues	Matrices in R	Strang, Chapters 1-3, 5 & 6
27 January	Applied linear algebra: MANOVA & discriminant analyses	MANOVA in R	Warner, Chapters 18 & 19 Field et al., Chapter 16
3 February	Factor analyses and principal components	Factor analysis in R	Warner, Chapter 20
10 February	Cluster analysis Distances (including Mahalanobis *D*)	Cluster analysis	Hair et al.: Cluster analyses
17 February	Landmark data Procrustes analysis	Euclidean distance analyses	*Selections from* Bookstein Claude, Chapter 3
24 February	Logistic regression (logit)	Logistic regression in R	Warner, Chapter 23 Field et al., Chapter 8
3 March	Maximum likelihood	Maximum likelihood in R	Eliason
10 March	Thinking like a Bayesian	An introduction to BUGS	Albert, Chapters 2 & 3 Buck et al., Chapters 1-4
24 March	Bayesian inference	An introduction to MCMCglmm	Buck et al., Chapters 7 & 8 Albert, Chapter 5
31 March	**CLASS CANCELLED (Dr. Auerbach at Anatomy Conference)**		
7 April	Model construction Information criteria (AIC, BIC, DIC)	MCMCglmm in R	McCarthy, Chapter 4 Albert, Chapters 6-8
14 April	Archaeological applications for Bayesian approaches: Spatial analysis & provenience analysis	Archaeological examples	Buck et al., Chapters 10 & 11
21 April	Biological applications for Bayesian approaches: Model-fitting and finite-mixture analysis	Biological anthropology examples	Roseman & Auerbach, 2015 Kramer & Konigsberg, 1999
28 April	**In-class presentations of term papers**		
1 May	**TERM PAPERS ARE DUE BY 5:00 P.M.!**		

Readings

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

Bookstein FL. 1992. *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge: Cambridge University Press. (ISBN: 9780521383851)

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

Eliason SR. 1993. *Maximum Likelihood Estimation: Logic and Practice*. Quantitative Applications in the Social Sciences No. 96. Los Angeles: SAGE Publications. (ISBN: 0-8039-4107-2)

Field A, Miles J, and Field Z. 2012. *Discovering Statistics Using R*. Los Angeles: SAGE Publications. (ISBN: 978-1446200469)

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)

Kramer A, and Konigsberg LW. 1999. Recognizing species diversity among large-bodied hominids: a simulation test using missing data finite mixture analysis. *Journal of Human Evolution* 36:409-421.

McCarthy, MA. 2007. *Bayesian Methods for Ecology*. Cambridge: Cambridge University Press.

Roseman CC, and Auerbach BM. 2015. Ecogeography, genetics, and the evolution of human body form. *Journal of Human Evolution* 78:80-90.

Strang G. 2009. *Introduction to Linear Algebra*. Fourth edition. Wellesley, MA: Wellesley Cambridge Press. (ISBN: 978-0980232714)

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

Zar JH. 2009. *Biostatistical Analysis*. Fifth edition. New York: Prentice Hall. (ISBN: 978-0-13-100846-5)