USING DATA TO MAKE SENSE OF STATISTICS: THE ROLE OF TECHNOLOGY IN SCAFFOLDING UNDERSTANDING

Gail Burrill
Michigan State University, United States of America
burrill@msu.edu

Research and classroom experience identify topics with which students in introductory statistics struggle such as interpreting box plots, standard deviation or z-scores and the normal curve. One reason is that many core statistical concepts are subtle and difficult to sort out. Dynamic interactive technology can provide opportunities for learners to begin to make sense of these concepts by enabling them to generate large amounts of data, explore distributions, examine probability models and investigate the nuances that often seem to obscure reasoning and sense making in statistics. Interactive technology allows learners, using real and motivating data that stem from questions about ways of reasoning in statistics, to move between representations, looking for patterns and generating models related to hypotheses and to informed decision making.

INTRODUCTION

In 1998 David Moore described statistics as the science of reasoning from data in the presence of variability. He went on to suggest that the advance in technology at that time forced us to focus on big ideas and general strategies for dealing with data rather than procedures and formulas. Nearly 12 years later, new and continually evolving technology has again raised the threshold of what is possible. The discussion below argues that one possibility is to use the technology to help us rethink how students learn statistical concepts.

Student misconceptions

Students often enter and leave introductory statistics courses–both secondary and post secondary courses–with misconceptions, and the literature is rife with examples. Researchers have identified school students' tendency to perceive data as individual points rather than as an aggregate whole with its own characteristics, and thus they have difficulty using descriptors of a data set that do not match any of the individual data points (i.e., Ben-Zvi & Arcavi, 2001; Bakker & Gravemeijer, 2004). Students at the school level often describe distributions using what Bakker (2004) calls "local views" on spread (i.e., "the dots are close together here and spread out there") and do not view spread as dispersion from a mean or median value. This view is still in evidence at the university level. Cooper and Shore (2008) suggest that students in introductory statistics courses have a "tenuous" understanding of variability. Chance, delMas and Garfield (2004) agree, noting students did not understand how key concepts such as variability and shape are integrated and were not able to reason about sampling distributions until they had a sound understanding of both variability and distribution. Lunsford, Rowell and Goodson-Espy (2006) found that students at the end of a post calculus statistics course still confused variability with frequency, had problems with the "averaging reduces variation" concept and did not fully understand that for a fixed sample size, the sample mean was a random variable with a distribution having a shape, center, and spread.

School students often struggle when using data analysis techniques to judge whether two groups are different (Watson & Moritz, 1999; Konold et al., 1997). Students at all levels believe the law of large numbers applies to small numbers: that a short sequence of random events will show the kind of average behavior that occurs only over the long run (Tversky & Kahneman, 1971; Innabi, 1999). According to Innabi, the believer in the law of small numbers has problems with the size of confidence intervals and determining significance in tests of hypothesis, is over-confident that the same results will be obtained in replicating the experiment and does not see sample size as affecting the validity of conclusions or generalizations. The authors of a review of misconceptions in inferential statistics provide a comprehensive list of students' misconceptions related to statistical inference and suggest the cumulative evidence indicates that, while students may be able to manipulate and carry out calculations with statistical data, they have difficulty interpreting the results from inferential techniques (Sotos et al., 2007).

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php [© 2010 ISI/IASE]

The research makes very visible the problems students at all levels have learning to think and reason statistically. However, student learning is clearly mediated by the instructor: the tasks instructors choose, how they frame the tasks and what they do with the responses to pave the way or inhibit the path for developing student understanding. The next section points to difficulties teachers themselves have with statistical concepts.

Teacher misconceptions

Moore (1998) suggested that intellectually sophisticated people are not automatically adept at statistical thinking, and many are more apt to believe an anecdote than the data. In fact, Haller and Krauss (2002) claimed a large number of statistics instructors share the misconceptions of their students and consequently have a large influence on fostering the misconceptions. Haller and Krauss's claim is supported by the discussions on a statistics teachers list serve sponsored by the College Board ("AP Statistics" ap-stat@lyris.collegeboard.com) and designed to provide support and advice for statistics teachers who are teaching an Advanced Placement program in which students take, and can receive university credit for, university level courses while still in secondary school. Questions raised on the list serve relate to nuances such as why divide by n-1 for the standard deviation of a sample; should np should be greater than 10 or 5 or…; why should a sample be less than 10% of the population (after all, the more you can ask the better). Issues of interpretation are raised almost daily: what can you say about a confidence interval, what is the difference between accepting and failing to reject the null hypothesis; what is the difference between practical and statistical significance?

Many ask questions about the design of experiments: what is blocking, how can you explain the difference between an outlier and an influential point; what is the difference in confounding, lurking and extraneous variables. One teacher commented, "I'm getting better at talking about experimental design–it took me years–but I still get things wrong." They ask questions to sort out the array of tests: why are assumptions necessary, which test should be used when and how do you know, why should you use a t-test? In most cases, the teachers are asking their peers and the professional statisticians on the list serve for clarification or to get a better understanding of what to say to their students. In some cases, they have things wrong–interpreting the fit of a model by the strength of the correlation coefficient, ignoring residuals, fitting a model to nonlinear data without transforming the data.

Sotos and colleagues (2007) suggested that their review of the literature around misconceptions related to inference made clear the need for more empirically based studies to shed light on the reasons for misconceptions and called for empirical studies to specifically address and help students overcome the misconceptions. The use of dynamic interactive technology seems to be a possibility for addressing some of the misconceptions described above, and the remainder of the paper describes this potential and offers some examples.

DYNAMIC INTERACTIVE TECHNOLOGY

Dynamic interactive technology adds new dimensions to what is possible to do in classrooms with students in an introductory statistics course. Using this technology, students can interact with data in two ways: 1) investigating data from a given contextual situation to make informed decisions related to that context and 2) generating their own data to develop and illuminate important statistical concepts. The most common usage is the first, where the technology is perceived as a toolbox to perform procedures and carry out calculations on sets of data. Dynamic interactive technology has opened up opportunities for the second use, where the technology is a tool for learning by enabling the creation of "environments" in which students can play with a statistical idea in a variety of ways but where the opportunity to go astray, both mathematically and operationally, is limited.

By imposing constraints on what is possible, teachers and students actually have more freedom to explore central statistical concepts in deeper ways. Such environments are similar to using applets, (see for example, the applets developed by Rossman and Chance at www.rossmanchance.com/applets/ or the Rice University Virtual Lab in Statistics at onlinestatbook.com/rvls.html) and have certain characteristics:

- Little knowledge about the operation of the handheld or software is required.

- The fundamental idea is simple and straightforward. The development has both statistical fidelity (is statistically sound and accurate) and pedagogical fidelity (does not present obstacles such as cluttered screens or too many decimal places that interfere with learning).
- The design is based on an action consequence principle, where students take an action on a statistical object, immediately see the consequences, and reflect on the implications of these consequences for a particular statistical objective (Dick, Burrill & Brady-Gill, 2007).
- The interaction is typically driven by one object such as a point, slider, shape or graph.
- The action/consequence document in which students operate is usually composed of two or three carefully sequenced pages designed to have students investigate a core statistical concept.

The following examples illustrate how data driven “action/consequence” documents might be used as tools for developing understanding of fundamental statistical concepts.

LEARNING THROUGH DATA

Normal Curve

How many normal curves are there? This question can produce surprising answers. Students often think there is one normal curve, with mean 0 and standard deviation 1. They have trouble understanding that, just as other functions have a basic structure with the characteristics determined by parameters, a family of normal curves is determined by the mean and standard deviation. Investigating the relationship between the graph, the mean and the standard deviation can help students identify defining characteristics of normal curves (area, symmetry, point of inflection) and recognize that normal curves form a family of curves that share these same characteristics (Figure 1). To understand how any member of the normal curve family behaves, students can consider questions such as “How is the curve the same or different from f(x) = ax^2+bx+c?”; “Look at the equation. What would you expect to happen to the distribution when the mean is changed? Why? To the standard deviation?” Answering questions such as these will help students connect the concept of distribution to other mathematics they have studied. Estimating the area for a variety of normal curves plotted on a grid as in Figure 2 where μ and σ can be changed sets the stage for understanding that the area under a normal curve is the same for any combination of standard deviations and means.

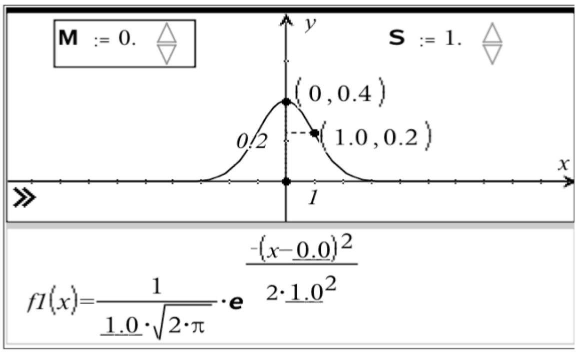

Figure 1. The Normal Curve

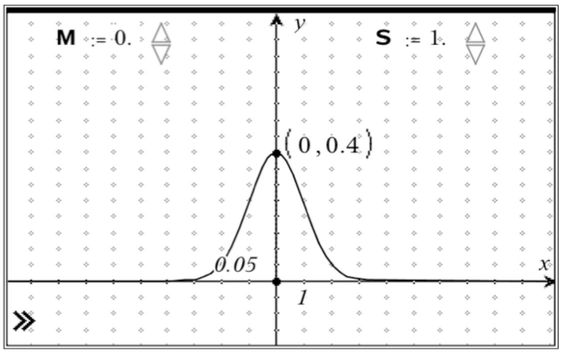

Figure 2. Area & the Normal Curve

Distributions

Students are not clear about the difference between a frequency distribution and a relative frequency distribution nor are they clear about the nature of the distribution with which they are working, often not specifying a distribution *of what*–the population, the elements of the sample from that population or the sampling distribution of a statistic calculated from samples drawn from the population. Chance, delMas and Garfield (2004) point out that it is hard for students to recognize that making inferences from one sample from an unknown population is based on understanding how samples from a known population behave. In a summary of the research and based on their own experience (as well as this author’s), their findings included the following misconceptions about sampling distributions: the sampling distributions have the same amount of variability for large and small sample sizes, large samples have more variability; one sample is confused with all possible samples for a distribution or potential samples; a sample is only representative if a large percentage of the population is in the sample.

An action/consequence document can allow students to visualize the population, the sample and the sampling distribution simultaneously, while changing the sample and generating the sampling distribution of the sample statistic (Figure 3). While others have suggested beginning with concrete experiences, then using applets (e.g., the Reese's Pieces in Garfield & Ben-Zvi, 2008), including sampling distributions of a variety of sample statistics such as the median, the maximum (Figure 4) or the standard deviation can not only help develop an understanding of what a sampling distribution of a sample statistic is but also help students get a better sense of what it means to be predictable.

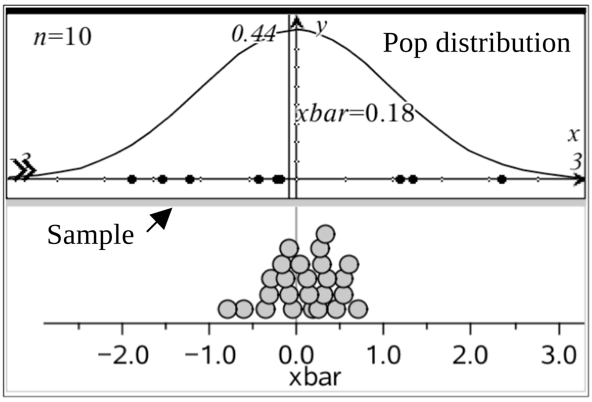

Figure 3. Sampling distribution of sample means

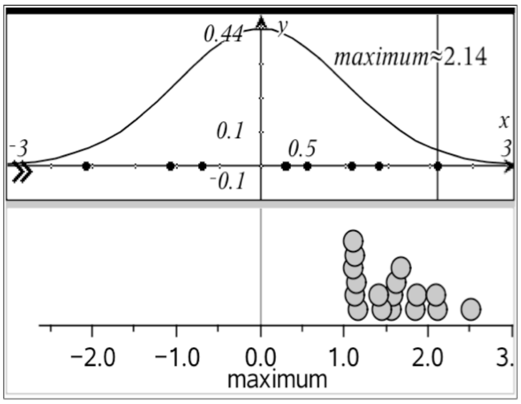

Figure 4. Sampling distribution of sample max

Students can be asked to predict the shape and characteristics of the sampling distribution for a particular statistic, then generate several, record the results and compare with classmates. Questions such as, "How do you think the sampling distribution of the median of the samples will compare to the sampling distribution of the mean of the samples?" can help students focus on the sampling distribution and on the sample statistic as differentiated from the population. The question "Do you think the sampling distribution of the maximum will be skewed as seems to be happening in Figure 4?" can help students review characteristics of the shape of a distribution. Students should discuss what is similar and what is different among populations, samples, and sampling distributions, justifying their reasoning in each case.

What are p-values and alpha levels?

The concepts of "*p*-value", "alpha level" and "significant" are often misinterpreted or confused by students in introductory courses. When the decision is to reject the null hypothesis, they often refer to *p*-value as the probability of making the wrong decision (Haller & Krauss, 2002). They also tend to define significance as the probability of being wrong when deciding to reject the null hypothesis (Vallecillos, 2002). Giving students explicit simulation experiences with these concepts can help them confront their misconceptions and ground their understanding in a visual way.

For example, assuming the null hypothesis $\mu = 10$ ($H_a : \mu < 10$), for a sample of fixed size students can generate an observed outcome, 9.26 in Figure 5, which determines the *p*-value, 0.88, represented by the shaded region in the sampling distribution for the sample means. Assuming it were possble to generate more samples, students then generate additional samples of the same size from the hypothesized population (Figure 6) and observe where the means of these samples fall. From the simulated sampling distribution of sample means, they can estimate the likelihood of getting by chance a sample mean at least as extreme as the original observed sample mean, 9.26, if the null hypothesis is true (0.13 in Figure 6). After students have repeated the simulation for the same observed outcome and corresponding *p*-value (Figure 7) and compared their results with others, questions such as "Assuming the mean is actually 10, are you very likely to get a sample mean at least as large as 9.26? Why or why not?" Changing the observed sample mean (Figure 8) and repeating the process allows instructors to ask where a *p*-value comes from or to give some *p*-values that seem to support the null hypothesis and some that do not, justifying their reasoning.

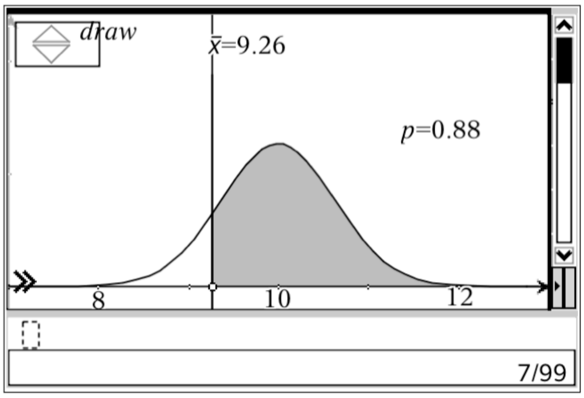

Figure 5. Observed sample mean of 9.26 & corresponding *p*-value

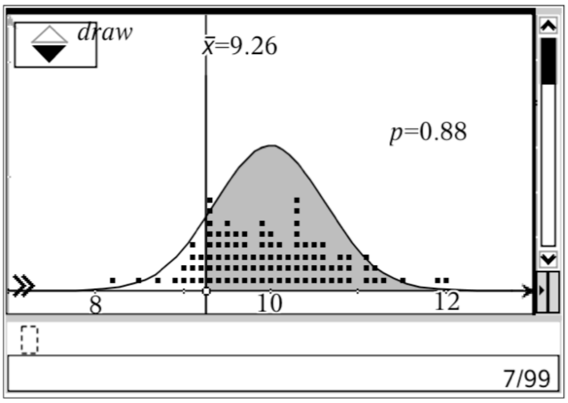

Figure 6. Chance of sample means at least as great as 0.88

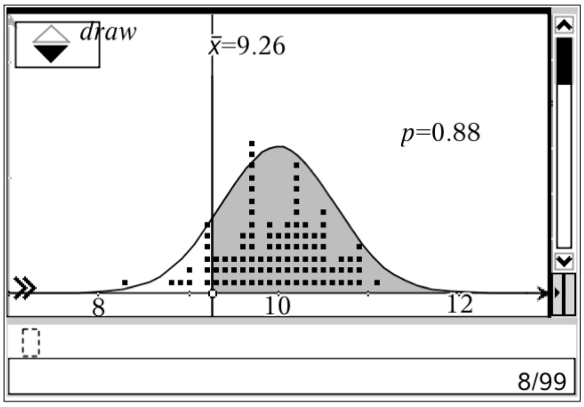

Figure 7. Chance of sample mean at least as great as 9.26

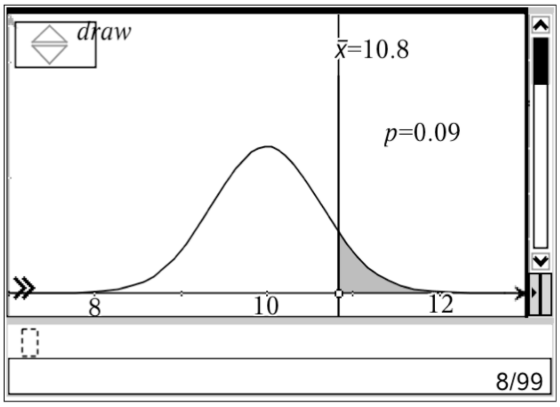

Figure 8. Observed sample mean of 10.8 & corresponding p-value

Going through a similar set of tasks related to alpha can help students see the subtle but key differences in reasoning in these two situations.

CONCLUSION

If teachers use technology only as a tool to do the same things they have already done, they will get the same results as they did before (Ehrmann, 1995; Belfort & Guimaraes, 2004). Moore (1998) suggested that, "To get different results, you must add new thinking to new technology. The reason: *Technology empowers. But thinking enables*." (p. 1258). The claim made in this paper is that, given the fragile understanding of many statistical concepts by students and often by their teachers, we need different results and that we now have new technology that can help us get those results. While a few researchers are using this technology in seemingly productive ways, we need to revisit core statistical concepts particularly with elementary and secondary teachers, including the nuances and subtle points that often elude students, with new thinking about how learning activities centered on action/consequence documents can make a difference in what students take from their study of statistics.

The examples above are only illustrative of a much larger possible set. Similar activities can be created around questions such as: What is power and how is it related to the types of error? Why do we restrict *np* to some minimum? What is a degree of freedom and why are degrees of freedom important? And the list goes on. Documents focused on such questions should become a core part of instruction in statistics classes–which means that statistics instructors must be familiar with the concepts and the questions that need to be asked to make sure that the ideas emerge from the explorations. This has serious implication for the statistics education community in terms of the design and delivery of professional development. In addition, we need to look across the research that already exists about promising uses of interactive dynamic technology to see what we can learn and begin to act on this knowledge. We also need to design and implement research to determine whether action/consequence or comparable documents do make a difference in what students learn and what factors might influence the results. It seems it is time for us to stop documenting the same errors and misconceptions related to student understanding of statistical concepts and begin to look for interventions that can prevent these from happening.

REFERENCES

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal, 3*(2), 64-83. International Association for Statistics Education.

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147-168). Dordrecht, The Netherlands: Kluwer.

Belfort, E., & Guimaraes, C. (2004). Teachers' practices and dynamic geometry. In M. Hoines & A. Fuglestad (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education*, *2*, 503-510.

Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics, 45*, 35-65.

Chance, B., delMas, R. C., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295-323). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Cooper, L., & Shore, F. (2008). Students' misconceptions in interpreting center and variability of data represented via histograms and stem-and-leaf plots. *Journal of Statistics Education.*

Dick, T., Burrill, G., & Brady Gill, L. (2007). New technologies offer new ways to engage students. *National Council of Supervisors of Mathematics Newsletter, 38*(1), 19-21.

Ehrmann, S. C. (1995). "Asking the right question: What does research tell us about technology and higher learning?" *Chance Magazine, March/April*, 20-27.

Garfield, J., & Ben-Zvi, D. (2008). *Developing Students Statistical Reasoning.* Springer Science+Business Media B.V.

Haller, H., & Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research*, *7*(1), 1-20.

Huck, S.W. (2009). *Statistical Misconceptions*. New York: Psychology Press, Taylor & Francis.

Innabi, H. (1999). Students' judgment of the validity of societal statistical generalization. In A. Rogerson (Ed.), *Proceedings of the international conference on mathematics education into the 21st Century: Societal challenges, issues and approaches.* Cairo, Egypt. math.unipa.it/~grim/egypt.pdf

Konold, C., Pollatsek, A., Well, A., & Gagnon, A. (1997). Students analyzing data: Research of critical barriers. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 IASE Round Table Conference* (pp. 151-167). Voorburg, The Netherlands: International Statistical Institute.

Lunsford, M. L., Rowell, G., & Goodson-Espy, T. (2006). Classroom research: Assessment of student understanding of sampling distributions of means and the central limit theorem in post-calculus probability and statistics classes. Online: amstat.org/publications/jse/jse_archive.html.

Moore, D. (1998). Statistics among the liberal arts. *Journal of the American Statistical Association, 93*(444), 1253-1259.

Tversky A., & Kahneman, D. (1971). Belief in the law of small numbers. *University of Jerusalem Psychological Bulletin, 76*(2), 105-110.

Sotos, A, Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review, 2*(2), 98-113.

Vallecillos, A. (2002). Empirical evidence about understanding of the level of significance concept in hypotheses testing by university students. *Themes in Education*, *3*(2), 183–198.

Watson, J. M., & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics*, *37*, 145-168.