DOCUMENT RESUME

ED 292 684 SE 049 067

AUTHOR Singer, Judith D.; Willett, John B.
TITLE Opening Up the Black Box of Recipe Statistics: Putting the Data Back into Data Analysis.
PUB DATE Apr 88
NOTE 41p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April, 1988).
PUB TYPE Viewpoints (120) -- Reference Materials - Bibliographies (131) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *College Mathematics; Data Collection; Higher Education; *Mathematical Applications; *Mathematics Curriculum; Mathematics Education; *Mathematics Instruction; *Statistical Analysis; *Statistical Data; Statistics
IDENTIFIERS Mathematics Education Research

ABSTRACT

Statistics tend to become interesting to non-methodologists when taught in a research context that is relevant to them. Real data sets supplemented by sufficient background information can provide just such a context. Despite this, many textbook authors and instructors of applied statistics rely on artificial data sets to illustrate statistical techniques. In this paper, it is argued that artificial data sets should be eliminated from the curriculum and that they should be replaced with real data sets. Towards this end, a rationale for using real data sets and the characteristics that make data sets particularly good for instructional use are described. The difficulties encountered when using real data and strategies for compensating for these drawbacks are also discussed. Two authentic data sets and an annotated bibliography of dozens of primary and secondary data sources are included. (Author/PK)

OPENING UP THE BLACK BOX OF RECIPE STATISTICS:

PUTTING THE DATA BACK INTO DATA ANALYSIS

Judith D. Singer
John B. Willett

Harvard University
Graduate School of Education

April 1988

Correspondence: Judith D. Singer, Ph. D.
Assistant Professor
Harvard University
Graduate School of Education
Larsen Hall, Seventh Floor
Cambridge, MA 02138
(617) 495-1961

Running Head: Data Sets to Teach Statistics

Paper presented at the annual meeting of the American Educational Research Association, April 1988, New Orleans, LA.

Abstract

Statistics becomes interesting to non-methodologists only when taught in a research context that is relevant to them. Real data sets supplemented by sufficient background information provide just such a context. Despite this, many textbook authors and instructors of applied statistics rely on artificial data sets to illustrate statistical techniques. In this paper, we argue that artificial data sets should be eliminated from the curriculum and that they should be replaced with real data sets. Towards this end, we describe the rationale for using real data sets and describe the characteristics that we have found make data sets particularly good for instructional use. Having learned that real data sets can present problems for instructors, we discuss the difficulties that we have encountered when using real data and some of our strategies for compensating for these drawbacks. We conclude by presenting two authentic data sets and an annotated bibliography of dozens of primary and secondary data sources.

Opening up the Black Box of Recipe Statistics:

Putting the Data Back into Data Analysis

Put yourself in your students shoes. It's Friday afternoon at 2:00 pm. You've survived another week of classes. By next Monday, you have to:

- Observe the interactions between another mother/preschooler dyad and write a three page paper on their use of language;
- Read a chapter from Alan Bloom's Closing of the American Mind and a chapter from E. D. Hirsch's Cultural Literacy and be prepared to discuss them during Monday's section; and,
- Do your statistics homework.

Hoping to complete the worst task first, you turn to your statistics homework and find the following problem.

> 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, ...
>
> Pedhazur (1982), p. 43

Would you want to do your statistics homework? Would you learn how regression models can help address interesting research questions? Would you be able to articulate what the F ratio really tells us? Would you learn how to use regression models to analyze data that might be of interest to you? Would you remember any of this two years from now when you have to analyze your dissertation data?

Now suppose you found the following problem.

> How attractive are America's most prestigious colleges to the high school seniors applying for admission? Is it as difficult as they say to turn Harvard down? How about Princeton and Yale? Is there evidence that some students apply to certain schools just because they are likely to be admitted, when they really have little intention of enrolling?
>
> Table 1 presents the 1986 admissions data for a random sample of 34 private colleges in the northeast. Two variables are given:
>
> ACCEPT: percent of applicants accepted
> YIELD: percent of accepted students actually enrolling
>
> You are going to examine the variable YIELD (Y), alone, and in relation to ACCEPT (X). ....

These data are real! Compared to them, Bloom and Hirsch seem abstract and theoretical. You might actually learn something interesting by doing this assignment. Which schools are hot? Which schools are safety schools? You might begin to understand the link between a research question and a statistical model. And you might even begin to think about how to use statistical models to examine the child language data that you have been collecting during the semester.

We believe that data sets of the first type do little to help our students become competent data analysts. Artificial data sets perpetuate the myth that statistics is dry and dull. After "analyzing" the data, students have not experienced the pleasure of doing research to investigate an interesting research question. Nor have they learned how statistical models can represent relationships between variables nor how statistical models can be interpreted in a real-life context. Most artificial data sets should be eliminated from the applied statistics curriculum.

In their place, we propose that statistics instructors use data sets of the second type, which enable students to learn analytic skills in a realistic research context. Real data sets provide a practical arena for learning how to link research questions to statistical models. Using real data sets helps us show our students how statistical analyses can inform current debates in educational research, thereby teaching students not only how to analyze data, but also why we analyze data. The use of real data sets helps us integrate statistics into the general education curriculum.

The purpose of this paper is to convince textbook authors and instructors of applied statistics to eliminate artificial data sets from the curriculum and replace them with real data sets. Towards this end, we describe the rationale for using real data sets and describe the characteristics that we have found make data sets particularly good for instructional use. Having learned (the hard way) that real data sets can present problems for instructors, we discuss the difficulties that we have encountered when using real data and some of our strategies for compensating for these drawbacks. We conclude by describing two real data sets that we have used in the classroom as well as an annotated bibliography of dozens of primary and secondary data sources.

The Rationale for Using Real Data

As applied statistics instructors, our mission is to teach our students the skills we believe necessary for conducting statistical analyses of high methodological quality. Although the techniques we cover during our four-semester sequence in quantitative methodology are diverse, our overarching goals are for students to be able to: (1) formulate interesting research questions; (2) select appropriate statistical techniques; (3) conduct all necessary calculations; (4) interpret the results of the analyses; (5) consider rival explanations of the results; and (6) summarize the findings in a cogent and convincing manner. The challenge for us is how best to achieve these goals.

Before the widespread availability of high-speed computing and pre-packaged statistics programs, the third goal--the computational aspects of data analysis--assumed priority over the other goals. After all, the success of any statistical analysis hinged upon the analyst's ability to perform the requisite calculations. Recognizing that the calculations could be time-consuming and tedious, many instructors and textbook authors tried to reduce student burden by using artificial data sets, constructed so as to simplify the arithmetic. For example, the observations in such data sets usually were integers, often chosen so that summary statistics, such as means, standard deviations and regression coefficients, also were integers. The American Statistician periodically published articles that described methods for constructing artificial data sets with specific characteristics (see, e.g., Edwards, 1959; Carmer & Cady, 1969; Dayton, 1972; Searle & Firey, 1980; Read & Riley, 1983; and Read, 1985) and artificial data sets were common fare in many popular applied statistics textbooks in education and the behavioral and social sciences (see, e.g., Hays, 1973, 1981; Winer 1962, 1973; and McCall, 1970, 1977).

Although the use of artificial data sets decreased the number of hours that students spent squaring and summing columns of numbers, it did not completely eliminate the drudgery of hand computation. The calculations were easier, but they still had to be performed. In the hope of keeping student attention focused on statistical concepts, not arithmetic details, many textbook authors provided step-by-step formulas ("recipes") designed to decrease the computational burden. By their very nature, these textbooks focused on confirmatory analyses, because only confirmatory analyses could be written as a sequence of specific steps, followed as one might follow a cookbook.

Unfortunately, although the rationale for using artificial data sets and cookbook strategies came from the desire to improve the quality of statistics instruction, the result usually fell far short of that goal. Artificial data sets perpetuated the myth that statistics is boring and unrelated to the students' substantive interests. Cookbook approaches to data analysis seduced students into believing that statistical analysis could be reduced to a set of predefined steps, conducted by a robot.

The widespread availability of high-speed computing has allowed statistics instructors to change the way in which they teach data analysis. Computers have eliminated the need for simplified arithmetic; the computer does not care if the observations are integers or if the summary statistics are integers. Most tedious calculations now can be relegated to a machine. No longer do students need to learn (let alone memorize) formulas whose sole purpose was to simplify computation. Exploratory and descriptive analyses, which previously were avoided, due in part to the time required to conduct them, can now be easily incorporated into the data analyst's tool kit. Just as computers have revolutionized the way in which we analyze data, so too should they revolutionize the way in which we <u>teach</u> how to analyze data.

One positive step in this direction is the increasing presence of computer output in statistics textbooks. In a review of 16 introductory statistics texts, Cobb (1987) noted that 8 included some computer output. But to our mind, the simple inclusion of computer output is not enough; <u>it is now time for teachers of applied statistics to change the data sets used to illustrate statistical techniques</u>.

What difference does the authenticity of a data set make? We have found that using real data sets has been a major factor in keeping our students--masters and doctoral candidates in education--motivated to learn statistical techniques. Although students report a diverse set of reasons for preferring authentic data sets, the major reason we hear is that the students find the real data we use to be intrinsically interesting--for example, most of them enjoy comparing the acceptance and yield rates at their school to the rates at other institutions. 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 our students report that some of the data sets themselves are memorable, thereby becoming mnemonics for recalling statistical techniques.

Over and above capturing the students' interest, we find real data sets to be particularly helpful instructional aids. Real data sets allow a 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 analysis problems that students encounter later in their own research. In essence, real data sets bring students as close as possible to an actual research experience.

But real data sets are helpful for another reason as well: They provide us an opportunity to teach students how to cope with many of the common problems that arise in real data, such as non-linearity, outliers, and missing values. These non-standard problems remind students of the need to investigate the tenability of assumptions; and the students respond by becoming interested in learning what they should do when the standard assumptions do not hold. Thus, the use of real data sets shows students that exploratory data analysis is an essential component of all statistical investigations.

Desired Characteristics of Real Data Sets

Not all real data sets are equally effective vehicles for teaching applied statistics. In this section, we discuss seven attributes that make a data set particularly well-suited for instructional use.

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 life-like variable names 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 a 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?

The problem with "life-like" data is that most students can easily see through the artifice. As a result, most students would not bother asking the questions raised above, because why should they care about how the data were "collected." Yet these questions reflect the very issues we would like students to raise when reviewing other people's research and when conducting their own. Because students can see through the ruse of "life like" data, we should not demean them by attempting to fool them.

<u>Background information</u>

A real data set should be accompanied by background information on the purpose and design of the research, the source of the data, measurement techniques, variable definition and so on. It is the provision of this information that allows students to fully assume the role of researcher.

As Cobb (1987) wrote when assessing the data examples used in 16 introductory textbooks:

> A data set is no longer alive if it is uprooted from its context like a pulled tooth. (What would you think of a dental school whose students only practiced drilling individual teeth that their instructor had already extracted?) To make a data set feel alive, the author must tell enough about what the numbers mean so that analysis is a search for meaning, not just an exercise in arithmetic. (pp. 331-332).

If the data come from a published paper or published tabulations, students should be given access to the original document. If the data are extracted from another source, the instructor must provide the background information. Many of the sources listed in the appendix are published papers; our students have found it interesting to read the original papers while conducting their analyses.

Interest and Relevance

Some of the best-selling statistics texts are filled with real data, but on topics of little interest to students in education and the social sciences. Snedecor and Cochran (1980) make 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 use real data, but on topics such as 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 of the classic statistics data sets, such as Fisher's iris data (1936) and Brownlee's stack loss data (1965) also fail to inspire our students.

Intrinsic interest is obviously in the eye of the beholder, but we can go a long way towards ensuring it by using data sets from our discipline. For example, the annual salary survey conducted by the American Association of University Professors (published annually in Academe), includes data of interest to most students: the average salaries of faculty members by institution and academic rank. The survey of school districts conducted by Education Resources Corporation is another useful source; it provides information on teacher's and administrator's salaries by district for a nationwide stratified random sample of districts. (The full citations for these sources are given in the appendix.)

Topicality often provokes student interest. Our students became very engaged in a data set recently reported in Chance (1988) on the relationship between race of victim, race of defendant and whether the defendant was given the death penalty. Although these data had little to do with education, students perceived it as quite relevant, especially in the current climate of racial tension on our nation's college campuses.

Controversy also has provoked student interest. Cyril Burt's data on the IQs of identical twins is interesting (Jensen, 1974), especially when analyzed in the context of Burt's views on the nature/nurture debate and Dorfman's (1979) and Kamin's (1976) 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 percent of students taking the test also arouses interest, especially when accompanied by newspaper accounts of Secretary Bennett's wall chart that ranks states according to these scores and critiques of all state comparisons of SAT scores given by Wainer, Holland, Swinton and Wang (1985), Rosenbaum and Rubin (1985), and Wainer (1986). By analyzing controversial data sets, students learn not just statistical techniques, but also how these techniques can support or undermine a hypothesis.

Historical data sets also have been an effective motivator for many of our students. The early volumes of journals such as Child Development, Journal of Educational Psychology, and Journal of Genetic Psychology, are filled with individual data. Although their topics are not always fascinating (e.g., four different types of math tests), their age often overcomes this gap. Moreover, it is interesting to compare modern's statistics more sophisticated analyses to the older tabular presentations given in the original sources.

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 from their data analyses that they did not know before, they discover just how useful statistical analysis can be. The substantive learning does not have to be on a grand scale relating to fundamental theories of education, but it should be real.

One of our most popular data examples is not from the research literature but from a local magazine. Every few years, Boston magazine conducts a survey of school districts in the local area and publishes data for each district on per-pupil expenditures, teacher salaries, student demographics and so on. The Boston Globe publishes similar data sets on a regular basis. When students analyze these data, they discover how their home town compares to others in the area and how district characteristics are related to each other. They gain new insight into the on-going political debate as to why some school districts are reported to be "better" than others. Substantive learning reinforces the reasons for conducting statistical analyses in the first place.

Availability of multiple analyses

As practicing statisticians, we often use more than one type of analysis to address a given research question. Different analyses provide different insights into the measures under study and when a data set is used in multiple analyses, the students learn that there may be more than one way to investigate a research question.

No experience reinforces the importance of multiple analyses as much as the discovery of previously unknown findings. For example, Scarcella (1984) published an analysis of the influence of language background and proficiency on choice of writing device (repetition, paraphrase, explanation). A re-analysis using log-linear modeling (given as a homework assignment to students) revealed previously unsuspected effects. In particular, it was possible to demonstrate that language proficiency, not language background, was a significant predictor of writing device.

The importance of raw data

We feel very strongly that the data must be given in raw form, not summarized using means and variance-covariance matrices. Rich information is lost when raw data are replaced by sufficient statistics and because students are one-step removed from the data they are seduced back into the cookbook approach fostered by hypothetical data. When raw data are available, students are free to adopt the data-analytic approach preferred by many practicing statisticians, be it the exploratory approach advocated by Tukey (1977) or the initial data examination approach advocated by Chatfield (1985). This allows students to look for high-leverage cases, heteroscedasticity, non-linearity and other non-standard problems that all too often arise in real data. The use of summary statistics may fool students into believing that such problems do not exist, or if they do, they are of little consequence.

Case identifiers

Many published data sets have case identifiers which allow students to bring background knowledge to their data analyses. State, school district and school identifiers have meaning for our students. If such identifiers are available, they should be provided with the data set so that students can use their background information about the cases to inform their statistical analyses. Case identifiers are particularly helpful for identifying outliers and high-leverage observations. When students analyze data on the citation frequencies of prominent researchers, for example, their knowledge about the researchers being studied helps them understand why Sigmund Freud and Jean Piaget might be outliers (Gordon, Nucci, West et al., 1984).

Drawbacks to Using Real Data

Using real data sets to teach statistics is not without shortcomings. Below we describe four problems that we have encountered and offer some remedies for overcoming these problems.

The Workload of Finding Data Sets

A major motivation for using artificial data is that an instructor can readily create any number of data sets with specific characteristics. For example, Dayton (1960) 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 student plagiarism by generating dozens of data sets and giving each student a different data set to "analyze." Producing a variable that is normally distributed, but with an outlier or two, is indeed a simple programming problem; identifying a real data set with the same features can take hours.

Having used real data sets for several years, we can clearly state that they increase the amount of time required to prepare classes, homeworks and exams. To identify a single data set to illustrate a specific technique we spend a great deal of time analyzing several data sets, some of which will not reveal interesting findings, others of which will present difficult analytic problems. This is especially true when developing materials for lower level courses, when students are still learning basic skills before learning to cope with non-standard problems.

These difficulties are, in fact, the major reason that we have written this paper. By including references to dozens of data sets that we have used, we hope that statistics instructors can gain access to a wide array of data sets that have many of the desired characteristics described in the previous section. Although it is still necessary to examine the data sets to determine which is most appropriate for introducing a specific concept, the availability of annotated bibliographies should facilitate the process.

<u>Small Data Sets and Statistical Power</u>

In our introductory and intermediate courses, we prefer small data sets, with sample sizes in the 35-75 range. Small data sets encourage the students to become intimately acquainted with each case, thereby fostering a more detailed understanding of the relationship between the data and the analyses. Once students have developed these skills, we introduce larger data sets in more advanced courses.

Unfortunately, small data sets create a false impression of how big effect sizes actually are in the real world. After all, null findings are not terribly interesting, so we tend to present data sets in which the effect sizes are large, yielding "statistically significant" results despite the small sample size. Although we, as instructors, know that these large effect sizes are not common in practice, the students do not see much evidence of this in their class problems or their homeworks. Thus, when we give them journal articles to read that report R^2's of 9%, many students conclude that the effect size is small, and it is, relative to <u>their</u> experience.

This problem is not unique to real data sets; most artificial data sets presented in applied statistics textbooks are also relatively small. The difference is that real data sets seem to reflect the larger class of statistical problems that arise in the real world. Because we see little means of eliminating this problem, we have chosen to specifically focus our students attention on it by discussing the concepts of statistical power, effect size, and the distinction between statistical significance and practical significance.

<u>Aggregate data and self-selected samples</u>

Easy access to information has often led us to use aggregate data or data on self-selected samples, such as mean SAT scores by state for the high school seniors who chose to take the test. In some of these data sets, the variables actually are measured at the aggregate level--for example, college tuition, student/faculty ratio, number of students enrolled--but many other data sets involve aggregate data, with all its attendant problems.

The question is whether the gains are worth the drawbacks, and in most instances, we believe they are. Aggregate data sets are some of the most readily available, intrinsically interesting data sets we use. The observations contained in aggregate data sets often have meaningful identifiers--names of towns, cities, counties, school districts, states or countries--thus enabling students to become more intimately associated with each individual data point. Nevertheless, in our more advanced classes we use these data sets to illustrate some of the problems involved in analyzing aggregate summaries or data on self-selected samples.

In-class testing

It is difficult, although not impossible, to test students in-class using real data. We do not use in-class exams, but rather multiple homework assignments and take-home exams. If you prefer to give an in-class exam, however, the solution is probably to hand out computer output and have the students interpret it. In doing so, though, note that the students are not choosing the analyses to be conducted--they are simply interpreting the output--and thus such an in-class exam may not be testing all of their analytic skills.

Two Examples

Perhaps the best way of discovering the advantages of authentic data sets is to try them in your classes. To assist in the search for real data, the appendix presents an annotated bibliography of primary and secondary sources. As illustrations of what you are likely to find in these sources, we present below two real data sets.

What does college tuition buy?

The cost of a college education has been rising rapidly during the past decade; at many schools in the northeast, it costs over $10,000 in tuition alone for a year at a private college. David W. Breneman, president of Kalamazoo College, has suggested that some colleges are simply raising their tuition to increase their prestige (President Says, 1988). With tuition at an all-time high, the question arises at to what tuition actually buy. Better trained faculty? Better student/faculty ratios? Better students?

Table 1 presents data on tuition and selected characteristics of the faculty and student body at 34 private colleges in the northeast, including:

TUITION	Total tuition for 1986-87 academic year
NAPPLY	Number of freshman applicants in Fall 1985
PCTADMIT	Percent of freshman applicants admitted in Spring 1986
PCTYIELD	Percent of admitted applicants who matriculated in Fall 1986
PCTDOC	Percent of faculty holding a doctorate or the highest degree in their field
PCTFIFTH	Percent of matriculating freshmen in 1986 who were in the top fifth of their high school graduating class
PT_FAC	Number of part-time faculty members
FT_FAC	Number of full-time faculty members
SFRATIO	Student/faculty ratio
PROFSAL	Average salary of full professors
ASSTSAL	Average salary of assistant professors

---

insert Table 1 here

We have used this data set to illustrate regression model building. Additional variables such as the school's endowment, mean financial aid award, and mean SAT scores of entering freshman could be easily added, as could additional colleges. In particular, we strongly recommend adding the school at which you teach so that the students can compare their institution to other schools.

What are the ratings rating?

In 1982, the National Academy of Sciences published a report rating "the scholarly quality" of research programs in the humanities, physical sciences and social sciences. The ratings were based upon rankings of quality and reputation made by senior faculty in the field who taught at institutions other than the one being rated. The report stirred much controversy, as most published ratings do. Critics argued that peer rating scales were not necessarily indices of quality, but may instead reflect the institution's prestige, reputation, productivity, or perhaps size. Thus, the question arises as to what the ratings rate.

Table 2 presents the quality ratings of 46 research doctorate programs in psychology, as well as six potential correlates of the quality ratings. The variables are:

QUALITY	Mean rating of scholarly quality of program faculty
NFACULTY	Number of faculty members in program as of December 1980
NGRADS	Number of program graduates from 1975 through 1980
PCTSUPP	Percentage of program graduates from 1975-1979 that received fellowships or training grant support during their graduate education
PCTGRANT	Percent of faculty members holding research grants from the Alcohol, Drug Abuse and Mental Health Administration,

the National Institute of Health or the National Science Foundation at any time during 1978-1980

NARTICLE Number of published articles attributed to program faculty members 1978-1980

PCTPUB Percent of faculty with one or more published articles from 1978-1980

----

insert Table 2 here

We have used this data set for teaching how to build regression models, but smaller portions of the data set could easily be used to illustrate other techniques. The data set could be modified by adding additional schools, additional predictors, or by choosing a sample from a different subject specialty.

Conclusion

Real data sets can be a statistics instructors strongest ally in motivating students to learn how to analyze data. Although the use of real data sets is not without problems, the strengths far outweigh the weaknesses. Moreover, the biggest drawback--the amount of time needed to identify data sets exhibiting specific statistical patterns and problems--can be overcome by communications in statistical journals that identify where such data sets can be found. The annotated bibliography in this paper is a first step. As more individuals identify data sources, we should be able to eliminate most artificial data sets from the applied statistics curriculum.

References

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

Carmer, S. G., & 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, Series 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., & Smith, H. (1981). Applied Regression Analysis, 2nd edition, (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. J., Nucci, L. P., West, C. K. et al. (1984) Productivity and citations of educational research: Using educational psychology as the data base. Educational Researcher, 13, 14-20.

Hays, W. L. (1981). Statistics, 3rd edition. (New York: Holt, Rinehart and Winston).

__________ (1973). Statistics, 2nd edition. (New York: Holt, Rinehart and Winston).

Jensen, A. R. (1974). Kinship correlations reported by Sir Cyril Burt. Behavioral Genetics, 4, 1-28.

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

__________ (1970). Fundamental Statistics for Psychology, 1st edition. (New York: Harcourt Brace Jovanovich).

Pedhazur, E. J. (1981). Multiple Regression in Behavioral Research, 2nd edition. (New York: Holt, Rinehart and Winston).

Powell, B. & Steelman, L. C. (1984). Variations in state SAT performance: Meaningful or misleading?. Harvard Educational Review, 54, 389-412.

President says 100 private colleges follow crowd: the higher their prices, the more students apply. The Chronicle of Higher Education, 2 March 1988, p. A29.

Read, K. L. Q. (1985). ANOVA problems with simple numbers. The American Statistician, 39, 107-111.

Read, K. L. Q., & Riley, I. S. (1983). Statistics problems with simple numbers, The American Statistician, 37, 229-231.

Rosenbaum, P. R. & Rubin, D. B. (1985). Discussion of "On State Education Statistics": A difficulty with regression analyses of regional test score averages. Journal of Educational Statistics, 10, 326-333.

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.

Searle, S. R., & Firey, P. A. (1980). Computer generation of data sets for homework exercises in simple regression, The American Statistician, 34, 51-54.

Snedecor, G. W., & Cochran, W. G. (1980). Statistical Methods, 6th edition. (Ames, Iowa: Iowa State Press).

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

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

Wainer, H. (1986). Five pitfalls encountered while trying to compare states on their SAT scores. Journal of Educational Measurement, 23, 69-81.

Wainer, H. ed. (1986). Drawing Inferences from Self-Selected Samples. (New York: Springer-Verlag).

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

__________ (1962). Statistical Principles in Experimental Design, 3rd edition. (New York: McGraw Hill).

Table 1. Characteristics of 34 private northeast colleges

OBS	NAME	TUIT	NAPPLY	%ADMIT	%YIELD	%DOC	%FIFTH	PT FAC	FT FAC	SFRAT	PROF$	ASST$
1	ADELPHI	5990	2855	66	39	75	91	500	350	12	39	25
2	ALBRIGHT	8245	1330	60	45	75	69	35	93	13	34	23
3	BABSON	9152	1723	42	48	91	37	50	94	23	46	30
4	BARD	11450	1000	60	33	80	36	20	65	9	34	21
5	BATES	11835	2808	42	37	85	75	38	122	13	39	24
6	BENTLEY	7300	3945	54	41	74	58	150	197	20	41	27
7	BOSTON COLLEGE	9120	16163	31	48	83	73	138	566	16	43	26
8	BOSTON UNIV	10950	18476	60	33	94	70	920	1511	15	42	26
9	BOWDOIN	10760	3555	23	50	78	90*	0	112	12	42	23
10	BRANDEIS	10950	3684	63	31	87	79	105	357	8	46	25
11	BROWN	11690	13630	19	52	95	93	60	475	11	45	25
12	BUCKNELL	10640	6279	43	33	90	74	20	213	14	41	25
13	CLARK	10200	2969	58	37	95	65	50	155	14	39	23
14	CONN. COLLEGE	11170	3456	46	28	75	56	94	136	12	35	22
15	FRANKLIN/MARSH	10200	3350	62	26	92	68	25	134	14	41	22
16	GETTYSBURG	9425	2640	66	32	78	53	36	132	13	37	22
17	HARVARD	11390	13614	16	73	99	98	198	674	10	59	28
18	HOFSTRA	6400	6229	68	39	85	55	325	400	18	39	25
19	IONA	5520	2940	74	39	56	13	130	247	22	38	26
20	ITHACA	7646	7306	65	31	73	14	104	362	15	33	22
21	LONG ISL UNIV	6350	4323	72	35	70	4	345	382	12	35	23
22	MARYMOUNT	6710	631	81	45	55	16	82	67	9	29	20
23	MIDDLEBURY	14500	3890	30	42	75	80	25	178	13	40	22
24	MONMOUTH	6662	2254	77	43	60	29	126	160	13	33	23
25	MT ST VINCENT	6150	617	76	35	75	27	10	64	12	29	20
26	NORTHEASTERN	6753	20901	57	33	80	30*	2108	784	20	43	26
27	RIVIER	5280	385	82	68	30	36	135	54	17	21	18
28	ST ROSE COLL	5570	672	73	47	60	45	100	108	12	29	19
29	SIMMONS	8992	1156	80	40	78	39	172	177	12	38	24
30	SWARTHMORE	11200	2590	32	41	90	92	15	135	10	44	24
31	SYRACUSE	7570	13654	69	32	71	43	405	898	16	40	23
32	TRINITY	10355	3388	40	33	90	46	14	121	15	42	23
33	VILLANOVA	6560	8488	50	39	73	61	306	518	14	40	26
34	WESLEYAN	10860	4365	38	42	84	82	63	251	11	43	24

* estimated from available data

Mazzari, L. (Ed.). (1986) College Admissions Data Handbook: 1986-87 Northeast Region (Concord, MA: Orchard House, Inc.), for all data except faculty salaries; and American Association of University Professors (1984). The annual report on the economic status of the profession. Academe, 70, for faculty salary data.

Table 2. Ratings of 46 research doctorate programs in psychology

OBS	NAME	QUALITY	NFACULTY	NGRADS	PCTSUPP	PCTGRANT	NARTICLE	PCTPUB
1	ADELPHI	12	13	19	16	8	14	39
2	ARIZONA--TUCSON	23	29	72	67	3	61	66
3	BOSTON UNIV	29	38	111	66	13	68	68
4	BROWN	36	16	28	52	63	49	75
5	U C BERKELEY	44	40	104	64	53	130	83
6	U C RIVERSIDE	21	14	28	59	29	65	79
7	CARNEGIE MELLON	40	44	16	81	35	79	82
8	UNIV OF CHICAGO	42	60	57	65	40	187	82
9	CLARK UNIV	24	16	18	87	19	32	75
10	COLUMBIA TEACHERS	30	37	41	43	8	50	54
11	DELAWARE,UNIV OF	20	20	45	26	25	49	50
12	DETROIT, UNIV OF	8	11	27	7	0	9	27
13	FLORIDA ST--TALAH	28	29	112	64	35	65	69
14	FULLER THEOL SEMIN	14	14	57	10	0	11	43
15	UNIV OF GEORGIA	27	38	167	28	13	196	84
16	HARVARD	46	27	113	62	52	173	85
17	HOUSTON,UNIV OF	29	32	122	51	19	79	69
18	UNIV ILLINOIS-CHAMP	42	56	116	56	32	208	73
19	IOWA,UNIV OF	33	32	54	49	19	120	69
20	KANSAS,UNIV OF	31	42	79	41	14	114	71
21	KENT STATE UNIV	23	30	76	22	20	87	67
22	LOUISIANA STATE	18	18	62	39	6	10	39
23	UNIV OF MARYLAND	29	41	98	41	12	101	66
24	MIAMI UNIV	21	23	52	33	4	59	78
25	U MICH--ANN ARB	45	111	222	64	32	274	70
26	U MISSOURI	25	26	63	39	23	160	89
27	U NEW HAMPSHIRE	18	16	24	4	31	39	63
28	NEW YORK UNIV	33	38	154	55	34	84	63
29	U NC -- GREENSBORO	21	19	40	7	5	60	84
30	NORTHEASTERN	24	16	18	25	63	31	63
31	NOTRE DAME	15	13	29	23	15	62	85
32	OKLA ST--STILLWATER	15	23	41	51	4	24	57
33	PENN STATE	36	32	69	65	16	122	75
34	PRINCETON	38	21	38	28	48	92	91
35	UNIV OF ROCHESTER	32	28	90	70	36	117	61
36	SUNY ALBANY	27	22	52	10	27	114	86

continued

Table 2. continued

OBS	NAME	QUALITY	NFACULTY	NGRADS	PCTSUPP	PCTGRANT	NARTICLE	PCTPUB
37	ST LOUIS UNIVERSITY	16	20	80	46	10	19	40
38	UNIV SOUTH FLORIDA	26	32	41	13	6	64	56
39	STANFORD	48	26	81	70	58	155	100
40	TEMPLE	26	40	81	42	10	70	68
41	TEXAS TECH LUBBOCK	14	19	87	15	5	72	79
42	UNIV OF TOLEDO	12	17	26	9	6	15	59
43	UNIV OF UTAH, SALT L	29	29	71	74	17	85	76
44	VIRGINIA POLYTECH	34	27	20	0	29	79	57
45	WASHINGTON UNIV-ST.L	28	26	70	68	27	84	73
46	UNIV WISC--MADISON	39	36	59	57	67	172	83

Jones, L. V., Lindzey, G., & Coggeshall, P. (Eds.) (1982). <u>An Assessment of Research-Doctorate Programs in the United States: Social and Behavioral Sciences</u>, (Washington, DC: National Academy Press).

APPENDIX

Annotated Bibliography of Published Data Sets

Afifi, A. A., & Azen, S. P. (1979). Statistical Analysis: A Computer Oriented Approach, 2nd edition, (New York: Academic Press).

Several extensive data sets describing the blood chemistry (cholesterol, blood pressure, etc.), cardiovascular state, socioeconomic status, and year of death. Some censored cases, could be used in the teaching of survival analysis. Other datasets 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., & Clark, V. (1984). Computer Aided Multivariate Analysis. (Belmont, CA: Lifetime Learning), p 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, person 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 marajuana, sudden infant death.

Aldrich, J. H., & Nelson, F. D. (1984). Linear Probability, Logit and Probit Models. Sage University Paper Number 45. (Beverly Hills, CA: Sage).

Data concerning the effect of the "Personalized System of Instruction" on course grades in an intermediate macroeconomics course, useful for logit analysis and log-linear modeling.

Allison, T. & Cicchetti, D. V. (1976). Sleep in mammals: Ecological and constitutional correlates. Science, 194, 732-734.

Average brain weights 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.

Andrews, D. F. & Herzberg, A. M. (1985). Data: A Collection of Problems from Many Fields for the Student and Research Worker. (New York: Springer-Verlag).

Raw data for 71 data sets. 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 (The Federalist papers data set and another on Platonic prose rhythm) and the birthday/deathday problem.

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. Institutions are categorized according to Carnegie classifications.

Aylward, G. P., Harcher, R. P., Leavitt, L. A., Rao, V., Bauer, C. R., Brennan, M. J., & 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 analysis of these data in Green, J. A. (1988). Loglinear analysis of cross-classified ordinal data: Applications in developmental research, Child Development, 59, 1-25.

Barrons' Publications (1987). Profiles of American Colleges, Sixteenth Edition, (New York: Barron's Publications).

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 freshman, 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, 467-170.

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).

Berenson, M. L., Levine, D. M., & Goldstein, M. (1983). Intermediate Statistical Methods and Applications. (Englewood Cliffs, NJ: Prentice Hall).

A large variety of non-educational data sets (lawn service, real estate market, professional sports, foreign food), with some educatonal data sets scattered here and there: categorical data on health issues in children by graduating class of pediatrician, starting salaries of MBA graduates, etc.

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 (4 grades) data on scaled vocabulary scores for boys and girls, and so on. (Data repeated in Finn, J. D., & Mattsson, I. (1978). Multivariate Analysis in Educational Research. (Chicago, IL: National Educational Resources).

Boli, J., Allen, M. L., & Payne, A. (1985). High ability women and men in undergraduate mathematics and chemistry courses. American Educational Research Journal.

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.

Chambers, J. M., Cleveland, W. S., Kleiner, B. & Tukey, P. A. (1983).

Graphical Methods for Data Analysis. (Belmont, CA: Wadsworth).

Raw data for 33 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, heights of singers in the New York Choral Society.

Chapman, J. C. (1914). Individual Differences in Ability and Improvement and their correlations. Teacher's College, 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 for 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).

Cooley, W. W., & 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. & Snell, E. J. (1981). Applied Statistics: Principles and Examples. (London: Chapman and Hall).

Raw data for 39 data sets. Relevant examples include: educational plans of Wisconsin school boys, statistical aspects of literary style, satisfaction with housing conditions.

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.

Devore, J. & 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., & Smith, H. (1981). Applied Regression Analysis, 2nd edition, (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, ailments of university alumni.

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 deviation) for total amount of calories eaten and minutes spent at table.

Dunteman, G. H. (1984). Introduction to Linear Models. (Beverly Hills, CA: Sage).

Data for 300 participants in the National Longitudinal Study on reading, math, gender, race, college status, SES, High school program, High school grades, creativity, stress avoidance, etc.

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.

Erickson, B. H., & 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, characteristics of social networks.

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.

Finn, J. D. (1974). A General Model for Multivariate Analysis. (New York: Holt, Rinehart and Winston).

Raw data for four studies of relevance to education: 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, causes of the 1907 Romanian Peasant rebellion.

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., & 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 (agressive or asocial). Authors explore relationship between IQ and all these predictors.

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.

Gordon, N. J., Nucci, L. P., West, C. K. et al. (1984) Productivity and citations of educational research: Using educational psychology as the data base. Educational Researcher, 13, 14-20.

Citation frequencies and dates of birth for 187 prominent educational researchers. Four sources of citations are given: AERJ & JEP, Review of Research in Education, selected Educational Psychology Texts, and Social Science Citation Index.

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, etc.

Hahn, H. H., & 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 pre-tests and post-tests for 167 students in grades 4, 5, 6, and 7.

Hand, D. J. & Taylor, C. C. (1987). Multivariate Analysis of Variance and Repeated Measures: A Practical Approach for Behavioral Scientists. (London: Chapman and Hall).

Raw data for 8 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., & Scammon, R. E. (1930). The Measurment of Man. (Minneapolis, MN: The University of Minnesota Press).

Many unique and interesting data sets on the link between physical and psychological characteristics. Do blonds have more fun? Do lunatics eyebrows join together in the middle? Are manic depressives thin or fat?

Harris, S., & 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.

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.

Hollander, M., & 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, 1970/1 draft lottery, promotion rates among male and female pharmacists, leisure time companions of black women, ranking of rum brands by different nationalities, preference for Charlie's Angel's actors, longevity and environment, color of canned tuna, etc.

Howard, G. S., Cole, D. A., & 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. Doctoral dissertation, Department of Statistics, UC Berkeley.

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. Behavioral Genetics, 4, 1-28.

"Burt's final assessments" of IQs of monozygotic 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, 245-246; Dorfman, D. D. (1979). Letter to the editor. Science, 204, 246-254. and Hearnshaw, L. S. (1979). Cyril Burt: Psychologist, (London: Hodder & Stoughton), Chapter Twelve.

___________ (1970). IQ's of identical twins reared apart. Behavioral Genetics, 1, 133-148.

Original data from four studies of IQs of identical twins reared apart: Newman, Freeman and Holzinger (1937), Shields (1962), Juel-Neilsen (1965) and Burt (1955).

Johnson, B. & 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.

Jones, L. V., Lindzey, G., & 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., Fisichelli, V. R., Costa, J., Karelitz, R., & 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.

Leinhardt, G., & 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. & Wasserman, S. S. (1979). Teaching regression: An exploratory approach. The American Statistician, 33(4), 196-203.

Life expectancy and per capita income for 105 nations divided into five national wealth classifications (industrialized, petroleum exporting, higher, middle and lower).

Maresh, M. M., & 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, or 5 occasions, by sex. The authors construct lots of individual growth curves.

Mason, T. J., & McKay, F. W. (1974). US Cancer Mortality by County: 1950-1969. (Washington, DC: US Government Printing Offi_e).

Lung cancer mortality by degree of urbanization and gender, in Louisiana.

Mickey, M. R., Dunn, O. J., & 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 high leverage cases.

Mosteller, F. & Tukey, J. W. (1977). Data Analysis and Regression: A Second Course in Statistics. (Reading, MA: Addison-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, municipal bond data for 20 US cities.

National Center for Education Statistics. (1987). The Condition of Education. (Washington, DC: US Department of Education).

____________ (1987). Digest of Education Statistics. (Washington, DC: US 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 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.

Opening lines. (1985, September). Harper's, pp. 29-30.

Selected results from a study of opening lines used in singles 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 Tanur, J.M., Mosteller, F., Kruskal, W. H., Link, R. F., Pieters, R. S., Rising, G. R., and Lehmann, E. L. (eds.) Statistics: A Guide to the Unknown, 2nd ed. (San Francisco, CA: Holden-Day). 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., & Carstensen (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., 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 cc ycat 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.

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, high school rank by gender and socioeconomic status.

Pcwell, B. & 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, H, Holland, P. W., Swinton, S., & Wang, M. H. (1985). On "State Education Statistics". Journal of Educational Statistics, 10, 293-325. Also see: Rosenbaum, P. R. & Rubin, D. B.

(1985). Discussion of "On State Education Statistics": A difficulty with regression analyses of regional test score averages. Journal of Educational Statistics, 10, 326-333. and Wainer, H. (1986). Five pitfalls encountered while trying to compare states on their SAT scores. Journal of Educational Measurement, 23, 69-81.

Rubin, E. (1972). Statistical exploration of a medieval household book, American Statistician, 26, 37-39.

Number of meals served, breads baked and ale brewed at the de Bryene household from October 1412-September 1413, by month. That's right, the fifteenth century. These data have nothing to do with education, but their age makes them intrinsically interesting.

Ryan, B. F., Joiner, B. L., Ryan, T. A. Jr. (1985). Minitab Manual, 2nd edition. (Boston, MA: Duxbury).

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.

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 non-native speakers and how this influences their choice of writing device.

Shearer, L. (1987). How will history rate Nancy Reagan? Parade Magazine, 14 June 1987, p. 8.

Rankings for 17 first ladies from Florence Harding through Nancy Reagan on 10 dimensions ranging from integrity, leadership and accomplishments.

Skodak, M., & 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.

Stevens, J. (1986). Applied Multivariate Statistics for the Social Sciences. (Hillsdale, NJ: Lawrence Erlbaum).

Approximately 20 interesting small to moderately sized educational (and other) data sets, including: pre/post data on the influence of

Sesame Street, risk of reading problems among kindergartners, behavior reversal, programmed music instruction of elementary school children, IQ testing, etc.

Stewart, L. H. (1955). The expression of personality in drawings and paintings. Genetic Psychology Monographs, 51, 45-103.

Data for 28 boys given including: 2 IQ scores (Terman and Stanford-Binet), 2 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.

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: eff .ts of delay in oral practice on second language learning (pp. 228-229), relationship between recall and sentence structure (p. 233), predictors of student performance on the Peabody Picture Vocabulary Test (p. 281).

Tufte, E. R. (1978). Registration and Voting. In Tanur, J. M. et al., Statistics: A Guide to the Unknown, 2nd edition. (San Francisco: Holden-Day), 195-204.

Data on percent on 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. & Bowen, W. G. (1967). Registration and Voting: Puting first things first. American Political Science Review. 61, 359-379.

United Nation's Children's Fund (1987). The State of the W. ld's Children. (New York: Oxford University Press).

Sociodemographic, education, health and economic indicators for 130 countries.

Walberg, H. J., & 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 are also present here, but it does make an interesting example.

Weisberg, S. (1980). Applied Linear Regression. (New York: Wiley).

Raw data for several interesting data sets including Cyril Burt's IQ data, Allison and Cicchetti's brain weight and body weight data, three time points for 26 boys and 32 girls who participated in the Berkeley Guidance study (anthropometric information only, however.)

Whiting, J. M. & Child, I. L. (1962) Child Training and Personality, (New Haven, CN: Yale University Press).

Many characteristics in dozens of societies around the world, including age at weaning, toilet training, fear of ghosts, rituals, etc.

Wilson, M. E., & Mather, L. E. (1974). Life expectancy [Letter to the editor]. Journal of the American Medical Association, 229(11) 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.

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.