# Advice to Mathematics Teachers on Evaluating Introductory Statistics Textbooks

**Robert W. Hayden**
*Plymouth State College*

## Introduction

Here's a little quiz for you. There are no right or wrong answers, and I won't even ask you to tally up a score at the end. It won't tell you anything really important, like whether you are compatible with the person you've been living with for 20 years, but your answers may help you to follow this essay.

> How much mathematical training should a person have in order to *teach* a non-remedial, first-year mathematics course?

> How much statistical training should a person have in order to *teach* a first-year statistics course?

Perhaps it's just paranoia, but many statisticians detect a discipline-centric bias in answers to such questions. As evidence, try to find a discipline in which more college students are taught by people with no degrees in the subject than they are in statistics. I do not raise this point to make you feel bad. The fact that you are reading this volume shows that you feel a need to learn more about the subject you are teaching, and statisticians welcome you with open arms. My real point is a corollary:

> How much mathematical training should a person have in order to *write* a textbook for a non-remedial, first-year mathematics course?

> How much statistical training should a person have in order to *write* a textbook for a first-year statistics course?

## Why do I need to read an essay on how to choose a textbook?

The following comments were made by "a mathematician who . . . got slung into the statistical pool by way of a swimming lesson" [7].

> While there are good and bad mathematics texts, I've never seen a pure math textbook with actual errors throughout. Maybe the occasional lapse from Bourbakiste rigor, or a folktale about Galois passed off as history—but that's about the worst. (This is not to say that some of them are not very bad indeed as textbooks.) However, in statistics, I have learned to treat the very factuality of textbooks with suspicion. Most of them are fine in this regard, and the good ones are pedagogically better than almost anything in pure mathematics. But in the three years in which I have taught stats, from three textbooks, I have had more occasions to tell students "The book says this

. . . please cross it out, and write in a correction, because the book is wrong" than in dozens of math courses. I do not do this on matters of taste, but only when there is an actual error.

The author goes on to wonder why this is so. Perhaps the reason is that introductory statistics textbooks are often written by people with little or no training in statistics. This is very different from the situation in mathematics, at least at the college level. You would expect the author of a calculus text to have a Ph.D. in mathematics and errors to be limited to typos. Failure of the author to grasp the content covered in the textbook would be most unusual. (But then, consider what calculus books would be like if most students took calculus outside the Mathematics Department.) Unfortunately, gross errors are not unusual in elementary statistics textbooks. What is even worse, these books sell. Indeed, they are often among the best selling texts! The reason textbooks written by people who do not know much about statistics sell is that many teachers of statistics do not know much about statistics either, and these two groups suffer a fatal attraction for one another. This explains why errors can go undetected for years: real statisticians do not read these books.

For example, consider this problem on the chi-squared goodness of fit test taken from a popular introductory statistics textbook of nearly twenty years ago.

> Jimmy Nut Company advertises that their nut mix contains 40% cashews, 15% brazil nuts, 20% almonds, and only 25% peanuts. The truth in advertising investigators took a random sample (of size 20 lb) of the nut mix and found the distribution to be as follows:
>
> | Cashews | Brazil Nuts | Almonds | Peanuts |
> |---------|-------------|---------|---------|
> | 6 lb    | 3 lb        | 5 lb    | 6 lb    |
>
> At the 0.01 level of significance, is the claim made by Jimmy Nuts true?

A student using this textbook observed that the calculated value of chi-squared for this problem, and thus its significance, could be made to take on any value whatever by a suitable choice of units. For example, converting the weights to metric tons makes Jimmy an honest man, while converting them to nanograms produces the largest chi-squared value ever seen by man or beast. The student's teacher posted a query on the internet asking for an explanation of this paradox. The quote you read above about errors in statistics textbooks was a part of the discussion.

The problem here is that the chi-squared goodness of fit test applies only to categorical (discrete) data. It compares the actually observed *counts* in each category to the counts we would expect if the hypothesis being tested were true. Counts are unitless, and the apparent paradox only arises if you misapply the technique to measurement (continuous) data. You might think that such a gross error would be an outlier, but there is another, similar problem in the same section. You might also expect such an error to be quickly spotted and removed, but the same two problems still appeared in subsequent editions of the same text, published years later. And, the existence of further editions suggests that this text has been successful in the marketplace.

## Look for an author who knows more than you do

Far from being an isolated error, Jimmy and his nut company is a symptom of a much deeper problem. Statistics is fundamentally and primarily concerned with analyzing real data. Yet many introductory statistics textbooks continue to be algorithmic cookbooks that mainly offer practice in plugging unreal numbers into formulas. Of course, if your goal is to train students to carry out the computations of statistics (without statistical software), then it really does not matter if the numbers you give them are made up or real. And if the context of the data is also a fiction, then it does not matter whether it is realistic, improbable, or absurd. However, if you are trying to teach the *concepts* of statistics and to help students learn to apply those concepts to real-world problems, asking students to apply a technique that works only for categorical data to measurement data is a mortal sin, a special case of the sin of teaching students to mindlessly apply computational algorithms to numbers. (Note the appropriate word is "numbers," not "data"!) Since computers can do this much more cheaply, and since this is not a task worth doing anyway, such training is a negative contribution to the human enterprise. The problem here is not one of isolated errors, but of a profound misunderstanding of the nature of statistics as a discipline.

While not all of the failings of such texts are as simple and clear cut as Jimmy and his nut company, the failings are widespread. The long-term solution to amateur statistics textbooks is adequate training for people (including textbook authors!) teaching statistics. In the meantime, we have to face the fact that many people who do not already have this training will be called upon to teach statistics. The goal of this paper is to help people "slung into the statistics pool" select a textbook that will be more like a life jacket than ballast.

So, let us suppose that your department chair told you that you will be teaching an introductory statistics class

next fall. It's often the case that you have been selected not because of your extensive training in statistics, but because you are at the bottom of the pecking order. This assignment algorithm sends a clear message about how your superiors view statistics, so you seek counsel elsewhere. Since you studied statistics long ago or not at all, you may not notice an error such as applying chi-squared to weights. How can you select a textbook that will lead you and your students down the path of truth rather than error? Statistician Paul Velleman [17] has said:

> If I found myself teaching a law course, I would certainly choose a text by the best lawyer I could find, not one by a friend who also was stuck in the same situation.

Although mathematicians tend to gravitate toward texts written by other mathematicians, such textbooks will reinforce both your strengths and your weaknesses. I suggest that you find a text that is strong in the areas where you are weak—probably in the actual use of statistics with real-world data. You can easily fill in any weaknesses in the mechanics of getting the calculations done. There are no sure ways of finding such a complementary textbook, but here are some pointers. They admit of exceptions, but they are a start.

First, it would be wonderful if a group of qualified statisticians, with an interest in good pedagogy, got together and evaluated the current crop of textbooks and offered their recommendations to the world. As a matter of fact, this has already happened. Such a group recommended textbooks for the initial offering of the Advanced Placement Test in Statistics [5, 6], and their recommendations are appended to this paper. It is interesting to note how short the list is! Although most "missing" textbooks are missing for good reasons, a few good books may be missing only because they were published after the list went to press, or because they did not match the AP syllabus.

You may also wish to read published textbook reviews. Unfortunately, textbook reviews vary in quality by almost as much as do statistics textbooks. Too many are just a routine recitation of the table of contents. Certainly you would want to look at reviews in *statistics* journals. One of these is George Cobb's classic, "Introductory Textbooks: A Framework for Evaluation (A Comparison of 16 Books)" [4]. I asked myself what made Cobb's paper a classic, and why would anyone feel a need to update it? Beyond its style and wit, the paper is a classic because it provides a framework for evaluating textbooks, rather than merely carrying out an evaluation. I think it is seen as dated only because it applied that framework to sixteen textbooks, most of which

are now either out of print or survive in much later editions. So, in this paper I will try to concentrate on the evaluation procedure, citing a few textbooks as examples, but not trying to offer thorough reviews of any. I will try not to overlap too much with the content of Cobb's paper, which is still essential reading. Finally, I will aim my comments not at Cobb's audience (professional statisticians), but rather at people who usually teach mathematics rather than statistics.

An evaluation of a textbook might start with an evaluation of the author's statistical expertise. The danger, of course, is that you may feel more of a sense of kinship with the fellow sufferer than with the expert! However, the less you know about statistics, the more essential it is that the textbook you choose be a reliable guide. The textbook (or its advertisements) may give an author profile. Does it concentrate on how qualified the author is, or on what a great person he or she is? Does it mention any degrees or training in statistics? Lacking that, the text may at least give an affiliation for the author. On the whole, a member of a Statistics Department at a large university is very likely to have solid credentials in statistics. Another indicator is whether the author is a member of the American Statistical Association. If you are an ASA member (and you should be), you can access the member list at `www.amstat.org`.

In addition to questions of how much training authors have, there is a question of the *quality* of the training. Did they study statistics recently in a leading statistics department? Or did they take such a course 20 years ago and keep up to date ever since? Or did they take a course 20 years ago that was itself already 20 years out of date, and learn nothing new since? The ASA membership list *sometimes* gives degrees and dates. The *Current Index to Statistics* can tell you if an author has published beyond the introductory statistics level. Again, many fine teachers do not publish much, but publication does suggest that the statistics community thinks the author's ideas are worthwhile.

## Look for data in the exposition

After evaluating the expertise of a textbook's author, turn to the text itself. A statistics textbook written by a working statistician is likely to reflect familiarity with current practice in statistics. Certainly your course should do so. For that reason, it may be helpful to have some benchmark items to look for. Fortunately, we have a landmark to guide us here: the revolution brought about in statistics during the 1960's by John Tukey and others. One part of this revolution was the development of what came to be known as exploratory data analysis (or EDA) [See 15 or

16 for background on EDA]. Often you can get an idea of up-to-dateness by seeing how a text handles EDA.

Some people characterize modern data-analytic statistics in terms of the four R's:

- residuals,
- robustness,
- resistance,
- re-expression.

Checking for these (or synonyms such as "transformations" for "re-expression") in the index can give an idea of the presence and pervasiveness of modern views in a text. To illustrate, Table 1 gives results for some textbooks on hand that show that even this simple technique decisively discriminates between different classes of textbooks. The books included are *not* a random sample of those available! The first four are the books used by the author in the past decade. However, the choice is not entirely idiosyncratic, as these are also (independently) on the recommended list [6] for the Advanced Placement Test in Statistics. Though they are not the only good textbooks available, they represent the kind of textbook I think you should be seeking. In contrast, the last row of Table 1 represents a composite of several textbooks I would *not* recommend. These lesser texts are here because they have sold well and been through many editions, and often end up on 'short lists' when mathematics teachers choose textbooks for statistics.

In constructing the table, I used the editions available to me at the time; questions of how these books have changed over the years will be discussed later. In the meantime, I will use these textbooks to illustrate a variety of points. After the reference number, the table gives the number of pages in each book's index. The next three numbers give the numbers of pages cited in the index for the EDA topics listed at the top of the table—robustness/resistance, transformations (or re-expressions), and residuals. Note that a book with a relatively short index might be expected to have fewer references to *any* topic.

I also counted index references to "outliers," since "resistance" means resistance to outliers. I also counted the

| Ref. | Pages | Rob/Res | Trans. | Resids. | Outliers | Par.BP | Norm |
|------|-------|---------|--------|---------|----------|--------|------|
| [11] | 6 | 6 | 2 | 13 | 24 | 6 | 28.7 |
| [12] | 6 | 21 | 12 | >19 | 16 | 4 | 34.9 |
| [13] | 13 | 13 | 92 | 15 | 19 | 6 | 96.2 |
| [14] | 7 | 6 | 18 | 7 | 13 | 11 | 26.4 |
| [??] | 3-5 | 0-1 | 0 | 0 | 0-1 | 0 | 0-1.4 |

**TABLE 1**
A comparison of textbooks on coverage of EDA topics.

number of sets of parallel boxplots in each book's chapter on Analysis of Variance to see if boxplots were actually used in the context where they are most appropriate. (This also indicates whether the book encourages the reader to LOOK AT THE DATA!) The last column gives the Euclidean norm for the vector whose components are the entries in the previous five columns, which provides a crude "modernity" index for statistics textbooks. The four exemplary texts are fairly similar—certainly in comparison with the texts in the final row. The apparent outlier [13] has an unusually long index and 92 references on transformations.

The counts in the table give a "quick and dirty" rating system meant for rapid screening. Still, you must use it with care. In the "new math" era, many college mathematics textbooks sprouted a "Chapter 0" on sets. This was supposed to integrate the material in the rest of the book, but as sets were never mentioned in the rest of the book, it is not clear how this could happen. Similarly, some of the weaker statistics textbooks have an early section on "current" topics, such as stem-and-leaf displays and boxplots, but then fail to use the tools developed there when they are needed later. For example, a chapter on analysis of variance should certainly include many examples of parallel boxplots, the most appropriate tool for visually comparing the centers of several groups. (This is why these were counted in the table above.) The newer exploratory techniques may also be used to check assumptions underlying an inferential technique. For example, parallel boxplots can also be used to check the ANOVA assumption of equal variances.

In general, make sure a textbook *mentions* assumptions and teaches students to *check* them rather than to *make* them. Cobb [4, p. 329] says he finds "it useful to distinguish *exploratory techniques*, such as stem-and-leaf diagrams and boxplots, from *exploratory attitudes*" such as looking at the data (or residuals), checking for outliers or violations of assumptions, or considering the possibility of transforming the data. "The techniques are relatively unimportant but the attitudes are essential [4, p. 329]."

To illustrate Cobb's distinction, let us take a stroll through specific sections of some of these texts. I will use the sections comparing the means of two groups as an example. This is elementary enough to be in virtually every textbook or course syllabus, but complicated enough that just about everything you would want to take into account in evaluating a text comes up. I strongly suggest that you grab copies of Moore and McCabe [12] and Siegel and Morgan [14], plus any other texts you wish to evaluate, and follow along.

Before undertaking any inferential procedures, an investigator should *look at the data*. Siegel and Morgan get the

prize for doing this. Their chapter on comparing two groups opens with an extensive analysis of several data sets. They point out that comparing the centers of two sets of numbers is much easier if they differ *only* in center. In particular, it is easier if the two data distributions have the same shape and the same variability. They give examples of a variety of situations with similar or dissimilar shapes and variability, and they show how transformations may be used to achieve similar shapes and/or variabilities. An even more extensive discussion of these issues opens their chapter on comparing several groups. These discussions are worth reading no matter what textbook you adopt.

After description comes inference. Older texts introduce "large sample" (based on the normal distribution) and "small sample" (based on the $t$ distribution) techniques. This terminology is a confusing anachronism. Historically, techniques for large samples based on the Central Limit Theorem and the normal distribution were developed first, while the $t$ distribution was developed later (1908) to correct for errors made when the sample standard deviation is used to estimate the population standard deviation. Although such estimation is usual, the errors introduced are greatest for small samples, and so statisticians trained before 1908 tended to see the $t$ methods as "corrections" to familiar techniques rather than a more exact replacement. Among the books discussed here, Moore and McCabe mention, but then dismiss, the "large sample" technique (without using that confusing name) while Siegel and Morgan discuss *only* the "small sample" procedure.

In carrying out inference comparing the means of two independent groups, the computations will be somewhat simpler if we assume that the two samples come from populations with the same variance, and pool data from the two samples to estimate this common variance. This was relevant in the days of hand calculations, but now most statisticians prefer *not* to make this assumption. From our list, only the Moore books [11,12] agree. (If the computations are too arduous, use statistical software. The preferred method has been the default in Minitab for many, many years.)

Whatever procedure you use, it will have been derived under certain hypotheses or "assumptions". Moore and McCabe [12, p. 509] are wonderfully direct about this:

> The results of $t$ procedures are exactly correct only when the population is normally distributed. Real populations are never exactly normal. The usefulness of the $t$ procedures in practice therefore depends on how strongly they are affected by nonnormality.

Because of this, we need to do two things when we make an inference. We need to assess possible violations of the assumptions, and we need to know how much we can get away with. Moore and McCabe probably do best at the latter [see 12, pp. 509–510, 538, 561–563].

The other three exemplary books in Table 1 also do well at assessing violations. The others gathered in the final row have not a single data display among them. They all use the hypothesis test technique that assumes equal variances in the two groups, and two recommend a preliminary $F$-test on the two variances. Moore and McCabe explain [12, pp. 557–558] why the $F$-test is *not* a good idea, and provide a wonderful quote on the subject from George Box.

Now that we have an idea of what should appear in these sections, we can look at how some of the weaker texts get updated. Later editions of weak texts usually tack on some exploratory techniques or terminology but refrain from adopting exploratory attitudes.

For example, a new edition of one text in the last row now has Euclidean norm of 3.3 as a result of some passing references in the index. However, it *still* recommends the preliminary $F$-test when comparing two independent means. The new edition contains a single set of parallel boxplots in the chapter on comparing two groups. The boxplots are of a simplified kind that does not explicitly flag outliers. However, it is quite clear that there is one—a point that falls eight standard deviations above the mean! The author effectively deletes this from the boxplots before analyzing them, but retains it in the data when doing the hypothesis test! (In addition, proper boxplots would have revealed that both groups have more modest low outliers as well, and a histogram of the data for one group that appears hundreds of pages earlier suggests the group is strongly skewed toward low values.) What purpose the displays serve is not made clear to the reader. As used here, they serve no purpose whatever. The many ways in which the data seem to fail to match the underlying assumptions are simply ignored.

A recent edition of another text included in the last row also adds a few passing references in the index and so now has a norm of 3.0. There is virtually no change in the sections on comparing two or more groups of measurements: variances are pooled (with no $F$-test), no data are plotted, and instead of checking the ANOVA assumptions they declare by fiat at the top of the problem set: "In each problem assume that the distributions are normal and have approximately the same population standard deviation." In doing a hypothesis test comparing two groups, the authors refer back to earlier discussions of the roles of $t$ and $z$ in testing hypotheses about the mean of a sample drawn from

a normal distribution. The earlier edition has a reasonable discussion of the fact that the test statistic has a normal distribution when calculated using the population standard deviation, but that replacing it with the sample standard deviation will give different numerical values with a different $t$ distribution. In the later edition, this is replaced with the incredible claim that it is the sampling distribution *of the mean* (or difference in means) that changes shape depending on whether we know the population standard deviation. I view this as an error comparable in seriousness to a calculus textbook that confuses function multiplication with function composition. (On a good note, I heard as this article went to press that an even more recent edition of this text has given Jimmy a well-deserved retirement from his nut company.)

The new edition of another last row text has exactly the same "length" as the old, though there are some improvements in details. The preliminary $F$-test is no longer recommended, and the preferred $t$-test is presented, but there are no data displays. The sections on Analysis of Variance include some displays to illustrate the concepts, but no parallel boxplots, no displays using real data, and no displays to check assumptions.

The need to present attitudes as well as techniques is one of many reasons that a good statistics text will include far more words than a typical mathematics textbook. It will give good advice on *when* to use a $t$-test as well as information on *how* to do a $t$-test. The sections on robustness cited in Moore and McCabe or the sections on visually comparing two sets of data from Siegel and Morgan are good examples of the type of discussions you should look for. Generally speaking, books that are not really teaching statistics are teaching arithmetic. You can scan a text quickly looking for a balance between concepts and calculation. An exposition that consists of little more than worked examples for the student to emulate is a bad sign. While this approach might be appropriate in a mechanical skills course such as remedial algebra, it is out of place in statistics because the skills being practiced have no value. In real life these computations are carried out by a computer, so there is no need to practice them until you get good at them. Of course, there are places where "hand" (i.e., calculator rather than computer) calculations help the student understand what is going on, but we must be realistic about our audience, and realize there is little evidence that carrying out a calculation adds to most people's understanding of the *meaning* of what was calculated.

For similar reasons, formulas should be few and far between. For many students, formulas are not a path to understanding, but a source of difficulty. While the formula

for the standard error of the sample mean may make it obvious *to you* that the error decreases with increasing sample size, this may not be obvious to your students, and plugging numbers into the formula is not likely to help them see that unless they do it in a very structured exercise directed specifically at that goal. Concrete examples or computer simulations are more likely to be convincing for many students. In any case, the focus should be on concepts rather than calculations. This can sometimes be a difficult change for mathematics teachers.

Another key feature of a text is its discussion of data production. For example, does the book discuss the difference between observational studies and experiments? Does it discuss randomization and sample selection? Measurement? These are key ideas in statistics that cannot be reduced to formulas. For just that reason, you need to see if the book's exercises (and your own exams) ask for verbal responses on such concepts as well as numerical results [9]. The free response sections of the Advanced Placement Test in Statistics provide some good examples of conceptual questions. Some of these as well as other sample problems are available at the College Board website.

It is worth noting that among the four exemplary books in our sample there is strong agreement between the outcomes of evaluating content and evaluating authors. All of the authors are at major universities. In 1989, about the time these books first appeared, all of the authors were ASA members. CIS cites three of the authors for more than the text at hand: Moore has 42 citations and McCabe and Siegel 25 each. The total for *all* the authors in the last row is zero.

## Look at the text at three levels

Once you have eliminated the questionable texts, you will find the stack of candidates has shrunken considerably. Next, you might want to check for any unusual constraints on topical coverage. Most of these texts cover pretty much the same topics, but you may have, for example, a commitment to another department to cover a topic that is missing in some of the candidate texts. After that, you may want to consider the *level* of the textbooks remaining.

For our purposes, there are three levels. The first is the *reading* level. Many introductory textbooks are above—sometimes far above—the reading level of their intended audience. Among our exemplary texts, Moore and McCabe [12] is at the highest reading level. I strongly recommend it as a reference for you, but it will be tough going for many students. Siegel's original text [13] is very simply and clearly written. It has a warm, friendly tone, most unusual

in a mathematics or statistics textbook, but not surprising if you have met the author. The second edition [14] is nearly as fine. Clarkson and Williams [3] give an empirical report on the reading level of statistics textbooks. Since the issues are not a great deal different from those for mathematics textbooks, I will not say more about them here.

Another kind of level is the *technical* level. This refers to how deeply the text delves into statistical details. For example, Moore and McCabe [12] goes more deeply into these matters than Moore [11]. Generally, the differences are not great among texts aimed at a first course for a general audience. One exception to be wary of is the cookbook in its umpteenth edition that has added every topic users of earlier editions have ever asked for. Some of these may cover far more topics than Moore and McCabe.

The third kind of level is the *mathematical* level. Here mathematicians are at a disadvantage. First, it is easy to underestimate mathematical level when you yourself are very fluent in mathematics. Second, if you have been previously teaching mathematics majors, you may also underestimate the mathematical skills of a typical statistics student. For a general introductory course, these skills tend to be below the skills one sees in a finite mathematics or precalculus student, and way below what one sees in a calculus student. While mathematics in which a student is fluent can be an aid to precise and rapid communication, mathematics a student has to labor through becomes an obstacle to learning rather than an aid. The good news here is that minimal mathematical skills are sufficient, providing your textbook gives non-algebraic explanations of statistics. The text by Siegel [13] uses little or no algebra. In areas where it cannot be avoided, such as working with straight lines and their equations in regression analysis, it reviews the needed algebra. Siegel and Morgan [14] combine the excellent verbal explanations in [13] with optional formulas.

## Look for data in the exercises

Even more important than a textbook's exposition are the problems it sets for students.

> Judge a statistics book by its exercises, and you cannot go far wrong! [4, p. 331]

For one thing, this may be the only part of the book that many students ever read. (One thing that good problems should do is encourage the student to read the rest of the text.)

> In judging the exercises, I think one good way to decide what to look for comes from thinking about

the kind of course that has helped to give our subject a bad name [especially among mathematics majors!]. Because statistics has too often been presented as a bag of specialized computational tools, with a morbid emphasis on calculation, it is no wonder that survivors of such courses regard their statistical tools more as instruments of torture than as diagnostic aids in the art and science of data analysis. All too many textbooks of the past have applied their tools to data sets that have little connection to the body of living knowledge—at worst the numbers have been total fictions, at best they have been dismembered fragments of some old scientific cadaver [4, p. 331].

We look for the same things in the problems that we look for in the body of the text. If the exercises do not ask students to look at the data and check assumptions, then anything we say about those issues will be empty preaching.

> How then do you read the vital signs of the exercises to distinguish the living text from the cold corpse? Look for three things: (a) Are the data sets real or fake? (Real statisticians don't analyze fake data.) (b) Does completing an exercise answer an interesting question, or is the number-crunching a dead end in itself? (Real statisticians don't stop with the arithmetic.) (c) What is the ratio of thinking to mere grinding? (Real statisticians think.) [6, p.331]

While there may be places (such as Anscombe's regression examples [1], succinctly discussed in [2, p.8]) in which a skillfully fabricated batch of numbers illustrates a pedagogical point, most made-up "data" serve only to insulate the student (and the author) from contact with real statistical applications. Be wary of an author who is not familiar with enough real data sets to illustrate a textbook.

Why are real data so important? First, it is the subject matter of statistics. A statistics book without data is like a calculus book without the real numbers. Another reason for using real data is to convince students that statistics is used in the real world.

> At the high end are the books in which you know a data set is authentic, and so do your students, because the author gives its source. There is really no excuse for using a data set without acknowledging the people who did the work, and enough authors are now citing sources that I think we should regard a data set as fake if no source is given. [4, p.331]

Devore and Peck [8, pp. 487-520 and 557-576] give 14 examples and 49 exercises involving real problem situations in their coverage of two-group comparisons, complete

with references to the scientific literature where the studies appeared. The many books authored or coauthored by Terry Sincich also have an abundance of examples citing real studies.

> Using real data sets (with sources) is important but only part of the story. A data set should not only be real, it should feel that way. [4, p.331]

At one extreme, consider these flagrantly unreal problem situations:

> Professor Roundhead claims that only 35% of the students at Flora College work while attending school. Dean Bigheart thinks the professor has underestimated the number of students with part-time or full-time jobs.

> The Big Break Moving Company claims a typical family moves every 5.2 yr.

These lead-ins are followed with summary statistics (no data, real or imagined) for which students are asked to test a hypothesis for a single mean or proportion. In the second problem, it would be more instructive to ask students to explain why they should not do the hypothesis test—because times between moves are almost certainly skewed toward high values.

Data that are real can seem unreal to the student if there is not an adequate explanation of the context of the data or the purpose for which it was gathered, or if it is too technical or far from the student's experience. Consider this pair from [8]:

> Here's one to sink your teeth into: The authors of the paper "Analysis of Food Crushing Sounds During Mastication: Total Sound Level Studies," *J. of Texture Studies* (1990): 165–178) studied the nature of sounds generated during eating. Peak loudness (in decibels at 20 cm away) was measured for both open-mouth and closed-mouth chewing of potato chips and tortilla chips. Forty students participated, with ten in each combination of conditions (such as closed-mouth, potato chip, and so on). We are not making this up! [p. 512]

> The effect of plant diversity on beetle density was examined in a series of experiments described in the paper "Effects of Plant Diversity, Host Density, and Host Size on Population Ecology of the Colorado Potato Beetle," *Environ. Entomology* (1987): 1019–1026). Potatoes grown in fallow plots and potatoes grown in plots that also included bean plants and

weeds (called *triculture*) were compared on the basis of the number of beetle eggs found on the plant leaves. [p. 517]

While even the authors seem to admit that students might question the authenticity of the chip munching situation, it is certainly a context that students can relate to, and they will know what the researchers were measuring, even if the units (or the motivation) remain unclear. In contrast, the beetle example seems a bit remote and abstruse, even to an author wearing bib overalls. Compare it with the wrenching realism of this example from Siegel and Morgan, which also illustrates the fact that outliers we might think are surely errors may be all too real.

> Display 2.30 shows the age at which a sample of runaway and homeless girls reported first having had sex. Exactly what "having sex" means was left to the individual respondent to decide. The data are part of a database assembled for the study of adolescent pregnancy, and four individuals who reported never having had sex were excluded. *Note:* Some individuals reported having first had sex as toddlers. In some cases, these instances of childhood abuse had been previously reported. [14, p.63]

Another reason for giving students real data is to encourage them to look at their data! In one text, we find fake data for all the exercises on paired data. What would motivate a student to look at fake data? If they found something unexpected in real data, their attention could profitably be focused on real issues, such as searching for influential data points, flaws in the design of the study, or some underlying reason why things might be different than they at first appear. If we find something unexpected in fake data, it merely suggests a sloppy forgery.

Real data also allow students to check to see if the assumptions underlying the techniques they apply have been met. For their exercises on independent samples, the same text gives no data at all, only summary statistics. This makes it impossible to check assumptions, and gives the impression that doing so is superfluous. These examples violate the dictum: "Check assumptions, don't make them." A similar cavalier attitude is shown toward issues of design and sampling, with random sampling decreed by fiat, as in: "a random sample of 10-year-old students with IQ scores below 80" or "a random sample of 15 U.S. adults." It would be much more enlightening to discuss why such samples are unlikely to have been taken. In addition, this text throws a red herring across the path of the unwary student: while no data are given for any independent situation, data (or numbers) are given for *every* paired sample situ-

ation. The obvious conclusion: use the paired technique when you have data, the independent samples technique when you don't. In contrast, Devore and Peck [8] give 38 examples of real problem situations involving independent samples. Unfortunately, they give raw data for only four of their problems. Hence, students can see that statistics is used in the real world, and see how it is used, but they cannot look at the data nor check inferential assumptions. In contrast, while Moore and McCabe [12, Ch. 7] give a mixture of real and unreal data and/or summary statistics, they do an outstanding job of asking students probing questions about the data or the design or the sampling procedure. Here are a few instructive examples:

> You should be hesitant to generalize these results to the population of all middle-aged men. Explain why.

> Explain in language that the manager can understand why he cannot be certain that sales rose by 6%, and that in fact sales may even have dropped.

> Examine each sample graphically with special attention to outliers and skewness. Is the use of a $t$ procedure acceptable for these data?

> What assumptions does your statistical procedure in (a) require? Which of these assumptions are justified or not important in this case? Are any of the assumptions doubtful in this case?

> The distribution of earnings is strongly skewed to the right. Nevertheless, use of $t$ procedures is justified. Why?

> Once the sample size was decided, the sample was chosen by taking every $k$-th name from an alphabetical list of undergraduates. Is it reasonable to consider the samples as [if they were] simple random samples from the male and female undergraduate populations?

> What other information about the study would you request before accepting the results as describing all undergraduates?

In contrast, nearly all the problems in the texts in the last row of Table 1 ask the student to merely grind out the computations and reject or fail to reject the null hypothesis.

I have already mentioned the use of real data to illustrate the type of data for which different statistical techniques are used. While it might seem unbelievable that anyone would present a technique without telling what it is used for, the discussion of boxplots in some books gives no clue as to their primary uses, which are to compare multiple groups and to provide a mechanism for flagging potential outliers. Instead, boxplots for a single batch of numbers

are introduced early in the text, but not used later when multiple groups are compared. One gets a sense of readers of earlier editions asking for boxplots and the authors trying to accommodate their wishes without understanding their reasons.

Contrast this lack of exploratory attitude with the books by Siegel, the master of using data to illustrate what a technique can do for you. In [14] he explains how to make a stem-and-leaf plot and a histogram. Then he looks at specific examples of data, with subsections on symmetric, Gaussian, skewed, long-tailed, rectangular and bimodal distributions as well as outliers. For bimodal distributions, he gives data on prize monies awarded in golf tournaments, and comments:

> When we see a data set with more than one mode, we should immediately consider the possibility that more than one group is being represented. We need to try to identify the groups, which may require creativity, imagination, and detective work. [14, p.42]

This presents a more attractive image of statistics than endless number-crunching exercises! (Note: Siegel finds the cause of the bimodality. What do you think it was?) Next he uses 45 histograms to show what happens when you take samples of various sizes from various theoretical distributions. For boxplots [14, pp. 87–96], he shows how to make one and then immediately uses six sets of parallel boxplots to compare gas mileage for different makes of car.

One objection to using real data is that the data sets are too large and the numbers too inconvenient for hand calculation. Making six boxplots for one data set will take a lot of time. That may be true, but the solution is to abandon hand calculation except for simple examples that help students to understand the ideas. This is but one of many reasons for using statistical software in a first course.

## Summary and Conclusion

People with little or no training in statistics often teach statistics. Some of these people write textbooks, which may then be adopted by others with similar backgrounds. A better alternative is to adopt textbooks that are strong in areas where the teachers are weak. Mathematicians are likely to be strong in computation and in stating and interpreting the theorems of mathematical statistics. They are likely to be weak in knowledge and experience in applying statistics to a real world in which the hypotheses of theorems are never met. Find textbooks that are strong in these areas, even if you have to supplement them with formulas and computational algorithms, or clean up occasional impre-

cise statements. A text that contains little or no real data is not likely to be much help in learning to work with and interpret real data. You can automate the computational aspects of analyzing real data with modern statistical computing software and focus students' attention on learning statistical concepts.

To make that job easier, seek a textbook that

- has a qualified author,
- reflects current statistical practice,
- includes real applications,
- includes real data,
- looks at the (real) data,
- explains how real data are produced,
- provides a context for the data it uses,
- includes more concepts than calculations,
- interprets the results of calculations in the context of the data,
- checks assumptions for every inference, and
- asks probing questions about the data.

This paper opened with a quiz and I'll close it with another.

What word appears most frequently in the list above? Why?

In recent years, statistics has infused much of the K–12 mathematics curriculum. One reason for this is that statistics offers mathematics teachers a much better opportunity than, say, factoring quadratics to convince students that quantitative reasoning is useful in the real world and relevant to the issues of our day. (For endless examples of the latter point, see the Chance website.) Statistics can also offer this advantage to college mathematics teachers, but only if it is taught in connection with real problems and real data.

**Note.**   Parts of this paper were presented at the Conference on Assessment in Statistics Courses sponsored by the Boston Chapter of the American Statistical Association, 19 April 1997. The author wishes to thank Don Burrill, Farid Kianifard and Paul Velleman for helpful comments on earlier drafts of this paper; George Cobb for comments on recent versions, and many wonderful quotes, both cited and stolen; Katherine Halvorsen, Farid Kianifard, and Joan Weinstein for assistance with the references; Chris Olsen for information on AP Statistics; and most of all Tom Moore, who nursed this article through an unexpectedly difficult path from a talk at a meeting to a paper suitable for inclusion in an MAA Notes volume.

## A list of resources for teaching statistics

### Textbooks Recommended for the Initial Offering of AP Statistics

This list comes from AP publications [5, 6] and the AP listserve. The list shifts slightly, but this is the core, as I see it.

Devore and Peck [8]
Iman [10]
Moore [11]
Moore and McCabe [12]
Siegel and Morgan [14]
Wardrop [18]

## Other Resources

The American Statistical Association has a web site at: `www.amstat.org`.

The *Current Index to Statistics* is published jointly by the American Statistical Association and the Institute of Mathematical Statistics. The paper version covers the preceding year. A CD-ROM version is cumulative over the past 20+ years.

The College Board has a website with information on AP Statistics at: `www.collegeboard.org`. It provides guidance to all aspects of the first course, and it is geared toward people whose normal assignment is teaching mathematics.

And the Chance Project has one at: `www.dartmouth.edu/~chance`.

## References

1. Anscombe, F. J., "Graphs in Statistical Analysis," *The American Statistician,* **27**(1), 1973, 17–21.
2. Chatterjee, Samprit, and Bertram Price, *Regression Analysis by Example,* 1977, John Wiley and Sons, New York.
3. Clarkson, Sandra, and William Williams, "The Readability of Some Popular Elementary Statistics Texts," *Proceedings of the Section on Statistical Education of the American Statistical Association,* 1996.
4. Cobb, George W., "Introductory Textbooks: A Framework for Evaluation (A Comparison of 16 Books)," *Journal of the American Statistical Association,* **82**(397), 1987, 321–339.
5. College Board, *Advanced Placement Course Description: Statistics,* 1996, The College Board, Princeton, NJ.

6. College Board, *Suggested Resources for Teaching Statistics,* 1996, The College Board, Princeton, NJ.

7. Dawson, Robert, email message, 1994.

8. Devore, Jay, and Roxy Peck, *Statistics: The Exploration and Analysis of Data,* 2nd edition, 1993, Duxbury Press, Belmont, CA.

9. Hayden, Robert W., "Using Writing to Improve Student Learning of Statistics," in *Using Writing to Teach Mathematics,* Andrew Sterrett, ed., 1990, Mathematical Association of America, Washington, DC.

10. Iman, Ronald J., *A Data-Based Approach to Statistics,* 1994, Duxbury Press, Belmont, CA.

11. Moore, David S. *The Basic Practice of Statistics,* 1994, W. H. Freeman, New York.

12. Moore, David S., and George P. McCabe, *Introduction to the Practice of Statistics,* 2nd edition, 1993, W. H. Freeman, New York.

13. Siegel, Andrew F., *Statistics and Data Analysis: An Introduction,* 1988, John Wiley and Sons, New York.

14. Siegel, Andrew F., and Charles J. Morgan, *Statistics and Data Analysis: An Introduction,* 2nd edition, 1996, John Wiley and Sons, New York.

15. Tukey, John W., *Exploratory Data Analysis,* 1977, Addison-Wesley, Reading, MA.

16. Velleman, Paul F., and David C. Hoaglin, *Applications, Basics, and Computing of Exploratory Data Analysis,* 1981, Duxbury Press, Boston.

17. Velleman, Paul F., email message, 1997.

18. Wardrop, Robert L., *Statistics: Learning in the Presence of Variation,* 1995, William C. Brown, Dubuque, IA.

**Note.** Wardrop's book is no longer published by the publisher listed above, but is available (at a much-reduced price) through the author's home address: 5573 Kupfer Road; Waunakee, WI 53597.