From **Gal, I. & Garfield, J. B. (editors).** **The Assessment Challenge in Statistics Education**. IOS Press, 1997 (on behalf of the ISI). Pages 85-90. ISBN 90 5199 333 1. Copyright holder: International Statistical Institute. Permission granted by ISI to the IASE to make the chapters from this book freely available on the Internet. The References section is a separate chapter and is available on the Internet at http://www.stat.auckland.ac.nz/~iase/publications/assessbkref.

7. Simple Approaches to Assessing Underlying Understanding of Statistical Concepts

Anthony E. Kelly
Finbarr Sloane
Andrea Whittaker

Purpose

Statistics should be introduced with clear linkages to the mathematics that students already understand and within contexts that students find meaningful. Otherwise, students may learn statistics in a rote fashion or apply statistics in a merely instrumental fashion and draw erroneous conclusions from data. In this chapter we present two examples of the use of simple assessment techniques that uncovered students' poor understanding of statistical concepts.

INTRODUCTION

Many people reason in ways that contradict accepted statistical models (Tversky & Kahneman, 1971; Fong, Krantz, & Nisbett, 1986; Konold, 1991a,b). Among 11- to 16- year-olds, statistical reasoning and routines are often viewed as arbitrary and inaccessible (Green, 1983). Problems in learning statistics persist even among researchers in the behavioral sciences (e.g., Greer & Semrau, 1984) and in medicine (e.g., Clayden & Croft, 1989). Medical researchers frequently confuse: histogram and barchart, correlation, standard deviation and variance (Clayden & Croft, 1989).

Completing a course in statistics does not inevitably lead to statistical insight. In a number of studies, students in statistics courses were found to: (a) describe rather than justify their statistical solutions (Allwood & Montgomery, 1982); (b) fail to establish a conceptual base for their solution strategies (Allwood & Montgomery, 1981); and, (c) when faced with errors in their statistical solutions misjudge their errors as correct (Montgomery & Allwood, 1978a, 1978b), or ignore their incorrect substeps when accounting for their solutions (Allwood & Montgomery, 1981).

Many students learn statistics as a set of rules without always learning the meaningful contexts in which they should be applied. Skemp (1979) describes mere rule-based learning as "instrumental understanding" consisting of recognizing a task for which one knows a particular rule. What we wish to strive for in our instruction is learning with meaning, what Skemp calls "relational understanding." In relational understanding, one relates a task to an appropriate schema or model; one does not blindly apply rules.

The difficulty for teachers of statistics is to recognise those cases in which instrumental understanding is being passed off as relational understanding. One way to guard against students' learning without understanding is to adopt assessment techniques that allow the teacher some insight into students' thinking. These techniques should be relatively easy to use, and economical of the teacher's time. This chapter will illustrate with two examples how assessment approaches that focus only on computational aspects of statistics may miss misunderstandings that can be exposed by judicious choice of tasks, and interview questions. In the first example, students' approaches to describing simple data were examined. In the second example, students' understanding of the analysis of variance technique were explored.

DESCRIBING SIMPLE DATA

Twenty-five graduate students took a non-compulsory, introductory-level statistics class. In the course, they were instructed in the use and value of plots from an exploratory data analysis (EDA) perspective, and were introduced to the SAS statistics package.

Assessing shape is an important first step in data description, whether the data summary is to be graphic, numeric, or verbal (see Moore, 1990). Data distribution can be assessed by drawing a bar chart (for a discretely measured variable), a histogram (for a continuously measured variable), or a stem-and-leaf plot. The students were asked to respond to the following question:

> The 'simplest' form of statistics is to summarize a set of univariate data. Summarize the following data in whatever way you think is appropriate. The data refer to the heights (in meters) of 20 women who are being investigated for a medical condition:

1.52	1.60	1.57	1.60	1.75	1.63	1.55	1.63	1.55
1.65	1.55	1.65	1.60	1.68	2.50	1.52	1.65	1.65

To test for mere instrumental understanding of statistical techniques, there are two deliberate misprints in this problem: (a) there are only 18 measures (not 20 as claimed); and (b) the observation of an 8-foot woman (2.50 meters) is almost certainly an error. The first misprint is designed to capture mechanical applications of procedure. The second is designed to capture mindless acceptance of data as "given ." A central danger in statistics education is that we neglect to tell our students that statistics deals not with numbers, but with numbers in context (Moore, 1990). For the latter error, a student could be expected to either note the data-entry error and provide a biased estimate of the mean, change it to 1.50 meters or other reasonable entry, and note the correction, or omit it and calculate a mean with the remaining 17 measures. In all cases some justificatory statement arguing that the 2.50 meters is a miskeying on the grounds that 8-foot tall females are not likely to exist would be expected.

Categories of responses

We have categorized the student responses in four ways: mindless reliance on statistics packages, generating a mean value without plotting the data, plotting the data and generating a mean value uninformed by the plot, and plotting the data and linking the choice of appropriate statistic to the distribution of the data.

Mindless reliance on statistics packages. One student responded, "I would enter these [data] into SAS and use the PROC UNIVARIATE function. This would give us all the info we needed and more: including mean, mode, median, standard deviation, variance, range, etc." The student failed to link the choice of appropriate statistic to any underlying distributional assumptions. Note that, ironically, the statistics package would generate the correct sample size.

Generating a mean value without plotting the data. The concept of the mean may appear so simple and unambiguous that adult students should not have any difficulties in understanding or using it. The mean can be seen frequently in everyday life (for example, in professional and college sports, in the construction of high school, college and graduate school GPA's). Most data reported in professional journals are means and inferential statistics that deal almost exclusively with means and mean differences, yet the mean causes problems for many learners (Pollatsek, Lima and Well, 1981). In this study, ten students provided a mean value without the use of a graphical method. Of these, four students trusted the claim that there were 20 measures in the study. These four students, consequently, reported a *mean value that lay outside the range of the data*— impossible for a correctly computed mean.

Plotting the data and generating a mean value uninformed by the plot. Students who plotted the data used a variety of graphical methods. These included frequency tables, stem-and-leaf plots and histograms. Six students constructed frequency tables of the sort displayed below:

Table 1: Example Student Frequency Table

Measures	Frequencies
1.52	ll
1.55	lll
1.57	l
1.60	lll
1.63	ll
1.65	llll
1.68	l
1.75	l
2.50	1

As the intervals between measures are uneven, tables of this type do not help the students see the effect of the outlying value. Moreover, two of the six students summed the number of frequencies to 20—the number provided in the problem statement—although their tallies indicate that there are 18 measures. The two students who constructed histograms failed to hold the areas, or the intervals constant across the x-axis. This resulted in their inability to see the measure 2.50 meters as being an outlier.

Plotting the data and linking the choice of statistic to distributional assumptions. Of the remaining eight students who constructed stem-and-leaf plots, five reported a biased mean value. One of these students, when probed, noted that, "graphing data is the right thing to do. You graph your data first, and then get your statistics." Only three of these eight noted 2.50 as an outlier when generating a mean value. These same students also realized that only 18 observations were available to them and moreover they based their choice of statistic on the distributions they had constructed. Student #17 constructed the following display, and noted, "Clearly woman 15 is very tall and should be considered an outlying value."

Table 2: Stem-and-leaf Display (student #17)

2.5	0
2.4	
2.3	
2.2	
2.1	
2.0	
1.9	
1.8	
1.7	5
1.6	0033550855
1.5	275552

From the interview data it became apparent that all of the students had at least an instrumental level of understanding. Most of the students described here were able to employ a rule (or procedure), but many saw little or no connection between the generated answers and their meaningfulness, or the meaningfulness of the original data themselves.

ANALYSIS OF VARIANCE

Graduate students who had scored either a "B" or better in a course on analysis of variance were interviewed. To earn a grade of "B," the students had to successfully complete periodic assignments in which statistics packages were used to analyze prepared data sets, and to complete a final exam involving computations. In this study, students were able to mask their conceptual confusion during the course when they worked on prepared data sets (on paper and on the computer), and when they took the final written exam. Here we characterize some of the misconceptions of students who had "successfully passed" the course.

The students were asked, individually, to: (a) talk about the new statistical concepts in their own words; (b) to apply the statistical concepts to data sets that were *unlike* the ones that they had met in the course; and (c) explain how the statistical concepts would help them interpret the results of experimental studies.

Eight of eleven students interviewed had memorized fragments of statistical knowledge that were sufficient to "earn" good grades on written examinations, but that did not form a solid basis for advanced statistical knowledge. Without concerning ourselves with the details of analysis of variance, we can say that the students' responses could be characterized in a number of distinct ways. We again noted evidence of merely instrumental understanding of statistics.

Four students knew the appropriate goal of the statistical technique, but were unable to discuss the technique with understanding (e.g., "ANOVA is for looking at means, so I would use ANOVA."). In response to questions, they parroted poorly remembered statistical fragments or labels (e.g., "I would get the 'mean square within' and the 'mean square between"). They were like tourists with a destination who could not locate the foreign language phrase book.

Two other students were like tourists who had mastered only the grammar of the language. They replied to conceptual questions with only mathematical statements or descriptions of computational routines. They were unable to connect the manipulations to word problems (e.g., "To solve this problem I would divide the 'mean square within' by . . .")

Two further students evidenced the opposite problem: they could interpret the terms in the statistical model using everyday vocabulary, but could not link that understanding to the underlying mathematics (e.g., "Well, you have a measure of the signal in the data and a measure of the noise in the data, and you place them in a ratio, but I am not sure which numbers to use"). They were like people who can drive, but who are powerless when the car malfunctions. When asked to "look under the hood" of the statistical techniques, they were unable to explain their higher-level reasoning in terms of the underlying equations.

One insight into instrumental understanding was the following. The probability of a Type I error—claiming there is a difference among groups when the null hypothesis is true— is conventionally set at $p < .05$. One of the students, who could apply this rule correctly on written assignments, when interviewed, argued that the probability of a Type I error should be large as possible because a p valueless than .05, was *such a small number* that it implied that the differences among the groups must be *insignificant:* the reverse, of course, being the case.

Clearly, for teaching and assessment purposes, a distinction must be drawn between instrumental and relational learning, between computational skill and statistical expertise. In both of these studies, computational assignments provided the students with the opportunity to mask poor relational understanding with (somewhat) effective instrumental understanding. When "traps" were set in test items, and the students were interviewed, their misconceptions came to light.

IMPLICATIONS FOR ASSESSMENT

Realize that the students must construct their own meaning for new concepts, but sometimes resign themselves to learning material mindlessly. If students do not make sense of what we teach both in terms of their own day-to-day language, and in terms of the language of mathematics, they may compartmentalize it, and respond to tasks instrumentally or robotically. As we have seen, both simple and advanced statistical tests can be calculated mindlessly. To guard against this, we must (in our teaching and assessment) explicitly link the modeling language of mathematics to the conceptual model we are developing (see, for example, Lovie, 1978 and Lovie & Lovie, 1976). If we do not link more familiar concepts with statistical routines, we risk teaching (and later assessing) an easily forgotten cryptology such as "PROC UNIVARIATE" or "mean square between divided by mean square within" whose symbolic mathematical explication is an even more alchemical incantation: "Sigma X sub i sub j..."

Don't just talk, listen. It does not profit our students to "talk at them" a subject matter that, if instrumentally learned, will have no value beyond the "final" for the course. Worse still, students may leave the course with an unjustified sense of mastery. We must design tasks that require students to be mindful, tasks that eschew the sanitized and "perfect" items of statistics textbooks. Then we must listen to their responses. Do not use the form: Here is the problem, calculate an answer. This approach failed in the written examinations described above. Rather, we should adopt the role of the naive listener who seeks continuous clarification: "You told me that I needed to look at the data. What does that imply here? I understand that a mean is a 'measure of central tendency.' What would it look like in this data set? Is it the same thing as a mode? Why not? Is it better than a mode? Can there be more than one mean in a data set? Explain your procedure to me in your own words," and so forth. Or more generally, we can ask, "Does the data make sense to you? Does your answer make sense to you?"

Listen not just to the content of the response; listen also for the affect and motivation of the student. Does the student sound confident when giving the response? As you listen, does the image come to mind of a traveler striding confidently along boulevards or a tourist fumbling sheepishly through an English-Statistics/Statistics-English dictionary? If it is the latter, remember that assessment is best when it is in the service of the student. Do not embarrass the lost traveler. Turn the student's bewilderment into an opportunity to revisit a concept that may be a concern for more than this one student.

Don't be misled by correct-sounding answers or flawless technique. As we listen to our students, we must guard against the error of assuming that, just because we hear the correct-sounding terminology, the students understand what they are talking about. Some students learn quickly how to manipulate mathematical symbols so that they get an "answer" (see Dallal, 1990). We must constantly look beyond the information given to us. Some high school students were asked by the first author to give an example of a "variable." They answered, "x." When asked what "x" was, they replied, correctly, but circularly, "A variable"! When asked for a *different* example of a variable, they replied, "y"!

Statistics is an interpretive science. It involves defensible descriptions, and defensible inferences, about samples and populations. Statistics is integrally related to models in the world and critical problem-solving skills. Statistical thinking involves understanding the characteristics of a problem well enough to be able to select and apply the appropriate tool to answer the problem. It is not enough to be able to "do" the routine (either by hand or on the computer); one must know why one has chosen it, what its application tells one, and what limitations one must place upon its conclusions.