BEYOND DATA ANALYSIS: STATISTICAL INFERENCE

Gail Burrill, University of Wisconsin-Madison, USA

Statistical education in schools in the United States has focused primarily on data analysis, with little attention to statistical reasoning. The addition of a statistics exam that can be used to earn university credit along with new curricula projects provided teachers the motivation and materials to begin including inferential techniques in their courses. The process of helping students, and eventually citizens and professionals, understand the power of statistics should ensure that they recognize the difference between reasoning from one, often personal, case and predictable behavior of a randomly selected group. Classroom research and experience reinforce the need to provide students with sufficient background and with opportunities to explore situations through simulation to understand that patterns exist and can be used as a basis to draw conclusions in a systematic way.

The emphasis on statistics in the K-12 curriculum in the United States has been increasing since the publication of the *Curriculum and Evaluation Standards for Teaching School Mathematics* (NCTM, 1989). In general, students are beginning to have experiences with simple data analysis techniques. The advent of the Advanced Placement (AP) Statistics Examination in 1997 was an impetus to include statistical inference in the curriculum. The AP Program is a cooperative effort of schools, universities, and the College Board, an independent testing agency, that offers students in high school university credit for passing an examination in subjects with the same content and degree of rigor as those at the university. Students completing the exam usually take a high school course designed from a prescribed syllabus with four major themes: exploratory analysis, planning a study, probability, and statistical inference. The AP test requires students to learn systematic methods to reason from data, and so, students need to: understand the interplay between data, the distribution and a summary statistic and recognize characteristics such as shape, location, variability and outliers; understand random behavior, the predictability of randomness, and the notion of risk, before they deal with the larger question of recognizing when an observation does not fit expected behavior; have a sense of experimental design and understand the implications the design may have on any inferences or conclusions that can be drawn; understand the concept of building a model to describe a situation or relationship.

The discussion that follows focuses on two issues: providing scaffolding for the conceptual development of statistical topics and the use of simulation as a way to increase understanding. The discussion is based on classroom work with secondary students and with teachers in professional development settings.

BUILDING A FOUNDATION

Much of what students are taught in the United States concentrates on definitions and procedures. (NCES, 1996) As a result students can find the mean or median but lack any sense of how to interpret the results. Confusion over the relation between mean and median is pervasive among the public. Wisconsin students were reported to be at the 80th percentile on a national test. When the state reported that the percent of students who achieved at a given academic level was 42%, there was widespread concern that test scores were declining. Students are not provided with opportunities to contrast different measures nor to think about what can and can not be learned from each measure. Intervention and scaffolding can help with this conceptual development, as illustrated below.

The relation between distributions and summary statistics is not well understood by students nor by some of their teachers. In one case, teachers were given three distributions, one skewed right, one left, and one normal, then asked to determine in which was the mean more likely to be greater than the median. The teachers confused the mean with the mode, and many clearly believed the mode was actually the mean. Discussion revealed that the relation between the median and the distribution was also not well understood and that teachers have trouble associating area with frequency. When students and teachers were given a histogram with values along the horizontal axis omitted, many balanced the blocks to create a mean, ignoring the fact that each block, while representing the same number of observations, was “worth” a different amount.

This same misunderstanding occurs when students are asked to make judgments about the likeliness of a result occurring in a region of a continuous distribution. Appropriate early experiences help students transfer from area to percent or relative frequency. When students are given a stem-and leaf plot or histogram and asked to estimate the proportion of the data in a given interval, they can use the actual data to reflect on their estimation. For example, the following problem came from sample AP questions.

“A sample of 56 college students was asked to report the number of hours devoted to study during a typical week,. The stem-and-leaf plot of the results is shown below.

STEM-AND-LEAF OF STUDY HOURS (n = 56)
0 | 3
0 | 555667777777899
1 | 000000000112222333344444
1 | 5555668
2 | 0
2 | 56
3 | 00
3 | 5
4 | 0 1 | 5 represents 15 study hours"

Students are asked whether the distribution is skewed or symmetric and about the relation between the mean and median. To develop this understanding students can begin with questions such as "Do more than half of the students study at least 20 hours? More than a fourth?" The exercise can be repeated from a histogram where students have to understand the frequency to verify their estimation. They learn to relate parts to an irregular whole rather than relating a part to a whole that is either circular or linear. With such scaffolding, students can make sense of questions about the likelihood of an outcome from a continuous distribution or comparing two distributions.

The notion of variability is not well understood by students nor is the curriculum structured to build this understanding. In particular, standard deviation is not at all well understood. Students need experiences that build an understanding that everything varies, recognition that important statistical objectives are to capture variability numerically and of the role of standard deviation, and the understanding of the importance of designing an experiment to reduce variability as much as possible. Simulation is a powerful tool that allows students to experience the variability inherent in situations and to understand that random events have a variability that can be predicted, a notion that contradicts the intuition of most students. When students are asked to estimate the probability that they will pass a true false test by guessing (60% pass rate), having the teacher write an estimate of the mean and standard deviation before simulating the situation lends credibility to the notion of stability in a random situation.

UNDERSTANDING A MODEL

School curriculum in the United States now addresses creating models for bivariate relationships, studying least squares regression and finding appropriate regression models. Very little is in the curriculum, however, with respect to other statistical models. Students need to understand how such models are generated and what they represent. The repeated simulation of a probability distribution or the comparison of many different sampling distributions for the same statistic helps students understand that over the long run, randomness has an order. Simulation allows students to recognize how the constraints of a situation, such as sample size or the number of samples generated, affects the model. Using a computer or calculator to create a distribution for the mean by simulating 100 samples; 200 samples; 500 samples gives students a sense of this long run stability. Systematically changing the sample size and investigating what happens to the distribution helps students understand the mathematical models they eventually must use to be efficient.

Students must understand how to generalize from a situation and how to formulate a statistical model. Consider the following: 1) The proportion of red cars sold is 0.4. Generate a distribution for the number of red cars in a sample of 30 randomly selected cars. 2) The proportion of students who ride the bus to school is 0.4. Generate a distribution for the number of students who would ride the bus given a random sample of 30 students. 3) 4 out of 10 people do not have adequate health care. Generate a distribution for the number of people in a random sample of 30 people who do not have adequate health care.

In repeated classroom situations, even the best students had to work their way through at least three such questions before they realized that the constraints were the same, that the model was independent of the words, and depended only on the population proportion and sample size. The ability to generalize from specific situations needs to be fostered in students by giving them a variety of experiences begin to generalize for themselves. The ability to abstract defining characteristics in a situation is not intuitive for most students. Helping them recognize the need to generalize gives students a background for making choices from what can be a confusing array of statistical tests.

Not only must students understand models, they should be able to reconcile different models for the same situation. Just as they must be able to understand the difference between an exponential model and a polynomial model for a given situation, they should understand the relation between the binomial distribution and the normal distribution. They must also understand the effect of extreme values in a population and on the sampling distribution. Simulation is a powerful way to demonstrate under different conditions a distribution of sample means from a skewed population.

DATA COLLECTION

One area in statistics that is not a part of the curriculum (and is relatively absent from most curriculum) is the design of experiments. The AP guidelines reflect the growing needs of the scientific and social science communities to design experiments that produce data that are useful and address the question of interest. “Data must be collected according to a well developed plan if valid information on a conjecture is to be obtained.” The AP guidelines go on to specify that a well designed plan must identify important variables, indicate how the data are to be measured and allow inferences to be drawn from a model based on the resulting data. The outline includes planning and conducting surveys with subtopics on sampling error, bias, stratification, and planning and conducting experiments with subtopics on confounding, controls, blinding, treatments, complete randomized design and blocking. Teachers and students struggle to understand these ideas, particularly the effect of randomization and its role in eliminating bias. A simulation comparing responses from self selected samples to randomly selected samples is an effective technique to confront student intuitions about sampling. Class discussions about experiments and surveys described in recent media articles provide good platforms on which to build an understanding of vocabulary and the importance of well designed data collection plans.

Blocking offers a special challenge. From the 1997 AP test: “A new type of fish food has become available for salmon raised on fish farms. Your task is to design an experiment to compare the weight gain of salmon raised over a six-month period on the new and the old types of food. The salmon you will use for this experiment have already been randomly placed in eight large tanks in a room that has a considerable temperature gradient. Specifically, tanks on the north side of the room tend to be much colder than those on the south side.” The arrangements of tanks on the diagram had two rows of four tanks, with windows and a door on the north and a heater on the south. Students were to design an experiment that accounts for the temperature gradient.

Many students did not know whether to block east-west or north-south. Simulations that explore variability in different designs, without blocking or with different blocks, help students understand that by dividing units into groups with respect to a given factor, it is possible to have less variability within a group than in the entire population.

STATISTICAL INFERENCE

Simulation can play a major role in setting the stage for understanding inferential techniques. Establishing confidence intervals through simulation provides concrete ways for students to visualize their reasoning. Many teachers and most students have trouble using statistical tests because they do not see the overall picture. Little attention is paid to the assumptions necessary before a given test can be applied, and there is a great deal of confusion over which test should be used when. Students rarely cite assumptions, and many have not moved beyond procedural knowledge . On the AP Test, students could push buttons on a calculator to use statistical tests, and a common tactic was to justify a solution by citing the calculator result. (Such answers were considered insufficient.) Students also have trouble recognizing that a problem can be approached several ways, and that, in fact, it might be wise to do so. Another question on the AP Test gave the number of defective computer chips produced from two ovens over a nine hour day. The question posed was whether there was a difference between the two ovens with respect to the mean percentages of defective chips. Few teachers studying the test recognized it would be useful to investigate the data for time trends, and many never even considered the paired data.

On the other hand, teachers and students do not recognize that when reasoning different ways about the same data leads to very different conclusions, they should investigate the problem, the assumptions or their work. A poll for *Life* magazine titled "'If Women Ran America' reports that two thirds of the women interviewed say that the problem of unequal pay for equal work is a serious one, while only half of the men have this opinion." (Gainesville Sun, Gainesville, Florida, May 15, 1992) Suppose 46 people responded as follows. Is there a gender difference?

	Yes	No	
Men	11	11	22
Women	16	8	24
	27	19	46

Teachers simulated the difference in proportions, or set up a hypothesis test, or used a chi square test. However, many responses included different approaches with different results and never raised a question about how such a difference could have occurred.

Students need experiences that will allow them to discriminate among statistical tests, recognize the impact of assumptions, and approach problems from a general knowledge base rather than an isolated application of technique. They need to have discussions about issues such as sample size -what is “reasonably large” and what happens if a sample is not large enough? Why does np have to be at least 15? What do degrees of freedom have to do with anything? Why can you estimate the population standard deviation with the sample standard deviation? How do I know which test to use when? Simulation can help students come to understand some of these issues; others follow from careful analysis of the concepts - but that happens only when the issues are confronted directly, not assumed as a consequence from another investigation.

CONCLUSION

Reasoning from data by the general public and even the science community is often based on inappropriate data or poorly designed experiments. In many cases, inferences go beyond what the design allows. Many reason from their own experience rather than from the data. “I was there and saw” is more readily believed than conclusions based on statistical evidence. To produce statistically literate citizens, students need to make reasonable conclusions from data and to understand the assumptions under which the conclusions are valid. Slow careful development of prerequisite understandings and the opportunity to investigate situations by simulation can help make this happen.

REFERENCES

College Entrance Examination Board and Educational Testing Service (1996). *Advanced Placement course description statistics May 1997, May 1998*. Princeton, NJ.: The College Board.

National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for teaching school mathematics*. Reston, VA.: The Council.

National Center for Educational Statistics (1996). *Pursuing excellence.* Washington D.C.: U.S. Printing Office.