

ADVANCED QUANTITATIVE METHODS

Course: ANG 7750-001 (Spring 2009)
Friday, 9:40 am – 12:30 pm in SOC 37

Instructor: Dr. E. Christian Wells
cwells@cas.usf.edu, 813/974.2337

Office hours: USF Anthropology Department
Friday, 1:00 pm – 3:00 pm in SOC 116

The world is fast becoming more complex. Drawing inferences and formulating explanations to account for variation in human behavior increasingly requires the use of quantitative approaches to data analysis. Such methods are indispensable for identifying and studying, in a systematic and organized fashion, patterns in empirical information composing numerous variables. This course is intended as an advanced consideration of quantitative methods and reasoning for anthropologists, although other social scientists make use of the same principles and procedures we will examine. *However, this is not a math class; there are no hand calculations.* Most of the techniques are fairly standard and are employed in the “exploratory data analysis” approach. Along with *Discovering Statistics using SPSS (and Sex, Drugs, and Rock ‘n’ Roll)* by Andy Field (Second edition, Sage Publications, 2005), we will examine a wide variety of professional journal articles representing anthropology and allied disciplines. In addition, we will use computers for data analysis with SPSS (v. 16.0), a packaged statistical software program.

Overall, this course is about learning how to use quantitative methods to solve research problems and to improve decision making. Therefore, the primary objectives of this course are: to introduce you to the quantitative analysis of multivariate data using computers; to help you to become a thoughtful and informed consumer of the quantitative anthropological literature; to discuss in detail the sorts of problems that can be addressed quantitatively in anthropology with multivariate data analysis; to provide you with the background necessary for designing research to yield multivariate data that may be easily and productively analyzed using quantitative methods; and to have *fun* learning statistics, oops, I mean statistics.

TOP 10 REASONS TO BECOME A QUANTITATIVE ANTHROPOLOGIST

10. Deviation is considered normal.
9. We feel complete and sufficient.
8. We are “mean” lovers.
7. Statisticians do it discretely and continuously.
6. We are right 95 percent of the time.
5. We can legally comment on someone’s posterior distribution.
4. We may not be normal but we are transformable.
3. We never have to say we are certain.
2. We are honestly significantly different.
1. No one wants our jobs!

-FORMAT, REQUIREMENTS, & EVALUATION-

The class will meet once a week and will consist of three consecutive 50-minute sessions (each separated by a five-minute break). Each session carries responsibilities for you that will be evaluated to compose your final grade for the course.

Session 1: Roundtable

The first session will encompass a “roundtable discussion” in which each student will report their progress on analyzing a multivariate dataset using the methods covered in the previous class. The Roundtable provides a venue for sharing your methods, results, successes, and failures with other students in a cooperative learning framework. Along with your active participation in the Roundtable, you will be responsible for turning in a brief report of your analysis and findings. The report must be typed, and the write-up normally should not exceed three double-spaced pages of text using 12-point TNR font and 1-inch margins; figures, tables, and any calculations may be appended as necessary. For each essay, briefly state the goal of the analysis, describe the methods used to achieve it, and then discuss exactly what of substance you conclude. Some degree of collaborative learning is encouraged, though write-ups should be independent. *You will complete a total of 10 essays, each worth five percent of your final grade.* The first essay is due on January 30 and the last one is due on April 10. Your report must be e-mailed to me before the start of the class in which it is due. If you do not receive an email from me confirming that I received the file, then you can assume that I did not receive it.

Session 2: Seminar

I will begin the second session with an illustrated lecture to provide you with an introduction to the weekly topic. This talk will cover the material in the required reading on method and theory. The presentation will be followed by an in-depth, critical discussion of two case studies that derive from the professional literature. Critical discussion entails evaluation of the strengths and weaknesses of an argument and the empirical data used to support it. For each class, you will read the articles listed under the “examples” section in the schedule and, before coming to class, you will type up and turn in a brief (ca. 250 words) précis for each article. These summaries will form the substance of class discussion, so it is imperative that each précis include: 1) a concise statement of the problem, 2) a well-reasoned critique of the analysis, and 3) a review of the author’s conclusions. *There will be a total of 25 summaries, each worth one percent of your final grade.* Your summaries must be e-mailed to me before the start of the class in which they are due. If you do not receive an email from me confirming that I received the file, then you can assume that I did not receive it.

Session 3: Workshop

The third session will be devoted to a workshop-style examination of student datasets using the methods we cover in class that day. For each class, you will need to bring a portable laptop computer with the latest versions of Excel and SPSS installed. Please let me know immediately if you do not have access to a computer or the appropriate software. Since adequate investigation of your dataset will require more time than we have scheduled for class, you will be responsible for completing the analysis on your own. A report of your findings will be due at the start of the next class when we discuss your work during the Roundtable. *Together, these analyses will provide the basis for a professional research poster (worth 25 percent of your final grade) that you will present in class on April 17.* The poster, which must adhere to standards for posters presented at professional conferences, should be designed to present the relevant analyses you conducted throughout the semester in such a way that, when combined, answer a question that is pertinent to your research.

-GRADING POLICY-

Performance evaluation includes the following grades: A (“excellent”), B (“good”), C (“average”), D (“poor”), and F (“failing”). These grades are earned based on the following scale of numerical scores: A+ = >98.0, A = 92.0-98.0, A- = 90.0-91.0, B+ = 88.0-89.0, B = 82.0-87.0, B- = 80.0-81.0, C+ = 78.0-79.0, C = 72.0-77.0, C- = 70.0-71.0, D = 60.0-69.0, F = <60.0. All scores will be rounded to the nearest whole integer (down when $x < 0.50$ or up when $x \geq 0.50$). Please note that incompletes will not be issued before April 3, 2009, and only then at my discretion. The College of Arts and Sciences dictates that incompletes should only be granted when, due to circumstances beyond the control of the student, only a small portion of the required work remains unfinished and the student is otherwise passing the course. Students seeking an incomplete will be required to sign a contract with me, specifying the work to be completed and the deadline for completion. I will not accept late or incomplete work, except under conditions of a documented medical or criminal emergency, a major religious observance that necessitates absence from class, or a significant professional conference (arranged in advance). Plagiarism of any kind will result in an F or FF in the course.

-WRITING ABOUT QUANTITATIVE ANALYSIS-

(Some Good Advice from George Cowgill, Keith Kintigh, Ken Kvamme, and Yours Truly)

- Look at your data first using simple tables and pictures. Often this tells you everything important. If not, it will tell you what is sensible or not sensible to do next.
- Statistical analysis is not a way to arrive at certainty; it is a powerful aid in discerning what your data suggest and how strongly they suggest it. This is often done better by an estimation approach than by hypothesis testing.
- If you must do a hypothesis test, report the actual probability level obtained and don't treat some arbitrary level, such as 5 percent, as a talisman that tells you what to think.
- It's not the sampling fraction that matters; it's the *size* of the sample.
- Proportions, percents, and ratios represent something relative to something else. They are fractions, with a numerator and a denominator. When you write, always report the denominator. When you read, always ask yourself whether you understand what denominator is implied.
- “Frequency” should always mean count of something, rather than ratio of something to something else; the latter should be referred to as “relative frequency.”
- “Data” is plural while “datum” is singular; thus, we say, “these data are” instead of “the data is.”
- The character, %, is usually spelled out, percent; thus, “10%” should be “10 percent.”
- Use a space to separate the unit of measure from the value of the case; “10ml” should be “10 ml.”
- Spell out whole integers less than 10; use Arabic numerals for whole integers greater than nine (unless they appear first in a sentence, then spell them out): zero, one, two, three, four, five, six, seven, eight, nine, 10, 11, 12, 13, 14, etc., but “Fifty-seven percent of the people are tall.” An important exception: do not spell out values that have measurements; thus, “five square meters” should be “5 m².”
- “Boxplot” is one word (not “box plot”) and is a noun, so there’s no need to qualify it, e.g., “boxplot chart.” An alternative reference is “box-and-whisker plot.” Similarly, “ANOVA Analysis” or “PCA Analysis” is unnecessarily repetitive to repeat.
- If most of the data are to the left in a histogram and there are outliers to the right, we say that the data are “skewed right,” that is, we note where the skew occurs.
- When comparing histograms, they must be based on the same scale for both X and Y axes.
- Lowercase “n” reports the sample size, while uppercase “N” reports the population size. There is no need to italicize “n” or “N” because they are not statistics (i.e., they are not parameter estimates).

-SCHEDULE-

January 9: Introduction to Multivariate Data (and Matrix Algebra)

The lottery is a tax on people who don't understand statistics.

Background

Field, A. P.

2005 Chapters 1 and 2. In *Discovering Statistics using SPSS*, by A. P. Field, pp. 1-62. Second edition. Sage, London.

Examples

Aunger, R.

1995 On Ethnography: Storytelling or Science? *Current Anthropology* 36(1):97-130.

Romney, A. K.

1999 Culture Consensus as a Statistical Model. *Current Anthropology* 40(Supplement):S103-S115.

January 16: Exploratory Data Analysis of Multiple Variables

"When she told me I was average, I wasn't upset because she was just being mean."

Background

Field, A. P.

2005 Chapter 3. In *Discovering Statistics using SPSS*, by A. P. Field, pp. 63-106. Second edition. Sage, London.

Examples

Dressler, W.

1993 Health in the African-American Community: Accounting for Health Inequalities. *Medical Anthropology Quarterly* 7(4):325-345.

Kamp, K. A., N. Timmerman, G. Lind, J. Graybill, and I. Natowsky

1999 Discovering Childhood: Using Fingerprints to Find Children in the Archaeological Record. *American Antiquity* 64(2):309-315.

January 23: Analysis of Variance and Covariance

What did the new statistics professor do when his lecture on analysis of variance flopped in front of a large class? He had to go over ANOVA ANOVA it again!!!

Background

Field, A. P.

2005 Chapters 8 and 9. In *Discovering Statistics using SPSS*, by A. P. Field, pp. 309-388. Second edition. Sage, London.

Examples

Guillette, E. A., M. M. Meza, M. G. Aquilar, A. D. Soto, and I. E. Garcia

1998 An Anthropological Approach to the Evaluation of Preschool Children Exposed to Pesticides in Mexico. *Environmental Health Perspectives* 106(6):347-353.

Gravlee, C. C., H. R. Bernard, and W. R. Leonard

2003 Heredity, Environment, and Cranial Form: A Reanalysis of Boas' Immigrant Data. *American Anthropologist* 105(1):125-138.

January 30: Multivariate Analysis of Variance

Why is a physician held in much higher esteem than a statistician? A physician makes an analysis of a complex illness, whereas a statistician makes you ill with a complex analysis!

Background

Field, A. P.

2005 Chapter 14. In *Discovering Statistics using SPSS*, by A. P. Field, pp. 571-606. Second edition. Sage, London.

Examples

Buss, D. M., T. K. Shackelford, L. A. Kirkpatrick, and R. J. Larsen

2001 A Half Century of Mate Preferences: The Cultural Evolution of Values. *Journal of Marriage and Family* 63(2):491-503.

Goodwin, L. D., and E. A. Stevens

1993 The Influence of Gender on University Faculty Members' Perceptions of "Good" Teaching. *The Journal of Higher Education* 64(2):166-185.

February 6: Hierarchical Cluster Analysis

If you laid all the statisticians on earth, end to end, you wouldn't reach a conclusion.

Background

Field, A. P.

2005 Cluster Analysis. Department of Psychology, University of Sussex, UK.

Examples

Adlaf, E. M., and Y. M. Zdanowicz

1999 A Cluster-analytic Study of Substance Problems and Mental Health among Street Youths. *American Journal of Drug and Alcohol Abuse* 25(4):639-660.

Sutton, M. Q.

1998 Cluster Analysis of Paleofecal Datasets: A Test of Late Prehistoric Settlement and Subsistence Patterns in the Northern Coachella Valley. *American Antiquity* 63(1):86-107.

February 13: Multivariate Correlation

It is proven that the celebration of birthdays is healthy. Statistics show that those people who celebrate the most birthdays become the oldest.

Background

Field, A. P.

2005 Chapter 4. In *Discovering Statistics using SPSS*, by A. P. Field, pp. 107-142. Second edition. Sage, London.

Examples

Hoare, R. E.

1999 Determinants of Human-Elephant Conflict in a Land-Use Mosaic. *Journal of Applied Ecology* 36(5):689-700.

Middleton, W. D., and T. D. Price

1996 Identification of Activity Areas by Multi-element Characterization of Sediments from Modern and Archaeological House Floors Using Inductively Coupled Plasma-atomic Emission Spectroscopy. *Journal of Archaeological Science* 23(5):673-687.

February 20: Multiple Regression

Statistics play an important role in genetics. For instance, statistics prove that numbers of offspring is an inherited trait. If your parents didn't have any kids, odds are you won't either.

Background

Field, A. P.

2005 Chapter 5. In *Discovering Statistics using SPSS*, by A. P. Field, pp. 143-217. Second edition. Sage, London.

Examples

Arlotti, J. P., B. H. Cottrell, S. H. Lee, and J. J. Curtin

1998 Breastfeeding among Low-Income Women with and without Peer Support. *Journal of Community Health Nursing* 15(3):163-178.

Fejgn, B.

1995 Factors Contributing to the Academic Excellence of American Jewish and Asian Students. *Sociology of Education* 58(1):18-30.

February 27: Logistic Regression

A statistician is skilled at drawing a line from an unwarranted assumption to a foregone conclusion.

Background

Field, A. P.

2005 Chapter 6. In *Discovering Statistics using SPSS*, by A. P. Field, pp. 218-268. Second edition. Sage, London.

Examples

Chavez, L. R.

1994 The Power of the Imagined Community: The Settlement of Undocumented Mexicans and Central Americans in the United States. *American Anthropologist* 96(1):52-73.

Strassmann, B. I.

1997 Polygyny as a Risk Factor for Child Mortality among the Dogon. *Current Anthropology* 38(4):688-695.

March 6: Principal Components Analysis

I'm not an outlier; I just haven't found my distribution yet!

Background

Field, A. P.

2005 Chapter 15. In *Discovering Statistics using SPSS*, by A. P. Field, pp. 619-680. Second edition. Sage, London.

Examples

D'Andrade, R., and P. A. Morin

1996 Chimpanzee and Human Mitochondrial DNA: A Principal Components and Individual-by-Site Analysis. *American Anthropologist* 98(2):352-370.

Kuhn, R. D., and M. L. Sempowski

2001 A New Approach to Dating the League of the Iroquois. *American Antiquity* 66(2):301-314.

March 13: Discriminant Function Analysis

*Statisticians are like the drunk leaning against the lamp pole:
they are there for support, not illumination.*

Background

Field, A. P.

2005 Chapter 14. In *Discovering Statistics using SPSS*, by A. P. Field, pp. 606-618. Second edition. Sage, London.

Examples

Stall, R., S. Heurtin-Roberts, L. McKusick, C. Hoff, and S. W. Lang

1990 Sexual Risk for HIV Transmission among Singles-Bar Patrons in San Francisco. *Medical Anthropology Quarterly* 4(1):115-128.

Wells, E. C.

2000 Pottery Production and Microcosmic Organization: The Residential Structure of La Quemada, Zacatecas. *Latin American Antiquity* 11(1):21-42.

March 27: Multidimensional Scaling

*Arguing with a statistician is a lot like wrestling with a pig.
After a few hours you begin to realize that the pig likes it.*

Background

Garson, G. D.

2008 Multidimensional Scaling. In *StatNotes: Topics in Multivariate Analysis*. Department of Public Administration, North Carolina State University, Raleigh.

Examples

Fry, R. E.

1979 The Economics of Pottery at Tikal, Guatemala: Models of Exchange for Serving Vessels. *American Antiquity* 44(3):494-512.

Garro, L. C.

1986 Intracultural Variation in Folk Medical Knowledge: A Comparison between Curers and Noncurers. *American Anthropologist* 88(2):351-370.

April 3: Correspondence Analysis

Did you know that if you torture the data long enough eventually they will confess?

Background

Garson, G. D.

2008 Correspondence Analysis. In *StatNotes: Topics in Multivariate Analysis*. Department of Public Administration, North Carolina State University, Raleigh.

Examples

Howell, T. L.

1995 Tracking Zuni Gender and Leadership Roles across the Contact Period. *Journal of Anthropological Research* 51(2):125-147.

Romney, A. K., and C. C. Moore

1998 Toward a Theory of Culture as Shared Cognitive Structures. *Ethos* 26(3):314-337.

April 10: How to Lie with Statistics

*A statistician always has something to say with numbers,
while a politician always has to say something with numbers...*

Background

Best, J.

2004 *More Damned Lies and Statistics: How Numbers Confuse Public Issues*. University of California Press, Berkeley.

Examples

Second Moment

2003 *Lies, Damned Lies, and Statistics*. Stone Capital, Dallas, Texas.

April 17: Poster Symposium

The student asked the professor, "How many problems will there be on the final exam?" The professor replied, "I think you will have lots of problems on the final."

Background

Miller, J. E., and J. K. Trainor

2005 *Creating Anthropology Conference Posters: A Guide for Beginners*. Department of Anthropology, University of South Florida, Tampa.

Examples

Hughes, S. A.

2007 AIDS Deaths and Structural Violence in Miami-Dade County, Florida. Poster presented in the ANG 7750 Poster Symposium. University of South Florida, Tampa.


Perrett, A. S.

2004 What Socioeconomic Characteristics Differentiate between the Voting Constituents in Districts 3 and 4 of Hillsborough County? Poster presented in the ANG 7750 Poster Symposium. University of South Florida, Tampa.

ANOVA: ANALYSIS OF VALUE

IS YOUR RESEARCH WORTH ANYTHING?

Developed in 1912 by geneticist R.A. Fisher, the Analysis of Value is a powerful statistical tool designed to test the significance of one's work.



am i wasting my time?

Significance is determined by comparing one's research with the **Dull Hypothesis**:

$$H_0 : \mu_1 = \mu_2 ?$$

where,

- H_0 : the Dull Hypothesis
- μ_1 : significance of your research
- μ_2 : significance of a monkey typing randomly on a typewriter in a forest where no one hears it.

The test involves computation of the $F'd$ ratio:

$$F'd = \frac{\text{sum(people who care about your research)}}{\text{world population}}$$

This ratio is compared to the F distribution with $I-1, N_I$ degrees of freedom to determine a p (in your pants) value. A low p (in your pants) value means you're on to something good (though statistically improbable).

Type I/II Errors

The Analysis of Value must be used carefully to avoid the following two types of errors:

- Type I: You incorrectly believe your research is not Dull.
- Type II: No conclusions can be made. Good luck graduating.

Of course, this test assumes both Independence and Normality on your part, neither of which is likely true, which means *it's not your problem*.

WWW.PHDCOMICS.COM
JORGE CHAM © 2007