Anthropology 269/469: Archaeological Data Sets.

University of Chicago Spring Quarter 2004
LAB: W 9:30-11:00 TTh:9:00-10:30 *H.Mezz. 102*

This course focuses on the analysis of previously collected archeological data. The goals of this course are twofold, first to provide you with an opportunity to examine research questions through the study of specific data sets, and second to allow you to evaluate evidential claims in light of analytical results. We will discuss analytical strategies in terms of their appropriateness to data, their methodological adequacy, and their logical connection with the proposed research problem.

The course will consider data collection, sampling and statistical populations, exploratory data analysis, and statistical inference. Graphical and numerical summaries of data and the importance of choosing appropriate analytical techniques will be emphasized. The course will cover elementary statistical analysis up to linear models.

The course is built around computer analysis and, thus, will also provide an introduction to computer analysis, data encoding, and data base structure. You will learn to use a number of data entry options, to transfer files between applications, and to restructure existing data sets for different types of analysis. By the end of the quarter, I will expect you to develop the skills necessary to write analysis programs using batch language commands as well as to conduct interactive analysis using menus.

Office Hours and Lab

I will be available to students in this class during the scheduled laboratory period. The purposes of the lab. section are consultation and collaboration rather than instruction *per se.* I do not anticipate that the scheduled lab time will be sufficient for you to complete all assignments for this class.

In addition to the lab., I will hold office hours on **M** 1-3 or by appointment. My office is in **Haskell Hall, room 224**. My office phone number is 2-6040. My email address is **m-lycett@uchicago.edu**.

Assignments

Grades will be based on the following:

(1) five of six laboratory assignments (50pts each).
(2) A final project (250 pts).
(3) presentations/ participation in sampling workshop (required for graduate students)

Each assignment is designed to develop specific skills as well as knowledge of a statistical procedure or concept. Assignments will consist of three parts: a one to two page discussion of the problem and explanation of the procedures employed; a copy of the computer program; and the results of the program. Assignments will be due approximately every week beginning with the third week of class. You must complete assignments 1-4 and either 5 or 6. The remaining assignment may be completed for extra credit.

You will also develop a pilot analysis that addresses a research question of their own choosing with an existing body of archaeological data. Specific analyses will depend on the research problem and data set chosen by the student. ***You must submit a one page description of the proposed data set, including a short discussion of its relevance to the research problem by the third week of class.*** If you need help finding a data set see me as soon as possible. In consultation with me, you will conduct an analysis of their chosen data set and prepare a paper on the basis of their results. The paper will present the

background to the research problem, a description of the data base and its relevance to the problem, a description of the analytical methods employed, and the results of the analysis.

For **graduate students** in the course, this project will entail two additional components. *First*, graduate students will prepare a 15 minute presentation on the basis of their project results. This presentation will include a critical assessment of the implications of their analysis for future research into this problem. Your assessment should address both the theoretical and methodological limitations of our current understanding, and suggest modifications to their problem statements and research designs.

Second, we will be holding a separate but related discussion of archaeological sampling design and the implications of data collection for analysis. This will occur sometime in mid-quarter at a place and time of mutual convenience. I anticipate two sessions, on the order of 1.5 to 3 hrs in length each. The point of this exercise is to discuss strategies for data collection, how to implement them, and how these strategies influence the possibilities for subsequent research. You will not receive a grade for this discussion but your participation is mandatory. These discussions will be open to interested grad. students from outside the class.

Readings: The following books have been ordered for this class and are available at the Seminary Co-op bookstore:

Drennan, Robert D.
1996 *Statistics for Archaeologists : A Commonsense Approach*. Plenum, New York.
Orton, C
2000 *Sampling in Archaeology*. Cambridge University Press, Cambridge.
Norusis, M.J.
2000 *SPSS 10.0 Guide to Data Analysis*. Prentice Hall, New York.

The following manuals, published by SPSS Inc. are widely available for those interested in more complete documentation:

SPSS Base 11.0 User's Guide Package
SPSS Base 11.0 Applications Guide
SPSS Base 11.0 Syntax Reference Guide

Recommended supplemental texts:

Siegal, S. and N.J. Castellan
1988 *Nonparametric Statistics for the Behavioral Sciences*, second edition. McGraw-Hill, New York.
Shennan, S.
1997 *Quantitative Archaeology*, 2nd edition. University of Iowa Press, Iowa City.

Recommended Web based resources:

http://davidmlane.com/hyperstat/index.html
http://vassun.vassar.edu/%7Elowry/webtext.html
http://www.execpc.com/~helberg/statistics.html#education

SCHEDULE OF LECTURES

NB: No LAB 3/31; no class 4/1

I. Introduction to Statistical Analysis. Overview of course: data, sampling and populations, descriptive and inferential statistics, statistical inference, nonparametric and parametric procedures. Creating data sets, manipulating data, transferring data between applications and formats.

II. Samples and Populations. Population parameters and sample statistics. Sampling from a population. The rules of probability.

READING: NORUSIS pp. 5-51; 525-545; 163-174..
DRENNAN pp. 79-96..

I. Data and variables. Quantifying observations and encoding data. Levels of measurement: nominal, ordinal, interval and ratio data.

II. Univariate Numerical Summaries. Measures of central tendency. Measures of dispersion. Measures of deviation from normality.

III. Univariate Graphical Summaries. Histograms, stem and leaf plots, box plots.

READING: NORUSIS pp. 59-72; 91-107.
DRENNAN pp. 3-73.

4/15 ***PROJECT PROPOSAL DUE***.

I. Normal curves and z-scores. Normally distributed populations and random sampling.

II. Sampling Distribution of the Mean. The central limit theorem.

III. Estimating Population Means. Confidence intervals. Student's-t.

READING: NORUSIS pp. 177-190; 200-208.
DRENNAN pp. 111-144.

4/15 ***ASSIGNMENT 1. DUE.***

I. Hypothesis testing: one sample tests. One tailed and two tailed tests. Problems of small samples. Nonparametric tests.

READING: NORUSIS pp. 208-213.
DRENNAN pp. 149-165.

I. Hypothesis testing: two sample tests. Difference of means. Difference of proportions.

Problems of small samples. Nonparametric tests for two independent samples. Nonparametric tests for two related samples.

4/29 ***ASSIGNMENT 2 DUE***.

READING: NORUSIS pp. 217-253; 325-334; 341-343.

I. Measures of association for nominal variables. Chi-square. Goodness of fit. Contingency table analysis.

READING: NORUSIS pp. 111-126; 301-332;337-339; 352-360.
DRENNAN pp. 185-201.

5/13 ***ASSIGNMENT 3 DUE***.

I. Hypothesis testing: multiple samples. Analysis of variance.

II. Hypothesis testing: multiple samples. Nonparametric tests for independent samples. Nonparametric tests for related samples.

READING: NORUSIS pp. 259-277; 334-337; 342-343.
DRENNAN pp. 167-182.

5/27 ***ASSIGNMENT 4 DUE.***

I. Correlation and regression. The correlation coefficient. Rank correlation.

II. The regression model. Analysis of residuals. Inference and prediction. Data transformations.

READING: NORUSIS pp. 133-159; 349-352; 360-369; 373-448.
DRENNAN pp. 203-233.

6/3 ***ASSIGNMENT 5 DUE.***

6/11 ***ASSIGNMENT 6 DUE. FINAL PROJECT DUE.***

Project Guidelines

In order to complete the final project, you must identify a research question and choose a data base appropriate to that question. **You should choose a question and data base appropriate to your own interests**. Any data are acceptable, so long as they can be analyzed using the techniques you have learned in this class. If you are having trouble finding a data set, try spending some time in the library looking through primary literature (journals, site reports, *et cetera*) or see me. Try to choose a data set that is manageable (neither very small nor very large). It would be helpful to you to know the methods that were used to collect the data in the first place. You must have your data base and research question approved by APRIL 12. You may consult any source material you wish; however, be sure to provide a complete bibliographic reference for each source you quote. Exclusive of statistical output and references, your project should not exceed 15-20 typed, double spaced pages. Include your SPSS program or programs in an appendix to the project. In a second appendix, include only that statistical output which is directly relevant to your project (graphic and numeric summaries and the output of testing procedures). **YOUR PROJECT IS DUE AT 5:00 PM ON FRIDAY JUNE 11.**

I. **Statement of the research problem**.

A. Describe the problem to be investigated, including a brief description of the relevant theoretical background of the problem. Describe why you think this research question is important and relevant to your interests. Discuss what you expect to learn from investigating this problem. Try to choose a question that is well defined with specific *statistical hypotheses*. Your problem must be specific enough to have empirical implications. Discuss these implications in terms of the data base you have chosen.

II. **Description of the data base.**

A. Discuss the source of the data. Where do they come from? How were they collected? What are the strengths and weaknesses of these data?

B. Discuss the relevance of these data for your research problem. Why are these data appropriate? What kinds of questions can you address with these data? What kinds of questions can't be addressed without new data?

III. **Research design**.

A. Outline the steps you will follow in analyzing these data. Be as specific as possible.

B. Data base design. What variables are included in the data base? How did you determine cases? Is this file design appropriate for the statistical procedures you plan to employ?

C. Discuss your plan for research including exploratory data analysis and hypothesis testing.

IV. **Statistical description**.

A. What univariate summaries did you use to describe the variables? Why did you choose these procedures? What did you learn from them?

B. Be sure to include a thorough discussion of the output for all graphic and numeric summaries.

V. **Hypothesis testing**.

A. Identify the statistical hypotheses to be tested. Decide which of the available tests is appropriate for each hypothesis test. Justify your choice of tests with reference to the assumptions required by that procedure.

B. For each hypothesis test, state the null hypothesis and the alternate hypothesis. Discuss the basis on which you will accept or reject the null hypothesis, including the value of alpha (pre-experiment error level) and the critical value of the test statistic where appropriate.

C. Describe and interpret the results of each statistical test. Be sure to report (1) the observed p-value, (2) whether or not the value of the test statistic computed by SPSS falls in the critical region for rejection of the null hypothesis, and (3) a confidence interval for the test statistic where appropriate.

VI. **Summary and conclusions**.

A. Restate your research problem and summarize the methods and results of your investigation.

B. Discuss your interpretation of the results of the quantitative analysis. What did you learn from the analysis? How useful were the quantitative techniques that you employed? Were there any discrepancies between what you expected to find and what you found? What might account for unexpected results? What new questions do you have about the data base or about your research question?

GRADUATE STUDENTS (NB: I will be particularly interested in your discussion of the following points):

C. Discuss the implications of your analysis for future research into this problem. What kinds of data would be appropriate for such studies, and how would you go about collecting it?

D. What do your results suggest about current theoretical and methodological approaches to your research problem?

E. Given your results, what modifications would you make to your problem statement and research design?