In F. Gordon and S. Gordon (eds.) *Statistics for the Twenty-first Century*, Washington, DC: Mathematical Association of America, MAA Notes, Number 26, 83-98.

# Providing a Statistical "Model": Teaching Applied Statistics using Real-World Data

**John B. Willett and Judith D. Singer**[1]
Harvard University Graduate School of Education

## Introduction

Service courses in applied statistics abound in college and graduate school curricula around the world [34]. In a recent survey of United States graduate programs in health professions, for example, Cockerill and Fried [12] found that *all* programs required their students to take at least one statistics and research methods course. In a similar survey of psychology departments, Aiken, West, Sechrest and Reno [1] found that the same was true for nine out of every ten doctoral programs.

The clear consensus, among students and professors alike, is that too many of these applied statistics courses are far from successful [3, 9, 25, 33]. In Dallal's [13] words, "[t]he field of statistics is littered with students who are frustrated by their courses, finish with no useful skills, and are turned off to the subject for life" (p. 266). Joiner [27] gave service courses in statistics "a grade of F" for being unmitigated failures (p. 53). These courses frequently receive the worst evaluations in a school. And is it any surprise? Most of them are abstract, mechanical and boring, with antiquated and formulaic pedagogy that is little more than a presentation of the "statistic of the day."

Many students enrolled in applied statistics courses are hamstrung further by false perceptions of their own inadequacy. They regard "stats" courses as a necessary evil, an unavoidable rite of passage, and they view data analysis with fear and trepidation. For those who last saw a logarithm or summation sign a decade earlier, such anxiety is understandable. But when it so cripples our students that they cannot learn, we need better methods for reaching them and teaching them as well.

How can learning applied statistics be made more interesting, more palatable, more successful? We believe that one approach is to capitalize on students' fascination, not for statistics itself, but for the *substantive problems that statistics can address*. College students are interested in discovering whether there is race or gender bias in achievement testing, whether children's pre-natal experiences influence their later inhibition, whether better economic incentives lengthen how long teachers stay in teaching. They take applied statistics courses not to become statisticians, but to learn how to address questions such as these or to learn how to

---

[1] The order of the authors has been determined by randomization. Earlier versions of parts of this paper were presented at the annual meetings of the American Educational Research Association, New Orleans, Louisiana (April, 1988) and Chicago, Illinois (April, 1991). An earlier version of the third and fourth sections was published in *The American Statistician*, 1990, 44(3), 223-230.

read others' research on such topics. They are less interested in the algebraic ins-and-outs of mathematical statistics than they are in learning how to *use* statistics. Our job, as educators, is to provide courses that meet their needs.

Put yourself in your students' shoes. It is Friday afternoon at 2:00 pm, and you have survived another week of classes. By next Monday, you must complete your statistics homework. You open your textbook to 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 of 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, ...*
> [37, p. 43]

As a student of psychology, business, the health sciences, education, or for that matter, mathematics or statistics, would you be motivated by this assignment or would you be turned off by the jargon and lack of context? Would completing your homework help you understand how regression analysis can answer interesting questions about relationships among substantively important variables? Would you remember any of this two years from now, when you have to analyze your thesis data or tackle a problem on the job?

Now suppose you opened your textbook and found a different sort of problem:

> *The cost of a college education has been rising rapidly during the past decade; at many private schools in the northeast, annual tuition now exceeds $10,000. David Breneman, president of Kalamazoo College, has suggested that some colleges are charging high tuition not just to raise revenues, but to create an aura of high prestige [39]. So with tuition at an all-time high, the question arises as to what the money actually buys. Better trained faculty? Better student/faculty ratios? Better students? Table 1 presents tuition rates and selected characteristics of faculty and students for a random sample of 34 private colleges in the northeast. Use multiple regression analysis to examine the relationship between tuition and two potential predictors:*
>
> *MEANSAT: Mean total SAT score for matriculating freshmen.*
> *PCTDOC: Percent of faculty holding a doctorate or the highest degree in their field.*
>
> *Report your findings using non-technical language in a letter to the editor of our school's Alumni Gazette. Organize the data-analytic evidence supporting your conclusions into a statistical appendix to be submitted with your letter.*

These data are real! You might actually learn something interesting by completing the assignment. Which schools are overpriced? Which schools are bargains? You would begin to see links between research questions and statistical models. You would begin to learn how to

translate statistical findings into readable prose. You might even begin to think about how to use regression analysis to examine the data you have been collecting all semester.

We believe that artificial data sets do little to help our students become competent data analysts. All they do is perpetuate the myth that statistics is dry and dull. After "analyzing" the data, students have not experienced the thrill of doing research, nor have they been challenged to express their results in non-technical terms. We believe such methods for teaching applied statistics should be purged from the pedagogic repertoire.

In their place, we recommend that instructors and textbook authors use real-world data, so that students can learn skills in a realistic and relevant context. In addition to being more interesting, real data sets provide a practical arena in which students can learn how to link research questions to statistical models. Real data sets illustrate how statistical methods can inform the current research debate. By using real data, we can teach not only *how* we analyze data, but also *why* we do so [4, 10, 11, 24, 35, 49, 50, 51, 53].

Use of real data sets has another advantage -- we can teach applied statistics in the way that statistics are applied. Real data allow instructors to "model" good data-analytic practice, making statistics more palatable to students and empowering them as well. In this paper, we describe how we use real data in the classroom and we identify characteristics of data sets that make them particularly good for teaching. We also identify advantages and disadvantages of this approach, and offer suggestions for overcoming the obstacles. In a separate section of this volume, we provide an annotated bibliography that lists several hundred primary and secondary data sources that teachers may use in their own courses (see also [15, 22]).

## How Can We Use Real-World Data to Teach Applied Statistics?

Before high-speed computing and statistical packages were widely available, computational burden assumed instructional priority in applied statistics classes. After all, analytic "success" hinged upon the analyst's ability to execute the requisite calculations. Because computation was time-consuming and tedious, many instructors and textbook authors reduced the burden by using arithmetically simple artificial data sets. They used observations that were integers, often chosen so that summary statistics were also integers. Articles describing methods for constructing such data sets were published periodically (e.g., [6, 14, 19, 40, 41, 45]), and they were common fare in statistics textbooks (e.g., [23, 32, 59]).

Although artificial data reduced the time spent manipulating numbers, the drudgery of hand computation remained. 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 and classroom teachers provided step-by-step formulae that decreased the computational burden. This inevitably led to an emphasis on confirmatory analyses that could be explicated as a rigid sequence of concrete steps, to be followed as one might follow a recipe. Applied statistics courses often became "cooking" classes in which students memorized the computations instead of learning the concepts.

Although use of artificial data sets and cookbook strategies stemmed from a desire to improve instructional quality, the result usually fell far short of that goal. This approach confirmed students' expectations that statistics was boring, unrelated to their substantive interests. Data analytic "cookbooks" seduced them into believing that analyses could be reduced to a set of predefined steps, conducted blindly by a robot. Concern for numerical accuracy took precedence over the acquisition of conceptual insight [11, 53]. In the end, researchers trained this way would often use methodology, rather than substance, to design their research. They

would ask, for example, how to design a project so that they could use analysis of variance rather than asking how best to address a specific research question.

In today's computer age, we can, and we should, change the way we teach applied statistics [15, 18, 54]. Computers eliminate the need for simplified arithmetic. Tedious calculations can be relegated to the machine. Students need not memorize formulae whose sole purpose is computational. Exploratory and descriptive methods, once avoided because they were messy and time-consuming, can be incorporated into the students' analytic repertoire. Data analysis can become a partnership of exploration and confirmation, of induction and deduction.

Relieving students of the computational burden helps us focus their energies on the tasks that only human beings can perform: stating research questions, selecting appropriate statistical models, interpreting parameter estimates, writing up results, contemplating the implications for policy and practice. Just as computers have revolutionized the way in which we *analyze* data, so, too, should they revolutionize the way in which we *teach* how to analyze data. We should not be training researchers who see themselves solely as technicians; we should be training researchers who can use their technical skills to address real-world problems [36].

But decades of educational research tells us that teachers tend to teach as they themselves were taught [47]. Teachers use as their models the teachers who impressed them the most when they were students. But the teachers who most impressed mathematicians and statisticians when they were students may be inappropriate models when teaching non-mathematicians and non-statisticians [27]. For statisticians in training, the methods are paramount; for applied researchers in training, the *application* of the methods should reign supreme.

How were most of us -- the mathematician and statistician teachers -- taught? Most of our teachers used the *theorem*, *proof*, and *worked example* method. Class began with a statement of a theorem or formula. The instructor presented a mathematical proof, or if time was short, the proof was assigned for homework. Then came a worked arithmetic or algebraic example that illustrated a specific property of the theorem or formula. There was little, if any, contact with real data. As Moore and Roberts [35] noted, this approach remains popular today because it is "seductively easy to teach" (p. 81).

The "theorem-proof-example" strategy may have worked fine for us, but it is a poor model for applied statistics instruction. Our students care little about methodology for its own sake, nor do they care for theorems, formulae and proofs. They are bored by variables called $X$ and $Y$. Their interests are substantive: they want to know what $X$ and $Y$ represent in the real world, why $X$ and $Y$ might be related, and what the implications are of any relationship that might exist between $X$ and $Y$. Our job, as teachers, is to convince them that statistics will help them address these substantive questions and to persuade them that without knowing how to address these questions, they are doomed to naively accept everything they read.

One particularly effective way to achieve this goal is to have the pedagogy of applied statistics courses simulate the practice of applied research. What do applied researchers do when they conduct their own research? Beginning with a substantive question or hypothesis, they obtain relevant data, either through primary data collection or by accessing an existing database. They then posit a statistical model whose parameters represent critical phenomena, or relationships, reflected in the research question or hypothesis. After selecting relevant statistical analyses, they implement them on the computer, fitting models, estimating parameters, examining diagnostics, as appropriate. In the end, they write up their findings for the research community, describing the entire enterprise from research questions and statistical results to implications for future research and they submit papers to journals or colleagues for peer review.

This process -- or any other conceptualization of the research process -- provides a wonderful framework for teaching applied statistics. We have yet to find a statistical concept or technique that cannot be taught in this way, in the context of a real research question that can be addressed with real data. By adopting this approach, the teaching process itself becomes a model of how research ought to be conducted.

Figure 1 illustrates how we use this paradigm to teach log-linear modeling. We offer it as an example, not the "last word" in pedagogy. We use a similar strategy, with the same data, to introduce classical contingency table analysis in introductory classes; indeed, we use this paradigm with all types of statistical content and all levels of student.

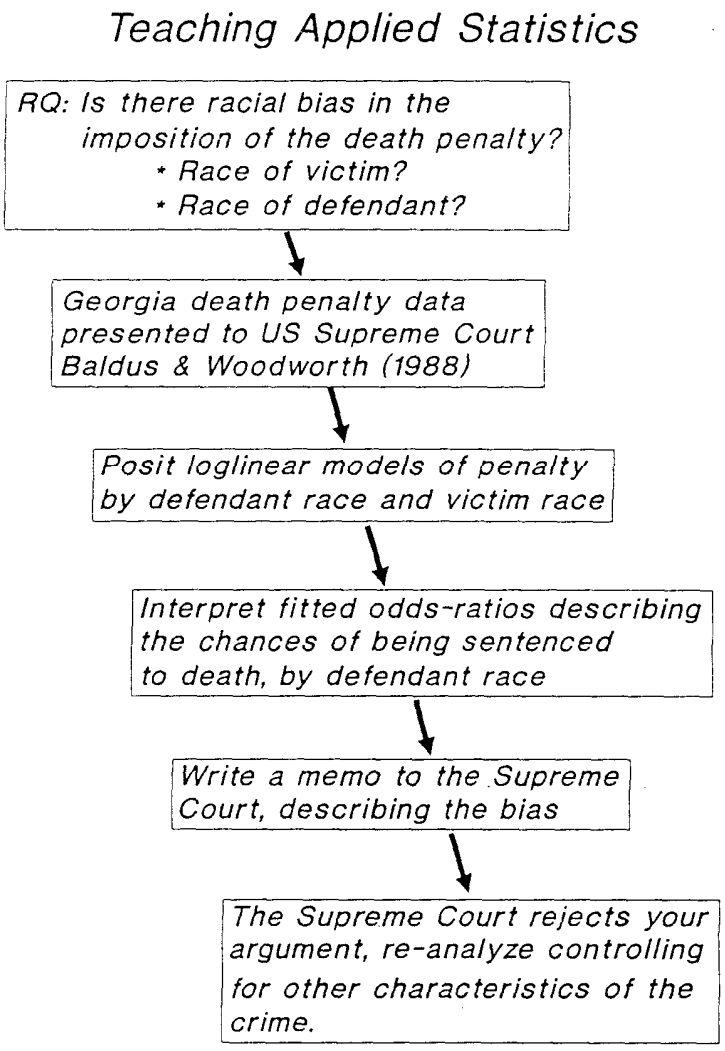


**Figure 1**

We always begin by introducing the *substantive* topic, not the statistical one. In this case, a black man named McClesky was sentenced to death in 1978 for killing a white policeman in Georgia. Defense lawyers appealed the case all the way to the U. S. Supreme Court, arguing that McClesky received the death penalty for reasons of race. We ask the class whether there might have been racial bias in the imposition of the death penalty in Georgia during and prior to the McClesky verdict and how they might go about detecting it. This topic never fails to engage students, and it often leads to heated and memorable argument -- particularly after the data analyses are conducted.

Contextual materials spark interest further. We provide students with newspaper articles, research reports, papers from scholarly journals, and in this case, transcripts of the Supreme Court testimony and opinions. We never reveal the findings in advance; they evolve over time as they do in real data analysis. Because lawyers for McClesky argued that there was a disparity in death-penalty sentencing in Georgia based on the race of the victim and, to a lesser extent, the race of the defendant, the question the class addresses is whether the data support this claim [2]. The data we used were published by *Chance* magazine in 1988.

Prior to class, we conduct a sequence of data analyses and prepare handouts that display the output (sometimes with annotations). During class, we distribute the handouts to students and display them on an overhead projector at pedagogically appropriate moments. The particular analyses vary, of course, but they typically include: (1) *exploratory analyses*, including graphical displays and descriptive statistics; (2) more focused *confirmatory analyses* intended to answer the research question; (3) *diagnostic analyses* that reveal model deficiencies and failure of assumptions; and (4) *follow-up analyses* rectifying any problems that arise.

Class presentations simulate the research environment as closely as possible. We spend most of the time discussing why we conducted the analyses we did and interpreting the analytic results. Student participation is encouraged, and we solicit suggestions for alternative analyses. When we have been prescient, we have supplementary handouts ready; when students offer a new suggestion, we prepare handouts (or overheads) for the next class. Underlying assumptions are studied seriously, especially when we analyze data sets selected because of their severe violations. We outline appropriate sensitivity analyses and, when possible, present fix-up strategies. With a helpful media laboratory, instructors can even conduct analyses interactively in front of the class.

The Georgia death-penalty data handouts begin with exploratory analyses, including univariate summaries of each variable and graphical displays of the relationships between penalty awarded (the conceptual "outcome") and the race of the victim and of the defendant (the two "predictors"). They then contain the results of fitting a taxonomy of competing hierarchical log-linear models each potentially capable of answering the research question. We always follow up with details on selected models -- residual plots, diagnostics, parameter estimates, standard errors, etc. -- so that assumptions can be checked and, if there is one, a final fitted model can be interpreted.

Class discussion focuses on the reasons why specific models were included in the taxonomy and which model, if any, could be considered "best-fitting." Discussion and debate among students reveals the competing constraints of parsimony and fit, substance and statistics. The pedagogic process can sometimes be circuitous, but is always profitable because students learn much more than the technique of "log-linear modeling" -- they learn how to use log-linear models to do research.

In the Georgia death-penalty data, the major finding -- summarized as odds-ratios obtained from parameters estimated in the "best-fitting" model -- is that a defendant was about

nine times more likely to be sentenced to death if s/he killed a white person rather than a black person. The effect was statistically significant at conventional levels and formed the statistical basis of the McClesky Appeal. The evidence of bias revealed in analysis leads to energetic discussion and permits us to ask the class to consider how they might report the findings authoritatively to a non-statistical audience (such as the Supreme Court). As an assignment, students might write a brief to the Supreme Court with statistical evidence attached. By submitting their work for grading and feedback (on the writing and on the content), students get exposure to the "peer review" process. They then read the Supreme Court decision in which Justice Powell wrote that "this study does not demonstrate a constitutionally significant risk of racial bias affecting the Georgia capital sentencing process" and in which Justice Brennan disagreed, writing about the role that statistical evidence plays in the courts [7].

This research-paradigm pedagogic approach allows the student to assume the role of researcher, exploring data that address a real research question. Class examples and homework exercises become "trial" runs in which students grapple with problems and anomalies that they will inevitably encounter in their own work. Real data sets and the research-paradigm pedagogy bring students close to an actual research experience -- warts and all.

We use this research-paradigm approach as a vehicle for teaching all types of statistical techniques. In introductory classes, for example, we acquaint students with univariate descriptive statistics by examining the distribution of governor's salaries across the fifty states (take a look at it; the variation will astound you!). We introduce simple correlation by looking at Cyril Burt's "data" on the IQs of identical twins reared apart. And we develop classical contingency table analysis through simpler views of the Georgia death penalty data.

The research-paradigm approach has at least four advantages over traditional pedagogic strategies. First, it reduces anxiety and empowers students because their initial energies focus on substance, not statistics. This allows easier access to the material, and once it is accessed, the students become engaged and motivated. Second, the realistic learning context shows students that statistical methods are relevant to their own work. The data sets themselves are memorable, and often they become the mnemonics for recalling techniques. Third, differing interpretations and misinterpretations allow the instructor to address a broad range of methodological issues dealing with research design, measurement and analysis. The student begins to realize not that one can "lie with statistics" but that it is much easier to lie *without* them. Fourth, the instructor can focus on the *why* of data analysis, not just the *how*. Extensive use of the computer frees up class time for understanding and interpretation -- it is not enough to learn how to do the computations, it is what the numbers mean that is important. Students begin to realize that the computer's output is only as the good as the instructions it has been given and that research findings can only make a difference if they can be communicated to others.

## Which Real Data are the Best, Pedagogically-Speaking?

Not all real data sets are equally effective vehicles for teaching applied statistics. In this section, we discuss eight attributes that we believe enhance a data set's instructional suitability. We have found that the best data sets come in raw form, are authentic, include background information, have case-identifying information, are intrinsically interesting or relevant, are topical or controversial, offer substantive learning, and lend themselves to a variety of statistical analyses.

### The importance of using raw data

Of these eight criteria, the most important is that the data be in raw form, not summarized using sufficient statistics. Rich information is lost when raw data are replaced by means and covariance matrices. Students further from data are less able to practice "real-life" data-management skills.

When raw data are available, students can adopt the data-analytic approach preferred by many practicing statisticians, be it the exploratory approach advocated by Tukey [55] or the initial data examination approach advocated by Chatfield [8]. This allows students to look for high-leverage cases, non-linearity, heteroscedasticity, and other problems that all too often arise in real data. For example, when missing values lead to data loss, instructors can introduce notions of data-imputation and sensitivity analysis [31]. Analyzing summary statistics may seduce students into believing that such problems do not exist, or if they do, that they are of little consequence.

### Authenticity

A "real" data set must be authentic; it must consist of real measurements taken on a real sample of cases. Attaching life-like variable names to artificial data will not do. Consider the following exercise from Hays [23]:

> *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 dexterity and muscular coordination measured? From what population was the sample drawn? Is the sample homogeneous with respect to age, physical development being a factor that might influence general muscular coordination and therefore the relationship between coordination and finger dexterity? Why is the investigator only interested in a linear relationship? Students easily see through the artifice of "life-like" data. As a result, they may not ask questions like those raised above, because it is clear that the data were not actually "collected." Yet these are precisely the questions we want thoughtful students to raise when reviewing other peoples' research and when conducting their own.

### Background information

Each data set should be accompanied by sufficient background information on the purpose and design of the research, the source of the data, measurement techniques, variable definitions and so on. This information enables the student to assume the role of researcher. If the data come from a published paper or published tabulations, students should be given access to the original document. As Cobb [11] 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).

### Case identifiers

When available, one particularly helpful piece of background information is the case identifier. Case identifiers allow students to use their own knowledge about the specific cases in their data analyses, thereby enriching the exercise. The data sets we use, for example, often include state, school district and school identifiers, all of which have meaning for students. Case identifiers are particularly helpful when detecting outliers and high-leverage observations. Students analyzing data on the citation frequencies of prominent psychologists, for example, can use their background knowledge to understand why Sigmund Freud might be an outlier [21].

### Interest and Relevance

Many statistics texts are brimming with real data, but on topics of no interest to students in a wide range of disciplines. Snedecor and Cochran [52] present data on the calcium concentration in turnip greens (p. 239) and the average daily weight gain of swine (p. 303). Draper and Smith [17] provide data on the viscosity of filled and plasticized elastomer compounds (p. 228) and on the effects of temperature on the growth rates of ice crystals (p. 66). "Classic" data sets, such as Fisher's iris data [20] and Brownlee's stack loss data [5] also fail to inspire most students today.

Intrinsic interest is obviously in the eye of the beholder, but we find it helpful to use data from the students' disciplines. For example, the annual salary survey conducted by the American Association of University Professors (published annually in *Academe*), includes data of interest to most of the students we teach: 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. (Full citations for these sources are given in the annotated bibliography at the end of this volume.)

### Topicality and controversy

Topicality and controversy can help motivate students. The Georgia death-penalty data always rouses interest in our students even though it has nothing to do with education, the subject they are studying. We have also found that Powell and Steelman's [38] analysis of the relationship between state SAT scores and the percent of students taking the test sparks interest, particularly when we provide newspaper accounts of Department of Education "wall charts" that rank states according to these scores and critiques of state comparisons of SAT scores [42, 56, 57, 58].

Older data sets *can* motivate students, especially if the topic is controversial enough. Burt's data on the IQs of identical twins provide a wonderful pedagogic vehicle [26], when analyzed in the context of Burt's views on the nature/nurture debate and Dorfman's [16] and Kamin's [28] evidence that Burt falsified data to support the nature argument. Analyzing controversial data sets show students not just the statistical techniques, but also how the techniques can support or undermine a hypothesis.

And very old data sets can sometimes be as interesting as up-to-the-minute ones. The early volumes of journals such as *Child Development*, *Journal of Educational Psychology*, and *Journal of Genetic Psychology*, are useful sources. Although they do not always address fascinating research questions, students are interested in seeing how researchers used to analyze

data. This provides an opportunity to compare findings using "modern methods" to those obtained under the older and simpler methods 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 analyze a real-world data set, they often "accidentally" learn something of substance, discovering just how useful statistics can be. The substantive learning involved does not have to be on a grand scale, but it must be real. One of our most popular data sets, for example, comes from a local magazine. Every few years, *Boston* magazine surveys local school districts and publishes district by district data on per-pupil expenditures, teacher salaries, student demographics and so on. *The Boston Globe* regularly publishes similar data sets. Students analyze these data and learn 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 perceived to be "better" than others.

### Possibility of multiple analyses

Empirical researchers often use more than one type of analysis to address their research questions; so, too, should instructors working with real-world data. When a data set is used in multiple analyses, students learn that research questions can be answered in many ways. Repeat analyses often come from showing that a key assumption may have been violated in an earlier presentation; revisiting the data facilitates the discussion of fix-ups and the notion of sensitivity analysis. This allows the teacher to stress that, in the real world, not all analyses of the same data will agree. The investigator must consider the nature of the question, the structure of the data and the suitability of the analytic method used. These are critical lessons for budding researchers.

Subsequent analyses of the same data allows a hierarchy of more complex questions to be answered -- perhaps questions suggested by earlier analyses. We might introduce simple linear regression, for example, by examining the relationship between state SAT scores and the percentage of students taking the test. Noting that the percentage-of-students-taking-the-test is nonlinearly related to the outcome introduces the notion of a polynomial regression model. Several weeks later, we might return to the data and use influence statistics to identify Alaska as an outlying and high leverage observation. The analytic sequence can be spread over many weeks, with additional findings being revealed over time.

No experience reinforces the importance of multiple analyses as much as the discovery of previously unknown findings. For example, in a course on categorical data analysis, we discuss a paper by Scarcella [43], who used classical techniques to examine the relationship between language background, language proficiency and an individual's choice of writing device (repetition, paraphrase, explanation). When students re-analyze the data using log-linear modeling, they discover a previously unnoticed effect -- that it is language proficiency, not language background, that predicts writing device.

## What Are the Drawbacks to Using Real-World Data?

Using real data and the research-process paradigm to teach applied statistics is not without shortcomings. Although we have not found that this approach uses more *class* time than traditional lecture methods, it does take more *preparation* time. Data sets tend to be small and lack statistical power. Aggregate data sets and data collected on self-selected samples are often the best we can do. In-class testing is more difficult. Vagaries of using the computer can overwhelm all other considerations. Below we discuss each of these problems and we offer some remedies for overcoming them.

### The workload of finding real data sets

A major motivation for using artificial data is that an instructor can readily create data sets with the requisite characteristics. Dayton [14], for example, showed how to construct data sets with suppressor variables. Searle and Firey [45] suggested that instructors could reduce student plagiarism by generating dozens of data sets and giving each student 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.

We have no doubt that using real data sets increases the amount of time required to prepare classes, homework and exams. To identify a single data set which permits illustration of a specific statistical technique, an instructor must spend hours analyzing different data sets, some of which do not support interesting findings, others of which present analytic problems out of line with the curriculum. This is especially true when developing materials for elementary courses, in which students are still learning basic skills, not how to cope with non-standard problems.

For these reasons we provide, in the annotated bibliography at the end of this volume, an extensive list of references to hundreds of data sets. Although instructors must still examine the data sets to determine which are best suited for teaching a specific concept with the particular types of students in the class, we hope that this annotated bibliography will facilitate the planning process.

### Small data sets and statistical power

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

The problem is that small data sets falsely represent the effect sizes typically found in the real world. Because null findings tend to be dull, we use data sets with large enough effect sizes that yield "statistically significant" results despite the sample size. Although we, as instructors, know that such effect sizes are rare in practice [30, Ch. 8], the students do not see much evidence of this in their class problems or their homework. Thus, when they read journal articles that report $R^2$ statistics with magnitudes of 20%, many students conclude that such effect sizes are small and rare -- and they are, relative to *their* in-class experience.

This problem is not unique to real data sets; most artificial data sets used by textbook authors and college instructors are also small. But real data sets *appear* representative of the larger class of statistical problems arising in the real world. Because we see little means of eliminating this problem, we 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.

### Aggregate data and self-selected samples

Aggregate data or data on self-selected samples, such as the SAT state data set, are among the easiest data sets to access. While some variables in these data sets are measured at the aggregate level -- college tuition, student/faculty ratio, number of students -- most are aggregates of lower levels of data, creating a host of problems.

The question is whether the gains are worth the drawbacks, and in most instances, we believe they are. Aggregate data sets are among the most readily available, intrinsically interesting data sets we use. The observations contained in such data sets often have meaningful identifiers -- names of towns, cities, counties, school districts, states or countries -- enabling students to become more intimately acquainted with each data point. And in more advanced classes, we return to these data sets and illustrate 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 unless computer terminals or personal computers are available for each student in the classroom. We use multiple homework assignments and take-home exams in place of in-class testing. In both cases, data sets are made available to students on the computer for analysis and students must write an account of their work in the form of a journal article or research paper.

Teachers who prefer in-class examinations might conduct a series of analyses on the computer in advance of the examination and distribute computer output to the students for interpretation. Contextual material, research questions and so forth could be provided during (or in advance of) the exam. In doing so, though, note that the students are not choosing the analyses to be conducted -- they are simply interpreting your output -- and thus such an exam may not be testing all the skills you have taught during the semester.

### The use and abuse of the computer

Our pedagogic approach relies heavily on the computer. This has advantages and disadvantages. Using the computer to shoulder the computational burden frees up time for thoughtful class activity. When students conduct their own analyses, however, computers can produce the reverse effect unless activities are carefully monitored. We have found that some students become so engrossed in "hacking" that their conceptual thinking suffers. Their attention and creative energies become devoted almost entirely to writing code, debugging and executing programs. The mindless, mechanical production of endless computer output becomes their sole objective [29, 44]. Instructors can avoid these problems, by crafting carefully-worded assignments and examinations that emphasize the importance of non-programming activities, including framing research questions carefully, choosing appropriate statistical models and methods of estimation, interpreting parameter estimates, and communicating findings.

## Postscript

Real-world data and an empirical research paradigm can be an applied statistics instructor's 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. Perhaps the only way of discovering the advantages of authentic data sets is to try one in your classes. We think you will see the difference.

## References


1. Aiken, L. S., West, S. G., Sechrest, L., & Reno, R. R. (1990). Graduate training in statistics, methodology, and measurement in Psychology. *American Psychologist*, **45**, 721-734.
2. Baldus, D., Pulaski, C., and Woodworth, G. (1983) Comparative review of death sentences: An empirical study of the Georgia experience, *Journal of Criminal Law and Criminology*, **74**, 661-753.
3. Brightman, H., & Broida, M. (1975). On problem solving, motivation and statistics. *The American Statistician*, **29**, 164-166.
4. Brogan, D. R. (1980). A program of teaching and consultation in research methods and statistics for graduate students in nursing. *The American Statistician*, **34**(1), 26-33.
5. Brownlee, K. A. (1965). *Statistical Theory and Methodology in Science and Engineering*. New York: John Wiley.
6. Carmer, S. G., & Cady, F. B. (1969). Computerized data generation for teaching statistics. *The American Statistician*, **23**, 33-35.
7. Chance (1988). Supreme Court Ruling on Death Penalty, *Chance: New Directions for Statistics and Computing*, **1**, 7-8.
8. Chatfield, C. (1985). The initial examination of data. *Journal of the Royal Statistical Society, Series A*, **148**, 214-253.
9. Chervany, N. L., Collier, R. O. Jr., Fienberg, S. E., Johnson, P. E. & Neter, J. (1977). A framework for the development of measurement instruments for evaluating the introductory statistics course. *The American Statistician*, **31**, 17-33.
10. Chottiner, S. (1991). Using real (intimate) data to teach applied statistics. *American Statistician*, **45**, 169.
11. Cobb, G. W. (1987). Introductory textbooks: A framework for evaluation. *Journal of the American Statistical Association*, **82**, 321-339.
12. Cockerill, R., & Fried, B. (1991). Increasing public awareness of statistics as a science and a preofession -- Reinforcing the message in universities. *The American Statistician*, **45**, 174-178.
13. Dallal, G. E. (1990). Statistical computing packages: Dare we abandon their teaching to others? *The American Statistician*, **44**, 265-269.
14. Dayton, C. M. (1972). A method for constructing data which illustrate a suppressor variable. *The American Statistician*, **26**, 36.

15. Dayton, C. M. (1988). Integrating analyses of data bases into statistical instruction. Paper presented at the annual meeting of the *American Educational Research Association*, New Orleans, April.
16. Dorfman, D. D. (1978). The Cyril Burt Question: New Findings. *Science*, **201**, 1177-1186.
17. Draper, N. R., & Smith, H. (1981). *Applied Regression Analysis*, 2nd edition. New York: John Wiley.
18. DuMouchel, W. H. (1979). Comment on Thisted. *The American Statistician*, **33**, 30-31.
19. Edwards, B. (1959). Constructing simple correlation problems with predetermined answers. *The American Statistician*, **12**, 25-27.
20. Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, 179-188.
21. Gordon, N. J., Nucci, L. P., West, C. K., Hoerr, W. A., Vguroglu, M., Vukosavich, P., & Tsai, S. L. (1984). Productivity and citations of educational research: Using educational psychology as the data base. *Educational Researcher*, **13**, 14-20.
22. Halperin, S. (1988). Real and contrived examples in statistics instruction. Paper presented at the annual meeting of the *American Educational Research Association*, New Orleans, April.
23. Hays, W. L. (1981). *Statistics*, 3rd edition. New York: Holt, Rinehart and Winston.
24. Herzberg, P. A. (1991). Comment on Singer & Willett. *American Statistician*, **45**(2), 169.
25. Hogg, R. V. (1972). On statistical education. *The American Statistician*, **26**, 8-11.
26. Jensen, A. R. (1974). Kinship correlations reported by Sir Cyril Burt. *Behavioral Genetics*, **4**, 1-28.
27. Joiner, B. L. (1988). Let's change how we teach statistics. *Chance: New Directions for Statistics and Computing*, **1**(1), 53-54.
28. Kamin, L. J. (1974). The Science and Politics of IQ. Potomac, MD: Erlbaum.
29. Levin, J. R. (1991). Teaching statistics conceptually: The case *against* technology. In J. P. Stevens (Chairperson), *On the teaching of applied statistics*. Symposium conducted at the annual meeting of the *American Educational Research Association*, April, Chicago.
30. Light, R. J., Singer, J. D., & Willett, J. B. (1990). *By Design: Planning Better Research in Higher Education*. Cambridge, MA: Harvard University Press.
31. Little, R. J.A., and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.
32. McCall, R. B. (1975). *Fundamental Statistics for Psychology*, 2nd edition. New York: Harcourt Brace Jovanovich.
33. Minton, P. D. (1983). The visibility of statistics as a discipline. *The American Statistician*, **37**, 284-289.
34. Minton, P. D. & Freund, R. J. (1977). Organization for the conduct of statistical activities in colleges and universities. *The American Statistician*, **31**, 113-117.
35. Moore, T. L. & Roberts, R. (1989). Statistics at liberal arts colleges. *The American Statistician*, **43**, 80-85.
36. Mosteller, F. M. (1988). Broadening the scope of statistics and statistics education. *The American Statistician*, **42**, 93-99.
37. Pedhazur, E. J. (1981). *Multiple Regression in Behavioral Research*, 2nd edition. New York: Holt, Rinehart and Winston.

38. Powell, B. & Steelman, L. C. (1984). Variations in state SAT performance: Meaningful or misleading? *Harvard Educational Review*, **54**, 389-412.
39. 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.
40. Read, K. L. Q. (1985). ANOVA problems with simple numbers. *The American Statistician*, **39**, 107-111.
41. Read, K. L. Q., & Riley, I. S. (1983). Statistics problems with simple numbers. *The American Statistician*, **37**, 229-231.
42. 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.
43. 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.
44. Searle, S. R. (1989). Statistical computing packages: Some words of caution. *The American Statistician*, **43**, 189-190.
45. Searle, S. R., & Firey, P. A. (1980). Computer generation of data sets for homework exercises in simple regression. *The American Statistician*, **34**, 51-54.
46. Sechrest, L. (1987). Data quality: The state of our journals. In L. S. Aiken & S. G. West (Chairpersons), *Adequacy of methodological and quantitative training: Perspectives of the disciplines*. Symposium conducted at the annual conference of the *American Psychological Association*, New York, April.
47. Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform, *Harvard Educational Review*, **57**, 1-22.
48. Singer, J. D., & Willett, J. B. (1988). Opening up the black box of recipe statistics: Putting the data back into data analysis. Paper presented at the annual meeting of the *American Educational Research Association*, New Orleans, April.
49. Singer, J. D., & Willett, J. B. (1990). Improving the teaching of applied statistics: Putting the data back into data analysis. *The American Statistician*, **44**(3), 223-230.
50. Singer, J. D., & Willett, J. B. (1991a). Providing a statistical model: Teaching applied statistics the way that "statistics" is applied. In J. P. Stevens (Chairperson), *On the teaching of applied statistics*. Symposium conducted at the annual meeting of the *American Educational Research Association*, April, Chicago.
51. Singer, J. D., & Willett, J. B. (1991b). Reply to Herzberg and Chottiner. *The American Statistician*, **45**(2), 170.
52. Snedecor, G. W., & Cochran, W. G. (1980). *Statistical Methods*, 6th edition. Ames, Iowa: Iowa State Press.
53. Stevens, J. P. (1990). On the teaching of applied statistics and applied statistics textbooks. Paper presented at the annual meeting of the *American Educational Research Association*, Boston, April.
54. Thisted, R. A. (1979). Teaching statistical computing using computer packages. *The American Statistician*, **33**(1), 27-30.
55. Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
56. Wainer, H. (1986a). Five pitfalls encountered while trying to compare states on their SAT scores. *Journal of Educational Measurement*, **23**, 69-81.
57. Wainer, H. ed. (1986b). *Drawing Inferences from Self-Selected Samples*. New York: Springer-Verlag.

58. Wainer, H., Holland, P. W., Swinton, S., & Wang, M. H. (1985). On "State Education Statistics". *Journal of Educational Statistics*, **10**, 293-325.
59. Winer, B. J. (1971). *Statistical Principles in Experimental Design*, 2nd edition. New York: McGraw Hill.

---

**John B. Willett and Judith D. Singer** are Associate Professors at the Harvard University Graduate School of Education specializing in quantitative methods. Collaborators since 1985, they have written and presented dozens of talks, workshops, and papers on the application of statistical methods in education and the social sciences. Together with other colleagues, they have written two books, *By Design?* and *Who Will Teach?*, both published by Harvard University Press. Their current research interests center on applications of survival analysis in the social sciences. They teach a five-semester graduate sequence in quantitative methods using the research process approach described in this paper. They have recently received the Raymond B. Cattell Award and the Palmer O. Johnson Award from the American Educational Research Association and an NSF Visiting Fellowship from the ASA.