Anth 6302
Statistics in Anthropology

B. Sunday Eiselt
Department of Anthropology
Heroy Hall Room #450
Spring 2008

Class Meets: Wednesday 2:00 4:50
Office Hours: Open door
Email: seiselt@smu.edu
Phone: (480) 755-7825

Course Description

Most social science research requires the recognition and analysis of *patterns* in empirical data. However, these are rarely immediately apparent in the large batches of numbers obtained during research. Quantitative methods of data analysis provide an indispensable means of detecting such patterning in a systematic and organized fashion.

This course is intended as an introduction to quantitative methods and reasoning for the archaeologist, although other anthropologists may make use of the same principles and techniques we will examine. Most of these are fairly standard, deriving from either classical statistics developed in the early 1900s or from the more recent exploratory data analysis approach. The course focuses on topics covered in *Quantifying Archaeology* (S. Shennan, 1997, 2nd edition, Edinburgh University Press) and *Statistics for Archaeologists* (R. Drennan, 1996, Plenum Press) occasionally supplemented by additional readings. In addition, we will cover the basics of using computers for the analysis of anthropological data with packaged statistical software, primarily JMP, SPSS, and R.

The major objectives of the course are:

- to provide the background necessary for informed, critical reading of quantitative archaeological literature
- to provide the background needed for designing research projects that will generate data that can be productively analyzed using quantitative methods
- to teach the use of computers in managing and analyzing archaeological data
- to teach methods for conveying quantitative arguments in scholarly publication

No prior knowledge or use of specific software packages or statistics is required and I anticipate that many students will have little or no such experience. However, the course will require a substantial commitment of time, including extensive use of computers on campus or at home.

The class meets once a week, but is divided into two parts. The first part, which lasts about an hour, consists of an illustrated lecture and group discussion of the weekly topic. The second part is devoted to a workshop-style

examination of an archaeological problem using the methods we have learned in class that day. Please be aware that you need to bring a portable laptop computer (preferably with a wireless Internet connection) with the latest versions of SPSS, JMP, and R installed (see below). Notify me if you do not have access to a laptop or software.

Course Requirements and Structure

Class Participation

Performance in class is based on attendance and engagement with the subject including coming to class prepared, asking relevant questions, and providing constructive comments. Class participation is worth 20 percent of your final grade.

Computer Assignments

The best way to learn about statistics and data analysis is to actually do it. Therefore the purpose of the computer assignments is to introduce you to statistical software that is essential for analyzing data. This does not mean you need to know less about the concepts studied in class but lets you use this energy to think about the problems rather than the computational details. There will be eight graded assignments - graded on a scale of 0-10 (each worth 30 percent of your final grade). Each assignment will be due at the start of the next class; late papers will not be accepted under any circumstances. Please submit your essays electronically prior to class and bring a hard copy to class. Grading for the paper and presentation will be based on your performance relative to the Standard Protocol for Computer Assignments and the Guidelines for Analyzing and Writing-up Quantitative Analyses (see below).

Final Project (Paper and Presentation)

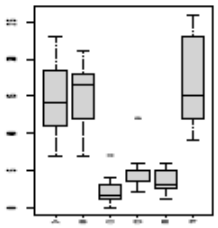
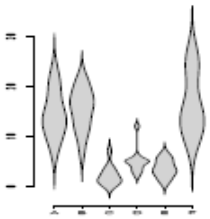
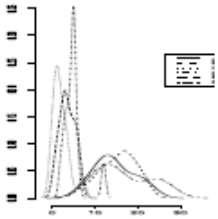
The final project for the class includes a paper and a presentation, due on December 3rd. The final paper will be graded on a scale of 0-20 and is worth 20 percent of your grade. The presentation also will be graded on a scale of 0-20 and will be worth 20 percent of your grade. Grading for the paper and presentation will be based on your performance relative to the Standard Protocol for Writing Reports, the Standard Protocol for Presenting Papers, and the Guidelines for PowerPoint Presentations and Guidelines for Analyzing and Writing-up Quantitative Analyses (see below). Over the first several weeks of the course, we will work together to identify suitable issues and datasets for everyone. The topic and dataset(s) for your project must be approved by the instructor.

Final Exam

The final exam will consist of a series of real-world archaeological questions for which you must identify the appropriate statistical techniques for analysis including data screening and relevant programs. No actual analyses of problem sets are required. This test will evaluate your ability to think critically about data analysis and presentation. The exam will be graded on a scale of 0-20 and will be worth 10 percent of your grade.

Grade Break-down

Class Participation:	20%
Computer Assignments:	30%
Final Paper:	20%
Final Presentation:	20%
Final Exam;	10%

Performance Evaluation includes the following grades: A (excellent), B (good), C (average), D (poor), F (failure), and I (incomplete). These grades are eared based on the following scale: A = 95-100, A- = 90-94, B+ = 86-89, B = 85-82, B- = 80-81, C+ = 76-79, C = 75-72, C- = 70-71, D = 69-60, F = 60 and below. Note that incompletes will only be given due to circumstances beyond the control of the student.

Resources

Readings

Readings for the course are divided into recommended and supplementary articles and textbook chapters. The recommended readings should be read prior to class. The supplementary readings provide examples and discussions that are relevant to the weekly topic. Supplementary readings include classic articles, debate articles, and recent applications that are intended to extend the basic readings and help to connect data analyses to real world problems and theoretical issues. Supplementary readings should be scanned/reviewed prior to class with a more careful reading of selected articles as necessary to assist with computer assignments and your final project. Nearly all of the supplementary readings (non-textbook articles) will be made available to you as PDFs on line through Bluehost.com (I will provide the instructions in class). Supplementary texts also are available on reserve in the ISEM library.

There are no required texts for this course, but I highly recommend that all students purchase at least two reference texts as resources. The course is based largely on material presented in two main texts; *Quantifying Archeology* [2nd edition] by Stephen Shennan (˜$35.00 through the University of Iowa Press or Amazon), and *Statistics for Archaeologists: A Commonsense Approach* by Robert Drennan (˜$40.00 through Amazon). Both texts are available on reserve in the ISEM library and in my office/lab. If you do purchase Shennan, make sure you get the 2nd edition.

Additional recommended texts are provided in the Additional Text Resources section below. Any of these references may be consulted for this class.

Statistical Packages (Required)

We will be using three statistical packages; JMP, SPSS, and R. These programs are required, and each does something slightly better than the others. JMP enables users to explore graphs and data interactively. SPSS is a powerful package for analyzing multivariate numerical summaries and has an excellent results coach that helps with data interpretation. R is complex, but provides the best defaults for graphic display (and it is free). You will benefit greatly from becoming familiar with all three during the course of this class and during your dissertation research. Some familiarity with Excel also is required for the preparation and screening of data.

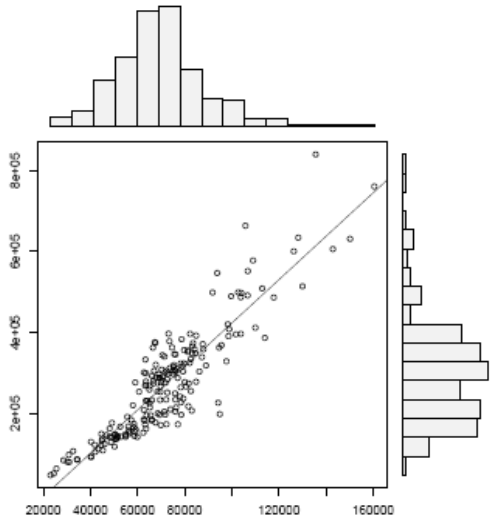
Please install each of the three following programs to your laptop prior to the first day of class.

1. JMP: A 6-month lease for the Student Educational version of JMP (for Windows and Mac) may be downloaded at: http://estore.e-academy.com/index.cfm?loc=jmp/main for a cost of $29.95. You will need to create a student account to download this software. Upon installation, follow the directions to obtain the certification (serial number) to unlock the program. Demos and tutorials for students of JMP are available on line http://www.jmp.com/academic/student.shtml (follow links to the right)
2. SPSS: A 6-month lease for the Grad Pack of SPSS 16 is available through e-academy.com at: http://estore.e-academy.com/index.cfm?loc=spss/main for a cost of $79.99. You will need to create an account with e-academy to download this software. The latest version of SPSS also is available in the Department Computer Lab.
3. *R* is an open-source (free) statistics package that can be downloaded from http://www.r-project.org/). Install the program by selecting a CRAN repository and operating system. Then choose the base folder and download the setup program. Run the installation program R-2.1.1-win32.exe. Initial tutorials will be provided in class.

Additional Statistical Packages (Not Required but Helpful)

1. SYSTAT: A little brother to SAS and SPSS, this program has a long tradition of supporting very flexible graphing output. There is a graphical user interface to begin the process of making graphics but you will probably need to use text commands to access all of the options.
2. SYSTAT has a new, free version called MyStat 12 that can be downloaded (for Windows only, from www.systat.com/products/mystat). It does not have all of the features of the full SYSTAT but it appears that some of the basic methods that archaeologists use are available and that it will handle problems with large data sets. MyStat is less cumbersome for producing graphics than R, so you may want to download it and experiment with the graphics interface.
3. Tools for Quantitative Archaeology: TFQA is a commercial package of more than 45 PC programs developed to satisfy the unusual analytical needs of archaeologists. The focus of the package is on methods developed for archaeology and not included in general-purpose statistical packages. While this package can perform many important analyses, it is not a complete substitute for a general purpose statistical package. TFQA was written by Keith Kintigh, Professor of Anthropology at Arizona State University. The program is available for a single user site license costing $150 at: http://tfqa.com/. The department has purchased a full site license that is available to students. Please see me for additional information if you wish to install this software (recommended).

Outline for Topics by Week



Week 1 (August 27) Introduction

Topics:

1. Introduction
2. Introduction to software and resources (Demos)

Recommended Readings:

1. Shennan Ch. 1
2. Aldenderfer 1998, 2005

Supplementary Readings:

1. Aldenderfer 1987a
2. Clark 1982

Computer Assignment (not graded)

1. JMP Tutorials:
    a. Save and Print the Quick Guide at: http://www.jmp.com/academic/pdf/421242_jmpse_guide.pdf
    b. View the Using JMP Student Edition Demo at: http://www.jmp.com/academic/se_demos.shtml
    c. Also go through the Beginners Tutorial from the Help Menu in the program
    d. View the Visualization Demo at: http://www.jmp.com/software/jmp7/tutorials/index.shtml
2. SPSS Tutorials:
    a. Familiarize yourself with SPSS by exploring the links provided in the On-line Statistical Program Tutorials section of the syllabus below.

Week 2 (September 3) Data, Coding, Forms, Data Manipulation, Databases

Topics:

1. Statistics Concepts and Terms:
    a. Aims of Statistics
    b. Population vs. Sample
    c. Batches, variables, cases
    d. Variable Types Scales of Measurement
        i. Nominal and Ordinal (categorical and qualitative)
        ii. Interval and Ratio (continuous and quantitative)
    e. Independent and Dependent Variables
    f. Parametric vs. Nonparametric data
    g. EDA and Formal Analysis
2. Data collection and Organization (coding and grouping)
3. Data presentation and graphics

> Statistics means never having to say you're certain.

Recommended Readings:

1. Shenan Ch. 2
2. Fletcher and Lock Ch. 1-2
3. Kintigh 2005
4. Aldenderfer 1987b
5. Kintigh 1987

Law of Probability Dispersal: Whatever hits the fan will not be evenly distributed.

Supplementary Readings:

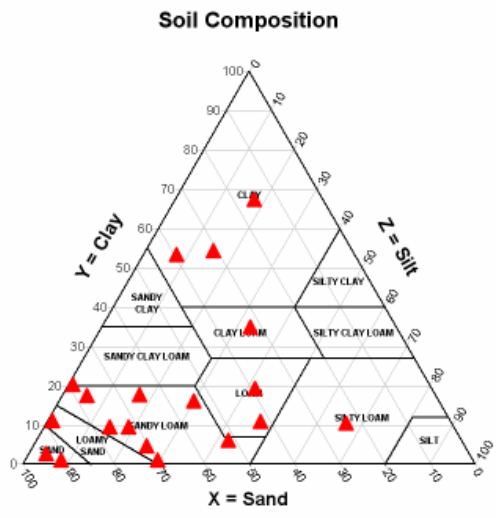
1. Cleveland 1994 Ch. 1-2
2. Read 1978

Computer Assignment: R Tutorial Exercises (not graded)

Week 3 (September 10) Univariate EDA and Descriptive Statistics and Probability Distributions

Topics:

1. Univariate Descriptive Statistics numerical summaries
    a. Measures of central tendency (mean, median, mode)
    b. Measures of dispersion (range, quartiles, and standard deviation)
    c. Measures of relative position (percentile rank, z-scores, T-scores)
    d. Comparing Batches - Transformations
2. Univariate Descriptive Statistics graphical summaries
    a. Histograms and bar charts
    b. Stem-and-leaf, box plots
    c. The shape of a batch: Skewness, kurtosis, multiple modes
3. Column and Row Proportions



Recommended Readings:

1. Whallon 1987
2. Shennan Ch. 3-4
3. Drennan Ch. 1-6
4. Fletcher and Lock Ch. 3-5

Supplementary Readings: None

1. Baxter and Beardah 1997

Computer Exercise #1: Univariate and Descriptive Statistics (graded)

Week 4 (September 17) Inferential Statistics: Sampling and Hypothesis Testing

Topics:

1. Samples and Populations
    a. Representativeness and sampling bias
    b. Distribution of means and standard error
2. Hypothesis testing
3. Parametric tests: assumptions and transformations

4. Comparing means (t-test, ANOVA, Kruskal Wallis)
5. Contingency Analysis (Chi-square, Freeman-Tukey Deviates)

Recommended Readings:

1. Shennan Ch. 5-7
2. Drennan Ch. 7-13
3. Fletcher and Lock Ch. 6-9, 11
4. Cowgill 1977

> In god we trust, all others must bring data…
> Edward Deming

Supplementary Readings:

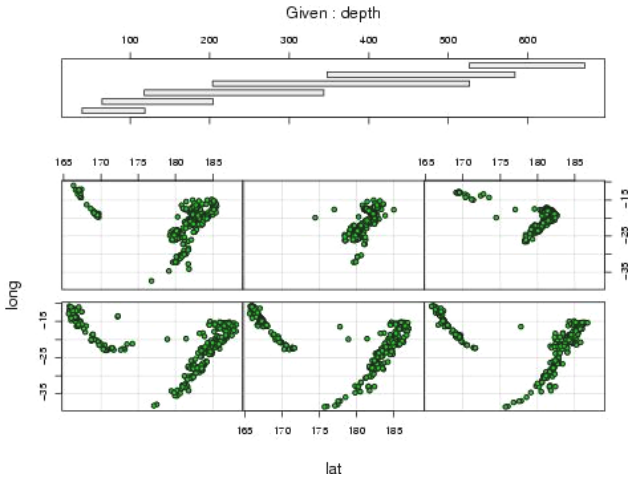
1. Doran 1986
2. Teltser 1991

Computer Assignment #2: Hypothesis testing and significance (graded)

Week 5 (September 24) Bivariate EDA, Correlation and Regression

Topics:

1. Bivariate EDA
2. Correlation and Regression
    a. Strength and direction: Pearsons r, Spearmans Rho, Taub
    b. The best-fit straight line
    c. Residuals
3. Graphical Summaries



Recommended Readings:

1. Shenan Ch. 8-9
2. Drennan Ch 14-15
3. Fletcher and Lock Ch. 10
4. Speth and Johnson 1976
5. Cowgill 1990a

Supplementary Readings:

1. Cleveland 1994 Ch. 3
2. Schriber and Kintigh 1996
3. Casteel 1974
4. Greenspan 1998
5. Steponaitis 1981

Computer Assignment #3: Correlation and Regression (graded)

Week 6 (October 1) Multivariate Analysis of Variance and Covariance, Regression

Topics:

1. Data Screening
2. Multiple Correlation and Regression MANOVA, MANCOVA
3. Supervised and Unsupervised Learning (basics and prep for subsequent classes)

Recommended Readings:

1. Shennan Ch. 10
2. Cowgill 1968
3. Baxter 2006 (outlines major supervised and unsupervised techniques)

Supplementary Readings

1. Thomas 1978a
2. Matson 1980
3. Thomas 1980
4. Simon and Coghalan 1989
5. Braun et al. 2008
6. Bronitsky and Hamer 1986

> In earlier times, they had no statistics and so they had to fall back on lies…
> Stephen Leacock

Computer Exercise #4: MANOVA (graded)

Week 7 (October 8) Numerical Classification and Cluster Analysis

Topics:

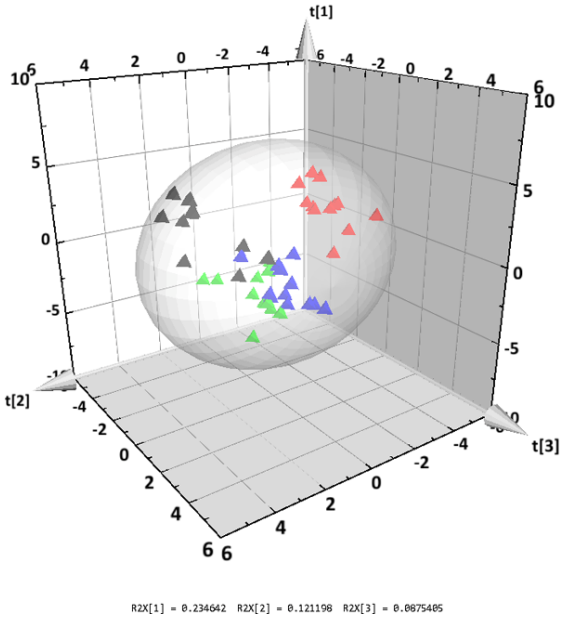
1. Numerical Classification
2. Cluster Analysis
3. Multidimensional Scaling

Recommended Readings:

1. Shenan Ch 11

Supplementary Readings:

1. Baxter 1993 Ch. 7-8
2. Riley 2008
3. Aldenderfer 1982
4. Cowgill 1982 , 1990b
5. Matson and True 1974
6. Aldenderfer and Blashfield 1978
7. Christenson and Read 1977
8. Hall 1982



Computer Exercise #5: Cluster Analysis (graded)

Week 8 (October 15) NO CLASS

Week 9 (October 22) Principal Components and Factor Analysis

Topics:

1. Factor Analysis
2. PCA

Recommended Readings:

1. Shenan Ch. 12

Supplementary Readings:

1. Baxter 1993 Ch. 3-4
2. Buchanan 2006
3. Abbott et al. 2008
4. Ericson and Glascock 2004
5. Vierra and Carlson 1981
6. Ammerman and Feldman 1974
7. Binford and Binford 1966

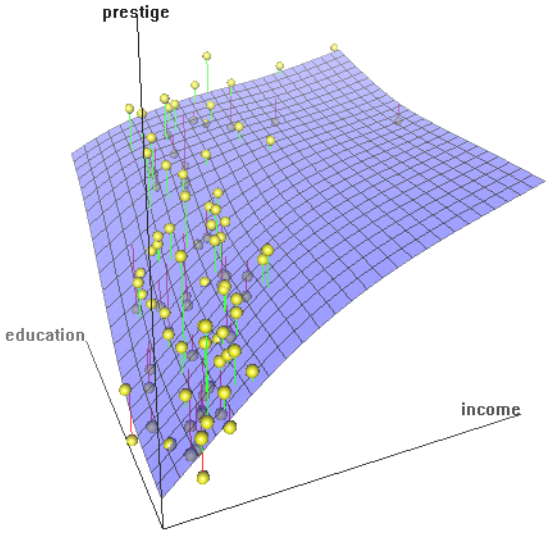
Fate laughs at probabilities.
Lynton E. G. Bulwer

Computer Exercise #6: PCA (graded)

Week 10 (October 29) Discriminant Analysis, Correspondence Analysis, and Logistic Regression

Topics:

1. Discriminant Analysis
2. Correspondence Analysis
3. Logistic Regression



Recommended Readings:

1. Shennan Ch. 13

Supplementary Readings:

1. Baxter 1993 Ch. 5-6, 9-10
2. Walker 2008
3. Bolviken 1982
4. Thomas 1978 b
5. Miguel-Angel et al. 2004
6. Konigsberg and Hens1998
7. McAnany 1989
8. Madsen 1988
9. Gebauer 1988

Computer Exercise #7: Discriminant and Correspondence Analysis (graded)

Week 11 (November 5) Sampling and Diversity

Topics:

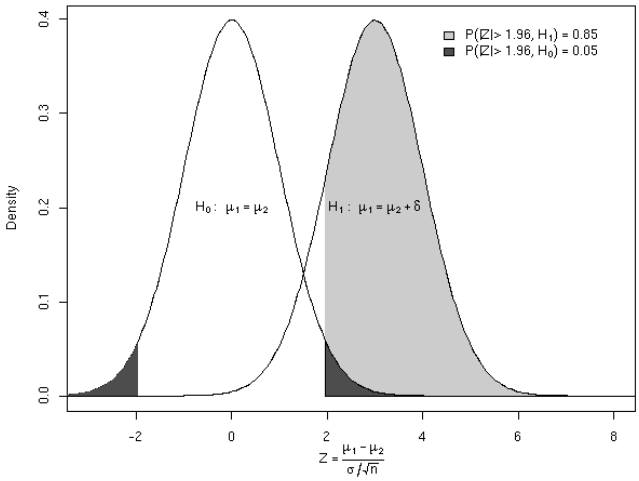
1. Probablistic and Random Sampling
2. Richness and diversity
3. Standardization Measures

Recommended Readings:

1. Shennan Ch 14
2. Drennan Ch 16-19
3. Jones, Grayson, and Beck 1983
4. Eerkens and Bettinger 2001

Supplementary Readings:

1. Nance 1990, 1981
2. Wandsnider and Camilli 1992
3. Robertson 1999
4. Meltzer et al. 1992
5. Rhode 1988
6. McCartney and Glass 1990
7. Kintigh 1984, 1989
8. Baxter 2001



Computer Exercise #8: Diversity and Standardization (graded)

Week 12 (November 12) Special Topics: Spatial Analysis

Topics:

1. Cluster and Nearest Neighbor Analysis
2. Logistic Regression

Recommended Readings:

1. Kintigh 1990

Supplementary Readings:

1. Kintigh and Ammerman 1982
2. Stark and Young 1981
3. Bettinger 1979
4. Cowgill, Altschul and Sload 1984
5. Whallon 1973, 1974, 1984
6. Dacy 1973
7. Pinder et al 1979
8. Relethford 2008

MCD = Σ(xifi)/ Σ fi

Computer Exercise: None (work on final project)

Week 13 (November 19) Special Topics: Chronological Ordering

Topics

1. Chronological Ordering and Seriation
2. Mean Ceramic Dating
3. Radiocarbon dates and calibration
4. Which statistical test should I use? (in class exam)

Recommended Readings:

1. Marquardt 1978
2. Samford 1997

> It's the mark of a truly intelligent person to be moved by statistics.
> George Bernard Shaw

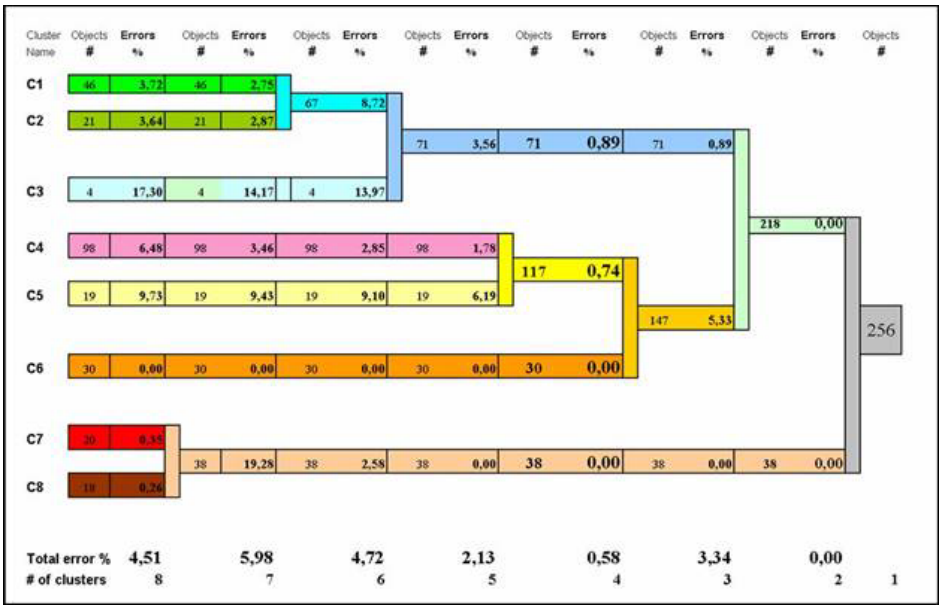
Supplementary Readings

1. Djindjian 1990
2. Drennan 1976
3. Duff 1996
4. LeBlanc 1975
5. South 1977 (209-218)

Computer Exercise: None (work on final project)

Week 14 (November 26) NO CLASS THANKSGIVING BREAK

Week 15 (December 3) FINAL MEETING PAPERS AND PRESENTATIONS DUE



Course Bibliography

Abbott, David R., Andrew D. Lack, and Mark R. Hackbarth

2008 Provenance and Microprobe Assays of Phyllite-Tempered Ceramics from the Uplands of Central Arizona. *Geoarchaeology* . 23,(2): 213242.

Aldenderfer, Mark S.

1982 Methods of Cluster Validation for Archaeology. *World Archaeology*, 14(1):61-72

1987a Assessing the Impact of Quantitative Thinking on Archaeological Research: Historical and Evolutionary Insights. In *Quantitative Research in Archaeology: Progress and Prospects*, edited by Mark S. Aldenderfer, pp. 9-29. Sage Publications, Newbury Park.

1987b On the Structure of Archaeological Data. In *Quantitative Research in Archaeology: Progress and Prospects*, edited by Mark S. Aldenderfer, pp. 89-113. Sage Publications, Newbury Park.

1998 Quantitative Methods in Archaeology: A Review of Recent Trends and Developments. *Journal of Archaeological Research* 6(2):91-120.

2005 Statistics for Archaeology. In *Handbook of archaeological methods,* edited by Herbert D.G. Maschner and Christopher Chippindale. Altamira Press, Lanham, MD.

Aldenderfer, M. S. and R. K. Blashfield

1978 Cluster Analysis and Archaeological Classification [comment on 'Numerical taxonomy, R-mode factor analysis, and archaeological classification', by A.L. Christenson and D.W. Read; with a reply by D.W. Read and A.L. Christenson]. *American Antiquity* 43(3):502-6.

Ammerman, Albert J. and Marcus W. Feldman

1974 On the "Making" of an Assemblage of Stone Tools. *American Antiquity*, 39(4): 610-616

Baxter, Michael J.

1993 *Exploratory Multivariate Analysis in Archaeology*. Edinburgh University Press, Edinburgh.

2001 Methodological Issues in the Study of Assemblage Diversity. *American Antiquity*, 66(4):715-725
Baxter, M. J.

2006 A Review of Supervised and Unsupervised Pattern Recognition in Archaeometry. *Archaeometry* 48(4):671-694

Baxter, Michael J. and C. C. Beardah

1997 Some Archaeological Applications of Kernal Density Estimates. *Journal of Archaeological Science,* 24:347-354.

Binford, Lewis R and Sally R. Binford

1966 A Preliminary Analysis of Functional Variability in the Mousterian of Levallois Facies. *American Anthropologist*, 68(2):238-295.

Bettinger, Robert L.
1979 Multivariate Statistical Analysis of a Regional Subsistence-Settlement Model for Owens Valley. *American Antiquity,* 44(3):455-470

Bolviken, Erik, Ericka Helskog, Knut Helskog, Inger Marie Holm-Olsen, Leiv Solheim, Reidar Bertelsen
1982 Correspondence Analysis: An Alternative to Principal Components. *World Archaeology* 14(1):41-60

Braun, David R., Joanne C. Tactikos, Joseph V. Ferraro, Shira L. Arnow, John W.K. Harris
2008 Oldowan Reduction Sequences: Methodological Considerations. Journal of Archaeological Science 35: 2153-2163

Bronitsky, Gordon and Robert Hamer
1986 Experiments in Ceramic Technology: The Effects of Various Tempering Materials on Impact and Thermal-Shock Resistance . *American Antiquity*, 51(1):89-101.

Buchanan, Briggs
2006 An analysis of Folsom Projectile Point Resharpening Using Quantitative Comparisons of Form and Allometry. Journal of Archaeological Science 33:185-199

Casteel, Richard
1974 A Method for Estimation of Live Weight of Fish from the Size of Skeletal Elements. *American Antiquity*, 39(1): 94-98.

Christenson, Andrew L. and Dwight W. Read
1977 Numerical Taxonomy, R-mode Factor Analysis and Archaeological Classification. *American Antiquity* 42(2):163-179.

Clark, Geoffrey A., and C. Russell Stafford
1982 Quantification in American Archaeology: A Historical Perspective. *World Archaeology*, Vol. 14(1): 98-119.

Cleveland, William S.
1994 *The Elements of Graphing Data*. Rev. ed. AT&T Bell Laboratories, Murray Hill, N.J.

Cowgill, George L.
1968 Archaeological Applications of Factor, Cluster, and Proximity Analysis. *American Antiquity,* 33(3): 367-375

1977 The Trouble with Significance Tests and What We Can Do About It. *American Antiquity* 42(3):350-368.

1982 Clusters of Objects and Associations between Variables: Two Approaches to Archaeological Classification. In *Essays on Archaeological Typology*, edited by R. Whallon and J. A. Brown, pp. 30-55. Center for American Archaeology Press, Evanston, Illinois.

1990a Why Pearson's r Is Not a Good Similarity Coefficient for Comparing Collections. *American Antiquity* 55(3):512-520.

1990b Artifact Classification and Archaeological Purposes. In *Mathematics and Information Science in Archaeology: A Flexible Framework*, edited by Albertus Voorrips, pp. 61-78.

Cowgill, George L., Jeffrey H. Altschul and Rebecca S. Sload
1984 Spatial Analysis of Teotihuacán: A Mesoamerican Metropolis. In *Intrasite Spatial Analysis in Archaeology*, edited by H. Hietala, pp. 154-195. Cambridge University Press, Cambridge.

Dacey, M. F.
1973 Statistical Tests of Spatial Association in the Locations of Tool Types. *American Antiquity* 38(3):320-328.

Drennan, Robert D.
1996 *Statistics for Archaeologists: A Commonsense Approach.* Plenum Press, New York.

Djindjian, François
1990 Ordering and Structuring in Archaeology. In *Mathematics and Information Science in Archaeology: A Flexible Framework*, edited by Albertus Voorrips , pp. 79-92.

Doran, James E.
1986 Formal Methods and Archaeological Theory: A Perspective. *World Archaeology* 18(1):21-37.
Drennan, Robert D.
1976 A Refinement of Chronological Seriation Using Nonmetric Multidimensional Scaling. *American Antiquity*, 41(3):290-302.

Duff, Andrew I.
1996 Ceramic micro-seriation: Types or Attributes? *American Antiquity* 61(1):89-101.

Eerkens, Jelner W. and Robert L. Bettinger
2001 Techniques for Assessing Standardization in Artifact Assemblages: Can We Scale Material Variability? *American Antiquity*, 66(3):493-504.

Ericson, Jonathon and Michael D. Glascock
2004 Subsource Characterization: Obsidian Utilization of Subsources of the Coso Volcanic Field, Coso Junction, California, USA. *Geoarchaeology* 19(8):779–805.

Fletcher, Mike and Gary L. Lock
1991 Digging Numbers: Elementary Statistics for Archeologists. Oxford University Committee for Archaeology Monograph 33. Oxbow Books, Oxford

Gebauer, Anne Birgitte
1988 Stylistic Variation in the Pottery of the Funnel Beaker Culture. In *Multivariate Archaeology: Numerical Approaches in Scandinavian Archaeology*, edited by Torsten Madsen, pp. 91-117.

Greenspan, Ruth L.
1998 Gear Selectivity Models, Mortality Profiles and the Interpretation of Archaeological Fish Remains: A Case Study from the Harney basin, Oregon. *Journal of Archaeological Science* 25:973-984.

Hall, Martin
1982 Quantifying Trends in Site Location with Multidimensional Scaling. *World Archaeology*, 14(1):131-152

Jones, G. T., D. K. Grayson, and C. Beck
1983 Artifact Class Richness and Sample Size in Archaeological Surface Assemblages. In *Lulu Linear Punctated: Essays in Honor of George Irving Quimby*, edited by R. C. Dunnell and D. K. Grayson, pp. 55-73. Anthropological Papers 72. Anthropological Papers, University of Michigan, Ann Arbor.

Kintigh, Keith W.
1984 Measuring Archaeological Diversity by Comparison with Simulated Assemblages. *American Antiquity,* 49(1): 44-54.

1987 Quantitative Methods Designed for Archaeological Problems. In *Quantitative Research in Archaeology: Progress and Prospects*, edited by Mark S. Aldenderfer, pp. 126-134. Sage Publications, Newbury Park.

1989 Sample size, significance, and measures of diversity. In *Quantifying Diversity in Archaeology*, edited by Robert D. Leonard and George T. Jones, pp. 25-36. Cambridge University Press, New York.

1990 Intrasite Spatial Analysis: A Commentary on Major Methods. In *Mathematics and Information Science in Archaeology: A Flexible Framework*, edited by A. Voorrips, pp. 165-200. Studies in Modern Archaeology. vol. 3. HOLOS-Verlag, Bonn.

2005 Writing Archaeology: Analyses and Archaeological Argumentation. *The SAA Archaeological Record,* September 2005.

Kintigh, Keith W. and Albert J. Ammerman
1982 Heuristic Approaches to Spatial Analysis in Archaeology. *American Antiquity*, 47(1):31-63

Konigsberg, Lyle W., and Samantha M. Hens
1998 Use of Ordinal Categorical Variables in Skeletal Assessment of Sex From the Cranium. *American Journal of Physical Anthropology* 107:97-112

LeBlanc, Steven A.
1975 Micro-Seriation: A Method for Fine Chronologic Differentiation. *American Antiquity*, 40(1):22-38

Madsen, Torsten
1988 Multivariate Statistics and Archaeology. In *Multivariate Archaeology: Numerical Approaches in Scandinavian Archaeology*, edited by Torsten Madsen, pp. 7-27.

Marquardt, W.
1978 Advances in Archaeological Seriation. *Advances in Archaeological Method and Theory* 1: 257-314.

Matson, R. G.
1980 The Proper Place of Multivariate Techniques in Archaeology. *American Antiquity*, 45(2):340-344

Matson, R. G., D. L. True
1974 Site Relationships at Quebrada Tarapaca, Chile: A Comparison of Clustering and Scaling Techniques. *American Antiquity*, 39(1):51-74.

McAnany, Patricia A.
1989 Stone-Tool Production and Exchange in the Eastern Maya Lowlands: The Consumer Perspective from Pulltrouser Swamp, Belize. *American Antiquity*, 54(2):332-346.

McCartney, Peter H. and Margaret F. Glass
1990 Simulation Models and the Interpretation of Archaeological Diversity. *American Antiquity* 55(1990):521-36.

Meltzer, David J., Robert D. Leonard, and Susan K. Stratton
1992 The Relationship Between Sample Size and Diversity in Archaeological Assemblages. *Journal of Archaeological Science* 19:375-387.

Miguel-Angel, Cau, Peter M. Day, Michael J. Baxter, Ioulia Papageorgiou, Ioannis Iliopoulos, Giuseppe Montana
2004 Exploring Automatic Grouping Procedures in Ceramic Petrology. Journal of Archaeological Science 31:13251338.

Nance, Jack D.
1981 Statistical Fact and Archaeological Faith: Two Models in Small-Sites Sampling. *Journal of Field Archaeology* 8:151-165.

1990 Statistical Sampling in Archaeology. In *Mathematics and Information Science in Archaeology: A Flexible Framework*, edited by Albertus Voorrips , pp. 135-163.

Pinder, D., I. Shimada and D. Gregory
1979 The Nearest-Neighbor Statistic: Archaeological Application and New Developments. *American Antiquity* 44(3):430-45.

Read, Dwight W.
1978 Some Comments on the Use of Mathematical Models in Anthropology. *American Antiquity*, 39(1):pp. 3-15

Relethford, John H.
2008 Geostatistics and Spatial Analysis in Biological Anthropology. *American Journal of Physical Anthropology* 136(1):1-10.

Riley, Tim
2008 Diet and Seasonality in the Lower Pecos: Evaluating Coprolite Data Sets with Cluster Analysis. *Journal of Archaeological Science* 3527262741

Rhode, David
1988 Measurement of Archaeological Diversity and the Sample-Size Effect. *American Antiquity*, 53(4):708-716

Robertson, Ian G.
1999 Spatial and Multivariate Analysis, Random Sampling Error, and Analytical Noise: Empirical Bayesian Methods at Teotihuacan, Mexico. *American Antiquity* 64(1):137-152.

Samford, Patricia M.
1997 Response to a Market : Dating English Underglaze Transfer-Printed Wares. *Historical Archaeology* Vol.31, Number 2, pp. 1-30.

Schreiber, Katharina J. and Keith W. Kintigh
1996 A Test of the Relationship Between Site Size and Population. *American Antiquity* 61(3):573-579.

Shennan, Stephen
1997 *Quantifying archaeology*. 2nd ed. Edinburgh University Press, Edinburgh.

Simon, Arleyn W, and William A. Coghlan
1989 The Use of Indentation Testing to Obtain Precise Hardness Measurements from Prehistoric Pottery *American Antiquity*, 54(1):107-122.

South, Stanley
1977 Method and Theory in Historical Archaeology. Academic Press, New York.

Speth, John and Gregory Johnson
1976 Problems in the Use of Correlation for the Investigation of Tool Kits and Activity Areas. In *Culture Change and Continuity: Essays in Honor of James Bennett Griffin*, edited by C. E. Cleland, pp. 35-58. Academic Press, New York.

Stark, Barbara L. and Dennis L. Young
1981 Linear Nearest Neighbor Analysis *American Antiquity*, 46(2):284-300.

Steponaitis, Vincas P.
1981 Settlement Hierarchies and Political Complexity in Nonmarket Societies: The Formative Period of the Valley of Mexico. *American Anthropologist* 83:320-363.

Teltser, Patrice A.
1991 Generalized Core Technology and Tool Use: A Mississippian Example. *Journal of Field Archaeology*, 18(3):363-375.

Thomas, David Hurst
1978a The Awful Truth about Statistics in Archaeology. *American Antiquity* 43(2):231-244.
1978b Arrowheads and Atlatl Darts: How the Stones Got the Shaft. *American Antiquity,* 43(3):461-472.
1980 The Gruesome Truth about Statistics in Archaeology. *American Antiquity*, 45(2):344-345.

Walker, Phillip L.
2008 Sexing Skulls Using Discriminant Function Analysis of Visually Assessed Traits. *American Journal of Physical Anthropology* 136:39-50.

Wandsnider, LuAnn and Eileen L. Canilli
1992 The Character of Surface Archaeological Deposits and Its Influence on Survey Accuracy. *Journal of Field Archaeology* 19(2):169-188

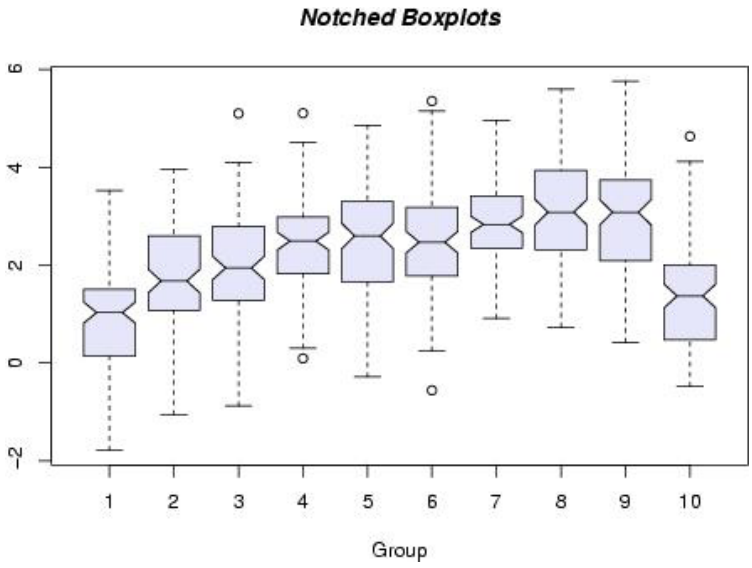
Whallon, Robert
1973 Spatial Analysis of Occupation Floors. 1: Application of Dimensional Analysis of Variance. *American Antiquity* 38(3):266-78.

1974 Spatial Analysis of Occupation Floors. 2, The Application of Nearest Neighbor Analysis. *American Antiquity* 39(1):16-34.

1984 Unconstrained Clustering for the Analysis of Spatial Distributions in Archaeology. In *Intrasite Spatial Analysis in Archaeology*, edited by H. Hietala, pp. 242-277. Cambridge University Press, Cambridge.

1987 Simple Statistics. In *Quantitative Research in Archaeology: Progress and Prospects*, edited by Mark S. Aldenderfer, pp. 137-150. Sage Publications, Newbury Park.

Vierra, Robert K and David L. Carlson
1981 Factor Analysis, Random Data, and Patterned Results *American Antiquity* 46(2):272-283.



Guidelines for Analyzing and Writing-up Quantitative Analyses

by George Cowgill

The Golden Rule of Data Analysis is this: Know exactly how you are going to analyze your data before you even begin to think of how to collect it. Ignoring this advice could lead to difficulties in any project.

1. The availability of powerful statistical packages such as SPSS, JMP, and SAS has made statistical data analysis very simple; making it easy and straightforward to subject a dataset to all manner of statistical analysis and tests of significance. It is, however, not advisable to proceed to formal statistical analysis without first exploring your data for transcription errors and the presence of outliers (extreme values). The importance of thorough preliminary examination of your data set before formal statistical analysis can not be overemphasized.
2. 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.
3. 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.
4. If you must do a hypothesis test, report the actual probability level obtained and don't treat some arbitrary level, such as 5%, as a talisman that tells you what to think.
5. It's not the sampling fraction that matters; it's the size of the sample.
6. 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.
7. When you read, always ask yourself whether you understand what denominator is implied.
8. "Frequency" should always mean count of something, rather than ratio of something to something else; the latter should be referred to as relative frequency.
9. Data is plural while datum is singular; thus, we say, these data are instead of the data is.
10. Dataset is one word (not data set or data sets).
11. The character, %, is usually spelled out, percent; thus, 10% should be 10 percent.
12. Use a space to separate the unit of measure from the value of the case; thus, 10ml should be 10 ml.
13. 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, 15, 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.
14. Boxplot is one word (not box plot) and is a noun, so there is no need to qualify it, i.e., boxplot chart; an alternative reference is box-and-whisker plot. Similarly, ANOVA Analysis or PCA Analysis is unnecessarily repetitive.
15. 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.
16. When comparing histograms, they must be based on the same scale for both X and Y axes.
17. Lowercase n reports the sample size, while uppercase N reports the population size; also, there is no need to italicize n or N because they are not statistics (i.e., they are not calculations).
18. All tables and figures should be numbered and have descriptive captions; the labels (e.g., Figure 1 or Table 2) should be referenced in the text when you discuss each table or figure.

Standard Protocol for Computer Assignments

Evaluation of your performance on each of the Computer Assignments is based on the quality of your analysis of the data and the quality of the argument you make concerning the archaeological question(s). You should be thorough in answering all questions posed for each problem, and present your results in a brief essay that references supporting figures and tables, as well as any relevant publications. Essays need to be typed, and the write-up is not to exceed five double-spaced pages of text using 12-point TNR font with 1-inch margins exclusive of figures, tables and any calculations. Longer papers will be down-graded. Figures and tables should always be clear and well-labeled using the techniques covered for graphic presentation covered in class and in Cleveland (1994). Generally follow the guidelines below.

1. Briefly state the problem (you may assume the reader has a reasonable knowledge of the methods we have discussed).
2. Present the results as you would in an article (other than giving somewhat less background on the methods).
3. Compactly and precisely state what the procedures that you employ actually do. Look at: Keith Kintigh's 2005 SAA Archaeological Record article Writing Archaeology including George Cowgill's side bar in this article List of the Things to Remember When You Forget Everything Else.
4. You must be very clear about what exactly procedures have been employed (don't say correlation if you mean the product moment correlation coefficient; don't say factor analysis if you mean principal components analysis), the set of data analyzed along with any transformations, *exactly what has been shown quantitatively, and exactly what of substance you conclude*.
5. You may wish to coordinate your work times with other members of the class so you can discuss any problems that arise. While I encourage you to discuss the methods and issues and help each other with computer problems, you will need to execute the procedures and write up the problems sets independently.

Standard Protocol for Writing Reports

The final project for the class is a publishable quality paper not to exceed 15 pages, exclusive of tables and figures (double-spaced, 1-inch margins, 12 pt. TNR font). Longer papers will be down-graded. Figures and tables should always be clear and well-labeled using the techniques covered for graphic presentation covered in class and in Cleveland (1994).

The final project is your opportunity to use the methods we have covered in the course to tease substantive meaning out of archaeological data. Note that the emphasis here is on meaning: the project should emphasize analytical methods as tools to make and evaluate defensible inferences about past dynamics from material evidence. You will need to identify an historical or anthropological issue and one or more sets of quantitative archaeological data that might be used to illuminate it. You are encouraged to choose issues and data in which you have a personal research interest and, therefore, basic familiarity with the current background literature. The data you use should contain information on several different variables that are of potential relevance to your problem. A key part of the project will be defending their relevance. The data should contain additional variables that independently document *either* the temporal *or* spatial contexts of the objects (assemblages, artifacts) characterized by the "relevant" variables.

Reports of experimental and analytical results, whether in the natural or social sciences, tend to use a different organizational structure from your average term paper. One goal of any analysis is to ensure its replicability. Therefore, these reports must include sufficient information about the objectives, procedures, and results such that the reader may compare, criticize, and replicate the study in an objective manner.

While you should situate your analysis with reference to relevant literature, the thrust of your paper should concern analysis. You may focus on any aspectbut you must define the problem that you wish to investigate. The following outline is suggested.

1. Introduction: What is the purpose of the analysis? Define a problem with specific questions to be translated into null hypotheses that you will analyze in the paper.
2. Background: You have some background information about your data. Use it. How does this analysis address a burning question? Where does it fit in with reference to what has been done up to this point?
3. Data: What are your data? How were they recorded? What is the unit of observation? What is the level of measurement of each variable? Do your data meet the necessary assumptions? Raw data goes in the appendix (not included in final page count). Please acknowledge the source of your data.
4. Procedure: Describe in detail. The reader must be able to replicate this experiment. Translate your problem into a statistical hypothesis. What statistical procedures will you use (correlation, ANOVA, regression)? Explain what the statistical procedures are designed to tell you. Discuss the requisite assumptions of the procedure (normality, lack of outliers, equal variance).
5. Calculations, Tables, and Graphs: This is the heart of your paper. Describe and illustrate patterns in your data using tables and graphs. I do not want raw output. Create your own tables from relevant output.
6. Results: Discuss and interpret the results obtained in the previous section. What are your conclusions with respect to your null hypotheses? Answer the question or questions posed in the introduction. Numbers and tables are not self-explanatory.
7. Discussion: Discuss your interpretations of the results in light of your introduction. How do your results compare with previous analyses of these data? How do these results advance our knowledge or put us back to the drawing board? What is explainable? What remains ambiguous or uninterpretable?
8. Conclusion---Wrap it up: Discuss the significance of your analysis, alluding to your introduction. Point out directions and questions for future work.
9. Appendix: Raw data and relevant output.

Standard Protocol for Presenting Papers

Your presentation must include a brief background, presentation of the data and results, and conclude with a short discussion and implications of the study. Presentations are limited to 15 minutes and must make full use of PowerPoint following the guidelines below.

1. Stressing What Is Important: The most important key to a successful presentation is to decide what your most important argument and conclusions are and to organize your presentation around these. It is generally best to state clearly at the beginning what you intend to demonstrate and why it is important, and to let the audience know how you will proceed through your argument. Each point raised in the course of the presentation should be clearly related to this argument. You should conclude with a restatement of your most important results, this time, allowing the audience to understand clearly that you

have indeed demonstrated what you set out to argue. Show your audience that you are interested in the essay! Use vocal inflection and be engaging. Remember to relax!!

2. Transitions: Transitions should be clear. It is almost impossible to be too obvious in an oral presentation. Obvious oral cues like, "I have three points. Number one will cover . . .," which sound wooden in writing, are helpful when read aloud.
3. Stay Focused: You will not be able to present everything from your paper. You do not need to provide all the background tracing how you reached this interpretation; present your point and back it up. You do not need to defend the validity of your idea. You also don't need to give a literature review. You want to make a clear, focused, and interesting argument that is backed up with a few interesting points of evidence, not give the entire content of your paper/study. Many conferences are intended for "works in progress" and expect presenters to bring up engaging questions and offer suggestions for future research, not give the final definitive word on a subject.
4. Visual aids: Make a note to yourself in your paper where you are going to use visual aids. Practice with your visual aids before you give your presentation.
5. Reading vs. Presenting: It is generally best not to read your paper from a prepared text unless you are a skillful reader or you have thoroughly practiced reading the paper, including ideally a run-through before a practice audience. When you read a paper without clarity of emphasis and dynamic intonation, it is almost inevitable that you will lose the attention of your audience. For this purpose, it is generally best to prepare an outline of the key points and to cover them in your presentation in natural, conversational speech.
6. Keeping to Allotted Time: This is extremely important. You are required to complete your presentation in the time allotted to you. I will end your presentation when your time has expired, regardless of whether you are finished. Therefore, it is crucial that you pace your presentation such that you are able to complete it and give a coherent ending within the time allowed. I strongly recommend that you practice your presentation -- if necessary, repeatedly -- until you can comfortably complete it in the available time.

Guidelines for PowerPoint Presentations

1. The text needs to be a simple, clear font like Arial or Helvetica, and large enough so it can be read from the back of the room. The rule of thumb is the smallest text should be 3% the height of the total height of the viewing area on the screen. Do not overwhelm your audience with a lot of text on a slide.
2. Don't try to cram two photos or graphs into the same screen - they get too small for the people in the back of the room to see the detail.
3. Be careful with the background colors - if the color darkness of the text is too close to that of the background, people will be trying to figure out what it says rather than listening to what you're saying. Use light on dark or dark on light, never dark on dark or light on light.
4. Special effects are acceptable - to a point. Use sparingly the gimmicks such as text sliding on and off the screen and fancy dissolving screens. No one would miss them if you didn't use them at all.
5. The picture files should be smaller than 1 megabyte in size. PowerPoint loads your whole show at the outset and keeps it in memory, so the total bulk of your slides must be less than the memory allocation for the PowerPoint program (usually around 20 megs). Resetting the memory allocation for PowerPoint can increase this limit, but demanding more memory than is available on the computer is a sure way to cause crashes. It's best to have the slides (including pictures) smaller than 200K. Using jpeg format facilitates this.

Some Common Statistical Terms

In order to use any statistical package (such as SPSS, Minitab, SAS, etc.) successfully, there are some common statistical terms that you should know. This document introduces the most commonly used statistical terms. These terms serve as a useful conceptual interface between methodology and any statistical data analysis technique. Irrespective of the statistical package that you are using, it is important that you understand the meaning of the following terms.

Most basic statistics techniques that you might learn in an introductory Stats 101 type class are called *parametric* statistics. Examples include a t-test, chi-square, and ANOVA. These tests all deal with estimating population parameters, such as determining the likelihood that two samples come from the same underlying population. Almost all of these test have assumptions such as the data is normally distributed and randomly selected (exact assumptions vary by technique). *Non-parametric* statistics avoid these assumptions, but can often perform some of the same types of calculations. For example, Kruskal-Wallis is a non-parametric form of an ANOVA and a Spearman correlation is a non-parametric equivalent of a Pearsons R correlation. *Exploratory Data Analysis* is the name given to the toolkit of techniques developed by John Tukey that try and deal with data sets for which the assumptions of parametric statistics may not hold. Many of these techniques are graphical in nature.

Variables: Most statistical data analysis involves the investigation of some supposed relationship among variables. A variable is therefore a feature or characteristic of a person, a place, an object or a situation which the experimenter wants to investigate. A variable comprises different values or categories and there are different types of variables.

Quantitative variables: Quantitative variables are possessed in degree. Some common examples of these types of variables are height, weight and temperature.
Qualitative variables: Qualitative variables are possessed in kind. Some common examples of these types of variables are sex, blood group, and nationality.

Hypotheses: Often, most statistical data analysis wants to test some sort of hypothesis. A hypothesis is therefore a provisional supposition among variables. It may be hypothesized, for example, that tall mothers give birth to tall children. The investigator will have to collect data to test the hypothesis. The collected data can confirmed or disproved the hypothesis.
Independent and dependent variables: The independent variable has a causal effect upon another, the dependent variable. In the example hypothesized above, the height of mothers is the independent variable while the height of children is the dependent variable. This is so because children heights are suppose to depend on the heights of their mothers.

Kinds of data: There are basically three kinds of data:

Interval data: These are data taken from an independent scale with units. Examples include height, weight and temperature.

Ordinal data: These are data collected from ranking variables on a given scale. For example, you may ask respondents to rank some variable based on their perceived level of importance of the variables.

Nominal data: Merely statements of qualitative category of membership. Example include sex (male or female), race (black or white), nationality (British, American, African, etc.).

It should be appreciated that both Interval and Ordinal data relate to quantitative variables while Nominal data refers to qualitative variables.

Overcoming Statistics Anxiety

The application of statistics to real archaeological problems will help you to overcome your anxiety of statistics and learn to love them. You may even feel empowered by the end of this class. Ideally, youll be addicted. In the meantime, it may be useful to review some of the common causes of math anxiety and some of the solutions to overcome it. There are five common factors and feelings that make us avoid math. When we avoid it, we lose confidence and then start building up dread and fear.

1. I'm Just Not Cut-out for Math

Sound familiar? Actually, there is no such thing as a brain type that makes one person better than another at math. Yes, studies show that there are different brain types, but those types just concern your *approach* at problem solving. Your approach can be different from anothers, but it can still be just as effective. One factor that affects math performance more than any other is confidence. Sometimes a stereotype can make us believe that we are naturally less capable than others. Studies have shown that math stereotypes are not true! Interestingly, studies do show that positive thinking can improve math performance. Basically, there are two things that you can do to improve your math performance: 1) Do not accept stereotypes about math, and 2) Think positive thoughts. If you are smart at any skill at all, then you can be smart at math. Also remember that if you were not smart at any skill at all, then you would not be here.

2. Building Blocks are Missing

This is a legitimate cause for anxiety. If you avoided math in middle school or college, you may be feeling stressed out because you know your background is weak. You can overcome this problem easily by skimming through a textbook that is written for a level slightly lower than the current class. First, you'll be surprised at how much you do know. Second, you'll find there are only a few skills you need to practice before you're completely caught up. And those skills will come easily!

3. It's Just So Boring!

This is a false accusation. There are many mysteries in archaeology and anthropology that require mathematical computations and statistics. If you embrace statistics, then you will embark on solving some of these great mysteries. Statistics poses challenges that can be amazingly gratifying to conquer.

4. Procrastination

It is true that many people suffer real anxiety when it comes to setting aside a certain span of time for statistics. This is one of the factors that often leads to procrastination, and it manifests in people of all ages. Realize that it's normal to resist devoting time to statistics. Then simply think your way through your fear. Think about the other things in your life that you'll need to set aside. With practice, you'll soon realize that statistics, although challenging, is a lot of fun and that hours go by effortlessly when youre lost in the details of a problem. You may even find that it is difficult to pull away. The hardest part is the learning curve associated with particular statistical packages, but even this can be fun once the basics are mastered. Be patient, resist frustration, and remember to talk to your

fellow students (and professor). The most important thing is not to procrastinate. Start your problems early so that you dont find yourself unable to solve a software issue prior to the due date of a project. This added stress will will only promote bad habits of procrastination

5. It's Too Complex to Understand

It is true that statistics involves some very complex formulas. Remember the process for overcoming any fear? Isolate it, examine it, and break it down into little parts. That's exactly what you have to do with statistics and (computer programs). Every formula is made of "little parts" or skills and steps that you've learned in the past. It's a matter of building blocks. When you come across a formula or process that seems too complex, just break it down. If you find that you're a little weak on some of the concepts or steps that make up one element of the formula, then just go back and work on your building blocks.

Some Helpful Suggestions

1. Buy used textbooks.

Sometimes we don't understand a concept because the explanation is just plain bad or it's not written in a way we can understand. It's good to have multiple texts that give alternate explanations and additional examples to work out. Many used book stores will have inexpensive texts and the library also has numerous texts. Suggestions for archaeologists are provided below.

2. Study actively.

Don't just work out a problem. Draw pictures and diagrams of a process and make up stories to go along with them. If you are an auditory learner you may want to define statistical terms or procedures and read textbook and article examples out loud.

3. Read actively.

Use sticky note flags to mark important things in chapters and articles or things you need to inquire about in class. Read the abstract (beginning) and conclusion (end) of assigned chapters and articles first. Take a look at the issues and problems that are being solved to get a preview of the goals. This gives your brain a framework to work with.

4. Make flashcards for terms.

Flashcards are good for visual and tactile learners. They reinforce information as you see it and as you create it with your own hand.

5. Take breaks.

If you come across a problem that you don't understand, read it over a few times and trybut then walk away from it and make a sandwich or do some other small task (not other homework). Your brain will continue to work on the problem subconsciously.

6. Talk to others:

Talking to others when you are stuck on a problem can be intimidating, especially if you dont feel confident about learning statistics. Two things can happen. You either come to the answer yourself by explaining what you dont understand, or you come to the answer through the assistance of someone else. Both options are better than crying yourself to sleep. Realize that everyone is wired differently. Other students may have problems understanding something you find elementary. Then it will be your turn to help them. What ever the case, we will

follow the Las Vegas Rule in and out of the class: *What happens in statistics, stays in statistics*. So dont let your ego get in the way of learning.

7. *Dont get lost in the formulas*

Some of the articles we will read have complex formulas and code that can be perplexing. In most cases, these formulas are easy to unpack by breaking down the individual parts, but do not spend a whole lot of time with this if it adds to your anxiety. Realize that the formulas are there to provide computations for new or untested techniques and protocols. They also allow other researchers to evaluate the assertions claimed by the author. This is an introductory stats class, so we will leave these formulas to the experts and focus on standard techniques and computer protocols for exploring data.

On-line Statistical Tutorials

1. *Statnotes: Topics in Multivariate Analysis:* http://www2.chass.ncsu.edu/garson/PA765/statnote.htm
   This site contains many topics such as ANOVA, ANCOVA, Cluster Analysis, Correlation, Correspondence Analysis, Logistic Regression, Logit, Probit models, etc. with SPSS related examples. Each topic includes a description of key concepts and term, assumptions, FAQ, and bibliography.
2. UCLA Academic Technology Services: http://www.ats.ucla.edu/stat/mult_pkg/whatstat/default.htm
   Presents a general guide for choosing the correct statistical analysis with instructions for SPSS
3. Winks, Statistics Tutorials for Statistical Data Analysis, SAS, SPSS, WINKS, Excel: http://www.stattutorials.com/
   Includes many tutorials for SPSS including logistical regression analysis
4. StatSoft, An Electronic Textbook: http://www.statsoft.com/textbook/stathome.html
5. Stat Trek: http://stattrek.com/
6. Concepts and Applications of Inferential Statistics: http://faculty.vassar.edu/lowry/webtext.html

On-line Statistical Program Tutorials

1. SPSS: SPSS Tools has numerous links at: http://www.spsstools.net/spss.htm. See also University of Toronto http://psych.utoronto.ca/courses/c1/spss/toc.htm, Central Michigan: http://calcnet.mth.cmich.edu/org/spss/toc.htm , and UCLA:

2. JMP: For Version 5 and lower: http://web.utk.edu/~leon/jmp/default.html. See also, Quick Reference Guide for Version Seven: http://www.jmp.com/academic/pdf/421242_jmpse_guide.pdf, and Beginning Demo Videos (Version 7): http://www.jmp.com/academic/se_demos.shtml

3. R: See the Documentation (Other) page for R at: http://www.r-project.org/: Especially useful are: Simple R, R for Beginners, Using R, and Introduction to R Graphics.

   The two best printed reference manuals on R are: Michael Crawley (2007) *The R Book*, Wiley Hoboken, N.J., and Paul Murrell (2006) *R graphics*, Chapman & Hall/CRC, Boca Raton.

Web Sites and Blogs

There is a Wiki on Quantitative archeology with some useful resources and sample R code at http://wiki.iosa.it/.

There is one great blog on statistics written by a statistician at Columbian University named Andrew Gellman: Statistical Modeling: http://www.stat.columbia.edu/~cook/movabletype/mlm/

The field of data visualization is a thriving field of both research and writing. The Cleveland and Tufte texts mentioned above are the most crucial, but it can also be fun to read bloggers that can share their informed opinions about charts they find in newspapers and journals. Andrew Gelman (above) sometimes discusses data visualization and this blog specializes in it: Junk Charts http://junkcharts.typepad.com/junk_charts/

Other blogs look at data visualization but focus mostly on infographics which are usually not desirable for academic presentations. Their comments can still be useful. I recommend:

Flowing Data	http://flowingdata.com/
Information Aestetics	http://infosthetics.com/
EagerEyes	http://eagereyes.org/
Statistical Graphics and Data Visualization	http://statisticalgraphics.blog.com/

Additional Text Resources

1. General Archaeological Resources

Shennan, Stephen

1997 *Quantifying archaeology*. 2nd ed. University of Iowa Press, Iowa City.

This is my most recommended single text. *Make sure and get the second edition* which is significantly revised from the first. Does a good job of discussing parametric and multivariate approaches and has a small section on sampling.

Drennan, Robert D.

1996 *Statistics for archaeologists : a commonsense approach*. Interdisciplinary contributions to archaeology. Plenum Press, New York.

An excellent well-written introduction to the basic concepts of statistics. Mostly deals with parametric statistics and the basics of sampling, populations, and randomness. Its best feature is that it discusses how to present the results of basic analyses.

Doran, J. E. and F. R. Hodson

1975 *Mathematics and computers in archaeology*. Edinburgh University Press, Edinburgh.

A classic introduction to many of the techniques important to archaeologists.

Orton, Clive
1982 *Mathematics in archaeology*. Cambridge University Press, Cambridge.

Good introduction organized around chapters like What is it? How old is it? Where does it come from? etc.

Thomas, David H.
1986 *Refiguring Anthropology: First Principles of Probability and Statistics*. Waveland Press, Prospect Heights, Illinois.

A canonical introduction to basic statistics. Covers more techniques than Madrigal but it is probably not as easy to read. Better as a reference book.

Fletcher, Mike and Gary R. Lock
2005 *Digging numbers : elementary statistics for archaeologists. Second Edition.* Oxford University Committee for Archaeology, Oxford.

An very basic, short, and easy to read summary of introductory statistics. The examples in the book use SPSS, but examples using R can be found at http://wiki.iosa.it/dokuwiki/diggingnumbers:start.

2. Advanced Archaeological Resources

Baxter, Michael J.
2003 *Statistics in archaeology*. Hodder Arnold, New York.

Could be considered the advanced methods appendix to the Shennan book. Examines the application of the most cutting edge techniques in archaeology, such as Baysian methods, spatial analysis, Monte Carlo techniques and others.

Baxter, M. J.
1993 *Exploratory multivariate analysis in archaeology*. Edinburgh University Press, Edinburgh.

The single most intensive resource on multivariate approaches in archaeology. Can be hard to find.

Buck, Caitlin E., William G. Cavanagh and Clifford D. Litton (editors)
1996 *Bayesian approach to interpreting archaeological data*. Wiley, Chichester, England.

Only book-length treatment of Bayesian methods in archaeology. In archaeology the most significant application has been to the treatment of chronometric dates.

3. Spatial Analysis

Clarke, David L.
1977 *Spatial archaeology*. Academic Press, London ; New York.

Hodder, Ian and Clive Orton
1976 *Spatial analysis in archaeology*. New studies in archaeology. 1. Cambridge University Press, Cambridge ; New York.

1978 *The Spatial organisation of culture*. University of Pittsburgh Press, Pittsburgh.

Kroll, Ellen M. and T. Douglas Price (editors)
1991 *The Interpretation of archaeological spatial patterning*. Plenum Press, New York.

Robertson, Elizabeth C., Jeffery D. Seibert, Deepika C. Fernandez and Mark U. Zender (editors)
2006 *Space and spatial analysis in archaeology*. University of Calgary Press, Calgary.

4. Typology

Whallon, Robert and James A. Brown (editors)
1982 *Essays on archaeological typology*. Kampsville seminars in archeology ; v. 1. Center for American Archeology Press, Evanston, Ill.

Faunal Analysis

Lyman, R. Lee
2008 *Quantitative Paleozoology*. Cambridge manuals in archaeology. Cambridge University Press, Cambridge; New York.

5. Sampling

Mueller, James W. (editor)
1975 *Sampling in archaeology*. University of Arizona Press, Tucson.

Orton, Clive
2000 *Sampling in archaeology*. Cambridge manuals in archaeology. Cambridge University Press, Cambridge.

6. Diversity

Leonard, Robert D. and George Thomas Jones (editors)
1989 *Quantifying diversity in archaeology*. Cambridge University Press, Cambridge.

7. Data Visualization

The best single resource on data visualization is

Cleveland, William S.
1994 *The elements of graphing data*. Rev. ed. AT&T Bell Laboratories, Murray Hill, N.J.

Other excellent canonical resources include:

Cleveland, William S.
1993 *Visualizing data*. AT&T Bell Laboratories; Hobart Press, Murray Hill, N.J.

Tufte, Edward R.

1997 *Visual explanations : images and quantities, evidence and narrative*. Graphics Press, Cheshire, Conn.

2001 *The visual display of quantitative information*. 2nd ed. Graphics Press, Cheshire, Conn.

2003 *Envisioning information*. 9th printing, Aug. 2003. ed. Graphics Press, Cheshire, Conn.

2006 *Beautiful evidence*. Graphics Press, Cheshire, Conn.

Andrienko, Natalia and Gennady Andrienko

2006 *Exploratory analysis of spatial and temporal data : a systematic approach*. Springer, Berlin; New York

(This book on data visualization is free electronically through SpringerLink)

8. EDA and Nonparametric methods

The father of Exploratory Data Analysis is John Tukey and his first book is a landmark in the field:

Tukey, John Wilder

1977 *Exploratory data analysis*. Addison-Wesley Pub. Co., Reading, Mass.

Unfortunately, it is also dense and challenging to interpret. The most readable treatment of the topic I know of can be found here:

Velleman, Paul F. and David C. Hoaglin

1981 *Applications, basics, and computing of exploratory data analysis*. Duxbury Press, Boston, Mass.

My favorite resource on nonparametric statistics is the comprehensible and readable:

Conover, W. J.

1999 *Practical nonparametric statistics. Third Edition.* Wiley, New York.

9. Cultural Anthropology

Most cultural anthropologists will find a good text on qualitative methods useful. For books on quantitative methods many will find the books written for Social Scientists useful. Two books that specifically discuss quantitative methods in the context of ethnographic data are:

Madrigal, Lorena

1998 *Statistics for Anthropology*. Cambridge University Press, Cambridge.

> Well written introduction to basic statistics using anthropological examples, primarily covers parametric techniques. This book does an excellent job stating the underlying assumptions behind each technique.

Johnson, Allen W.
1978 *Quantification in cultural anthropology : an introduction to research design*. Stanford University Press, Stanford, Calif.

A discussion about how to think about quantitative procedures for cultural anthropologists. This is NOT an introduction to statistical methods such as how to do a t-test or a chi-square.

Bernard, H. Russell
2005 *Research methods in anthropology : qualitative and quantitative methods. Fourth Edition.* AltaMira Press, Walnut Creek, CA.

The single most frequently used resource for cultural anthropological methods. Contains some discussion of quantitative approaches.

10. Biological Statistics

Sokal, Robert R. and F. James Rohlf
1995 *Biometry : the principles and practice of statistics in biological research. Third Edition*. W.H. Freeman, New York.

This appears to be the most frequently used introductory statistics text for the biological fields. Covers parametric and non-parametric techniques. A list of statistical tables that accompanies the book is sold separately.

Zar, Jerrold H.
1999 *Biostatistical analysis. Fourth Edition.* Prentice Hall, Upper Saddle River, N.J.

Another popular introduction to statistics with biological examples. Primarily deals with parametric techniques.

11. Social Science Statistics

There are many books in this category as an Amazon search of social science and statistics will reveal. Some examples include:

Aron, Arthur, Elaine Aron and Elliot J. Coups
2008 *Statistics for the behavioural and social sciences : a brief course. Fourth Edition*. Pearson Prentice Hall, London.

Cohen, Barry H. and R. Brooke Lea
2004 *Essentials of statistics for the social and behavioral sciences*. Essentials of behavioral science series. Wiley, Hoboken, NJ.

Levin, Jack
2007 *Elementary statistics in social research : the essentials*. Allyn And Bacon, New York.

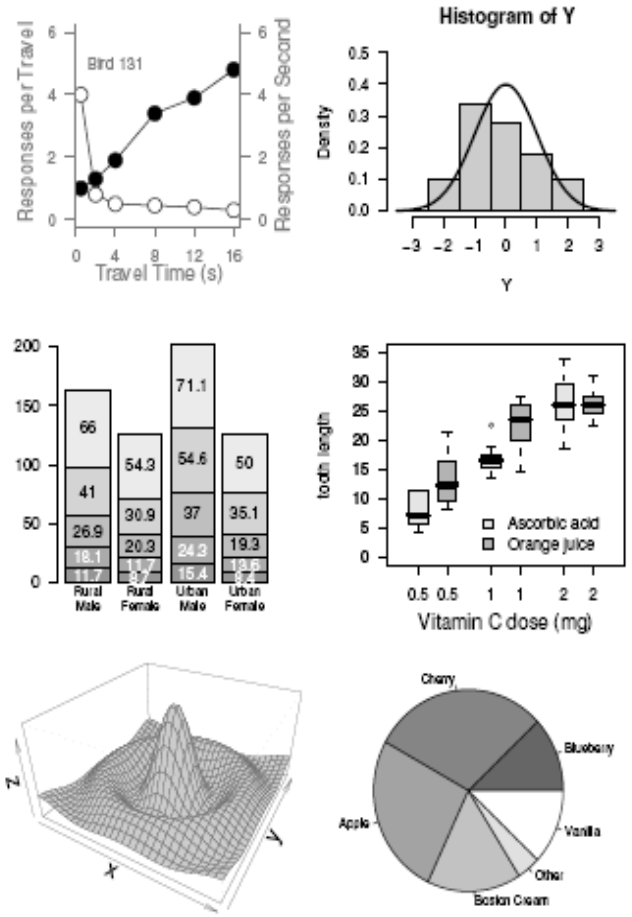
12. General Statistics

Freedman, David, Robert Pisani and Roger Purves
2007 *Statistics, Fourth Edition*. W. W. Norton, New York.

This is the single most frequently recommended general introductory statistics text. Used copies of the third edition, published in 1998, are relatively inexpensive.

Gonick, Larry and Woollcott Smith
1993 *The cartoon guide to statistics*. Collins Reference, New York.

This is actually a reasonably clear and concise description of the most basic concepts of probability. Dont select this as a reference work, but it might be fun to peruse a copy if you have high math anxiety.



Which Statistical Test should I Use?

Number of Dependent* Variables	Number of Independent** Variables	Type of Dependent Variable(s)	Type of Independent Variable(s)	Measure	Test(s)
1	0 (1 population)	continuous normal	not applicable (none)	mean	one-sample t-test
		continuous non-normal		median	one-sample median
		categorical		proportions	Chi Square goodness-of-fit, binomial test
	1 (2 independent populations)	normal	2 categories	mean	2 independent sample t-test
		non-normal		medians	Mann Whitney, Wilcoxon rank sum test
		categorical		proportions	Chi square test Fisher's Exact test
	0 (1 population measured twice) *or* 1 (2 matched populations)	normal	not applicable/ categorical	means	paired t-test
		non-normal		medians	Wilcoxon signed ranks test
		categorical		proportions	McNemar, Chi-square test
	1 (3 or more populations)	normal	categorical	means	one-way ANOVA
		non-normal		medians	Kruskal Wallis
		categorical		proportions	Chi square test
	2 or more (e.g., 2-way ANOVA)	normal	categorical	means	Factorial ANOVA
		non-normal		medians	Friedman test
		categorical		proportions	log-linear, logistic regression
	0 (1 population measured 3 or more times)	normal	not applicable	means	Repeated measures ANOVA
	1	normal	continuous		correlation simple linear regression
		non-normal			non-parametric correlation
		categorical	categorical or continuous		logistic regression
			continuous		discriminant analysis
	2 or more	normal	continuous		multiple linear regression
		non-normal			
		categorical			logistic regression
		normal	mixed categorical and continuous		Analysis of Covariance General Linear Models (regression)
		non-normal			
		categorical			logistic regression
2	2 or more	normal	categorical		MANOVA
2 or more	2 or more	normal	continuous		multivariate multiple linear regression
2 sets of 2 or more	0	normal	not applicable		canonical correlation
2 or more	0	normal	not applicable		factor analysis

* outcome

** predictor