Anthropology 304:
Data Analysis in the Anthropological Sciences

Instructor: James Holland Jones

March 28, 2013

Friday, 9:00-11:50
Spatial Lab, Building 50
Spring Quarter 2013

1 Course Description

In this course, we will develop a statistical toolkit appropriate for biosocial scientists. The emphasis will be on practical data analysis and the development of a problem-solving approach to inference. We will focus on a particular set of models, known collectively as “event history analysis,” of tremendous utility to demography, ecology, and population biology. Extensive prior knowledge of statistics or software is not assumed, but a reasonable background in anthropology or human ecology is. Introductory statistics at the level Stat 60 or its equivalent should be sufficient preparation. The course will require a substantial commitment of time, including extensive use of computers for analysis and presentation work.

2 Software

Statistical Software  We will use a freely-available software package called R. R has a number of advantages: (1) it is free, (2) it is essentially platform independent, (3) R represents the cutting-edge in statistical computation, (4) there is extensive (free) documentation and a large online community of users to provide support, (5) it is used overwhelmingly by professional statisticians and therefore learning R facilitates conversations with statisticians for consulting and collaborative purposes. Did I mention it’s free?

R may be downloaded from one of the CRAN mirror website. The closest mirror lives across the Bay: http://cran.cnr.berkeley.edu/

If you find going to a Berkeley website for the software too distasteful, perhaps you would prefer the UCLA mirror: http://cran.stat.ucla.edu/

Text Editors  In addition to R, students should ensure that they have access to a text editor. Note that this is quite different from a word processor! A text editor works with plain text and is an essential tool for the aspiring data analyst. Choice of text editor is largely a matter of
personal preference. Some suggestions are provided on the course website. I will use an editor called Aquamacs for in-class demonstrations.

3 Problem Sets

Much of the practical benefit from this course will come from working through problem sets; these will be assigned almost every week and will be due in class on the assigned date. Since we will be discussing the problems in class, late assignments will only be accepted under unusual circumstances.

Problem sets will be of three general types: (1) practice problems, (2) vignettes, (3) paper reviews. Vignettes require a more involved write-up. The goal of these vignette problem sets is to get students to think about integrating quantitative and inferential thinking into their scholarly argumentation. Grades for vignette problem sets will reflect the quality of the analysis and the quality of the arguments made concerning relevant anthropological questions. Be thorough (but to the point) in answering all questions posed in the problem-set handout. Results of analyses should be presented in a brief, typed essay that references supporting tables and figures and does not exceed three double-spaced pages of 11-point or larger font, and 1-inch margins (tables, figures, citations are not included in page limit). Especially in the case of non-computer based calculations, show clearly the steps that were followed in arriving at the answer. Figures and tables do not have to be of publication quality, but they should also be clear, neat and appropriately labeled. Practice problem sets do not need to be so formally presented, but should similarly fulfill the spirit of clarity, brevity, and explication of the steps leading to the final answer (i.e., show your work!).

In your problem set write-ups, briefly state the problem at hand. You may assume a reader’s knowledge of methods that have been discussed or presented in class, but beyond that, you need to summarize the results and general argument as you would in an article for publication. Be exact about procedures that have been used—don’t say “correlation” if you mean “product-moment correlation coefficient;” don’t say “factor analysis” if you have just done a principal components analysis. Be clear about the set of data you have analyzed, any transformations you have used, what has been shown quantitatively, and any substantive conclusions you arrive at. With respect to conclusions, be careful not to overstate things or to claim too much—learning when to rein yourself in is part of what we want to teach here. Finally, you should adopt a style of presentation that is relatively serious and formal—again, the model you should have in mind is an article for a professional journal.

You may find it is helpful to coordinate your computing time with other students so that you can assist each other in solving technical problems associated with the use of R. Discussing the methods and issues with each other is fine, but you need to execute procedures and write up problem sets independently—much of the pay-off in this course lies in eventually struggling though difficult concepts on your own.

In addition to the problem sets, every other week there will be a paper summary due. The idea of this summary is to read the paper, critically analyze the methods used, summarize the results and explain how the statistical results bear on the scientific hypotheses.
4 Final Project

This is a graduate class, so the ideal final project involves some data analysis relevant to your own research. If this is not a possibility, then we can find a project that is effectively a longer vignette.

5 Course Website

The course website will be a repository for R code, data sets, problem set answers, links to useful online resources, etc. The url for the course website is: http://anthro304.stanford.edu

The website also contains a blog that we will use for posting results, vignettes, sample code, and paper critiques.

6 Grading

The breakdown of grading for this class will be as follows:

50% Weekly Problem Sets
25% Paper Summaries
25% Final Project

7 Readings

There are two required texts for this class:


In addition, we will make extensive use of the primary scientific literature. These readings are typically provided electronically either through hyperlink or via a secure section of the course website.
8  Course Outline (Subject to Change)

Week 1  Introduction to R, Graphics
Week 2  Probability and Likelihood
Week 3  Hazards, Probabilities, and Survival
Week 4  Linear Models, Mostly About Regression
Week 5  Generalized Linear Models I
Week 6  Generalized Linear Models II
Week 7  Regression Models for Survival Data
Week 8  The Duality of Events
Week 9  Time-Varying Covariates
Week 10 Frailty and Heterogeneity