Eight Features of an Ideal Introductory Statistics Course

Donald B. Macnaughton*

This paper discusses the following features of the author's ideal introductory statistics course: (1) a clear statement of the goals of the course, (2) a careful discussion of the fundamental concept of 'variable', (3) a unification of statistical methods under the concept of a relationship between variables, (4) a characterization of hypothesis testing that is consistent with standard empirical research, (5) the use of practical examples, (6) the right mix of pedagogical techniques: lectures, readings, discussions, exercises, activities, group work, multimedia, (7) a proper choice of computational technology, and (8) a de-emphasis of less important topics such as univariate distributions, probability theory, and the mathematical theory of statistics. The appendices contain (*a*) recommendations for research to test different approaches to the introductory course and (*b*) discussion of thought-provoking criticisms of the recommended approach.

KEY WORDS: Teaching; Statistical education; Statistics education; Pedagogy; Research in statistical education.

1. INTRODUCTION

In his excellent review of the state of statistical education, David Moore separates current issues into those of content, pedagogy, and technology (1997). In this paper I discuss six content issues, a broad pedagogy issue, and a technology issue. I have structured the material in terms of features I envision in an ideal introductory statistics course.

2. A CLEAR STATEMENT OF THE GOALS OF THE COURSE

2.1 The Value of Emphasizing Goals

Emphasizing the goals of any undertaking forces us to define and focus on what is most important. Otherwise, we may take a scattershot approach, which is generally less efficient. Thus, following Hogg (1990), I believe the first feature of an ideal introductory statistics course is that it have a clear statement of its goals.

I have observed goal-setting exercises in which the goals were given much attention for a brief period and then forgotten. But the exercise of setting goals can be much more fruitful if we *keep* our attention focused on the goals across time. We must review the goals at regular intervals to see (*a*) if the goals are still valid and (*b*) if our day-to-day operations are properly serving the goals.

If you teach an introductory course, and if you have not already done so, I urge you to

- define the goals of your course
- publish the goals to your students as a sign of your commitment, and
- regularly ask yourself how you can improve the course to better satisfy the goals.

This cannot but help improve any course.

2.2 A Definition of "Empirical Research"

Consider a definition:

> *Empirical research* is any research in which data are gathered from the "external world" and then conclusions are drawn from the data about the external world.

Empirical research plays an essential role in many areas of human endeavor, including science, government, business, and industry. In particular, no statement of fact in any branch of science is accepted until it has been verified through careful empirical research.

2.3 Recommended Goals

I recommend the following two goals for the introductory statistics course:

Goal 1: to give students a lasting appreciation of the vital role of the field of statistics in empirical research

Goal 2: to teach students to understand and use some useful statistical methods in empirical research.

Note how the goals imply a commitment to the practical application of statistics in empirical research. Emphasis on practical applications is important because the social value of statistics is through its practical application, and not (directly) through theory or mathematics. Thus, as many teachers now agree, we stand a much better chance of impressing students if we discuss practical applications.

(We must also discuss generalizations that link statistical ideas together, but the practical applications should come first. More on generalizations below.)

Note that the recommended goals make no mention of specific statistical topics, such as box plots, the normal

---

*Donald B. Macnaughton is president of MatStat Research Consulting Inc, 246 Cortleigh Blvd., Toronto Ontario, Canada M5N 1P7. E-mail: *donmac@matstat.com* The author thanks the following teachers for valuable comments and criticisms: Carol J. Blumberg, Donald F. Burrill, Christopher Chatfield, George W. Cobb, Sir David Cox, David J. Finney, Robert W. Frick, Robert W. Hayden, Robert V. Hogg, Gudmund R. Iversen, William H. Kruskal, Christine E. McLaren, David S. Moore, Thomas L. Moore, John A. Nelder, Allan J. Rossman, Richard L. Scheaffer, Milo A. Schield, Stephen Senn, Paul F. Velleman, four correspondents quoted anonymously in the appendices, and several subscribers to the ApStat-L e-mail list who commented on a draft of section 8.

distribution, or *t*-tests. This is because I believe that giving students a lasting appreciation of the role of statistics is much more important than any specific statistical topic.

I appeal several times below to the first goal above. I propose a definition of the role of statistics in empirical research in section 4. I further discuss goal setting and the concept of "measurable" goals in appendix A.

3. A CAREFUL DISCUSSION OF THE FUNDAMENTAL CONCEPT OF 'VARIABLE'

If we examine every lesson and indeed every sentence used in an introductory statistics course, we find one concept is almost always present—the concept of 'variable'. Obviously, this concept is very important in statistics. If you ask your students to explain the concept of 'variable', how do they answer? How do you explain the concept of 'variable' to yourself?

Since many students have trouble understanding statistics, and since many students lack a good understanding of the concept of 'variable', I suggest an important feature of an ideal introductory course is a careful discussion of this fundamental concept. In particular, I recommend that we break the concept into its constituent parts and characterize a variable as

> A *variable* is a formal representation of a property of entities (things).

Implementing this approach is straightforward—we begin the course by discussing entities and properties, presenting numerous examples from the students' own experience. Then we introduce the concept of a variable as a formal representation of a property of entities, again with numerous examples.

(All statistical "populations" are populations of entities. I discuss the importance of developing the concepts of entities, properties, and variables in a bottom-up fashion in terms of concrete cases and experiences in appendix C.)

4. A UNIFICATION OF STATISTICAL METHODS UNDER THE CONCEPT OF A RELATIONSHIP BETWEEN VARIABLES

By introducing variables in terms of properties of entities we set the stage for introduction of the concept of a *relationship between* properties (relationship between variables). I suggest an important feature of an ideal introductory course is a unification of the field of statistics under the concept of a relationship between variables.

By a "relationship between variables" I mean the standard statistical idea that one variable (called the *response* variable) "depends" on one or more other variables (called the *predictor* variable[s]). All statisticians are familiar with this unifying idea, although it is generally not emphasized. I give a formal definition of the concept of a relationship between variables in a paper for students (1996a, sec. 7.10).

In order to cement the concept of a relationship between variables in students' minds, we must discuss many examples of relationships, emphasizing relationships that are of practical value, as I discuss below.

After presenting the concept of a relationship between variables, we can develop the field of statistics for students as a set of techniques for studying properties and relationships between properties of entities (relationships between variables). I shall refer to this approach as the entity-property-relationship (EPR) approach. I give details of the approach in two papers (1996b, 1996a).

* * *

In an earlier essay (1997) I state that almost all empirical research projects that use statistical methods in all fields of empirical research can be usefully interpreted by filling in the blanks in the following general schema:

Population of Entities: ______________________________
Response Variable: ______________________________
Predictor Variable(s): ______________________________
Statistical Questions:
1. Is there a relationship between the response variable and the predictor variable(s) in the entities in the population?
2. If there is a relationship, how can we best predict or control the values of the response variable in new entities from the population on the basis of the relationship?
3. How accurate will the prediction or control be?

In a paper I state that almost all the standard statistical techniques can be usefully viewed as techniques for studying relationships between variables (1996b, sec. 4.2).

So far, no reader has suggested that either of the above statements is incorrect. If the statements are correct, they provide a formal basis for unifying the field of statistics under the concept of a relationship between variables.

* * *

There are three important benefits of the unification. The first is that we can interpret each new statistical method we teach to students in terms of the same set of fundamental concepts: entities, properties, variables, and relationships between variables. This unifies students' understanding of statistical methods because (at the highest level) almost every statistical method is interpreted the same way. This makes the field of statistics substantially easier for students to understand.

A second benefit of unifying the statistical methods relates to the goal of giving students a lasting appreciation of the *role* of statistics in empirical research. By unifying the methods, we are able to define the role of the field of statistics.

To define the role of statistics in empirical research, let us step back a step and ask

What is the goal of empirical research?

I suggest that the goal of all empirical research can be usefully viewed as being to obtain knowledge about entities, relationships between entities, properties of entities, and relationships between properties. I suggest that all empirical research projects (including all empirical scientific research) can be easily and usefully characterized in these terms.

In particular, empirical researchers are interested in studying relationships between properties because knowledge of such relationships enables us to predict and control the values of properties—an ability that often has substantial social value. For example, as medical researchers learn more about how to predict and control the amount of cancer in people, this knowledge has the social benefit of reducing suffering and saving lives. Similarly, as physicists and engineers discover how to predict and control the amount of energy coming from a fusion reactor, this knowledge will likely provide inexpensive safe and clean energy.

Almost all prediction and control in science (and in all other areas of empirical research) is done on the basis of relationships between variables.

The foregoing ideas suggest that we can characterize the role of statistics in empirical research as follows:

> The main role of the field of statistics in empirical research is to provide an efficient set of techniques to help researchers study variables and relationships between variables (relationships between properties of entities), mainly as a means to predicting and controlling the values of variables.

A third benefit of the unification is that by emphasizing the wide applicability of the concept of a relationship between variables and by demonstrating that the field of statistics has broad and versatile techniques for studying relationships between variables, we demonstrate that our field plays a central role in many fields of human endeavor. This helps students to develop a lasting appreciation of the field.

Appendices C through G discuss comments by statistics teachers about the EPR approach.

5. A CHARACTERIZATION OF HYPOTHESIS TESTING THAT IS CONSISTENT WITH STANDARD EMPIRICAL RESEARCH

What do the various statistical tests of hypotheses have in common? That is, if we perform a *t*-test, or a test in regression, or a test in some other statistical method, what are we testing?

There are three valid points of view of statistical tests as follows:

1. statistical tests are tests of the existence of relationships between variables
2. statistical tests are tests of differences between groups
3. statistical tests are tests of the values of parameters.

Which point of view should we emphasize in the introductory course? Let me first consider each point of view in turn.

If we survey a large set of randomly selected journal articles in any field of empirical research (including any field of science), it turns out that almost all statistical tests that are used in actual empirical research can be viewed as tests of the existence of a relationship between a single response variable and one or more predictor variables. For example, if a researcher performs a two-sample *t*-test, we can view this as a test of the existence of a relationship between two variables. The response variable is the continuous variable measured in each of the entities in the two groups. The predictor variable is the discrete variable that reflects the property that distinguishes the two groups of entities from each other. Thus we can view the *t*-test (and its generalization, analysis of variance) as helping to answer the following question: "Is compelling evidence available that the value of the response variable *depends* on the value of the predictor variable(s) for entities in the population? That is, is compelling evidence available of a relationship between the variables?"

The second common interpretation of some statistical tests is to view them as tests of "differences between groups". For example, we can view the two-sample *t*-test as a test of whether a difference exists between two groups of entities on a continuous variable. Under this second approach we do not appeal to the concept of a relationship between variables, and we may not even acknowledge that more than one variable is present. That is, we concentrate on the response variable, and we say that we are testing whether compelling evidence is available that the population mean of the values of the response variable in one of the groups *differs* from the population mean of the values of the response variable in the other group.

The third common interpretation of statistical tests is to view them as tests of the values of parameters—parameters of populations, or parameters of distributions, or parameters of relationships between variables. We view tests as testing whether compelling evidence is available that a particular parameter (possibly a vector) has a particular value, or (more frequently) testing whether a parameter's value likely lies within some range (or set) of values. This is a very general way of viewing statistical tests because

1. most statistical tests of the existence of relationships between variables can be, at a mathematical level, viewed as tests of the values of parameters (some

"nonparametric" tests of relationships are arguably excluded)
2. most statistical tests of differences between groups can also be viewed as tests of the values of parameters
3. most (all?) statistical tests that are not in one of categories 1 or 2 can also be viewed as tests of the values of parameters.

* * *

Which point of view of statistical tests should we emphasize in the introductory course? Let me first compare the view that tests are tests of relationships between variables with the view that tests are tests of differences between groups.

Viewing statistical tests as tests of differences between groups lacks the unifying power provided by the concept of a relationship between variables because only a limited number of statistical tests can be viewed as testing for differences between groups. For example, the tests in regression and time series analysis cannot generally be viewed as testing for differences between groups because often when these statistical procedures are used no "groups" are present in the research. On the other hand, most (but not all) statistical tests used in real empirical research can be viewed as tests of the existence of relationships between variables. (This includes *all* tests of differences between groups.) Thus (assuming both concepts can be taught equally easily to students) the concept of a relationship between variables is preferred to the concept of a difference between groups, because the concept of a relationship between variables is more general.

Can the concept of a relationship between variables be taught to students as easily as the concept of a difference between groups? I have no hard evidence here, but consider the following simple view of a relationship between variables:

> When the value of *x* changes in members of a certain population of entities, the value of *y* also tends to change in the entities in synchrony with the changes in *x*.

Intuitively it seems that we can easily cultivate this view in student's minds if we nourish it with many practical examples.

* * *

Let me now compare the view that hypothesis tests are tests of relationships between variables with the view that hypothesis tests are tests of the values of parameters. To help decide which of these views is more reasonable for the introductory course, recall the first goal I propose above—the goal of giving students a lasting appreciation of the vital role of the field of statistics in empirical research. Which of the two views of hypothesis tests gives students a better appreciation of the role?

In section 4 I discuss how we can view the role of the field of statistics in empirical research as being to help researchers study variables and (more importantly) relationships between variables (relationships between properties of entities) as a means to prediction and control. To maximize students' appreciation of the role we should characterize statistical tests in terms that are *closest* to the role. Thus in the introductory course it makes sense to characterize statistical tests as tests of the existence of relationships between variables.

The "relationship between variables" characterization of statistical tests is supported by the fact that most empirical researchers are generally not (directly) interested in testing the values of parameters. Instead, they are interested in knowing whether certain relationships exist between variables. They are interested because if relationships are found, we (as society) may be able to use the knowledge of the relationships to predict or control the values of the response variables. Since researchers using statistical tests are mainly interested in detecting relationships, and they are usually not (directly) interested in testing the values of parameters, it makes sense to characterize statistical tests as tests of the existence of relationships between variables.

If we characterize statistical tests as tests of the existence of relationships between variables, we lose the generality we obtain if we characterize them as tests of the values of parameters. Thus it is important to ask if we are overlooking important cases if we use the "relationship between variables" characterization. Examination of actual empirical research projects in many fields of research suggests we are not. Can you think of significant examples of statistical tests in *actual practical empirical research* (i.e., not made-up "science fiction" examples) that cannot be reasonably characterized as tests of relationships between variables?

(Some such examples can be found, although they do not appear to be significant: Some tests are not direct tests of the existence of relationships between variables but are instead tests of *extensions* to relationships between variables. Also, a small proportion of real empirical research is aimed directly at determining the values of certain parameters, and this research cannot easily be interpreted in terms of relationships between variables, but see appendix H.)

In recommending that we de-emphasize testing the values of parameters, am I recommending that the general concept of a statistical test of the value of a parameter be banished? Certainly not—the concept is extremely important, since it reflects the mathematical basis of almost all statistical hypothesis testing. However, I do suggest that we defer detailed discussion of this mathematical concept until later courses. This allows students to first master the more practical concept of a test of the existence of a relationship between variables.

* * *

The above discussion takes for granted the idea that hypothesis testing should be taught in the introductory statistics course. Cobb notes that some statistics teachers feel that hypothesis testing should *not* be taught or should be de-emphasized (1992, p. 4).

This movement away from hypothesis testing may be due to

- the view (which I argue against above) that hypothesis testing is testing the values of parameters and
- a recognition that many empirical researchers are uninterested in testing the values of parameters.

Cobb also notes a movement *toward* model building (which he calls "model choosing" and "model checking" 1992, p. 5). I believe model building is important both in statistics and in empirical research because I believe relationships between variables are important, and statistical models are simply succinct statements of relationships between variables.

(Although I believe model building is important in statistics, I do not recommend a heavy dose of model building in the introductory statistics course because the algebra of statistical models can intimidate some students.)

If we value model building, and if we value objectivity, it follows that hypothesis testing is also important. Hypothesis testing is important because hypothesis tests are the most efficient objective way to determine (in typical noisy data) whether a research project provides compelling evidence of a relationship in the population between the response variable and the predictor variable(s). Only if we have compelling objective evidence of a relationship can we confidently build a predictor variable into a statistical model.

6. THE USE OF PRACTICAL EXAMPLES

We can make the field of statistics come alive for students if we crown our lessons with practical examples. Thus an important feature of an ideal introductory course is that it be rich in appropriate examples. But good examples are hard to find (Singer and Willett 1990; Cobb 1992). In this section I propose some guidelines for choosing examples for use in an introductory course.

6.1 Use Real Or Realistic Data

One widely-accepted guideline for choosing examples is

> Use examples that have *real* data or at least *realistic* data.

This guideline goes almost without saying because most teachers agree that real or realistic data make the material more concrete for students, thereby facilitating understanding, and thereby increasing the likelihood that students will obtain a lasting appreciation of our field.

6.2 Use Response Variables That Students Can See Value In Predicting or Controlling

I propose in section 4 that the main role of the field of statistics is to help empirical researchers discover how to predict and control the values of variables. In light of this role, a second useful guideline for choosing examples is

> Use examples with response variables that students can see clear *value* in predicting or controlling. (Typically the value comes through providing a *basis for action.*)

This guideline is important because if students see *value* in being able to predict or control the response variable in an example, they are more likely to see value in the field of statistics. That is, they are more likely to obtain a lasting appreciation of our field.

To illustrate what happens if we violate this guideline, suppose a teacher chooses to demonstrate the concept of a relationship between variables by discussing the relationship between "forearm length" and "foot length" in a sample of, say, one hundred people. Using this example, with real or realistic data, the teacher can show the class how one can infer a fairly strong increasing relationship between "forearm length" and "foot length" in the population from which the sample was drawn. Thus if we know a person's forearm length, we can predict his or her foot length.

After being shown this example, students can rub their chins and say, "Hmmm, that's very interesting." But in the end, the intelligent student is going to say: "So what? Who cares? What is the value of being able to predict a person's foot length from their forearm length? What is the value of studying this relationship between variables? What does this example have to do with *anything* that's important in life?"

The correctness of the preceding statements is underscored by the fact that no serious researcher would dream of performing this kind of (essentially pointless) study and no serious medical or health journal would nowadays consider publishing (without further redeeming qualities) a report of this type of study.

Similarly, if we ask students to collect data about themselves, this will clearly raise student interest. But unless the response variables are important, I believe this use of personal data will *not* do much to raise student *appreciation of statistics.* For example, students might collect data about the amount of money spent on haircuts by male and female students, and they might detect the following relationship between gender and spending: females tend to spend more on haircuts than males. But this relationship *provides no obvious basis for action.* Thus the intelligent student can say, "So what? Who really cares if this (not surprising) fact is true? What of actual value has the field of statistics brought to the table by formally demonstrating this fact?"

(Furthermore, sociology is the scientific discipline that is probably closest to studying the relationship between gender and haircut spending, but it seems unlikely that any sociologist would view this relationship as particularly important.)

Thus I believe that a teacher who uses this type of example does the field of statistics a disservice because the teacher is associating the field with examples that are (or appear to be) *frivolous.* These examples can lead students to conclude that the field of statistics specializes in dealing with frivolous problems or, worse, that the field of statistics is itself frivolous.

On the other hand, if we present students with the results of a research project that measured "good manners" and "happiness" in students, and if the results show that students with good manners tend to be happier, these results are of obvious practical value to students—they suggest a clear basis for action. This is because this research project uses a response variable ("happiness") that students can see clear value in predicting and controlling.

(These results also provide a good starting point for discussion of measurement issues ["How can we measure 'happiness' and 'good manners'?"] and the relationship between correlation and causation ["Is there a way of demonstrating that good manners *cause* happiness?"]. I discuss this type of example further in the following subsections.)

6.3 Finding Important Response Variables

I shall refer to a response variable that students can see clear value in predicting and controlling as an "important" response variable. In light of the discussion in the previous subsection, a key question is How should we go about *finding* important response variables?

For students who are majoring in a particular discipline, it is most reasonable to choose important response variables and research examples from that discipline. For other students, one effective way of finding important response variables is to ask the question: "What do students want in life?"

Some answers are good grades, good health, good interpersonal relationships, happiness, freedom, peace, good weather, money, and so on. Each of these answers suggests fertile areas for finding important response variables and important relationships between variables for study by students.

For example, hundreds of research projects in the literature of education research use the variable "grade" or "mark" in a course of study as the response variable. In particular, an interesting question is whether a relationship exists between "hours of study done in the course" and "grade" for students in a high school, college, or university course. Most of us have intuitions about this relationship, but what do data say about it? This example can be easily studied close to home with data collected from previous years' classes.

Another interesting research topic for students is a study of the relationship between the *method* students use to study and their grades. (Interestingly, after the variable "hours of study" is taken into account, the last time I looked in this area, learning researchers had found no evidence of a relationship between "method of study" and "grade".)

Similarly, because most students are interested in good health, many meaningful examples can be found in the field of medicine. Examples are best if they relate to student health or to matters of general social concern, such as (*a*) vitamin C and the common cold, (*b*) treatments for AIDS, or (*c*) relationships between the response variable "longevity" and predictor variables that reflect lifestyle or diet.

When discussing statistical examples it is important to tie them together with the common thread of the concept of a relationship between variables as a means to accurate prediction and control. Earlier compilations of good examples (e.g., Tanur et al. 1989, Freedman et al. 1991) can be easily tied together with this common thread, but this fact is not pointed out to readers. Thus readers fail to obtain a sense of the overall value of statistics, because they are buried in details.

6.4 Finding Data

Statisticians have recently recognized the lack of good data sets for use in the introductory course and have created libraries of interesting data that are readily accessible over the Internet (DASL Project 1998, EESEE 1998, StatLib 1998). Despite the availability of these libraries, one sometimes needs to obtain new data sets. Two problems arise:

1. It is hard to find real research projects with important response variables that are simple enough for discussion and analysis in an introductory statistics course.
2. When an appropriate real research project *is* found, it is often hard to actually obtain the data from the research project, or the data set may be obtainable but is too large to be easily used by students.

In view of these problems, some teachers obtain data from almanacs or other publications that have many small tables of numbers. Unfortunately, when such examples are chosen, one often finds that students are highly *un*interested in predicting or controlling the response variables associated with the examples. As suggested in subsection 6.2, these uninteresting response variables help to make students uninterested in our field.

Fortunately, there is an easy solution to this problem: It is not necessary to use *real* data to make points in a statistics course—it is only necessary to use *realistic* data. Therefore, as some teachers know, it is fair ball to make up data for use in examples. I have spent many interesting hours analyzing realistic data I made up with the help of random number generators in statistical packages.

Some authors (e.g., Singer and Willett 1990; Witmer 1997) insist that the data used in statistics courses must be real, not just realistic. I suggest that this requirement is too constraining. Clearly "real data" does not imply "good data" since many real data sets lack important response variables and are therefore unlikely to impress students. On the other hand, if we are careful in how we generate and describe them, made-up realistic data can imply excellent data. That is, as experts in data analysis, statistics teachers are well qualified to make up interesting realistic data. Such data can be generated in minutes, rather than in hours or days, and such an approach allows us to show any effects (or lack of effects, or anomalies) in the data we wish.

Of course, if we present made-up data to students, we must ensure that the data *appear* real, because that is part of what makes data interesting. Thus in describing the data we must include details that illustrate the issues and concerns of real empirical research. One easy way of accomplishing this is to pattern our examples after real research projects, even though the data we use may be made up.

One reason some authors insist that statistics teachers use real data is that these authors recognize that students will obtain a greater appreciation of the field of statistics if we use real data. However, I believe the incremental increase in appreciation achieved through real data (with an important response variable) in an example over using *the same example with made-up data* is small—small enough to be often outweighed by the advantages to the teacher of using made-up data. (The existence of the small advantage does, however, imply that we should use real data whenever the cost of obtaining real data can be easily borne.)

A second reason some authors insist that statistics teachers use real data is that these authors believe the insistence will encourage teachers to read journal articles about empirical research. I agree that it is very helpful in understanding the use of statistics in empirical research to read journal articles about empirical research. Thus I heartily recommend a subscription to *Science* or *Nature* or a similar journal as a way of becoming exposed to a broad range of empirical research. (Unfortunately, some statistics teachers have little or no experience with real empirical research—they are only familiar with empirical research described in statistics textbooks, which is sometimes sterile or impractical.)

If we use made-up data, and if (as I recommend) our examples are socially significant, we should warn students that the data are made up, perhaps at the beginning of each set of exercises. This reduces the chance of students drawing incorrect social implications from our examples.

6.5 Activities With Important Response Variables Are Hard To Find

Consider two types of student assignment in an introductory statistics course:

- *exercise:* an assignment in which students analyze and interpret data *provided by the teacher*
- *activity:* an assignment in which students analyze and interpret data that they themselves have *collected* through performing an empirical research project.

(If an activity is carried on outside of class, it is sometimes called a "project".)

Note that the only difference between exercises and activities is the source of the data: in an exercise the data are provided by the teacher, but in an activity the data are collected by the students in an actual empirical research project.

(Exercises and activities can be performed either by individual students or they can be performed by student groups. Garfield presents evidence that students obtain a better appreciation of statistics if they work in groups [1995].)

Recently there has been increased interest in using activities to help teach statistics (Rossman 1996; Scheaffer, Gnanadesikan, Watkins, and Witmer 1996). Although it is clear that activities reduce the number of topics that can be covered in a course (because time must be borrowed from other course work for the students to actually collect the necessary data), activities have the advantage of making the ideas that *are* covered clearer, because the students are involved in hands-on collection of data.

However, there is a serious problem with activities: It is hard to find activities that can be reasonably performed by students and that also have an *important response variable*—that is, a response variable that students can see clear value in predicting or controlling. Generally, activities with important response variables are hard to find due to limitations of (*a*) funding (for necessary apparatus and other resources) and (*b*) time.

This leads to the following question:

> What proportions of the introductory course should be devoted to the following two modes of operation:
>
> - mode 1: study data sets with important response variables
> - mode 2: study data sets with response variables that are *not* important (but which have other benefits, such as student collection of the data)?

I propose above in section 2 that the first goal of the introductory statistics course be to give students a lasting appreciation of the vital role of the field of statistics in empirical research. Given this goal, the proportions of time allocated to the above two modes of operation should be chosen so as to maximize students' appreciation of the role of statistics.

Here we have a relationship between three variables. The response variable is "appreciation of statistics" and the two predictor variables are "importance of the response variables in examples" and "amount of student collection of data". We would like to find the particular values of "importance" and "amount of student data collection" that maximize "appreciation". That is, we would like to find the point on the response surface generated by the three variables at which "appreciation of statistics" is highest.

Of course, finding the point of maximum appreciation is an empirical problem, so we cannot find it through speculation. However, we can note some likely facts about the relationship. First, with other things held equal, "appreciation" is almost certainly a monotone increasing function of "importance"—the more we use response variables that are important to students, the more students will appreciate statistics.

Second, as noted above, if we use student collection of data, we are usually constrained to use response variables that are *not* important. I suggest that if we use unimportant response variables, students will not appreciate the use of statistics whether they collect the data themselves or not. Students will not appreciate statistics with unimportant response variables because in this case students cannot easily see any *value* in what they are studying. Thus the point of maximum appreciation occurs where the use of important response variables is high and the use of student collection of data with unimportant response variables is medium or (more likely it seems) low.

(Of course, if we can find workable activities with *important* response variables, this is ideal. But, as noted, activities with important response variables are hard to find.)

This suggests that teachers should de-emphasize dealing with impractical data, and emphasize exercises (and, when possible, activities or projects) that use important response variables—i.e., response variables that students can see clear value in predicting or controlling.

Although it is difficult to find workable activities with important response variables, it is sometimes possible to find activities with response variables that are *akin* to important response variables. Such activities can be effective because students quickly see the link between the activities and practical empirical research. For example, Bisgaard describes an activity involving factorial experiments to maximize the flying time of paper "helicopters" (1991). Similarly, Scheaffer, Gnanadesikan, Watkins, and Witmer describe an activity involving a factorial experiment to maximize the shooting distance of a catapult made from Popsicle sticks and rubber bands (1996 pp. 289-293). These activities are effective because students can easily see how the principles used to maximize "flying time" and "distance" readily apply to optimizing the values of important response variables in the real world.

If it is deemed helpful to use an activity with an unimportant response variable, I recommend that the teacher explain to students how the approach used in the activity also applies to examples in practical empirical research. This raises the likelihood that students will see value of the material they have learned.

6.6 Summary

In lectures, exercises, activities, and projects in the introductory statistics course we should use realistic examples that have response variables that students can see clear value in predicting or controlling. (Typically the value comes through providing a basis for action.) This helps give students a lasting appreciation of the vital role of the field of statistics in empirical research.

7. THE RIGHT MIX OF PEDAGOGICAL TECHNIQUES

There is a presently a justified focus on various pedagogical techniques for teaching statistical concepts. An ideal introductory statistics course will have the right mix of the available techniques.

It is important to note that the focus on pedagogical techniques is (at a basic level) independent of the statistical concepts (i.e., the "content") we teach in a course. Thus (at a basic level) we can discuss pedagogical techniques quite independently of statistical concepts (e.g., concepts used in the EPR approach or the concepts used in other more traditional approaches).

(Although at a basic level pedagogical techniques and content are independent, at a higher level some pedagogical techniques may be better for teaching some forms of content under a given set of course goals and other constraints.)

Following are seven pedagogical techniques that are currently used in introductory courses, in rough order of increasing newness:

- *lectures:* the traditional approach; one-way flow of information from the teacher to the students
- *readings:* one-way flow of information from the author to the students
- *discussions:* two-way flow of information between the teacher and the students
- *exercises:* in which the students prepare answers to questions, using no resources beyond various textual materials
- *activities:* sometimes called "laboratories" or "projects", in which students prepare answers to questions by first collecting necessary data; e.g., students may roll dice or flip coins, or collect data about some real world entities to which they have access, such as themselves
- *group work:* in which two or more students perform exercises or activities together
- *multimedia courseware:* in which students use a statistical "textbook" that is written in computer software, which displays information on the computer screen instead of on the printed page, with sound and (possibly) with graphics-intensive narrated lectures. The courseware has built-in exercises and activities, and a two-way flow of information between the software and the students.

Multimedia courseware has just begun to appear (Velleman 1998; Cobb 1997, Cryer and Cobb 1997).

What is the optimal mix of the above seven pedagogical techniques for an introductory statistics course? As suggested above, the answer to this question depends on the goals of the course, the choice of course material (content), the nature of the students in the course, the available resources (including time), and the teacher's skills at various teaching tasks. Furthermore, the true optimal mix of the techniques can only be determined through empirical research, not by speculation.

However, it is possible to venture some generalizations about the techniques. I begin by listing ten properties on which one can evaluate any pedagogical technique:

- ability to quickly transfer information to students
- ability to provide a proper general conceptual foundation for further work in statistics and empirical research (what Moore calls the "big picture" [1997, 125])
- ability to capture students with limited interest or time (e.g., engineering students or medical students)
- ability to stimulate students to participate
- ability to capture students with limited mathematical ability or with "math anxiety" (e.g., some students in general arts or in the social sciences)
- ability to reinforce student understanding
- ability to present students with many forms of information
- reproducibility
- teacher cost (in terms of time)
- student cost.

It is of interest to ask how each of the seven pedagogical techniques for teaching statistics rates on the ten properties listed above. Table 1 summarizes how the techniques rate on the properties.

Table 1
Ratings of Seven Pedagogical Techniques for the Introductory Course on Ten Properties

Technique	Info. Transfer	Provide General Foundation	Capture Limited Interest	Stimulate Participation	Capture Limited Math	Reinf. Student Understand.	Many Forms of Info.	Reproducibility	Teacher Cost (time)	Student Cost ($)
Lecture	High	High	low	low		low	High		High	low
Reading	High	High	low	low				High	low	medium
Discussion			medium	medium		medium			medium	low
Exercise						High			low	low
Activity	low	low	High	High	High?	High?			medium	low
Group Work	low	low	High	High					medium	low
Multimedia	High	High	High	High	High?	High	High	High	medium	High?

Note that I have put entries in only some of the cells in the table, namely, in those cells that I view as reflecting important properties of each approach. All the entries in the table reflect my own subjective judgments. Readers are encouraged to mark their own revisions to the table.

Let us consider some of the rows in the table. In the first row, "Lecture", I imply that the amount of information transfer in lectures is high. Actually, the "high" rating is only a potential rating, since the amount of information transfer depends on several other variables, such as the students' motivation to learn the material and the teacher's ability to present the material in an understandable way. The table states that lectures have a low ability to capture students with limited interest, and a (generally) low ability to stimulate students to participate.

In contrast to lectures, well-designed activities are excellent at capturing students with limited interest, and they are excellent at stimulating students to participate. On the other hand, activities have a low amount of information transfer because students must spend significant amounts of time setting up each activity and collecting the data—work that teaches relatively few statistical concepts per unit of time. Also, activities provide a low amount of a general conceptual foundation for further work in statistics or empirical research.

In brief, lectures teach the ideas faster, but are poor at capturing weaker students. Activities teach the ideas slower, but are better at capturing weaker students.

Note the many "high" ratings for multimedia teaching in the bottom row of the table. The information transfer for multimedia teaching is high because, when necessary, multimedia teaching can actually present spoken "lectures", except that multimedia lectures can be accompanied by more sophisticated and attractive animated graphics than can be presented in most standard lectures. Also, in order to reinforce student understanding, multimedia lectures can be punctuated with exercises or activities at each appropriate point.

Multimedia teaching has a high potential for capturing students with limited interest because it can present many varieties of interesting material (e.g., videos) to students and because it can branch down different paths according to student responses. If carefully designed, the branching can ensure that most students are successful since the software can tailor the course for each student, and can have the "patience" to ensure that students make the appropriate conceptual links as the statistical ideas are developed.

A disadvantage of multimedia teaching is the cost—substantial computer access must be available for students taking multimedia lessons. However, this disadvantage will probably vanish when it becomes the norm for students to have their own computers. (Current prices of multimedia textbooks suggest that the cost of a multimedia textbook will be about the same as the cost of a good paper-based textbook.)

A final important property of multimedia teaching is that multimedia lesson designers can easily perform field trials of the lessons. That is, a part of the multimedia software can record the students' successes and failures in a lesson, including the time spent at different points in the lesson (with a screensaver to allow determination whether the student walked away). The software can also present students with a questionnaire soliciting comments about a lesson. With proper incentives (e.g., draws for prizes) and privacy guarantees, students will transmit this information over the Internet back to the lesson designer. This allows forward-looking designers to perform experiments to zero in on the multimedia teaching methods that work best.

I discuss three problems with current multimedia teaching in a review of ActivStats, an innovative new multimedia textbook [1998].

* * *

What is the best mix of the seven pedagogical techniques? Clearly, the best mix is the one that best satisfies the goals of the introductory course, subject to the constraints a teacher has in his or her teaching environment. Consider

- the goals of the introductory course I propose in section 2
- the properties of a teaching approach I feel are important, as shown by my choice of the ten columns in the body of table 1 and
- the values of the properties I postulate in the body of table 1

In view of these considerations, I believe that in situations in which the higher cost is not a problem, empirical research in statistical education will demonstrate that emphasis on multimedia teaching (using courses developed by master statistics teachers) is the most effective way to teach the introductory statistics course.

I also believe research will demonstrate that use of a small number of activities (especially activities involving designed experiments) helps substantially to increase student appreciation of statistics *if* care is taken to show students how the activities relate to practical empirical research.

I discuss empirical research in methods of statistical education in appendices A and B.

8. A PROPER CHOICE OF COMPUTATIONAL TECHNOLOGY

Nowadays most teachers agree that computers or hand calculators are important aids in the introductory statistics course. These aids relieve teachers and students of the drudgery of statistical computations, enabling them to focus instead on statistical concepts. However, controversy exists among teachers (especially high school teachers) about which is preferred—computers or hand calculators.

Note that computers and calculators are essentially the same devices (i.e., computing devices), although they differ widely on various properties. Thus it is helpful to make a list of relevant properties on which computers and calculators differ as an aid to understanding the differences and as an aid to choosing the best device for the introductory course. Let us compare properties of the popular TI-83 calculator from Texas Instruments with properties of a typical generic low-end non-portable computer system running, say, either ActivStats or the student version of Minitab. Table 2 lists some relevant properties and their values.

Note that computers lead the TI-83 on all but the last three properties, which I now discuss in turn.

First, (assuming that students have computer access at the location where they study, a cost issue—see below) the low portability of fixed computers is not a serious disadvantage because it is not necessary that students be able to perform statistical computations "on the street".

Second, the fact that calculators are permitted in some examinations is not a significant argument in favor of using calculators during the term because the use of calculators in examinations has little to do with the practical use of computers or calculators to solve real statistical problems—it has only to do with the decisions of exami-

Table 2
Values of Properties of Two Computing Devices

Property	TI-83	Computer
extensive online help available?	no	yes
screen size (pixels), which speaks to the ability of the device to display large bodies of text and detailed statistical graphics	96 × 64	800 × 600
number of statistical operations available without manually loading external programs	lower	higher
ability to run integrated software that comprehensively *teaches* statistics	low	high
main statistical software in device upgradeable with new releases?	no	yes
printing immediately available?	usually no	usually yes
ability for students to use the device for other functions (e.g., writing essays or Internet access)	low	high
color screen?	no	yes
space available for program and data storage*	.028 MB	300 MB+
programming language generations available**	2nd	2nd, 3rd, and 4th
portability	high	low
permitted in examinations?	sometimes yes	usually no
cost	low	high

* MB = million bytes. The 300 MB is an estimate and varies from computer to computer.

** 2nd generation = assembly language; 3rd generation = standard high-level programming language (e.g., C, Fortran), 4th generation = e.g., statistical package with stored user programs.

nation designers. (That is, the choice of examination designers should not drive the choice of computing device used during the term—rather, the preferred device should drive the choice of examination designers.)

Thus it appears that computers are superior to the TI-83 on all important properties except the property of cost. Thus it appears that computers are preferred to the TI-83 in all situations in which the cost problem can be solved.

Finally, it is important to note that in situations in which the cost problem *cannot* be solved, hand calculators such as the TI-83 are a practical alternative since, although they are inferior to computers, calculators are far better than nothing. I look forward to the day when all students and teachers can have the substantial benefits of full access to computers.

9. A DE-EMPHASIS OF LESS IMPORTANT TOPICS

Once we have formulated the goals of an introductory statistics course, it is important to consider each topic proposed for the course and to ask whether the topic helps to satisfy the goals. Here, while tradition has the right to speak, the choice of topics for a course should be dictated not by tradition, but by the effectiveness of the topics in satisfying the goals. Thus we must be prepared to jettison any topic that does not help to satisfy our goals.

9.1 Univariate Distributions

Many introductory statistics courses spend significant amounts of time discussing univariate distributions, including measures of center and spread and various graphical techniques for illustrating distributions, such as dot plots and box plots. Some courses also discuss the mathematical aspects of univariate distributions, focusing on density functions, distribution functions, and related ideas. Univariate distributions are emphasized for two reasons

- univariate distributions give us a simple picture of data
- the mathematical theory of univariate distributions underlies much of statistical theory.

Although many introductory statistics courses emphasize univariate distributions, I recommend that teachers de-emphasize this topic at the beginning of the course. Instead, I recommend beginning the course by introducing the concepts of entities, properties, and variables—topics that clearly precede univariate distributions since univariate distributions are distributions of the values of *variables.* After introducing entities, properties, and variables, I recommend that teachers go directly to discussing relationships between variables and prediction and control on the basis of relationships between variables.

Before I discuss why I recommend deferring treatment of univariate distributions, I should make two points:

1. An understanding of univariate distributions is *mandatory* for full understanding of the field of statistics. Therefore, I am not suggesting that the topic of univariate distributions be removed from the curriculum—I am only suggesting that it be moved later (either near the end of the introductory course or into a later course).

2. The concept of a univariate distribution is clearly *logically prior* to the concept of a relationship between variables because as soon as we have a relationship between variables we have at least two variables while, of course, univariate distributions deal with only one variable.

Although univariate distributions are important and logically prior to relationships between variables, I see four reasons why we should defer discussing univariate distributions in favor of discussing relationships between variables and prediction and control on the basis of relationships:

1. The concept of a univariate distribution plays only a *peripheral conceptual role* in real empirical research. On the other hand, the concept of prediction and control on the basis of a relationship between variables plays a pivotal role in almost all empirical research. One can see this by noting that it is almost impossible to get an empirical research paper published in a respectable journal if the paper merely reports the univariate distributions of one or more variables. On the other hand, almost all published reports of empirical research can be easily characterized as reporting information about one or more relationships between variables.
2. The concept of a univariate distribution is *not necessary* for initial study and understanding of the concept of a relationship between variables.
3. The concept of a univariate distribution is *boring* for most students because they see no practical use of the concept. On the other hand, methods for making accurate predictions on the basis of relationships between variables are of broad practical use and are therefore (when taught with important response variables) fascinating.
4. Although we can make predictions on the basis of the study of univariate distributions, these predictions are never more accurate and are usually *less accurate* than predictions made on the basis of comparable study of relationships between variables.

Some teachers have already independently adopted the approach of emphasizing relationships between variables. For example, using an idea developed by Gudmund Iversen, George Cobb teaches two introductory courses, both of which start with relationships—one devoted to experimental design and applied analysis of variance and the other devoted to applied regression (Cobb 1993, sec. 3.1). Similarly, Robin Lock teaches an introductory course devoted to time series analysis—i.e., methods for studying relationships between variables when an important predictor variable is "time" (Cobb 1993, sec. 3.1).

Some teachers may feel that certain studies of univariate distributions *are* of practical use, and thus are not boring. I discuss in appendix G how some (perhaps many) such studies can be better viewed as studies of relationships between variables.

(Empirical researchers do directly study univariate distributions when they wish to determine norms. However, even here other variables and relationships between variables play a critical role because researchers determining norms usually hold other important variables at specific constant values. Otherwise, the norms may be muddied by variation in these other variables causing [through a relationship between the variables] extra variation in the values of the variable being normed. Thus it is reasonable to defer discussing norms [= univariate distributions] until students have a good understanding of the concept of a relationship between variables.

(Another situation where univariate distributions are of interest is in sample surveys, in which the distribution of the responses on a surveyed variable is, of course, a univariate distribution. People are sometimes interested in seeing a summary of the univariate distributions of important surveyed variables. However, although the distributions of the values of individual surveyed variables are sometimes of interest, such distributions do not generally provide a focused basis for action for the body that commissioned the survey. On the other hand, *relationships between* surveyed variables on carefully designed surveys often do provide a focused basis for action, and thus are generally of significantly greater value than the simple univariate distributions of the surveyed variables.)

In summary, we should defer discussing univariate distributions in an introductory statistics course until students have a good understanding of the more interesting and more important concept of prediction and control on the basis of relationships between variables.

9.2 Probability Theory

Probability theory has traditionally played a prominent role in the introductory statistics course for at least two reasons

- probability theory is necessary to make mathematical sense of univariate distributions
- probability theory is necessary to make mathematical sense of relationships between variables.

However, I agree with Moore (1997, 128) that most discussion of probability theory should be postponed until a later course. I believe this postponement is appropriate because, as noted above, I believe discussion of univariate distributions should be downplayed in the introductory course, and thus we do not need probability theory to support univariate distributions. Also, I believe we can effectively describe relationships between variables (including *p*-values in statistical tests) for students without having to appeal to the details of probability theory. (I illustrate such an approach in the paper for students [1996a].)

9.3 Mathematical Theory of Statistics

Finally, as also suggested by Moore (1997, 127), I agree that there is no need to discuss the underlying mathematical theory of statistics in the introductory course. If we want students to develop a lasting appreciation of the role of statistics, it is much more important to directly discuss the role in non-mathematical practical terms.

10. SUMMARY

The first important feature of an ideal introductory statistics course is that it have a clear statement of the course goals, since then all course design decisions can be efficiently driven by the goals. I recommend that the main goal of an introductory course be to give students a lasting appreciation of the vital role of the field of statistics in empirical research. Such an appreciation can be obtained if we unify the field of statistics under the concepts of entities, properties, variables, and relationships between variables. It is reasonable to describe statistical tests of hypotheses as tests of the existence of relationships between variables. We can make the field of statistics come alive for students by choosing examples with response variables that students can see clear value in predicting or controlling. Multimedia teaching has significant potential for improving the introductory course. Computers are superior to hand calculators in many areas and thus are preferred if they can be afforded. To maximize students' appreciation of the role of statistics, discussion of univariate distributions, probability theory, and the mathematical theory of statistics should be de-emphasized in deference to the ideas of prediction and control on the basis of relationships between variables.

APPENDIX A: MEASURABLE GOALS

Recently the practice of goal setting in serious endeavors has seen a new emphasis—an emphasis on setting *measurable* goals. For example, at the beginning of the year an employee in a company may formally set herself the goal of earning a score of 85 percent or more on a customer satisfaction survey to be administered at the end of the year. Using measurable goals has the following advantages:

- Measurable goals are highly motivating to a person trying to meet the goals (assuming the person accepts the goals) because the person knows that a measurement will be made, and then recognition will be apportioned depending on how well the goals are met.
- Measurable goals, when properly implemented, give us a proper (i.e., valid and reliable) indication of how well the goals are actually being met.

How can we set measurable goals for the introductory statistics course? I propose in section 2 that the first goal of the introductory course be to give students a lasting appreciation of the vital role of the field of statistics in empirical research. Clearly it would be useful if we could reliably measure students' appreciation of the role of statistics.

We can measure students' appreciation with an appropriate (brief) multiple choice test, with questions like

- On a scale of one to seven, how do you rate the social importance of the field of statistics?
- On a scale of one to seven, how do you rate your interest in the field of statistics?

The test should assume a univariate property of students' appreciation of the role of statistics and should ask enough questions to obtain a reliable measure of the property. (In order to keep the test as general as possible, perhaps a definition of the role of statistics should not be stated or assumed by the test.)

The test should be constructed using the principles of psychological test design and should be proven valid and reliable using the standard methods of evaluating test designs (Cohen, Swerdik, and Phillips 1996, Gregory 1996, Kaplan 1997, Kline 1993).

The test might also ask demographic questions (including questions about students' prior education) since this information will likely help to account for some of the variation in "appreciation" from student to student.

The test should *not* test students' knowledge of statistical topics per se since students' appreciation of the role of statistics can be viewed as being mostly independent of knowledge of specific statistical topics. (However, again, questions testing knowledge of statistical topics might be used as auxiliary questions on the test to help explain some of the variation in the appreciation scores.)

The test should be designed so that it can be administered twice to students—once at the very beginning of an introductory course, and once after the course has ended. That way, we can compute and analyze within-student measures of the amount of improvement. Such measures give substantially more information about relationships between appreciation of statistics and teaching approaches than if we collect only one score from each student.

The test could be short enough that the second administration could take place as part of the final exam, yielding substantial logistics advantages. However, it must be clear to students that (1) their grade in the course is completely independent of their responses on the test, and (2) frankness in test responses is desirable if statistics courses are to be improved.

Students might also be tested (by the test developer) over the Internet, since this will be more convenient for some teachers. This also facilitates generation of a large database of test results, allowing identification of centers of excellence.

The option of scheduling the second test several months after completion of the course would also be useful, because we are interested in giving students a *lasting* appreciation of the role of statistics, not appreciation that evaporates shortly after the final exam. (Students can be encouraged to participate in the later test by an e-mail reminder system and a chance to win a prize, such as a computer, for participating.)

The score a student obtains on the test is a measure of the student's appreciation of the field of statistics. A reasonable measurable goal of an introductory course is to maximize the difference between each students' score on the second test and his or her score on the first test. I hope that statisticians familiar with educational test development will develop a valid and reliable measure of appreciation of the field of statistics that can be used in many courses.

If we develop a measure of the ability of a course to meet its goals, this measure is an obvious candidate for the response variable for experimental research in statistical education. I discuss such research in appendix B.

APPENDIX B: EXPERIMENTAL RESEARCH IN STATISTICAL EDUCATION

It is interesting that we statisticians, who are the keepers of the keys to empirical (scientific) research, perform relatively little empirical research in statistical education. Indeed, at the conference on Assessment in Statistics Courses in Boston in 1997, one speaker went so far as to speak out *against* doing experimental research in statistical education, giving several reasons why he believed such research is ill-advised. (I cover his points in the discussion below.)

Clearly, the process of finding the best approach to the introductory statistics course is in large part an *empirical* process. Thus if we have a reliable measure of the ability of a course to meet its goals, we can carry out this empirical process *formally.* Formal experimental research (when carefully performed) is more efficient at optimizing the value of a variable than a hit-and-miss approach.

The discussion in appendix A proposes a measure of the ability of a course to meet the goal of giving students a lasting appreciation of the role of statistics. This measure is thus an obvious candidate for the response variable in research in statistical education. Obvious candidates for predictor variables are variables that reflect different approaches to teaching statistics, such as the EPR approach, multimedia teaching, activity-based teaching, and traditional approaches. (Mixtures of the various approaches are easy to envision.) If we perform properly designed experiments to manipulate variables that reflect different approaches to teaching statistics, we can (in a response-surface fashion) determine the particular approach that maximizes students' appreciation of the role of statistics.

* * *

In performing experimental research in statistical education, we would like our research findings to be validly generalizable across broad groups of teachers and students. Thus we must ensure that the effects of good or bad teachers and other confounding factors are removed from the analysis.

To remove teacher and institution effects, we must perform the research across multiple teachers and multiple institutions. Thus the research is similar to a multi-center clinical trial, as used in the investigation of new medical treatments. Participating students at each institution should be randomly assigned to one of two (or possibly more) groups and each group should receive a different one of the teaching approaches under study. To maximize statistical power, steps must be taken to ensure that students do not change groups or attend the class sessions of an unassigned group. To give each approach the best opportunity for success, each approach should be taught by a teacher who is reasonably well committed to the approach.

Two challenges in experimental research in statistical education relate to (*a*) recruiting participating institutions and teachers, and (*b*) ensuring protocol adherence.

An apparent problem with experimental research in statistical education is that we cannot achieve full double-blindness, such as can be achieved in some clinical trials. In particular, we cannot prevent students from discerning which of the various treatments they are receiving. However, this is not a serious problem because none of the treatments is a placebo (i.e., the absence of treatment) and students do not generally carry preconceived opinions about approaches to teaching statistics. Furthermore, we can describe the experiment to students as part of the course, and show students how the preferred attitude is to have an open mind about which teaching approach is better until the data on the various approaches are in.

Another apparent problem with experimental research in statistical education arises with the definition of the predictor variable(s). Since the various approaches to statistical education are generally only loosely defined, perhaps the loose definitions will make it difficult or impossible to define the predictor variables in the research. However, this is not a serious problem as long as the researcher carefully defines what *he or she* means by the various predictor variables used in the research, and as long as these are reasonable definitions, and as long as protocol adherence is obtained. If those conditions are satisfied, it will matter little whether the (careful) definitions used in the research are in *exact* agreement with the common notions attached to the labels for the different approaches.

Another apparent problem arises through the analogy with clinical trials: In clinical trials much "background" research is done on a new medical treatment before it is actually used in human experiments. Perhaps similar background research must be done on new approaches to statistical education before it is reasonable to experimentally compare the approaches.

But this argument by analogy breaks down: Background research is needed in medical research

- for safety reasons and
- because inexpensive background animal experimentation is an efficient screening prelude to expensive human experimentation.

But generally we need not be concerned with the "safety" of methods of teaching statistics since the safety of approaches to teaching statistics (or an analogue of safety) is not an issue. Also, it is not possible to do preliminary experiments of approaches to teaching statistics on animals. Thus as soon as we have reasonable response and predictor variables and as soon as we have reasonable ways of dealing with other potential problems discussed above, "background" research on different teaching methods can provide relatively little assistance in experiments in statistical education.

* * *

In view of the high cost of experimental research in statistical education, we must choose the response and predictor variables carefully so that we maximize the chance of obtaining useful results. In particular I suggest it is wasteful to spend resources studying effects that most teachers already believe, such as the claim that students appreciate statistics more if concrete examples are used. Almost nobody will be surprised if research demonstrates that this claim is true. Instead, we should focus on testing teaching approaches that seem to have promise but about which there is disagreement among statistical educators. This focus will yield the highest payoff.

Can experimental research in statistical education obtain funding? Decidedly yes. A carefully presented proposal for such research is almost guaranteed to obtain funding since many statisticians and statistical organizations strongly support the need for statistical education reform, and thus will support reasonable research proposals. Furthermore, since statistics is a general but broadly misunderstood tool supporting empirical research, funding agencies will see the natural appeal of using carefully designed empirical research to study the teaching of statistics itself.

I hope that statisticians familiar with experimental research will turn their attention to experimental research in statistical education. I look forward to the day when we can make empirically supported statements that a particular approach to teaching statistics is preferred.

APPENDIX C: IS THE EPR APPROACH TOO ABSTRACT OR TOO GENERAL?

A statistics teacher whom I shall call Dr. A sent me a thought-provoking criticism of the EPR approach. He refers to another statistics teacher whom I shall call Dr. B. Dr. A writes

> Your approach seems to me to smack of what … might be called "the [Dr. B] syndrome." [Dr. B] thinks that people learn top-down, so that instruction should start with general (and necessarily abstract) notions. I think (and I think much education research suggests) that most people learn bottom-up, from concrete cases and experiences to generalizations based on them. So I prefer to start with hands-on data work and to delay (maybe until a course on epistemology) abstract notions such as entities and their characteristics.

I fully agree with Dr. A that most students learn bottom-up. That is, students find course material easiest to understand if the course begins with the simplest and most concrete ideas and builds generalization and detail atop these ideas.

Thus if the EPR approach is to be successful, it must be taught in a bottom-up fashion. This immediately raises the following question:

> Is it *possible* to teach the EPR approach to students in a bottom-up fashion beginning with concrete cases and experiences? Or are the concepts of entities, properties, and relationships too abstract or too general to be taught in this fashion?

To answer this question, let us first ask whether there are concrete examples of entities in students' everyday experience that we can use to help familiarize students with the concept of 'entity'. Obviously, the answer is Yes, because *everything* (every thing) in human reality is an example of an entity. Perhaps the easiest type of entity to understand is the type we call "physical objects". Pencils, books, and overhead projectors are all concrete examples of entities that are physical objects.

From the concept of 'physical object' we can (in a bottom-up fashion) generalize to other types of "concrete" entities present in students' reality, such as people, examinations, educational institutions, and ideas—all "concrete" examples of types of entities.

Similarly, we can easily teach the concept of a property to students using concrete cases and experiences if we begin by concentrating on properties of physical objects. We can discuss instances of properties of physical objects until students are comfortable with the concept. Then we can again generalize in a bottom-up fashion to show students that all types of entities have properties. We can also show students how it is useful to list the important properties of any type of entity we wish to study, since we can only know an entity through knowledge of its properties and knowledge of the values of the properties.

Similarly, we can teach the concept of a relationship between properties to students in a bottom-up fashion using many concrete examples. Important examples can be found in all areas of human experience. For example

- aphorisms (haste makes waste; a penny saved is a penny earned)
- education research (longer hours of study are associated with higher grades)
- medical science (drug cocktail A retards the progress of AIDS better than drug cocktail B)
- physics ($E = mc^2$).

Each of these examples (and tens of thousands like them) can be interpreted for students as a statement of a relationship between properties of entities.

Thus it is possible to teach the EPR approach to students in a bottom-up fashion, beginning with concrete cases and experiences. I illustrate steps in the paper for students (1996a).

* * *

Dr. A suggests that the concepts of the EPR approach are "abstract", and (with qualifications I discuss below) I agree. However, the abstractness of the concepts of 'entity', 'property' and 'relationship' is counterbalanced by students' intuitive familiarity with the concepts (especially the first two concepts). This familiarity comes through daily (in fact minute-by-minute) use of the concepts. That is, students (and all other humans) appear to organize their entire external reality in terms of entities and properties. And whenever students need to describe an entity (of any type) in their everyday lives, they invariably describe it in terms of its properties and the values of the properties.

It is important to note here that although we humans use the concepts of 'entity' and 'property' to organize our thinking, these concepts are usually not *conscious* in our thinking. This is because in everyday life it is almost never necessary to treat entities and properties in their full generality. Instead, in most practical situations we are interested in a particular *type* (or types) of entities, and these are best referred to by their type names, rather than by treating them in their full generality as entities. Likewise, we almost never need to treat properties in their full generality, but instead we treat them as specific instances of properties. Thus it takes some initial effort to get students (and some teachers) to recognize that the whole of human external reality can be readily viewed as consisting of a set of many types of entities, with each type having a set of properties. It is clear, however, that students can learn these concepts because the concepts merely reflect how (at a deep level) the students already organize (all?) their thinking.

* * *

Since the concepts of entities, properties, and relationships are "abstract", it is important to ask whether the abstractness is an advantage or a disadvantage. To answer this question, it is helpful to note different senses of the adjective "abstract", seven of which are given in a popular dictionary (Merriam-Webster 1993). I believe the concepts of entities, properties, and relationships are very abstract in five of the seven senses, but not in the other two.

In particular, I do not believe that the three concepts are abstract in the sense of "difficult to understand : abstruse". Instead, I suggest that students can easily understand the concepts because (as I discuss above) the concepts are already ubiquitous in students' thought, just below the surface of consciousness.

Similarly, I do not believe that the concepts of entities, properties, and relationships are abstract in the sense of "insufficiently factual : formal". Instead, whenever the three concepts are used in empirical research they inherit the factualness of the specific entities, properties, and relationships that are being studied in the research.

The dictionary gives the other five senses of the adjective "abstract" as follows:

- disassociated from any specific instance <*abstract* entity>
- expressing a quality apart from an object <the word *poem* is concrete, *poetry* is *abstract*>
- dealing with a subject in its abstract aspects : theoretical <*abstract* science>
- impersonal, detached <the *abstract* compassion of a surgeon—*Time*>
- having only intrinsic form with little or no attempt at pictorial representation or narrative content <*abstract* painting>.

For these five senses of the word "abstract", I believe that the concepts of entities, properties, and relationships are very abstract. However, instead of being a disadvantage, this abstractness is a significant *advantage* because it enables the EPR approach to make true broad unifying generalizations. As discussed in section 4, the main generalization is

> The field of statistics can be usefully viewed as a set of optimal techniques to help empirical researchers study variables and relationships between variables (relationships between properties of entities) mainly as a means to predicting and controlling the values of variables.

* * *

In the last sentence of the quotation, Dr. A states that he prefers to omit discussing entities and properties in the introductory course. Instead, he prefers to start with "hands-on data work". Here, he is saying that he prefers to start with hands-on data work with *variables,* since any work with data is work with variables.

But if students must work with "variables", should we not tell them what a "variable" is? After all, if we gloss over the important concept of 'variable', how can we expect students to understand what the "hands-on data work" is about? What is a variable if not a formal representation of a property of entities?

* * *

Dr. A's principle that we should start from concrete cases and experiences raises another interesting question

> Which concept is more concrete and fundamental in human thought—the concept of 'property' or the concept of 'variable'?

It is easy to see that the concepts of 'entity' and 'property' are more concrete and fundamental than the concept of 'variable'. In particular, it is clear that children attain (unconscious) mastery of the concepts of 'entity' and 'property' before they enter kindergarten. That is, before children enter kindergarten they have already developed the ability to understand and (in a limited way) use nouns, adjectives, and adverbs, which are parts of speech that denote entities (nouns) and values of properties (adjectives and adverbs). On the other hand, children do not master the concept of 'variable' (if they ever do) until it is taught to them in school (after they have first learned arithmetic). Thus the concepts of 'entity' and 'property' precede the concept of 'variable' in human thought. Thus the concepts of 'entity' and 'property' are more concrete and more fundamental.

* * *

Another important question is

> How much time should a teacher spend discussing the concepts of entities, properties, variables, and relationships at the beginning of an introductory course?

The answer to this question depends on the type of students in the course. My experience suggests that it takes between fifteen minutes and three class sessions to introduce students to the concepts of entities, properties, variables, and relationships. (Longer discussions use more examples or discuss philosophical ramifications of the concepts.)

A problem with the EPR approach occurs if a teacher covers the concepts of entities, properties, variables, or relationships too quickly. Thus teachers who are new to the material must carefully assess students' understanding before moving from one topic to the next. As the song says, "House built on a weak foundation will not stand, oh no."

A second problem with the EPR approach occurs if a teacher covers the concepts of entities, properties, variables, and relationships at the beginning of the course, but then forgets to link the various concepts covered later in the course back to the earlier foundational concepts.

* * *

Dr. A continues (and concludes) his comments as follows:

> I doubt that many students are prepared for the larger questions until they have worked informally for quite a while.

Perhaps Dr. A's reasoning here can be expanded as follows:

> Students frequently view statistics as the worst course taken in college (Hogg 1991). That is, many students have trouble with the various standard topics discussed in the typical introductory course. Since students have trouble with the standard topics of statistics, how could we ever expect them to understand the unifying generalizations (which Dr. A calls the "larger questions")? Therefore, we should omit discussing the unifying generalizations and instead concentrate for quite a while on the standard topics of statistics.

I suggest that the foregoing reasoning is backwards. I believe that if students do not understand the unifying generalizations, it is much harder for them to understand the standard topics of statistics. This is so because the unifying generalizations appear to unify everything, including the standard topics. I believe that trying to teach an introductory statistics course without careful discussion of entities, properties, variables, and relationships at the beginning is like trying to teach arithmetic to grade school students before they have learned to count.

APPENDIX D: THE EFFECTS OF DIFFERENT TYPES OF STUDENTS ON THE DESIGN OF THE INTRODUCTORY STATISTICS COURSE

The question arises whether the EPR approach is appropriate for all types of introductory statistics courses, or only for some. To address this question, let us first consider how we might categorize a group of students in an introductory statistics course in terms of properties of the students. Following are some properties on which we can categorize students:

- age
- general intelligence
- degree of interest in empirical research
- level of mathematical achievement
- level of mathematical interest
- time available for homework
- likelihood that the student will engage in later empirical research
- and so on.

As well as categorizing the students, we can also categorize courses according to various properties of the course, such as the number of class hours in the term, special topics that the syllabus dictates must be covered, and so on.

If we consider various categories of students and courses according to one or more of the above properties, do we find that we should teach the EPR approach in some introductory statistics courses, but we should teach another approach or approaches in other introductory statistics courses? For example, a statistics teacher whom I shall call Dr. C wrote to me that for most students the EPR approach

> ... would be very appropriate; an exception might be students with a strong interest in mathematics who would need emphasis first on interesting mathematical formulations leading on to more important conceptual issues.

To evaluate Dr. C's point, let us evaluate the importance of the concept of a relationship between variables. I believe this concept is a candidate for the most important concept in statistics (and perhaps the most important concept in science), since (as I discuss in section 4) much of scientific research and much of statistics can be characterized as study of relationships between variables.

In view of the high importance of the concept of a relationship between variables, and because no other statistical concept appears to be of greater importance in empirical research, I suggest that all students in introductory statistics courses should be taught the concept of a relationship between variables at the beginning of an introductory course. Furthermore, since the EPR approach is simply a logical breaking apart of the concepts that lead up to the concept of a relationship between variables, the EPR approach would appear to be the most reasonable approach to use in the beginning parts of the introductory statistics course for all types of students.

Once students have properly learned the concept of a relationship between variables, it makes sense to choose the next topics according to student and course properties. For example, if it is likely that the students in a course will later engage in empirical research, it makes sense to orient the course toward the design and analysis of empirical research projects. On the other hand, if it is unlikely that the students will engage in empirical research, it makes sense to orient the course toward understanding and interpreting the results of empirical research done by others as, for example, exemplified by Snell (1998). Similarly, if the students in a course are highly interested in mathematics, then after they have learned the empirical (as opposed to mathematical) concept of a relationship between variables it makes sense to introduce some of the mathematical ideas that underlie the concept.

APPENDIX E: LANGUAGE ISSUES

A statistics teacher whom I shall call Dr. D wrote to me that

> …The language of cases and variables offers an already-established alternative to entities and properties. I think it's a very hard argument to persuade people that the language of entities and properties offers enough advantages to compensate for giving up what's become a standard language.

Dr. D raises the question: Do the terms "entity" and "property" have any significant advantages over the terms "case" and "variable"? Let me split this language question and address it in the next two subsections.

E.1 "Entity" Versus "Case"

I believe a large part of any individual human reality can be usefully viewed as a mass of knowledge and belief about

- entities
- relationships between entities
- properties of entities
- preferred values of properties
- actual values of properties and
- relationships between properties.

Of course, wherever the term "entity" appears in the preceding sentence, we might reasonably replace it with the term "case". So how shall we decide which term is preferable—"entity" or "case"?

I suggest our decision should be driven by a single simple criterion: Use whichever term gives better overall student understanding. I believe the term "entity" gives substantially better understanding than the term "case" for four reasons

1. The term "entity" has a concrete and tangible connotation—an entity is a *thing* out there in the external world. On the other hand, the term "case" has an abstract and intangible connotation—a case is an *instance* of the members of some class of things. I believe the abstract connotation of the term "case" makes the term more difficult for students to understand than the more concrete term "entity".
2. Cases often seem to be something invented by humans (e.g., court cases), while entities are the actual "real things" in each person's external reality, whether invented by humans or not. Thus entities have a greater perceived realism (and generality) in the external world than cases.
3. The term "case" is presently in use in a different sense in some introductory statistics courses—especially introductory business statistics courses. This sense is as in "case study", in which a particular entire problem and its solution are referred to as a "case".
4. If we consider various practical situations, the term "entity" seems a more natural name for the things under study. For example, consider physical objects—in particular, consider bicycles. Are bicycles better viewed as being entities (things) or are they better viewed as being cases—possibly instances of the platonic concept of 'bicycle'? For me, if I need a very general classification term to refer to bicycles, it seems more natural to refer to them as entities (things) rather than as cases (instances).

(I view the words "entity", "thing", "object", "unit", and "item" as being synonymous. I prefer the word "entity" in formal discussion because the words "thing", "object", "unit", and "item" have a nebulous ring to them—they are names we use for something when the correct name is unavailable, or when the thing being referenced is not the center of attention. Also, the words "thing", "object", "unit", and "item" seem less appropriate if the entities we are studying are living organisms, especially people. On the other hand, the word "entity" has a specific, concrete, and attention-grabbing ring to it, and seems appropriate for all types of things, including people.)

Try reading the following paragraph substituting the word "entity" for the blanks:

> What word should we use for the _____s that populate human reality? By _____s I don't just mean physical objects, although physical objects are an important type of _____. But there are many other types of _____s, for example, universities, songs, and mathematical vectors. If we want people to understand the role that _____s play in statistical thinking, we should choose a word that
> 1. has a concrete connotation
> 2. catches people's attention
> 3. is appropriate for every type of _____ in human reality, including living _____s and
> 4. is consistent with everyday use.

Try reading the paragraph substituting other words for the blanks, such as "thing", "object", "unit", "item", "case", and "instance". Although "thing", "object", "unit", and "item" work satisfactorily, "entity" works best overall for me.

In view of the preceding points, I believe students obtain substantially better understanding if we use the term "entity" (thing, object, unit, item) instead of the term "case" (instance).

I give some historical facts about the statistical use of the term "case" in appendix F.

E.2 "Property" Versus "Variable"

Consider the second part of Dr. D's question:

> Does the term "property" have any significant advantages over the term "variable"?

Unlike the situation with "entity" and "case", I believe the ideas behind the terms "property" and "variable", although closely related, are *not* (theoretically) the same. I reason as follows:

Consider the concept of 'variable'. As noted in section 3, since this fundamental concept is involved in almost all discussions in the field of statistics, we must provide students with a careful definition and discussion of the concept at the beginning of the introductory course. As also noted in section 3, I believe a reasonable definition of "variable" is

> A *variable* is a formal representation of a property of entities.

This definition appeals to the concepts of 'entity' and 'property'. Therefore, a teacher using the definition must first introduce students to those two concepts. Thus if we use the above definition of "variable", the concepts of 'property' and 'variable' are *not* interchangeable.

Thus, returning to Dr. D's suggestion that it is a hard argument to persuade people to use the term "property" instead of the term "variable", fortunately it is not necessary to attempt this argument. Instead, we must use *both* terms, with the concept of 'variable' being *defined* in terms of the concept of 'property'.

APPENDIX F: THE STATISTICAL USE OF THE TERM "CASE"

Dr. D's comment in appendix E about the term "case" being part of a "standard language" suggests that it is of interest to consider the history of the term "case" in statistical discussion. The term received substantial impetus from three ground-breaking statistical software packages—BMDP, SPSS, and Data Desk. The designers of these packages chose the term "case" to denote the objects (entities) associated with the rows in the standard statistical data table.

BMD (later to be BMDP) used the term "case" to name these objects in the first edition of the BMD manual (Dixon 1964). When SPSS (Nie, Bent, and Hull 1970) and Data Desk (Velleman and Lefkowitz 1985) were introduced, they also used the term "case" to name the objects associated with the data-table rows and, as far as I understand, they have all used this term ever since. Other packages, however, use other terms: GENSTAT and SAS use "observation" (a less abstract term than "case"); Minitab and S use "row" (perhaps the most abstract of the three terms).

Often a one-to-one mapping exists between the main entities studied in an empirical research project and the rows in the data table obtained in the research project. However, this mapping does not always occur. For example, the rows in the data table may be associated with *trials* in the research project, instead of with the main entities studied in the research project, e.g., human subjects. Because the main entities in a research project are not always associated with the data-table rows, we cannot in general refer to the rows as "entities". Hence the terms "case", and "observation" are useful general names for the objects associated with the rows.

* * *

In a statistical dictionary for social scientists, Vogt (1993) defines the term "case" as follows:

> **Cases** The subjects, whether persons or things, from which data are gathered. A

> case is the smallest unit from which the researcher collects data. Compare "unit of analysis".

Note the difference between Vogt's definition and the data-table-row definition used by BMDP, SPSS, and Data Desk, which Vogt has missed. The publication dates and similarity of meaning imply that Vogt's definition is the etymological child of the data-table-row definition.

Vogt also gives the following definition:

> **Units of Analysis** The persons or things being studied in a research work. Units of analysis in research in the social and behavioral sciences are often individual persons but may be groups, political parties, newspaper editorials, ..., and so on. A particular unit of analysis from which data are gathered is called a case.

Thus for Vogt, cases and units of analysis are synonymous. Note that Vogt defines both cases and units of analysis in terms of two other more fundamental concepts: 'persons' and 'things', both of which are types of entities.

The term "case" is not defined by Marriott (1990), nor by Kotz and Johnson (1982), nor by Kruskal and Tanur (1978), so the term is not strongly established in the statistical lexicon.

APPENDIX G: RELATIONSHIPS BETWEEN VARIABLES VERSUS UNIVARIATE DISTRIBUTIONS

After hearing my comment that univariate distributions are boring, a statistics teacher whom I shall call Dr. E wrote to me with two apparent counterexamples. Dr. E writes

> … there are some, perhaps many, instances where univariate distributions are both interesting and illuminating. For example, my students seem to enjoy looking at distributions of tuition charges in colleges across [a certain geographical area]. They see distinct clusters corresponding to public and private schools,

It is reasonable to view this example as a study of a univariate distribution. However, it is also reasonable to view it as a study of a relationship between two variables (i.e., a relationship between two properties of entities) as follows:

Population of Entities:	colleges in a certain geographical area
Response Variable:	tuition fee charged by a college
Predictor Variable:	sector of a college (i.e., public or private)
Statistical Questions:	1. Is there a relationship in this population of colleges between the tuition fee that a college charges and the sector to which a college belongs? 2. If there is a relationship, how can we best predict (or control) the value of the response variable (tuition fees) on the basis of the relationship? 3. How accurate will the prediction (or control) be?

Thus we can view the example as being simply a study of a univariate distribution, with no need to invoke the concept of a relationship between variables. Alternatively, we can view the example as being a study of a relationship between two variables.

Dr. E next writes

> and they [students] like seeing where their school fits into the distribution.

This clause underscores the fact that Dr. E has chosen an important response variable—a response variable that students are highly interested in predicting and controlling. Therefore, students are interested in any relationships between tuition fees and other variables (such as "sector" or other properties of the colleges including the variable "college name").

Dr. E continues

> I'm also not sure where *comparisons* between distributions fits into your thinking. This is a form of prediction, I suppose. Again, many problems are both interesting to students and illustrative of the power of statistics: comparing male and female salaries, for instance.

Dr. E views the salary example as a "comparison" between two distributions, again with no reference to the concept of a relationship between variables. However, it is also possible to view this example as a study of a relationship between two variables as follows:

Population of Entities:	members of some specific group of people
Response Variable:	a person's salary
Predictor Variable:	a person's gender
Statistical Questions:	1. Is there a relationship in this population of people between a person's salary and a person's gender? 2. If there is a relationship, how can we best predict the value of the response variable (salary) on the basis of the relationship? 3. How accurate will the predictions be?

Thus we can view both of Dr. E's examples as studies of univariate distributions or we can view them both as studies of relationships between variables. Which point of view is preferred?

Note that the "univariate distribution" point of view is somewhat vague. Only one variable is present—the response variable. And the predictor variables in the "relationship between variables" point of view (i.e., "sector" and "gender" respectively in Dr. E's examples) fade into the background, thereby hiding the commonality between these research projects and other research projects that *are* viewed as studying relationships between variables.

On the other hand, the "relationship between variables" point of view is less vague and is fully consistent across both Dr. E's examples and across a broad range of other research projects in the sense that each research project can be interpreted in terms of the schema I discuss in section 4.

Since the "relationship between variables" point of view subsumes Dr. E's "univariate distribution" point of view, and since the "relationship between variables" point of view has substantially broader applicability than the "univariate distribution" point of view, I suggest that the "relationship between variables" point of view is preferred.

I give a proof that relationships between variables subsume univariate distributions in appendix H.

APPENDIX H: PROOFS OF TWO THEOREMS (RELATIONSHIPS SUBSUME DISTRIBUTIONS)

This appendix gives proofs of the two simple theorems that illustrate how the concept of a relationship between variables subsumes the concept of a univariate distribution.

Theorem: All studies of univariate distributions can be subsumed under the concept of a relationship between variables.

Proof: Any example that purports to be a study of a univariate distribution can be assigned to one of two categories

1. those examples (like Dr. E's in appendix G) that have other variables present and which can thus also be viewed as studies of relationships between variables
2. those examples that do not have other variables present and thus are true studies of univariate distributions.

For any example in the first category, we can view the example as a study of a relationship between variables, and thus the theorem is satisfied for this category.

For any example in the second category, we have a *limiting case* of a study of a relationship between variables. Specifically, this is the case in which the number of predictor variables is reduced to zero. This limiting case idea applies tightly both from the empirical research point of view and from the formal mathematical statistics point of view.

The limiting case aspect becomes clearer if we consider the schema for characterizing empirical research projects I discuss in section 4. Research projects that are true studies of univariate distributions are easily interpreted in terms of the schema in the sense that they fully satisfy the schema except that

- the set of predictor variables is empty
- the statistical questions are reduced to:
  1. How can we best predict the value of the response variable in new entities from the population?
  2. How accurate will the predictions be?

Thus the theorem is satisfied for the second group of examples, and thus the theorem is proved.

* * *

Theorem: No studies of relationships between variables can be subsumed under the concept of a univariate distribution.

Proof: Whenever we have a study of a relationship between variables we have at least two variables, but a univariate distribution takes account of only a single variable. Therefore, a univariate distribution cannot take account of a relationship between variables, and thus the theorem is proved.

REFERENCES

Bisgaard, S. 1991. Teaching statistics to engineers. *American Statistician.* 45:274-283.

Cobb, G. W. 1992. Teaching statistics. In *Heeding the Call for Change: Suggestions for Curricular Action,* edited by L. A. Steen. Washington, D.C.: Mathematical Association of America.

Cobb, G. W. 1993. Reconsidering statistics education: A National Science Foundation conference. *Journal of Statistics Education* 1(1). Available at http://www.stat.ncsu.edu/info/jse/

Cobb, G. W. 1997. *An Electronic Companion to Statistics* (computer software). New York: Cogito Learning Media.

Cohen, R. J., M. E. Swerdik, and S. M. Phillips. 1996. *Psychological Testing and Assessment: An Introduction to Tests and Measurement.* 3d ed. Mountain View, CA: Mayfield.

Cryer, J. D. and G. W. Cobb. 1997. *An Electronic Companion to Business Statistics* (computer software). New York: Cogito Learning Media.

DASL Project. 1998. This library of data sets is available at http://www.stat.cmu.edu/DASL/

Dixon, W. J. ed. 1964. *BMD Biomedical Computer Programs.* Los Angeles: Health Sciences Computing Facility, University of California.

EESEE. 1998. This library of data sets is planned to be released in the fall of 1998. It will be available at http://stat.mps.ohio-state.edu/projects/eesee/

Freedman, D., R. Pisani, R. Purves, and A. Adhikari. 1991. *Statistics.* 2d ed. New York: Norton.

Garfield, J. 1995. How students learn statistics. *International Statistical Review,* 63:25-34.

Gregory, R. J. 1996. *Psychological Testing: History, Principles, and Applications.* 2d ed. Boston: Allyn and Bacon.

Hogg, R. V. 1990. Statisticians gather to discuss statistical education. *Amstat News,* no. 169 (November):19-20.

Hogg, R. V. 1991. Statistical education: Improvements are badly needed. *The American Statistician* 45:342-343.

Kaplan, R. M. 1997. *Psychological Testing: Principles, Applications and Issues.* 4th ed. Pacific Grove, CA: Brooks/Cole.

Kline, P. 1993. *Handbook of Psychological Testing.* London: Routledge.

Kotz, S. and N. L. Johnson, eds. 1982-1988. *Encyclopedia of Statistical Sciences.* 9 vols. New York: John Wiley.

Kruskal, W. H. and J. M. Tanur, eds. 1978. *International Encyclopedia of Statistics.* 2 vols. New York: Free Press.

Macnaughton, D. B. 1996a. The entity-property-relationship approach to statistics: An introduction for students. Available at http://www.matstat.com/teach/

Macnaughton, D. B. 1996b. The introductory statistics course: A new approach. Available at http://www.matstat.com/teach/

Macnaughton, D. B. 1997. Response to comments by Samuel M. Scheiner. Published in sci.stat.edu and EdStat-L on February 24, 1997. Available at http://www.matstat.com/teach/

Macnaughton, D. B. 1998. Review of ActivStats 2.0. Available at http://www.matstat.com/teach/

Marriott, F. H. C. 1990. *A Dictionary of Statistical Terms.* Harlow, England: Longman.

Merriam-Webster, Incorporated. 1993. *Merriam-Webster's Collegiate Dictionary.* 10th ed. Springfield, MA: author.

Moore, D. S. 1997. New pedagogy and new content: The case of statistics (with discussion). *International Statistical Review* 65:123-165.

Nie, N., D. H. Bent, and C. H. Hull. 1970. *SPSS: Statistical Package for the Social Sciences.* New York: McGraw-Hill.

Rossman, A. J. 1996. *Workshop Statistics: Discovery with Data.* New York: Springer-Verlag.

Scheaffer, R. L., M. Gnanadesikan, A. Watkins, and J. A. Witmer. 1996. *Activity Based Statistics: Instructor Resources.* New York: Springer-Verlag.

Singer, J. D., and J. B. Willett. 1990. Improving the teaching of applied statistics: Putting the data back into data analysis. *The American Statistician* 44:223-230.

Snell, J. L. 1998. Chance Project. Information is available at http://www.geom.umn.edu/locate/chance/

StatLib. 1998. This library of data sets and other statistical information is available at http://lib.stat.cmu.edu/

Tanur, J. M., F. Mosteller, W. H. Kruskal, E. L. Lehmann, R. F. Link, R. S. Pieters, and G. R. Rising, eds. 1989. *Statistics: A Guide to the Unknown.* 3d ed. Pacific Grove, CA: Wadsworth and Brooks/Cole.

Velleman, P. F. 1998. *ActivStats* (computer software). Reading, MA: Addison Wesley Longman.

Velleman, P. F. and J. M. Lefkowitz. 1985. *The Data Desk Handbook.* Ithaca, NY: Data Desk, Inc.

Vogt, W. P. 1993. *Dictionary of Statistics and Methodology.* Newbury Park, CA: Sage.

Witmer, J. A. 1997. DASL—The data and story library. *American Statistician* 51:97-98.

(version of July 7, 1998)

CONTENTS