Improving the Teaching of Applied Statistics: Putting the Data Back Into Data Analysis

JUDITH D. SINGER and JOHN B. WILLETT*

Artificial data sets are often used to demonstrate statistical methods in applied statistics courses and textbooks. We believe that this practice removes much of the intrinsic interest in learning to do good data analysis and contributes to the myth that statistics is dry and dull. In this article, we argue that artificial data sets should be eliminated from the curriculum and that they should be replaced with real data sets. Real data supplemented by suitable background material enable students to acquire analytic skills in an authentic research context and enable instructors to demonstrate how statistical analysis is used to model real world phenomena. To facilitate the integration of real data into applied statistics curricula, we identify seven characteristics that make data sets particularly good for instructional use and present an annotated bibliography of more than 100 primary and secondary data sources.

KEY WORDS: Artificial data; Real data; Statistical education.

1. INTRODUCTION

When students enroll in service courses to learn statistical methods, they do so because they are fascinated by some critical issue in psychology, education, communication, economics, or other applied area. They must learn to be good and careful data analysts so that they can return to their substantive arenas and answer the research questions that interest them. Although statistical methodology is usually taught through data-based examples, the data used rarely originate in practice. Most data sets are expedient, artificially created lists of numbers that give computational practice, as typified by the following problem from Pedhazur (1982):

Here are a set of X and a set of Y scores . . .

X: 2 2 1 1 3 4 5 5 7 6 4 3 6 6 8 9 10 9 4 4
Y: 2 1 1 1 5 4 7 6 7 8 3 3 6 6 10 9 6 6 9 10

Calculate:

(a) The means, sums of squares and cross products, standard deviations, and the correlation between X and Y.
(b) The regression of Y on X.
(c) Regression and residual sums of squares.
(d) The F ratio for the test of significance of the regression of Y on X. (p. 43)

Although data sets like these are mathematically tractable and may help students learn the algebraic underpinnings of statistical methods, we believe that such data sets do little to help students become competent and invigorated data analysts. Artificial data sets help perpetuate the myth, prevalent among nonstatisticians, that statistics is dull and boring. After analyzing the "data," students have not experienced the thrill of doing research to answer an interesting research question, nor have they learned how statistical models can be interpreted in the context of real life.

We believe that real data sets provide a more meaningful and effective vehicle for the teaching of applied statistics. Such data enable students to acquire analytic skills in a realistic research context. They provide a legitimate arena in which the linking of research questions and statistical models can be practiced. Real data help students see how statistics can inform topical debate, enabling us to teach not only *how* data are analyzed, but also *why* they are analyzed. Using real data helps integrate statistics into the general education curriculum.

In this article we further explore the rationale for replacing artificial data by real data in the applied statistics classroom and provide an extensive annotated bibliography of primary and secondary data sources that instructors might find useful. In addition, we describe some of the characteristics of a pedagogically suitable data set and outline some of the difficulties that we have encountered when using real data sets.

2. THE RATIONALE FOR USING REAL DATA

Good data analysts require diverse skills if they are to conduct analyses of high methodological quality. They must be able to formulate interesting research questions, select appropriate statistical techniques, conduct the necessary calculations, interpret the analytic results, identify rival explanations of the results, and summarize their findings in a cogent and convincing manner. The challenge for the teacher is how best to enhance these skills.

Before the widespread availability of high-speed computing and prepackaged statistics programs, the computational aspects of data analysis assumed priority over the other skills. After all, the success of the analysis hinged upon the analyst's ability to perform the requisite calculations correctly. Recognizing that calculations could be time-consuming and tedious, instructional settings were designed so that the student burden was reduced by the use of arithmetically simple artificial data sets. Observations in such data sets were usually integers, often chosen so than hand-computed summary statistics were also integers. *The American Statistician* periodically published articles describing methods for constructing such data sets (see, e.g., Carmer and Cady 1969; Dayton 1972; Edwards 1959; Read 1985; Read and Riley 1983; Searle and Firey 1980), and artificial data sets were standard fare in the popular applied statistics textbooks in the social sciences (see, e.g., Hays 1981; McCall 1975; Winer 1971).

*Judith D. Singer and John B. Willett are Associate Professors of Education, Harvard University, Graduate School of Education, Cambridge, MA 02138. An earlier version of this article was presented at the annual meeting of the American Educational Research Association, April 1988, New Orleans, Louisiana. The authors would like to thank Howard Wainer and several anonymous referees for their helpful comments on an earlier version. The order of the authors was determined by randomization.

Although the use of artificial data shortened the time spent squaring and summing columns of numbers, some of the drudgery of hand computation remained. Calculations became easier, but still had to be carried out. This led to the rise of a cookbook approach to data analysis, based on step-by-step computational formulas ("recipes") given to decrease the computational burden.

Although using artificial data sets and cookbook strategies were intended to improve the quality of statistics instruction, the result usually fell far short of that goal. Artificial data sets confirmed students' expectations that statistics was boring and unrelated to their own substantive interests. Cookbook approaches to data analysis seduced them into believing that statistical analysis was always confirmatory, never exploratory, and that it could be reduced to a set of predefined steps conducted by a robot.

Today, computers have revolutionized the way we analyze data and they should be permitted to revolutionize the way we teach data analysis. They have eliminated the need for simplified arithmetic; the computer does not care if the observations and summary statistics are integers. Tedious calculations should be relegated to the machine. Students need not memorize formulas whose sole purpose is computational simplification. Exploratory analyses, which previously were avoided because of the time required to conduct them, can now be appropriately incorporated into the data-analytic repertoire.

One positive step in the new direction is the increasing presence of computer output in statistics textbooks. In a review of 16 introductory statistics textbooks, Cobb (1987) noted that eight included some computer output. As teachers, we should build on this trend by using real data and real data analyses in applied statistics classrooms.

What difference does the authenticity of a data set make? A principal difference lies in the level of student motivation generated. Students find real data to be intrinsically interesting. With real data, their efforts are rewarded not only with information on how to use statistics to conduct research, but also with information on an interesting research question. And because well-chosen data sets can, themselves, be memorable, they become mnemonics for recalling statistical techniques later.

In addition to capturing the students' interest, real data sets can be helpful instructional aids. Real data sets allow the student to assume the role of researcher, exploring data in the hopes of addressing a specific set of research questions. Class examples and homework exercises become "trial" runs for data-analytic crises that will be encountered later in empirical research. Real data sets bring students as close as possible to the actual research experience.

Real data are helpful for another reason as well: They provide an opportunity for teaching students to cope with many common problems as they arise in practice—problems of nonlinearity, outliers, and missing values. That these nonstandard problems are ubiquitous reminds students to constantly investigate the tenability of their assumptions. They become interested in learning what to do when standard assumptions do not hold. Again, the use of real data validates the importance of exploratory data analysis.

3. DESIRED PEDAGOGIC CHARACTERISTICS OF REAL DATA

Authenticity. First, and foremost, a real data set must be authentic. The data given must be actual measurements taken on an actual sample of cases. Attaching lifelike variable means to artificial data is not an acceptable substitute. Consider the following exercise from Hays (1981):

> An experimenter was interested in the possible linear relationship between the measure of finger dexterity X, and another measure representing general muscular coordination Y. A random sample of 25 persons showed the following scores: . . . Compute the correlation coefficient, and test its significance. (p. 490)

Why should a student believe that these data are real? How were finger dexterity and general muscular coordination measured? From what population was the random sample chosen? Is the sample homogeneous with respect to age, a factor that might influence general muscular coordination and, perhaps, the relationship between coordination and finger dexterity? Is the experimenter only interested in a linear relationship? Most students can easily see through the artifice of "lifelike" data. As a result, they may not bother asking the questions raised here because there is no reason to worry about how the data were "collected." Yet these questions are precisely those that they ought to ask when reviewing research and conducting their own.

Background Information. Real data should be accomplished by background information on the purpose and design of the research, the source of the data, the measurement techniques, and the definitions of the variables. As noted by Cobb (1987), Hunter (1976), and Joiner (1988), background information brings a data set to life and allows the student to fully assume the role of researcher. If the data come from published papers or tabulations, students should be given access to the original document. If the data are extracted from another source, the instructor should provide the citation. Students may find it interesting to read the original papers while conducting their own analyses.

Interest and Relevance. Some of the best-selling statistics textbooks are filled with real data, but on topics of little interest to students in the social sciences. Snedecor and Cochran (1980) made ample use of real data, but on topics such as the calcium concentration in turnip greens (p. 239) and the average daily weight gain of swine (p. 303). Draper and Smith (1981) also used real data, but on the viscosity of filled and plasticized elastomer compounds (p. 228) and the effects of temperature on the growth rates of ice crystals (p. 66). Most classic statistics data sets, such as Fisher's iris data (1936) and Brownlee's stack loss data (1965), fail to inspire students in psychology and sociology (although, of course, interest lies in the eye of the beholder).

Topicality often provokes student interest. Students easily become very engaged in a data set recently reported in *Chance* (Supreme Court Ruling 1988) on the relationship between race of victim, race of defendant, and whether the defendant was given the death penalty. Regardless of substantive background, most students perceive these data as relevant and interesting, especially in the current climate of racial tension on our nation's college campuses.

Controversy can also provoke student interest. Cyril Burt's data on the IQ's of identical twins is controversial (Jensen

1974), especially when analyzed in the context of Burt's views on the nature/nurture debate and Dorfman's (1978) evidence that Burt falsified data to support the nature argument. Powell and Steelman's (1984) analysis of the relationship between state SAT scores and the percentage of students taking the test also arouses interest, especially when accompanied by newspaper accounts of politicians' claims and critiques of state-level comparisons of SAT scores (Rosenbaum and Rubin 1985; Wainer 1986; Wainer, Holland, Swinton, and Wang 1985). Analyzing controversial data sets shows students how statistical methodology can support or undermine a particular position.

Historical data sets may also be an effective motivator. The early volumes of journals such as *Child Development*, *Journal of Educational Psychology*, and *Journal of Genetic Psychology* are brimming with raw data. Although these data sets are not always intrinsically interesting, their history may compensate for this lack. Moreover, comparing a modern reanalysis with the older, less sophisticated tabular presentations given in the original sources can be absolutely fascinating.

Substantive Learning. Empirical researchers analyze data because they want to learn something about the way the world works, not because they want to conduct statistical analyses for their own sake. When students learn something new from a data analysis, they discover just how useful statistical analysis can be. The substantive learning does not have to be on a grand scale, but it should be real. Even an analysis of data on per-pupil expenditures, teacher salaries, and student demographics obtained from the local superintendent's office can be stimulating. Students can learn how their home town compares with others and how district characteristics are interrelated. They gain insight into the ongoing political debate as to why some school districts are reputed to be "better" than others. Substantive learning reinforces the reasons for conducting statistical analysis.

Availability of Multiple Analyses. Thoughtful data analysts often use several types of analysis to address a research question. Students need to learn that different analyses provide different insights. No experience reinforces the importance of multiple analyses as much as the discovery of previously unknown findings. For instance, in courses on categorical data analysis, it is often the case that using a more powerful log-linear modeling approach to reevaluate a classical contingency table analysis will bring out a previously unnoticed effect.

The Importance of Raw Data. We believe that data must be presented in raw form, not summarized by means and variance–covariance matrices. Rich information is lost when raw data are replaced by sufficient statistics and, because students are one step further away from the actual data, they are more likely to be seduced by a cookbook approach. When raw data are available, students can begin by exploring their data (Chatfield 1985; Tukey 1977), facing high-leverage cases, nonlinearly, heteroscedasticity, and many other problems that arise in real data. Being asked to analyze summary statistics rather than raw data may fool students into believing that such problems do not exist or that, if they do, they are of little consequence.

Case Identifiers. Many published data sets include case identifiers that allow students to bring their own background knowledge to bear. State, school district, and school identifiers may have meaning for the students. If such identifiers are available, they should be provided with the data set so that background information about the cases can inform the statistical analyses. Case identifiers are particularly useful when outliers and high-leverage observations have been identified. When data on the citation frequencies of prominent researchers are being analyzed, for example, personal knowledge of the authors helps students conjecture as to why Sigmund Freud and Jean Piaget might be outliers (Gordon, Nucci, West, et al. 1984).

4. DRAWBACKS TO USING REAL DATA IN TEACHING

The Workload of Finding Real Data Sets. A major motivation for the instructional use of artificial data is that many data sets with specific characteristics can be created. For example, Dayton (1972) presented a simple method for constructing a data set illustrating the effects of suppressor variables. Searle and Firey (1980) suggested that an instructor could reduce plagiarism by generating enough data sets to give each student a different one to "analyze." Producing a variable that is normally distributed, but with an outlier or two, is a simple programming problem; identifying a real data set with the same features can take hours.

Using real data sets vastly increases the time required to prepare classes, homework, and exams. For each data set that clearly illustrates a specific technique, several hours must be spent in preliminary analysis of several data sets, of which some turn out to be inapplicable and others present nontrivial analytic problems. This latter difficulty is especially problematic when developing elementary course materials, and students cannot be expected to make nonstandard decisions.

Small Data Sets and Statistical Power. Pedagogically, small data sets—with sample size in the 35–75 range—are probably preferable. With a small data set, the student can become intimately acquainted with each case, learning the relationship between data and analysis. Once students have developed these skills, larger data sets can be introduced. Unfortunately, using small data sets creates a false impression as to what constitutes adequate sample size in practice. In addition, because null findings are less interesting, the data sets chosen usually have large effect sizes so that "statistically significant" results can be obtained despite the small sample. Although large effect sizes are rare in practice, classwork builds a false anticipation of their practical magnitudes.

Of course, this problem is not unique to real data; most artificial data sets are also small. The issue is that real data sets seem representative of the real world. Although we have no easy solution to this problem, we deal with it by discussing concepts of statistical power, effect size, and the distinction between statistical significance and practical significance.

Aggregate Data and Self-Selected Samples. The difficulties of finding high-quality real data may provoke the teacher into being satisfied with data that are less than adequate. Easy access to information sometimes leads to the use of aggregate data or data collected from self-selected samples (such as mean SAT scores by state for the high school seniors who chose to take the test) (Wainer 1986). In some of these data sets, variables are actually measured at the aggregate level—for example, college tuition, student/faculty ratio, number of students enrolled—and can be appropriately used for class and home analysis. But other data sets contain aggregated individual-level data, with all of their attendant problems. The question is whether the pedagogic gains are worth the statistical drawbacks, and, in most instances, we believe that they are. Aggregate data sets contain some of the most readily accessible, intrinsically interesting data available and, in addition, they often contain meaningful identifiers—names of towns, cities, counties, school districts, states, or countries—connecting students more intimately with each data point.

In-Class Testing. It is difficult, although not impossible, to test students in class using real data. However, multiple homework assignments and take-home exams—both making use of real data—can be used in place of in-class exams. Or, if in-class exams cannot be avoided, preprinted computer output can be distributed for student interpretation. If this is done, though, students have not planned and carried out the analyses and, therefore, all of their analytic skills may not be tested.

5. CLOSING COMMENT

Because of the difficulties associated with locating real data sets for use in the teaching of applied statistics, we have included annotated references to usable real data sets in the Appendix. Most of the data sets are from subject-matter journals, although we have also included citations to many statistics textbooks. Other sources not listed may suggest themselves to the reader: local dissertations are a gold mine of real data; documents available from microfiched data bases, such as ERIC, often include data listings in their appendixes; the Census Bureau's publications and other government reports are replete with fascinating data; and, last but not least, an instructor can generate real data by distributing questionnaires at the beginning of the course.

In our annotated bibliography, we have tried to present data sets from a variety of substantive literatures to appeal to students from different backgrounds. We have annotated each entry briefly to provide a flavor of the citation and, although the reader will need to examine selected data sets analytically to determine their utility for his or her specific purposes, we hope that our annotations will facilitate the selection process.

APPENDIX: ANNOTATED BIBLIOGRAPHY OF PUBLISHED DATA SETS

Journal Articles and Reports

Adkins, P. G. (1969), "Deficiency in Comprehension in Non-native Speakers," *TESOL Quarterly*, 3, 197–201. Ninth-grade student achievement (40 subjects) on tests of seventh-grade material, with and without idioms and figures of speech. Useful for bivariate analysis.

Allison, T., and Cicchetti, D. V. (1976), "Sleep in Mammals: Ecological and Constitutional Correlates," *Science*, 194, 732–734. Average brain and body weights for 62 species of mammals. Both variables are very skewed, but logarithmic transformations alleviate the skewness and improve the linearity of the scatterplot.

Angell, R. C. (1951), "The Moral Integration of American Cities," *American Journal of Sociology*, 53, 1–140. Measures of the moral integration, ethnic heterogeneity, crime, welfare effort, integration, and mobility of residents in 43 American cities.

American Association of University Professors (1987), "The Annual Report on the Economic Status of the Profession, 1986–1987," *Academe*, 73, 1–88. Salary data by rank, sex, and tenure status for faculty at 1,901 colleges and universities.

Aylward, G. P., Harcher, R. P., Leavitt, L. A., Rao, V., Bauer, C. R., Brennan, M. J., and Gustafson, N. F. (1984), "Factors Affecting Neobehavioral Responses of Preterm Infants at Term Conceptual Age," *Child Development*, 55, 1155–1165. Contingency table of the relationship between gestational age and neurological status for 505 babies. Also see detailed log-linear analyses of these data in: Green, J. A. (1988), "Loglinear Analysis of Cross-Classified Ordinal Data: Applications in Developmental Research, *Child Development*, 59, 1–25.

Bain, R. (1939), "Verbal Stereotypes and Social Control," *Sociology and Social Research*, 23, 431–446. Multiple multiway contingency data sets on proverbs and mechanical cliches used by 133 college freshmen in 1937. Data are broken down by urban/rural, gender, etcetera.

Barrons' Publications (1987), *Profiles of American Colleges* (16th ed.), New York: Author. One of many sources describing the more than 1,500 four-year colleges in this country. Relevant data include number of applicants, number of students accepted, number of students enrolling, mean SAT scores of incoming freshmen, mean class rank of incoming freshmen, faculty/student ratios, financial aid available, number of part-time students and faculty, percent of faculty with doctorates, sex composition of student body. Can be supplemented with information from American Association of University Professors salary survey and endowment data given in the *Digest of Education Statistics*.

Bell, J. C. (1914), "A Class Experiment in Arithmetic," *Journal of Educational Psychology*, 5, 467–470. Individual data for 25 college sophomores at the University of Texas on the speed and accuracy with which they solved four types of arithmetic problems (addition, subtraction, multiplication, and division).

——— (1916), "Mental Tests and College Freshmen," *Journal of Educational Psychology*, 7, 381–399. Scores on nine tests for 37 of the "best" students and 37 of the "worst" students, with notations of class rank, designed "to be of assistance to college authorities in aiding freshmen to adjust themselves to their environment" (p. 381).

Bishop, D. W., and Ideda, M. (1970), "Status and Role Factors in the Leisure Behavior of Different Occupations," *Sociology and Social Research*, 54, 190–208. Multidimensional data on leisure activities by occupation. Useful for multidimensional scaling, cluster analysis, factor analysis, principal components analysis, discriminant analysis, etcetera.

Boli, J., Allen, M. L., and Payne, A. (1985), "High Ability Women and Men in Undergraduate Mathematics and Chemistry Courses," *American Educational Research Journal*, 22, 605–626. Perceptions of course performance among high-ability men and women in physics and chemistry courses at Stanford.

Bullen, A. K. (1945), "A Cross-Cultural Approach to the Problem of Stuttering," *Child Development*, 16, 1–88. Raw data for 46 children divided into four groups—stutterers, well-adjusted, medium-adjusted, and poorly-adjusted. Measures include age, achievement, receptivity to education, physical condition, social-personality traits, insightfulness, family background, somotype, and anthropometrics.

Chapman, J. C. (1914), *Individual Differences in Ability and Improvement and Their Correlations*. Columbia University Contributions to Education Number 63, New York: Teacher's College. Six to ten longitudinal data sets (10 waves) on measures of computation, color-naming, and opposites-naming for 22 college-age males in New York at the turn of the century. Suitable for growth curve analysis.

Cobb, M. C. (1917), "A Preliminary Study of the Inheritance of Arithmetic Abilities," *Journal of Educational Psychology*, 8, 1–20. Data on the mother, father, and children in eight families, with the age of each family member and their scores on five tests (addition, subtraction, multiplication, division, and copying figures). The author concludes that "it is difficult to avoid the conclusion that . . . likeness is due to heredity" (p. 16).

Council of Great City Schools (1983), *Statistical Profiles of the Great City Schools*, Philadelphia, PA: Author. Educational and demographic descriptors for 32 large urban school districts, including data on how these characteristics have changed over time.

Crymes, R. (1971), "The Relation of Study About Language to Language

Performance With Special Reference to Nominalization," *TESOL Quarterly*, 5, 217–230. Categorical and continuous data on the number of nominals produced by native and nonnative speakers. Data for experimental and control groups on pretest and posttest measures.

Dunshee, M. E. (1931), "A Study of Factors Affecting the Amount and Kind of Food Eaten by Nursery School Children," *Child Development*, 2, 163–183. Data on the eating habits of 37 children, including age, sex, means (and standard deviations) for total amount of calories eaten and minutes spent at table.

Eash, M. (1983), "Educational Research Productivity of Institutions of Higher Education," *American Educational Research Journal*, 20, 5–12. Research productivity rankings for 25 schools of education. Measures are based on presentations at AERA and articles published in AERA journals. The paper also presents data comparing AERA program contributions in 1975 and 1976 with those in 1980 and 1981. Compare with smaller data sets presented in: West, C. K. (1978), "Productivity Ratings of Institutions Based on Publication in the Journals of the American Educational Research Association," *Educational Researcher*, 7, 13–14; and Schubert, W. H. (1979), "Contributions to AERA Annual Programs as an Indicator of Institutional Productivity," *Educational Researcher*, 8, 13–17.

Educational Research Service (1987), *Scheduled Salaries for Professional Personnel in Public Schools 1986–1987*, Arlington, VA: Author. Raw data for 1,031 school districts on enrollment, per-pupil expenditures, and salaries for superintendents, central office administrators, principals, teachers, staff, and support personnel.

Fales, E. (1933), "A Comparison of the Vigorousness of Play Activities of Preschool Boys and Girls," *Child Development*, 4, 144–157. Age, IQ scores, and activity ratings for 16 boys and 16 girls. Two activity ratings are available for each child.

Ferdinand, T. N., and Luchterhand, E. G. (1970), "Inner-City Youth, the Police, the Juvenile Court, and Justice," *Social Problems*, 17, 510–527. Multiple multiway categorical data sets on the crimes and police disposition of inner-city youth, by race, gender, type of crime, age, etcetera.

Fleming, J. T. (1967), "The Measurement of Children's Perception of Difficulty in Reading Materials," *Research in the Teaching of English*, 1, 136–156. Individual data on 60 subjects. Data include background characteristics (gender, age, IQ, etc.) and achievement (comprehension, reading, word knowledge, spelling, language, study skills) for school-age children. Suitable for multiple regression, multivariate analysis, etcetera.

Fraumeni, J. F., Jr. (1968), "Cigarette Smoking and Cancers of the Urinary Tract: Geographic Variation in the United States," *Journal of the National Cancer Institute*, 41, 1205–1211. Aggregate cigarette smoking and cancer death rates, by type of cancer and state.

Garwood, A. N. (ed.) (1986), *Massachusetts Municipal Profiles*, Wellesley Hills, MA: Information Publications. Sociodemographic characteristics for 353 Massachusetts towns and cities, including data on age, race, sex, income, labor force participation, voter registration, police, fire, crime, taxation, libraries, and schools. The company publishes similar books for other states; write to them at Box 356, Wellesley Hills, MA 02181.

Gelb, S. A., and Mizokawa, D. T. (1986), "Special Education and Social Structure: The Commonality of 'Exceptionality,' " *American Educational Research Journal*, 23, 543–557. State-level data on percentage of children in each category of special education and sociodemographic composition of the states. Washington, DC is a high-leverage outlier for the relationship between percent of students classified as educably mentally retarded and percent of population that is black.

Gerlach, M. (1939), "A Study of the Relationship Between Psychometric Patterns and Personality Types," *Child Development*, 10, 269–278. Raw data for 61 maladjusted children on two IQ tests, as well as information on their sex, age, parentage (both foreign, both American, mixed) and maladjustment type (aggressive or asocial).

Gordon, N. J., Nucci, L. P., West, C. K., Hoerr, W. A., Uguraglu, M., Vukosavich, P., and Tsai, S. L. (1984), "Productivity and Citations of Educational Research: Using Educational Psychology as the Data Base," *Educational Researcher*, 13(7), 14–20. Citation frequencies and dates of birth for 187 prominent educational researchers. Four sources of citations are given: *AERJ and JEP*, *Review of Research in Education*, selected educational psychology textbooks, and *Social Science Citation Index*.

Gulliksen, H. (1934), "A Rationale Equation of the Learning Curve, Based on Thorndike's Law of Effect," *Journal of General Psychology*, 11, 395–434. Longitudinal data on rats—some brighter than others—trained to discriminate between 18-cm and 12-cm circles.

Hahn, H. H., and Thorndike, E. L. (1914), "Some Results of Practice in Addition Under School Conditions," *Journal of Educational Psychology*, 5, 65–83. Individual data from an experiment on the effects of time lapsed between pretests and posttests for 167 students in grades 4, 5, 6, and 7.

Harris, S., and Harris, L. B. (1986), *The Teacher's Almanac*, New York: Facts on File. Assorted education data by state and school district, including teacher salaries, high school graduation rates, functional illiteracy rates, and presence of computers in schools.

Heyneman, S. P. (1976), "Influences on Academic Achievement: A Comparison of Results from Uganda and More Industrialized Societies," *Sociology of Education*, 49, 200–211. Aggregate data on school achievement and economic development for 18 nations. Includes measures of preschool influence, GNP, and percent enrollment in primary and secondary school.

Hirsch, N. D. M. (1928), "An Experimental Study of the East Kentucky Mountaineers: A Study in Heredity and Environment," *Genetic Psychology Monographs*, 3, 183–244. Ages and IQ scores of siblings in 44 families.

——— (1930), "An Experimental Study Upon Three Hundred School Children Over a Six-Year Period," *Genetic Psychology Monographs*, 7, 487–546. IQ scores for a six-year period for 343 children.

Howard, G. S., Cole, D. A., and Maxwell, S. E. (1987), "Research Productivity in Psychology Based on Publication in the Journals of the American Psychological Association," *American Psychologist*, 42, 975–986. Productivity, reputation, and size of psychology departments at 75 universities.

Izenman, A. J. (1972), "Reduced Rank Regression for the Multivariate Linear Model," unpublished doctoral dissertation, University of California–Berkeley, Dept. of Statistics. Body length of crickets as a function of geographical location and weather throughout the USA.

Jensen, A. R. (1974), "Kinship Correlations Reported by Sir Cyril Burt," *Behavior Genetics*, 4, 1–28. "Burt's final assessments" of IQ's of monzygotic twins reared apart, with "social class" ratings of the homes. For information on Burt's falsification of the data, see Dorfman, D. D. (1978), "The Cyril Burt Question: New Findings," *Science*, 201, 1177–1186. Other sources include correspondence related to Dorfman's article: Stigler, S. M. (1979), Letter to the editor, *Science*, 204, 242–245; Rubin, D. B. (1979), Letter to the editor, *Science*, 204, 246–254, and Hearnshaw, L. S. (1979), *Cyril Burt: Psychologist*, London: Hodder and Stroughton, Chapter Twelve.

——— (1970), "IQ's of Identical Twins Reared Apart," *Behavioral Genetics*, 1, 133–148. Original data from four studies of IQ's of identical twins reared apart.

Johnson, B., and Courtney, D. M. (1931), "Tower Building," *Child Development*, 3, 161–162. Twenty-five children were asked to build towers on each of two occasions. Each time they were given: (a) a set of cubes; and (b) a set of cylinders. Raw data are given on the number of blocks of each type used each time, and how many minutes it took to construct the tower.

Johnson, W. D., and Koch, G. G. (1971), "A Note on the Weighted Least Squares Analysis of the Ries–Smith Contingency Table Data," *Technometrics*, 13, 438–447. Large four-way contingency table derived from a randomized experiment in which several brands of detergent were compared under a variety of user-related conditions.

Jones, L. V., Lindzey, G., and Coggeshall, P. E. (1982), *An Assessment of Research Doctorate Programs in the United States: Social Sciences*, Washington, DC: National Academy Press. "Quality" rankings and characteristics of university departments in the social sciences, by discipline. Data include number of faculty, number of students, productivity of faculty, number of grants awarded, follow-up placement of doctoral students.

Karelitz, S., Fischelli, V. R., Costa, J., Karelitz, R., and Rosenfeld, L. (1964), "Relation of Crying in Early Infancy to Speech and Intellectual Development at Age Three Years, *Child Development*, 35, 769–777. Data for 38 infants on their crying activity in early infancy and later measures of IQ.

Koch, H. L. (1933), "Popularity in Preschool Children: Some Related Factors and a Technique for Its Measurement," *Child Development*, 4, 164–175. Popularity scores for 17 children: percent each child was named first, percent each child name last, effects of ordering, and the effects of sex.

Kroc, R. J. (1984), "Using Citation Analysis to Assess Scholarly Productivity," *Educational Researcher*, 13, 17–22. Citation frequency data for 51 schools of education. Measures include mean citation rate, percentage of faculty with 10–100 citations and percentage of faculty with no citations. Kroc correlates these measures with five rankings of schools of education; these analyses could be recreated by abstracting data from the five sets of rankings cited in his bibliography.

Leinhardt, G., and Leinhardt, S. (1980), "Exploratory Data Analysis: New Tools for the Analysis of Empirical Data," *Review of Research in Education*, 8, 85–157. Three measures of reading instruction for a sample of 53 learning disabled students, by curricular approach and school.

Leinhardt, S., and Wasserman, S. S. (1979), "Teaching Regression: An Exploratory Approach," *The American Statistician*, 33, 196–203. Life expectancy and per capita income for 105 nations divided into five national wealth classifications (industrialized, petroleum exporting, higher, middle, and lower).

Louano-Kerr, J., Semles, V., and Zimmerman, E. (1977), "A Profile of Art Educators in Higher Education: Male/female Comparative Data," *Studies in Art Education*, 18, 21–37. Multiway contingency table data on the qualifications (master's vs. doctorate) of art educators, by year and gender.

Maresh, M. M., and Deming, J. (1939), "The Growth of the Leg Bones in 80 Infants: Roentgenograms Versus Anthropometry," *Child Development*, 10, 91–106. Individual data for 80 children on the sizes of 10 bones, measured by both x-rays and anthropometry at each of 3, 4, and 5 occasions, by sex. The authors construct lots of individual growth curves.

Mason, T. J., and McKay, F. W. (1974), *US Cancer Mortality by County: 1950–1969*, Washington, DC: U.S. Government Printing Office. Lung cancer mortality by degree of urbanization and gender, in Louisiana.

Mickey, M. R., Dunn, O. J., and Clark, V. (1967), "Note on the Use of Stepwise Regression in Detecting Outliers," *Computers and Biomedical Research*, 1, 105–109. Gesell adaptive scores and age at first word (in months) for 21 children with cyanotic heart disease. The data set contains some interesting outliers and hgh leverage cases.

Mosteller, F., Hyman, H., McCarthy, P. J., Marks, E. S., and Truman, D. B. (1949), *The Pre-Election Polls of 1948*, New York: Social Sciences Research Council. "Dewey beats Truman" is one of the classic headlines of the twentieth century. The data contained in this book provide an opportunity to examine the pre-election polls to ask why this headline seemed plausible. Data are provided by the respondents' social class, as well as the month in which the pre-election poll was conducted. Also see related analyses presented in: Baker, S. G., and Laird, N. M. (1988), "Regression Analysis for Categorical Variables With Outcome Subject to Nonignorable Nonresponse," *Journal of the American Statistical Association*, 83, 62–69.

National Center for Education Statistics (1987), *The Condition of Education*, Washington, DC: U.S. Department of Education, ——— (1987), *Digest of Education Statistics*, Washington, DC: U.S. Department of Education. Annual reports issued by the Department of Education providing descriptive information on education, often over time, sometimes by state, occasionally by school district. The data on university endowments can be used in conjunction with other university level data, such as that given in Barron's (1987).

National Education Association (1926), "The Ability of the States to Support Education," *Research Bulletin of the National Education Association*, 4, 12–22. Several chapters of state-level data documenting wealth, per capita income, educational expenditure, population size, number of adults per child, and so on. Of historical interest.

——— (1965), "Annual Report on Public-School Financing," *NEA Research Bulletin*, 43, 90–92. State-by-state listing of per-pupil expenditures (in 1955–56 and 1964–65) and percent of school revenues from state and local sources. The entire journal series contains a wealth of information on class size, salaries, public expenditures for education, etcetera.

National School Boards Association (1986), *A Survey of Public Education in the Nation's Urban School Districts*, Alexandria, VA: Author. Data for 61 school districts on educational policies and practices, as well as selected education and economic descriptors.

Neuman, S., and Ziderman, A. (1985), "Do Universities Maintain Common Standards in Awarding First Degrees With Distinction?: The Case of Israel," *Higher Education*, 14, 447–459. Number of first degree graduates with and without distinction, by school and faculty within school. Suitable for log-linear modeling.

"Opening Lines," (1985), *Harper's*, September, 29–30. Selected results from a study of opening lines used in single's bars in the St. Louis area. Two-way contingency table describing the relationship between type of opening line (compliments, propositions, etc.) and time of evening.

Phillips, D. P. (1978), "Deathday and Birthday: An Unexpected Connection," in *Statistics: A Guide to the Unknown* (2nd ed.), J. M. Tanur, F. Mosteller, W. H. Kruskal, R. F. Link, R. S. Pieters, G. R. Rising, and E. L. Lehmann, eds. San Francisco, CA: Holden-Day, pp. 71–85. See also: Phillips, D. P. (1977), "Motor Vehicle Fatalities Increase Just After Publicized Suicide Stories," *Science*, 196, 1464–1465; Phillips, D. P. (1978), "Airplane Accident Fatalities Increase Just After Newspaper Stories About Murder and Suicide," *Science*, 201, 748–750; Phillips, D. P., and Carstensen, L. L. (1986), "Clustering of Teenage Suicides After Television News Stories About Suicide," *The New England Journal of Medicine*, 315, 685–689 (and related articles in this issue); and Schultz, R., and Bazerman, M. (1980), "Ceremonial Occasions and Mortality: A Second Look," *American Psychologist*, 35, 253–261. David Phillips has made a cottage industry of looking at what many might term coincidences—birthdays and deathdays and copycat suicides after popularized accounts in the media. These are but a handful of articles, each listing the detailed raw data on deaths following these events that led him to his conclusions.

Powell, B., and Steelman, L. C. (1984), "Variations in State SAT Performance: Meaningful or Misleading?," *Harvard Educational Review*, 54, 389–412. Mean SAT scores and percent of high school seniors taking the SAT, by state for 1982. For additional data and a critique of their analyses, see Wainer et al. (1985). Also see Rosenbaum and Rubin (1985).

Rubin, E. (1972), "Statistical Exploration of a Medieval Household Book," *The American Statistician*, 26, 37–39. Number of meals served, breads baked, and ale brewed at the de Bryene household from October 1412 to September 1413, by month. That's right, the fifteenth century.

Sandvern, J. (1971), "Causes of Lacking Sense of Well-Being in School," *Scandinavian Journal of Educational Research*, 15, 21–60. Many multiway contingency tables dealing with students' liking of school in Norway, broken down by type of school, socioeconomic status, etcetera.

Scarcella, R. C. (1984), "How Writers Orient Their Readers in Expository Essays: A Comparative Study of Native and Non-native English Writers, *TESOL Quarterly*, 671–688. Categorical data on the language background and language proficiency of native and nonnative speakers and how this influences their choice of writing device.

Shearer, L. (1987), "How Will History Rate Nancy Reagan?," *Parade Magazine*, June 14, p. 8. Ranking for 17 first ladies from Florence Harding through Nancy Reagan on 10 dimensions ranging from integrity, leadership, and accomplishments.

Skodak, M., and Skeels, H. M. (1949), "A Final Follow-up Study of One Hundred Adopted Children," *Journal of Genetic Psychology*, 75, 85–125. Raw data for 100 children who were adopted at birth. Measures include natural mother's IQ and education level, foster mother's IQ and education level, foster father's occupation, and child's IQ on each of 5 occasions, from infancy through pre-adolescence.

Stewart, L. H. (1955), "The Expression of Personality in Drawings and Paintings," *Genetic Psychology Monographs*, 51, 45–103. Data for 28 boys given, including two IQ scores (Terman and Standford-Binet), two socioeconomic status measures (parents' education and father's occupation), somotypes (endomorph, mesomorph, ectomorph), and drawing type.

"Supreme Court Ruling on Death Penalty," *Chance*, 1, 7–8. Three-way contingency table on the relationship between race of victim, race of defendant, and use of the death penalty, showing that the death penalty is not uniformly applied.

Taagepera, R. (1972), "The Size of National Assemblies," *Social Science Research*, 1, 385–401. Aggregate data on the size of the national assemblies in more than 100 countries, with data on population, literacy, working age, etcetera.

Tufte, E. R. (1978), "Registration and Voting," in *Statistics: A Guide to the Unknown* (2nd ed.), J. M. Tanur, F. Mosteller, W. H. Kruskal, R. F. Link, R. S. Peters, F. R. Rising, and L. L. Lehmann, San Francisco: Holden-Day, pp. 195–204. Data on percent of population registered and percent of population voting in the 1960 election for 104 cities. Data are analyzed in greater detail in: Kelly, S., Jr., Ayres, R. E., and Bowen, W. G. (1967), "Registration and Voting: Putting First Things First," *American Political Science Review*, 61, 359–379.

United Nation's Children's Fund (1987), *The State of the World's Children*, New York: Oxford University Press. Sociodemographic, education, health, and economic indicators for 130 countries.

Walberg, H. J., and Rasher, S. P. (1974), "Public School Effectiveness and Equality: New Evidence and Its Implications," *Phi Delta Kappan*, 66, 3–9.

——— (1976), "Improving Regression Models," *Journal of Educational Statistics*, 1, 253–277. The authors analyze data for the 50 states on the relationship between failure on the selective service exam administered during 1969–1970 and contextual and education descriptors of the states. The selection bias inherent in analyses of state-level SAT scores is present here, but the data set is interesting.

Whiting, J. M., and Child, I. L. (1962), *Child Training and Personality*, New Haven, CT: Yale University Press. Many characteristics in dozens of societies around the world, including age at weaning, toilet training, fear of ghosts, rituals, etcetera.

Wilson, M. E., and Mather, L. E. (1974), "Life Expectancy" (letter to the editor), *Journal of the American Medical Association*, 229, 1421–1422. Age of person at death (in years) and the length of the person's lifeline (in centimeters) for 50 individuals. Not surprisingly, the test of $H_0 : r = 0$ cannot be rejected.

Zeligs, R. (1938), "Tracing Racial Attitudes Through Adolescence," *Sociology and Social Research*, 23, 45–54. Chronological and mental age, social distance scores, and gender for 12 children tested three times during adolescence. In addition, indices for racial tolerance over three occasions of measurement for 39 racial groups.

Zimmerman, J. (1917), "The Binet–Simon Scale and Yerkes Point Scale: A Comparative Examination of 100 Cases, *Journal of Educational Psychology*, 8, 551–558. Individual data for 100 students on these two IQ tests, with information on student, sex, age, and native language.

Textbooks

Afifi, A. A., and Azen, S. P. (1979), *Statistical Analysis: A Computer Oriented Approach* (2nd ed.), New York: Academic Press. Several extensive data sets describing blood chemistry, cardiovascular state, socioeconomic status, and year of death. Some censored cases could be used in the teaching of survival analysis. Other data sets include body flexibility, diet, testosterone levels in right and left testes of mice (!), weaning of rats. Some educational data sets on infant cognitive development.

Afifi, A. A., and Clark, V. (1984), *Computer Aided Multivariate Analysis*, Belmont, CA: Lifetime Learning, pp. 30–39. Depression scores and selected covariates for 294 participants in the Los Angeles Depression Study. Data set includes individual item responses for a 20-question depression scale, personal background characteristics, and selected health variables.

Aickin, M. (1983), *Linear Statistical Analysis of Discrete Data*, New York: John Wiley. A large variety of categorical data sets, including tenure in American universities, dolphin sightings, transitions between Piagetian stages, college expectations and participation in high school athletics, political preferences, religion and marijuana, and sudden infant death.

Aldrich, J. H., and Nelson, F. D. (1984), *Linear Probability, Logit and Probit Models*, Sage University Paper 45, Beverly Hills, CA: Sage Publications. Data on "Personalized System of Instruction" and course grades in an intermediate macroeconomics course, useful for logit analysis and log-linear modeling.

Andrews, D. F., and Herzberg, A. M. (1985), *Data: A Collection of Problems From Many Fields for the Student and Research Worker*, New York: Springer-Verlag. 71 sets of raw data. Many substantive areas are included, but the emphasis is generally on the physical and natural sciences. Several interesting social science examples are given, including unemployment statistics, insurance rate information, literary data sets, and the birthday/deathday problem.

Berenson, M. L., Levine, D. M., and Goldstein, M. (1983), *Intermediate Statistical Methods and Applications*, Englewood Cliffs, NJ: Prentice-Hall. A large variety of nonsocial science data sets (lawn service, real estate market, professional sports, foreign food), with some social science data sets scattered here and there: categorical data on health issues in children by graduating class of pediatrician, starting salaries of MBA graduates, etcetera.

Bock, R. D. (1975), *Multivariate Statistical Methods in Behavioral Research*, New York: McGraw-Hill. A variety of educational growth data sets suitable for repeated measures/MANOVA analysis, including data on responses to inkblot plates by grade and IQ over time, longitudinal (four grades) data on scaled vocabulary scores for boys and girls, and so on. Data repeated in: Finn, J. D., and Mattson, I. (1978), *Multivariate Analysis in Educational Research*, Chicago, IL: National Educational Resources.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983), *Graphical Methods for Data Analysis*, Belmont, CA: Wadsworth. 33 raw data sets. Many substantive areas are included, and many of these data sets are just plain interesting, such as the ages of signers of the Declaration of Independence, murder/suicides by crashing private airplanes, and heights of singers in the New York Choral Society.

Cooley, W. W., and Lohnes, P. R. (1985), *Multivariate Data Analysis*, Malabar, FL: Robert E. Kreiger. Two large data sets: (1) a 20-variable subset of the PROJECT TALENT data (234 males, 271 females); and (2) the RECTANGLES data set on the physical dimensions of 100 rectangles, useful for factor analysis and principal components analysis.

Cox, D. R., and Snell, E. J. (1981), *Applied Statistics: Principles and Examples*, London: Chapman & Hall. 39 raw data sets. Relevant examples include educational plans of Wisconsin school boys, statistical aspects of literary style, and satisfaction with housing conditions.

Devore, J., and Peck, R. (1986), *Statistics: The Exploration and Analysis of Data*, (St. Paul, MN: West Publishing). The data sets tend to be small, and many are from the sciences, but there are dozens of them. One interesting example is the movie production and promotion costs for "dumb movies," such as "Revenge of the Nerds" and "Police Academy."

Draper, N. R., and Smith, H. (1981), *Applied Regression Analysis*, 2nd ed., New York: John Wiley. Some educational data sets are submerged among the many others, including sex differentials in teacher pay, aptitude and age of first word, nutrition of preschoolers, and ailments of university alumni.

Dunteman, G. H. (1984), *Introduction to Linear Models*, Beverly Hills, CA: Sage Publications. Data for 300 participants in the National Longitudinal Study on reading, math, gender, race, college status, socioeconomic status, high school program, high school grades, creativity, stress avoidance, etcetera.

Erickson, B. H., and Nosanchuck, T. A. (1977), *Understanding Data*, Toronto, Canada: McGraw-Hill Ryerson. An introductory textbook that melds together Tukey's exploratory data analysis and the more traditional confirmatory approaches. Many interesting data sets, including frequency of teacher criticism by student IQ, sex differences in reactions to hostile treatment by an experimenter, experimenter artifacts in social psychology research, and characteristics of social networks.

Finn, J. D. (1974), *A General Model for Multivariate Analysis*, New York: Holt, Rinehart & Winston. Raw data from four studies of relevance to the social sciences: creativity and achievement, memory for words, essay grading practices, and effects of programmed instruction.

Fox, J. (1984), *Linear Statistical Models and Related Methods With Applications to Social Research*, New York: John Wiley. Several interesting data sets including relationship between status, authoritarianism and conformity, methods to enhance recall of words, and causes of the 1907 Romanian peasant rebellion.

Freedman, D., Pisani, R., and Purves, R. (1978), *Statistics*, New York: W. W. Norton. Several small data sets on topics such as the gender of people selected in the Health Examination Survey and the levels of confidence people have in different U.S. institutions (ranging from the FBI to organized religion).

Gnanadesikan, R. (1977), *Methods for Statistical Data Analysis of Multivariate Observations*, New York: John Wiley. Several multivariate data sets from a variety of disciplines including engineering, manufacturing, biology, and mining. The volume includes several well-known data sets, such as Fisher's iris data (1936) and Rothkopf's Morse-code confusion data (1957), which have utility for the teaching of principal components analysis, factor analysis, multidimensional scaling, and cluster analysis.

Haberman, S. J. (1978), *Analysis of Qualitative Data*, New York: Academic Press. Several categorical data sets of wide interest: suicides by

day of the week, homicides by month, stressful events, etcetera.

Hand, D. J., and Taylor, C. C. (1987), *Multivariate Analysis of Variance and Repeated Measures: A Practical Approach for Behavioral Scientists*, London: Chapman & Hall. Raw data from eight studies in psychology and psychiatry, on topics as diverse as headaches, smoking, and Alzheimer's disease.

Harris, J. A., Jackson, C. M., Paterson, D. G., and Scammon, R. E. (1930), *The Measurement of Man*, Minneapolis, MN: The University of Minnesota Press. Many unique and interesting data sets on the link between physical and psychological characteristics. Do blondes have more fun? Do lunatics' eyebrows join together in the middle? Are manic-depressives thin or fat?

Hollander, M., and Proschan, F. (1984), *The Statistical Exorcist: Dispelling Statistics Anxiety*, New York: Marcel Dekker. A host of data sets of different sizes on many different topics including blood pressure and obesity of Mexican-Americans, baseball data, 1970/1971 draft lottery, promotion rates among male and female pharmacists, leisure time companies of black women, ranking of rum brands by different nationalities, preference for Charlie's Angels' actresses, longevity and environment, color of canned tuna, etcetera.

Koopmans, L. H. (1987), *Introduction to Contemporary Statistical Methods* 2nd ed., Boston, MA: PWS-Kent. Many interesting data sets on topics such as the number of armed robberies in Albuquerque per month between 1970 and 1974, gender differences in juvenile court sentences, growth rates for children treated with differing doses of human growth hormone, and differences in drug prices between brand names and generic versions.

Moore, D. S., and McCabe, G. P. (1989), *Introduction to the Practice of Statistics*, New York: W. H. Freeman. Several real data sets (mostly small) on topics such as the taste of cheddar cheese (p. 703) and salaries of men and women scientists (p. 761). Two additional larger data sets on computer science and reading comprehension.

Mosteller, F., and Tukey, J. W. (1977), *Data Analysis and Regression: A Second Course in Statistics*, Reading, MA: Addision-Wesley. Raw data for 13 data sets across several disciplines. Relevant examples include a subset of 20 from the Coleman Report, educational expenditures for Massachusetts school districts, and municipal bond data for 20 U.S. cities.

Moses, L. E. (1986), *Think and Explain With Statistics*, Reading, MA: Addison-Wesley. Raw data covering a variety of topics, including the heights of students taking statistics, highway fatalities, and football scores.

Ott, L. (1988), *An Introduction to Statistical Methods and Data Analysis* 3rd ed., Boston, MA: PWS-Kent. Many real data sets on topics including rates of violent crime in 90 U.S. cities, relationship between unemployment rate and crime rate in 20 U.S. cities, and employee turnover.

Plackett, R. L. (1981), *The Analysis of Categorical Data*, New York: MacMillan. A large variety of categorical data sets, including fingerprints, family size, work conditions and work quality, behavioral problems and birth order, and high school rank by gender and socioeconomic status.

Ryan, B. F., Joiner, B. L., and Ryan, T. A., Jr. (1985), *Minitab Manual* 2nd ed., Boston, MA: Duxbury Press. Thirty data sets of small to moderate sizes, on topics ranging from education to cartoons. The educational data sets include information on school strikes and freshman SAT verbal and math scores.

Stevens, J. (1986), *Applied Multivariate Statistics for the Social Sciences*, Hillsdale, NJ: Lawrence Erlbaum. Many artificial data sets as well as approximately 20 interesting small to moderately sized real data sets, including pre/post data on the influence of Sesame Street, risk of reading problems among kindergarteners, behavior reversal, programmed music instruction of elementary school children, IQ, testing, etcetera.

Timm, N. H. (1975), *Multivariate Analysis With Applications in Education and Psychology*, Belmont, CA: Wadsworth. Raw data for a handful of data sets gathered in educational settings, including effects of delay in oral practice on second language learning (pp. 228–229), relationship between recall and sentence structure (p. 233), and predictors of student performance on the Peabody Picture Vocabulary Test (p. 281).

Weisberg, S. (1980), *Applied Linear Regression*, New York: John Wiley. Raw data for several interesting data sets, including Cyril Burt's IQ data, Allison and Cicchetti's brain weight and body weight data, and three time points for 26 boys and 32 girls who participated in the Berkeley Guidance Study (anthropometric information only, however.)

[*Received May 1988. Revised November 1989.*]

REFERENCES

Brownlee, K. A. (1965), *Statistical Theory and Methodology in Science and Engineering*, New York: John Wiley.

Carmer, S. G., and Cady, F. B. (1969), "Computerized Data Generation for Teaching Statistics," *The American Statistician*, 23, 33–35.

Chatfield, C. (1985), "The Initial Examination of Data," *Journal of the Royal Statistical Society, Ser. A*, 148, 214–253.

Cobb, G. W. (1987), "Introductory Textbooks: A Framework for Evaluation," *Journal of the American Statistical Association*, 82, 321–339.

Dayton, C. M. (1972), "A Method for Constructing Data Which Illustrate a Suppressor Variable," *The American Statistician*, 26, 36.

Dorfman, D. D. (1978), "The Cyril Burt Question: New Findings. *Science*, 201, 1177–1186.

Draper, N. R., and Smith, H. (1981), *Applied Regression Analysis* (2nd ed.), New York: John Wiley.

Edwards, B. (1959), "Constructing Simple Correlation Problems With Predetermined Answers, *The American Statistician*, 12, 25–27.

Fisher, R. A. (1936), "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, 7, 179–188.

Gordon, N., Nucci, L., West, C. K., Hoerr, W. A., Vguroglu, M., Vukosavich, P., and Tsai, S. L. (1984), "Productivity and Citations of Educational Research: Using Educational Psychology as the Data Base," *Educational Researcher*, 13(7), 14–20.

Hays, W. L. (1981), *Statistics* (3rd ed.), New York: Holt, Rinehart & Winston.

Hunter, W. G. (1976), "Some Ideas About Teaching Design Experiments With 2 Examples of Experiments Conducted by Students," *The American Statistician*, 35, 227–233.

Jensen, A. R. (1974), "Kinship Correlations Reported by Sir Cyril Burt," *Behavioral Genetics*, 4, 1–28.

Joiner, B. L. (1988), "Let's Change How We Teach Statistics," *Chance*, 1, 53–54.

McCall, R. B. (1975), *Fundamental Statistics for Psychology* (2nd ed.), New York: Harcourt, Brace & Jovanovich.

Pedhazur, E. J. (1982), *Multiple Regression in Behavioral Research* (2nd ed.), New York: Holt, Rinehart & Winston.

Powell, B., and Steelman, L. C. (1984), "Variations in State SAT Performance: Meaningful or Misleading?," *Harvard Educational Review*, 54, 389–412.

Read, K. L. Q. (1985), "ANOVA Problems With Simple Numbers," *The American Statistician*, 39, 107–111.

Read, K. L. Q., and Riley, I. S. (1983), "Statistics Problems With Simple Numbers," *The American Statistician*, 37, 229–231.

Rosenbaum, P. R., and Rubin, D. B. (1985), "Discussion of On State Education Statistics: A Difficulty With Regression Analysis of Regional Test Score Averages," *Journal of Educational Statistics*, 10, 326–333.

Searle, S. R., and Firey, P. A. (1980), "Computer Generation of Data Sets for Homework Exercise in Simple Regression," *The American Statistician*, 34, 51–54.

Snedecor, G. W., and Cochran, W. G. (1980), *Statistical Methods* (6th ed.), Ames, IA: Iowa State University Press.

Supreme Court Ruling on Death Penalty (1988), *Chance: New Directions for Statistics and Computing*, 1, 7–8.

Tukey, J. W. (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley.

Wainer, H. (1986), "Five Pitfalls Encountered While Trying to Compare States on Their SAT Scores," *Journal of Educational Measurement*, 23, 69–81.

——— (ed.) (1986), *Drawing Inferences From Self-Selected Samples*, New York: Springer-Verlag.

Wainer, H., Holland, P. W., Swinton, S., and Wang, M. H. (1985), "State Education Statistics," *Journal of Educational Statistics*, 10, 293–325.

Winer, B. J. (1971), *Statistical Principles in Experimental Design* (2nd ed.), New York: McGraw-Hill.