Teaching American History:
Data Analysis Resources

Robert S Michael, Ph.D.
Center for Evaluation and Education Policy

I. Introduction

The goal of this teleconference is to provide a conceptual overview of data analysis. The assumptions underlying selected statistical tests and research designs receive special attention. (Click on the urls below to activate the link).

II. Data preparation and summary

A. Data screening/editing: How clean is your data? Codebooks.

a) Israel & Moore: Coding Techniques
http://edis.ifas.ufl.edu/pdffiles/PD/PD00400.pdf

b) Creating a Codebook
http://bss.sfsu.edu/tygiel/hist661/codebook/default_files/frame.htm

c) Gonzále: Creating a Codebook
http://www.csusm.edu/gonzalez/psyc402/Creating%20a%20Codebook%20Spring%202002.htm

d) Respect for Data — Sage advice for handling your data.
http://www.indiana.edu/~educy520/sec6342/week_08/respect_for_data.pdf

III. Descriptive statistics: Why bother?

A. Identify data problems (e.g., missing values, impossible values, outliers).

B. Summarize and describe data succinctly.

a) David Lane. HyperStat text. See chapters 2, 3, 4
http://davidmlane.com/hyperstat/

b) WebStat: Descriptives
http://www.une.edu.au/WebStat/unit_materials/c4_descriptive_statistics/index.html

c) Michael. Descriptive Statistics and Exploratory Data Analysis
http://www.indiana.edu/~educy520/sec6342/week_08/descriptive_stats_outline.pdf

d) Exploratory Data Analysis. cf. *The Role of Graphics* and *Graphical Techniques: Alphabetical*
http://www.itl.nist.gov/div898/handbook/eda/eda.htm

e) Classic Anscombe example:
http://www.itl.nist.gov/div898/handbook/eda/section1/eda16.htm

C. Check assumptions required for inferential statistics.

1. Assumptions for parametric tests include: a) Random sampling. b) Scores in sample are independent of one another. c) Sample comes from normal distribution. d) Variance in different groups is approximately equal. e) Measurement scale is at least interval level.
2. How to test assumptions:
   a) Garson. Testing of Assumptions
   http://www2.chass.ncsu.edu/garson/pa765/assumpt.htm
   b) Prophet StatGuide: Distribution Tests
   http://www.basic.nwu.edu/statguidefiles/dist.html
   c) Indiana University. Testing Normality
   http://www.indiana.edu/~statmath/stat/all/normality/Testing_Normality.pdf
   d) Micceri (1989). The Unicorn, The Normal Curve, and Other Improbable Creatures.
   http://www.indiana.edu/~educy520/sec6342/week_08/micceri89.pdf

IV. Inferential Statistics

A. Statistical tests based on distributions (z, t, chi-square, F): Which to use?

a) Meyer (1976).
http://www.indiana.edu/~educy520/sec6342/week_10/meyer76.pdf
b) Motulsky (1995). How to choose a statistical test.
http://www.graphpad.com/www/book/Choose.htm

B. When assumptions are violated: Try a transformation first.

a) Michael. Transformations
http://www.indiana.edu/~educy520/sec6342/week_08/transformations.pdf

C. Non-parametric statistical tests: What are they and when to use

a) Ask Professor Mean
http://www.cmh.edu/stats/ask/parametric.asp
b) Prophet StatGuide: Nonparametric assumptions
http://www.basic.nwu.edu/statguidefiles/nonpar.html
c) Halperin. Table of corresponding non-parametric tests
http://www.ns.purchase.edu/psych/statistics/handouts/nonparametric.htm
d) Yu, Alex. Nonparametric Caveats
http://seamonkey.ed.asu.edu/~alex/teaching/WBI/parametric_test.html
e) StatSoft. Nonparametric Statistics
http://www.statsoft.com/textbook/stnonpar.html

V. Hypothesis Testing & Statistical Power

A. Small sample size — implications and consequences

1. The variance in small samples is unreliable. We cannot be confident that the variance observed is representative of the variance in the population, and it is the variance that we see when we look at a graph of a distribution.
2. Sample size calculation
   a) Prof. Mean: Selecting an appropriate sample size
   http://www.cmh.edu/stats/size.asp
   b) Lenth. Comments about sample size
   http://www.stat.uiowa.edu/techrep/tr303.pdf
   c) Sample size calculator
   http://www.stat.uiowa.edu/~rlenth/Power/

B. Statistical significance and practical significance: Are they the same?

a) Michael. Statistical Inference Summary
http://www.indiana.edu/~educy520/sec6342/week_10/hypothesis_test_summary021126.pdf

b) Glaser. The Statistical Inference Controversy.
http://www.aacn.org/AACN/jrnlajcc.nsf/GetArticle/ArticleThree85?OpenDocument

c) Kirk, Roger (1996). Practical Significance: A Concept whose time has come. *Educational and Psychological Measurement, 56*, 746-759.

C. Effect size: What is it and how to calculate

a) Thompson, Bruce. (1997). Computing Effect Sizes
http://www.coe.tamu.edu/~bthompson/effect.html

b) Becker, Lee. (1997). Effect Sizes
http://www.uccs.edu/~lbecker/psy590/es.htm

c) Effect Size Calculators
http://www.uccs.edu/~lbecker/psy590/escalc3.htm
http://davidmlane.com/hyperstat/effect_size.html

D. Hypothesis Testing, Statistical Power, Effect Size

a) Statistical Power & Effect Size
http://www2.msstate.edu/~dmorse/8214_power_es.pdf

b) Barrett, Paul. (2000). Hypothesis Testing and Power Analysis
http://www.pbarrett.net/hyptest.pdf

c) Statistical Power Calculators
http://www.uky.edu/~ldesh2/power.htm
http://www.dssresearch.com/toolkit/spcalc/power.asp
http://calculators.stat.ucla.edu/powercalc/

VI. Review of selected designs

A. Overview of designs: *"Issues of design always trump issues of analysis."* — G.E. Dallal

a) Cobb, George W. (2002). *Introduction to Design and Analysis of Experiments.* New York: Springer-Verling.

b) Keppel, Geoffrey. (2004). *Design and Analysis: A Researcher's Handbook*. New York: Prentice-Hall.

c) Michael. Experimental Designs
http://www.indiana.edu/~educy520/sec6342/week_05/experimental_designs2004handout.pdf
http://www.indiana.edu/~educy520/sec6342/week_05/exp_designs_2up.pdf

d) US DOE. *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User-Friendly Guide.*
http://www.excelgov.org/usermedia/images/uploads/PDFs/User-Friendly_Guide_12.2.03.pdf

B. Pre-test—Post-test, Matched Group Design

1. This is a quasi-experimental design in which existing, intact groups are used. Participants are matched on one or more variables; thus, we can't be certain if groups are equivalent on the remaining unmatched variables. Matching is never a substitute for random sampling and random assignment to groups (viz., for justifying statements of causality).

2. Gain Scores: The difference between the the pre-test score and the post-test score is known as the "gain." Different classes begin with different pre-test scores. Classes beginning with a low pre-test score have more room for improvement than do classes beginning with a high pre-test score. A class that begins low might show more gain and yet its post-test score may still be lower than the pre-test score of another class. The normalized gain score corrects for the initial pre-test level by expressing gain as the ratio of actual gain to possible gain:

   normalized gain = (posttest - pretest)/(total - pretest)

   See Metzler for another apporach to normalizing gain scores:

   a) http://www.physics.iastate.edu/per/docs/Addendum_on_normalized_gain.pdf

3. Analysis of Pre-test — Post-test data.

   a) Paired t-tests for the pre- and post-scores of the dependent groups. Also compare post-test score of treatment group to score of comparison group.

   b) Repeated Measures. Avoids the danger of significant results due to chance that can arise from performing multiple t-tests.

   c) Individual Growth Curves: Willett, John B., Singer, Judith D., & Martin, Nina C. (1998). The design and analysis of longitudinal studies of development and psychopathy in context: Statistical Models and methodological recommendations. *Development and Psychology, 10,* 395-426.
   http://gseweb.harvard.edu/~faculty/singer/devpsych.pdf

C. True experimental designs (random selection and random assignment)

a) One factor designs: Traditional one-factor-at-a-time approach. (Use t-test or one-way anova).

b) Factorial Designs. Several factors are varied simultaneously in an experimental layout (known as the "factorial design"). Advantages include the ability to detect interactions among the factors under study and reduction in the number of subjects needed. (e.g., look at teachers and students).

c) Crossed vs nested designs. Factors are crossed if every combination of levels of factors occurs. That is, every level of one factor is repesented at every level of the other factors. In a nested design, one factor is nested within another, meaning each level of one factor (e.g., students) occurs with one level of the higher factor (e.g., teachers). Note that nested designs violate the assumption of independence and so we must adjust the error term in our statistical test.
Crossed example: http://bill.psyc.anderson.edu/exdes/ex5.htm
Nested example: http://bill.psyc.anderson.edu/exdes/ex6.html

d) Nested Designs: Hierarchical Linear Modeling, Multilevel Models, Random Coefficient models:
Kaz www.estat.us/sas/KazHLM.ppt
Singer Textbook http://gseacademic.harvard.edu/alda/
Raudenbush, Steven W. (1995). Hierarchical models: The case of school effects on literacy. in Marilyn Binkley, Keith Rust, & Marianne Wingless (Eds.). *Methodological Issues in Comparative Educational Studies: The Case of the IEA Reading Literacy Study.* Washington, DC: USDOE.

e) Hlm and effect size
Castro, Maria & Gavirla, Jose-luis. (2000). Application of hierarchical linear models to meta-analysis: Study of the Monte Carlo Simulation on the functioning of traditional and empirical-Bayes effect size. *Quality and Quantity, 34*, 33-50.

D. Quasi-experimental designs (problems and commonly used alternatives)

1. Blocking
   a) Michael. Experimental Designs
   b) Becker
   http://www.uccs.edu/~lbecker/psy590/blocking.htm
   c) Cobb, p. 247, p. 677

2. Matching
   a) Michael. Experimental Designs
   b) Cobb, p. 325
   c) Becker
   http://www.uccs.edu/~lbecker/psy590/matching.htm

3. Correlation & Regression
   a) Cobb, p. 339 — Recognizing Alternatives to ANOVA
   b) Becker. Correlation and Regression.
   http://www.uccs.edu/~lbecker/psy590/ancova.htm

4. Analysis of Covariance
   a) Prophet StatGuide
   http://www.basic.nwu.edu/statguidefiles/ancova.html

b) Cobb, p. 673-683. Analysis of Covariance

c) Becker
http://www.uccs.edu/~lbecker/psy590/ancova2.htm

d) Loftin, L., & Madison, S. (1991). The extreme dangers of covariance corrections. In B. Thompson (Ed.), (1991). *Advances in educational research: Substantive findings, methodological developments* (Vol. 1, pp. 133-148). Greenwich, CT: JAI Press. (IBSN: 1-55938-316-X)

e) Thompson, B. (1992). Misuse of ANCOVA and related “statistical control” procedures. *Reading Psychology, 13*, iii-xviii.