10. TOWARD A THEORY AND PRACTICE OF USING INTERACTIVE GRAPHICS IN STATISTICS EDUCATION

John T. Behrens
Arizona State University

INTRODUCTION

Graphs are good (Loftus, 1993)

Graphic approaches to data analysis have been gaining ground since the work of Feinberg (1979), Tukey (1977), Wainer (1992; Wainer & Thissen 1981) and others. This has been spurred on in part by great advances in computing machinery. Although a number of aspects of graphics lend themselves to statistical instruction, some of the value of graphics can be demonstrated using data presented by Anscombe (1973). Anscombe presented bivariate data in which the *x* variable had a mean of 9 and a standard deviation (SD) of 3.3, and the *y* variable had a mean of 7.5 and a SD of 2.0. For this data, the slope was 3 and the *r* was .82.

Picture how data looks in such a form and how you would draw the dataset if you were explaining a correlation of this size to your students. One of the most likely forms to imagine is shown in Figure 1, which is a plot presented by Anscombe.

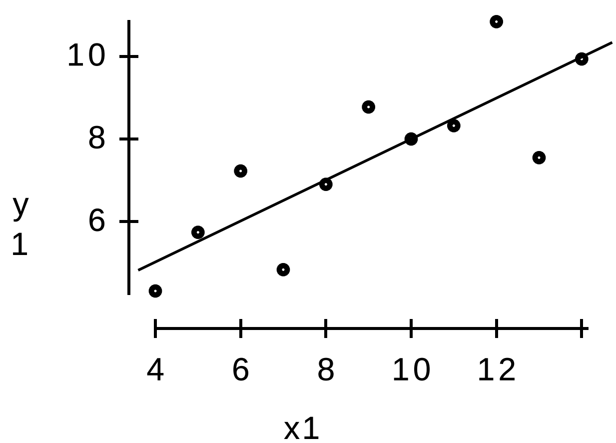

Figure 1: One set of Anscombe (1973) data

Although this is one possible scenario, Anscombe (1973) presented three other datasets that have exactly the same values for the mean, SD, slope, and correlation (see Figure 2).

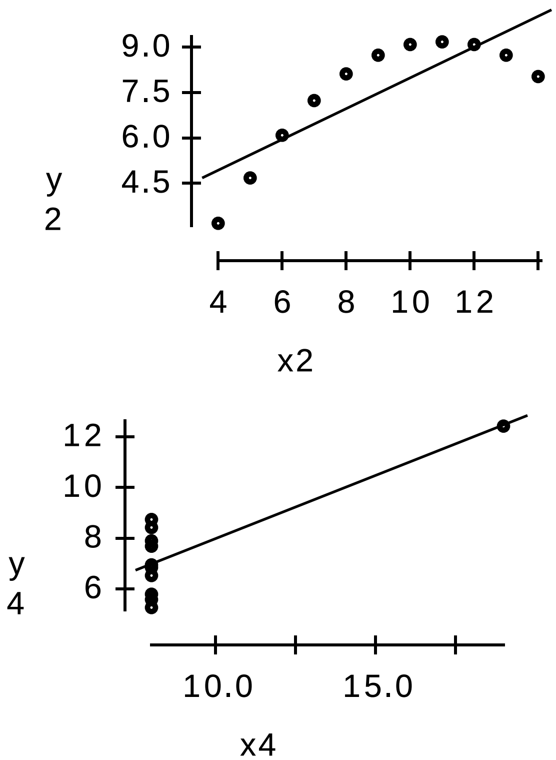

Figure 2: Other possible configurations of data with summary statistics

These plots illustrate the idea that graphics are helpful in data analysis because they allow the perception of numerous pieces of information simultaneously, and disambiguate the algebraic description. Anscombe's (1973) graphics force us to rethink our assumptions about regression and how data occur in the world. For this same reason, graphics are valuable aids in teachings statistics.

From Anscombe's (1973) simple demonstration, we are reminded of a number of important aspects of data analysis that are easily forgotten:

- Always plot the data.
- Each algebraic summary can correspond to numerous patterns in the data.
- Graphics allow us to see many aspects of the data simultaneously (Kosslyn, 1985).
- Graphics allow us to see what we did not expect (Tukey, 1977).
- Graphics (sometimes) help us to solve problems by organizing the information (Larkin & Simon, 1987).

Each of these reasons make graphics an important part of data analysis. Accordingly, it is important to teach our students to think graphically and for instructors to take advantage of the power of graphics in instruction. Teaching graphics is of value because graphics are a parsimonious and effective way to store and organize information (Kosslyn, 1985; Larkin & Simon, 1987) and because it provides good exposure to current data analytic methods (Behrens, 1997; Behrens & Smith, 1996; Cleveland, 1993; Velleman & Hoaglin, 1992; Wainer, 1992). The importance of graphics can also be argued from a philosophy of science approach, which advocates the exploratory data analysis (Tukey, 1977) approach; however, this view is not necessary in order to accept the other arguments presented above.

TWO GOALS AND TWO METHODS

This discussion leads to two goals: to teach students how to use and interpret graphics, and to use graphics to bring specificity to the abstract language of statistics. To reach these goals we use two methods. First, we expose students to numerous examples of real data and real graphs. This has led to the development of Dr. B’s Data Gallery [http://research.ed.asu.edu/siip], which is composed of a collection of one-, two-, and three-dimensional graphic displays Most statistics classes expose students to very few examples of histograms or scatterplots from real data (usually r = -1, -.3, 0, .3, and +1 and the normality and homogeneity are also unrealistic). The Data Gallery provides a place where students can explore the variety of graphical patterns that exists in the real world and where instructors can find examples to use in their classes.

A second method of teaching with graphics is using interactive dynamic graphics. These graphics are dynamic because they are represented on the computer screen and change over time, unlike a simple plot on a piece of paper. They are interactive because they change in response to the actions of the user. For the remainder of the paper they will be called interactive graphics. The interactive graphics assembled are collectively called the Graphical Environment for Exploring Statistical Concepts (GEESC). Although the name is not the most elegant, it clearly expresses the goal, which is to provide students with tools to assist their own learning. GEESC also provides instructors with examples to use in their own expository materials.

This work began in 1991 (see Lackey, Vice, & Behrens, 1992) and has been developed more slowly than originally anticipated. One of the most costly aspects has been the time required to evaluate the effectiveness of the graphics and to determine what conditions make them most useful. It was naively believed that the errors students make and students’ thinking processes were well understood. Time and time again it has been shown the misconceptions supposed by the instructors and graduate students are only a small part of the set of all misconceptions held by students.

Interactive graphics have been available for more than two decades (e.g., Cleveland & McGill, 1988), although it is only in the last few years that they have been widely available and inexpensive enough for general users. One impetus for the increased availability of interactive graphics was the creation of the LISP-STAT programming language (Tierney, 1990). This system allows the (relatively) easy creation of interactive graphical objects; however, it requires that programmers understand the object organization of LISP-STAT. We have found the LISP-STAT environment especially conducive to rapid prototyping for research purposes. However, as an interpreted language it lacks some advantages of more portable languages such as JAVA, which we are using for the implementation of tools prototyped in LISP-STAT.

FRAMEWORKS FOR UNDERSTANDING THE PSYCHOLOGICAL PROBLEMS ADDRESSED BY INTERACTIVE GRAPHICS

The information processing advantages that make graphics so useful for data analysis also make them useful for instruction. In this respect, two cognitive theories of learning and understanding have motivated many of the design considerations in developing GEESC.

The first framework is the theory of mental models put forward by Johnson-Laird and his associates (e.g., Bauer & Johnson-Laird, 1993; Johnson-Laird, 1983; Johnson-Laird, Byrne, & Schaeken, 1992). This theory of mental models argues that individuals construct models comprised of states of affairs that are consistent with the information given to the individual, and that people use these models to make inferences and decisions. These models may be graphical, or may consist of a number of propositions that are experienced as ideas. Johnson-Laird argued that this process of extrapolating consistent states of affairs is ubiquitous in cognition and that mental models are used to fill in the gaps of the information we receive. These fillers serve as stepping-stones for filling in subsequent gaps. This theory accounts for the common instructional experience in which an instructor provides the formula for the SD, has students explain it, questions them, and all students nod politely to show that they understand the material. The students then go home with numerous erroneous beliefs that they logically derived from the ambiguous statements that the instructor thought were completely unambiguous. In this case, each student had constructed a mental model that was coherent given their interpretations of the lecture, but not necessarily what the instructor wanted them to have learned.

Discussing a series of experiments on the use of diagrams to aid reasoning, Bauer and Johnson-Laird (1993) argued that diagrams reduced load on working memory by making key relationships explicit. They concluded that “as a result, reasoners are much less likely to overlook possible configurations, and so they tend to draw more accurate conclusions” (pp. 377-378). Their interpretation of years of experimental data closely matches my clinical experience in the classroom and laboratory. In sum, we value graphics, especially dynamic graphics, because when they are properly constructed, they help make aspects of the data explicit so the students can compare their mental models with the detail of the graphics.

A second theory of learning that has influenced this work is the idea of impasse-driven learning articulated by VanLehn (1990). This theory stresses that individuals build representations of the world (like mental models) that have untested assumptions and bugs. In this view, learning is most effective when an individual’s representation cannot solve the problem they are addressing. This situation is an impasse that requires the learner to stop and rethink their beliefs. This theory is well aligned with instructional experience. It seems that the greatest learning sometimes comes from people realizing that they “had it all wrong” and that they needed to work hard at rethinking their underlying assumptions. These two theories account for much of what is seen in less than optimal statistics education. Students learn the definitions (e.g., the mean is the arithmetic average) and rules (e.g., use the median if the data is skewed) and maybe even how to recognize a few patterns. However, when they face the complexity of the real world they find they did not really understand what we meant in the first place (e.g., “I think it is skewed, but not as much as it was in class -- what do I do?”).

As noted above, this problem can be addressed in part by providing students many examples of patterns that occur in real data. Of course, this should be balanced with an emphasiss on problem solving and working with real life problems. In addition, the use of properly constructed interactive graphics can create impasses and challenge student's mental models, which will lead them to greater understanding.

INTERACTIVE GRAPHICS: THREE EXAMPLES FROM GEESC

Three examples of interactive graphics for exploring statistical concepts from GEESC are presented. Each example grew out of a desire to let students explore the structure of statistical concepts.

Example 1: Changing histograms

The histogram appears to be a simple device that needs little or no introduction. Nevertheless, I sometimes encounter students with excellent skills at hand-computation of analysis of variance procedures who improperly interpreting histograms. Even in my own classes, I encounter students who can recognize patterns of positive and negative skew, yet are unable to translate such patterns into substantive conclusions concerning the phenomena under investigation.

To address this problem, a graphic representation was designed in which the user can interactively add data and respond to questions that require a deep structure interpretation of histograms and related concepts. The goal was to create tasks that could only be accomplished with a rich understanding and for which incomplete or erroneous interpretations would lead to failure.

The graphic that resulted from this effort was an interactive histogram (see Figure 3). Students can add data to a bin in the histogram by moving the cursor over the bin and clicking the mouse button. The location of the cursor specifies the value to be added. Clicking the mouse button adds the data. Recently added points are highlighted.

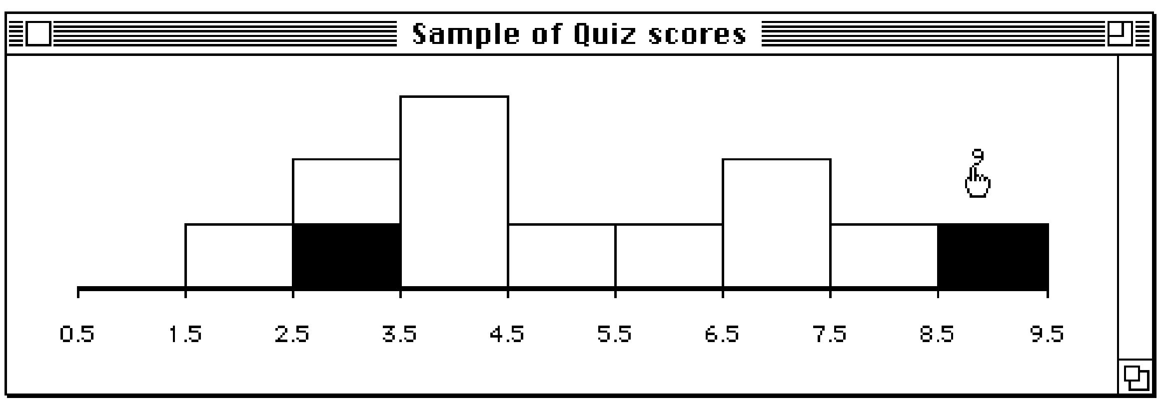

Figure 3: Interactive histogram

In this simulation, students add five points of data and then are asked (by a dialog box) to estimate the mean and SD (see Figures 4). The goal is not to focus specifically on estimation skills, but rather to help the students see the relationship between the shape of the data and the mean and SD. This is demonstrated in class where students are given goals such as "make the SD bigger," "make the mean higher," and other variations of these tasks. This simulation is not simply about histograms; it also concerns the mean and SD. For example, the mean is less influenced by outliers as sample size increases. This is something most students do not recognize until they see the mean “behaving” differently as they work through the histogram session.

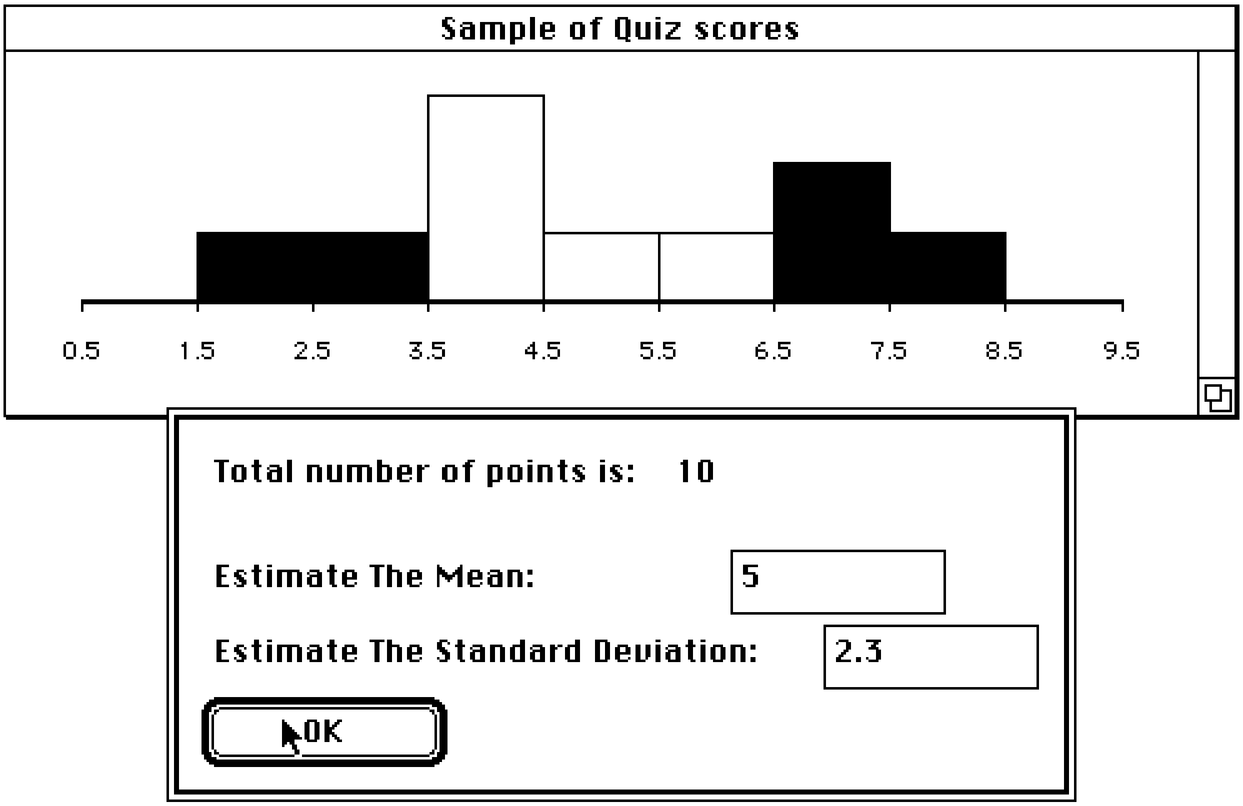

Figure 4: Dialog box prompting estimation.

After students have entered their estimations, they are given feedback concerning the accuracy of their estimations (see Figure 5). Estimation is encouraged because it forces them to consider the relationship between the pattern of the data and the algebraic summary statistics, and forces them to make their beliefs explicit. The tasks given to the students working with the interactive graphics largely determine the success of the experience. If students are simply given the simulation without specific tasks, they may undertake simple tasks that fail to produce impasses or may even appear to confirm their misconceptions. The technology of the graphic is nearly useless outside of well-structured instructional tasks.

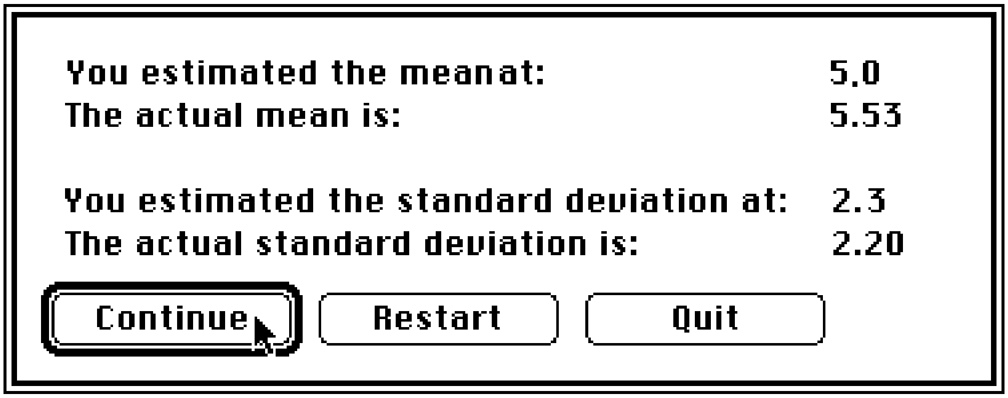

Figure 5: Dialog box giving feedback on estimations

Example 2: Sampling distribution of slopes

Although the sampling distribution of the mean and other common summary statistics (e.g., median, proportion) are sometimes simulated to illustrate the law of large numbers or the central limit theorem, these basic concepts are often lost for students when they try to generalize this thinking to bivariate sampling. To illustrate the sampling distribution of slopes, a simulation with three windows is presented. The first window shows the population of data with $\rho = 0$. Because the notion of a theoretically infinite population can be confusing, and because this version of sampling is based on the physical analogy of “drawing out” or “picking up,” it is helpful to show the data of the population from which the samples will be drawn. A second window is used to briefly display the bivariate data from each sample as they are drawn from the population. A third window consists of a histogram of the size of slopes from the random samples. Pictures of the windows before and after beginning the simulation are presented in Figure 6.

For most of our students this is the only time they will see a true random sample of bivariate data. Most students are surprised at how different the sample data looks from the population. It is this graphic explicitness that forces the students to rethink their assumption that the sample would look like the population. For each random sample, a regression line is fit on the sample data and a copy of the regression line is put on top of the population graphic. After the data for a sample is plotted, a new regression line is drawn over the data. The current sample regression line is also plotted on the population window. This window, however, accumulates the regression lines so their overall variability can be assessed. The histogram of slopes is also updated with each sample.

The histogram is an essential aspect of this simulation because it forms a sampling distributions of slopes. It also allows a direct visual comparison between the set of all slopes (upper left) and the individual slopes of the histogram. In the absence of such graphics, students must remember possibly meaningless rules about the type of distributions that match some statistics. This simulation is completed in class using several different sample-sizes. Students discuss their expectations and predict the differences that will occur. What counts as a rare slope and the issues of Type I and Type II error are easy to illustrate with concrete examples from this simulation.

Example 3: The many sides of statistical power

If students are not completely befuddled after hearing the typical instruction regarding sampling distributions, they still have the opportunity to gain confusion while learning about statistical power. Yu and Behrens (1995) developed a graphical power simulation that is part of the GEESC collection. Close study of students using this simulation led to the identification of a number of debilitating misconceptions and revision of the simulation to force impasses at these misconceptions.

Difficulties appear to arise from two characteristics of the system. First, the system is multidimensional. Effect size, power, alpha, and sample size combine to make a four-variable function that is not easily comprehended. Second, many students have incomplete or erroneous conceptualizations of sampling distributions and of the hypothesis testing model of accept-reject rules. Statistical and historical reasons for some of these misconceptions are discussed by Behrens and Smith (1996).

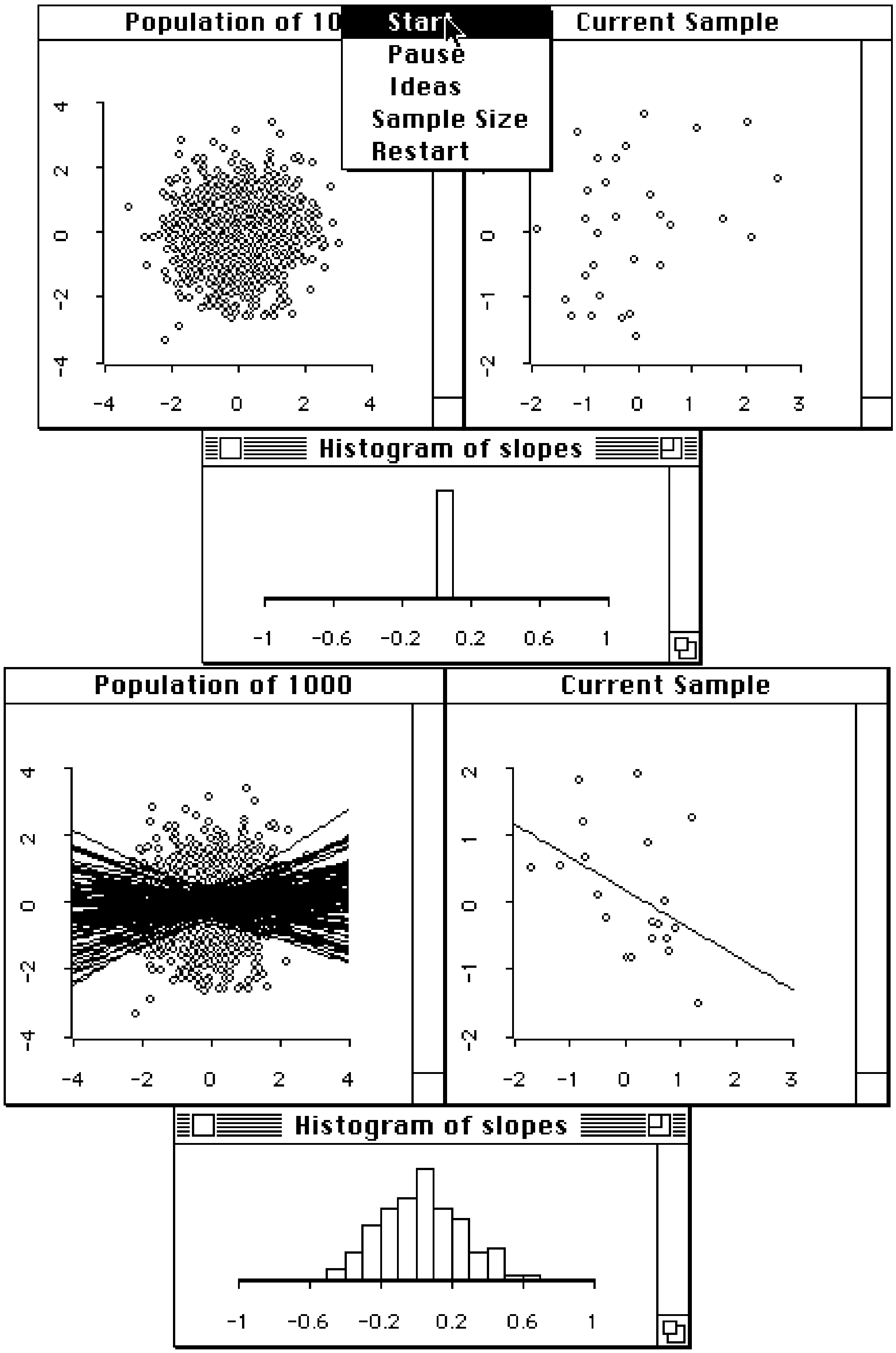

Figure 6: Slope sampling simulation, when ready to start and after a number of iterations

To address these difficulties, the power simulation consists of a graphic display of two sampling distributions with effect size and sample size controlled by slider bars. Alpha level and the shading of various reject-accept regions are controlled by buttons on the display (see Figure 7). The power of the test is displayed and updated as the user interacts with the system. Increasing the effect size increases the separation between the distributions; increasing the sample size decreases the variance of the sampling distributions. These changes occur in a mathematically accurate manner with a minor short-cut (using central distributions).

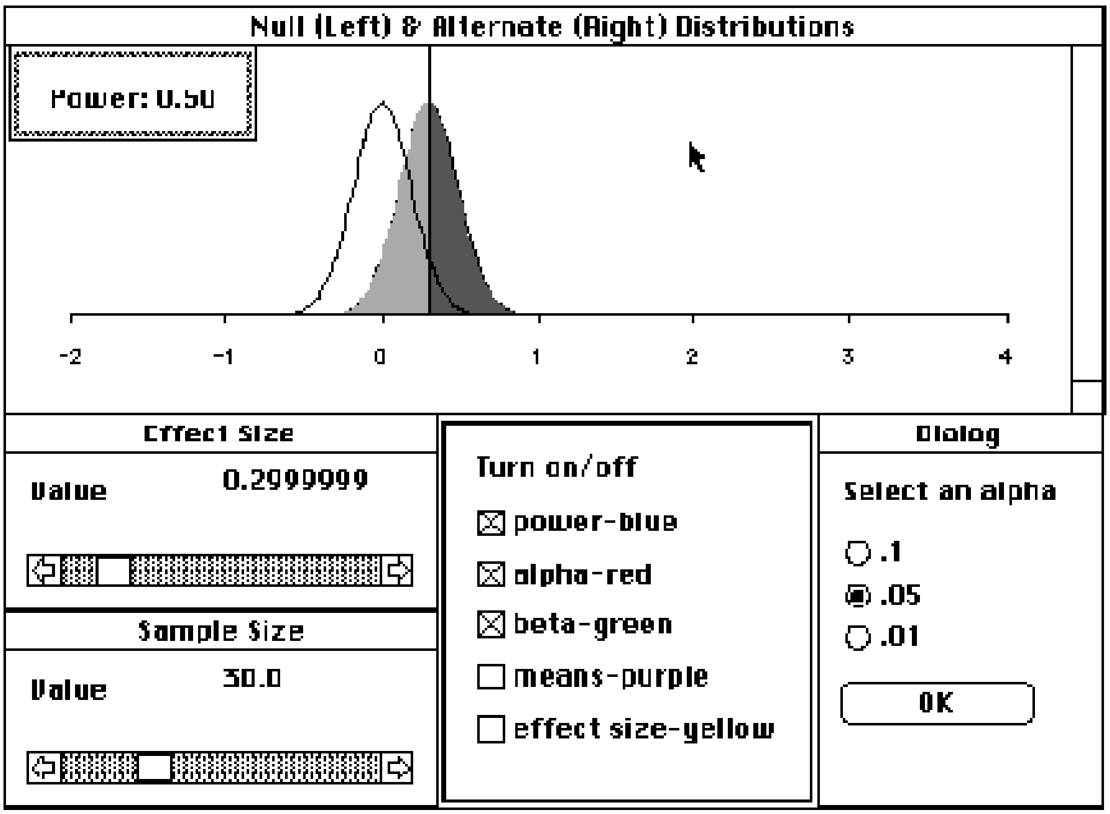

Figure 7: Power analysis simulator

There are many aspects about this graphic display that conflict with students' expectations. Yu and Behrens (1995) reported that students were subtracting the shaded area of the critical region of the null distribution from their estimate of the size of the alternative distribution when the distributions overlapped. Students confused the overlap with intersection because they failed to recognize that the two distributions were distinct and the coloring was only heuristic. The critical region is only outlined (not shaded) in the current version of the graphic so that power and beta are left undisturbed by graphic overlap with the null distribution. Yu and Behrens also found that students frequently overlooked the theoretical status of the distributions and asked why the null distribution never moved. In each of these cases, we achieved the desired goal of having students find their erroneous beliefs and address them. Tasks for this simulation include obtaining specific levels of power with fixed effect sizes or sample sizes and examining the effect sample size has when effect size is fixed. Experience has shown that, without clear tasks, students simply move the sliders without carefully weighing the effect of each variable. Under such conditions the multivariate nature of the system cannot be determined.

SUMMARY: THE POWER OF GRAPHICS

Statistical graphics are a powerful tool for illustrating concepts at varying levels of abstractness. This paper has described how interactive graphics have been used to help students become aware of their misconceptions and confront them by predicting and checking their predictions. These simulations are not a complete answer to improving statistical education, but they are quite valuable for dealing with abstract concepts, especially when students have poor mathematical training.

Experience with these graphics over a number of years suggests they are most effective when their component parts are well understood and when they are used with well-specified tasks that lead to mental-model checking and impasse creation. Coupling the student with technology alone is generally insufficient to reach the desired effect.

When demonstrated by the instructor and used in class by individual students, the student evaluations have been uniformly positive. In fact, the only negative comment that has come from consistently using these simulations over a three-year period is that not enough simulations were used in class and that the students want more. Each semester a number of students comment that the simulations are "the best part of the class." Additional in-depth user analysis of the type reported by Yu and Behrens (1995) is planned, as well as an expanded evaluation with comparison group studies. The GEESC materials are currently being updated and revised to serve as part of the Computing Studio in the Statistical Instruction Internet Palette at http://research.ed.asu.edu/siip.

REFERENCES

Anscombe, F. J. (1973). Graphs in statistical analysis. *American Statistician, 27,* 17-21.

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods, 2,* 131-160.

Behrens, J. T., & Smith., M. L. (1996). Data and data analysis. In R. C. Berliner & D. C. Calfee (Eds.), *Handbook of educational psychology* (p. 9450). New York: Macmillan.

Bauer, M. I., & Johnson-Laird, P. N. (1993). How diagrams can improve reasoning. *Psychological Science, 4,* 372-378.

Cleveland, W. S. (1993). *Visualizing data*. Summit, NJ: Hobart Press.

Cleveland, W. S., & McGill, M. E. (Ed.). (1988). *Dynamic graphics for statistics*. New York: Chapman and Hall.

Feinberg, S. E. (1979). Graphical methods in statistics. *The American Statistician, 33,* 165-178.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review, 33*, 418-439.

Kosslyn, S. M. (1985). Graphics and human information processing. *Journal of the American Statistical Association, 80,* 499-512.

Lackey, J. R., Vice, L., & Behrens, J. T. (1992, April). *Adapting LISP-STAT: A dynamic computing environment for data analysis and instruction*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Larkin, J. H., & Simon, H. (1987). Why a diagram is (sometimes) worth a thousand words. *Cognitive Science, 11,* 65-99.

Loftus, G. R. (1993). Editorial comment. *Memory and Cognition, 21*, 1-3.

Tierney, L. (1990). *LISP-STAT: An object-oriented environment for statistical computing and dynamic graphics.* New York: Wiley.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison Wesley.

VanLehn, K. (1990). *Mind bugs: The origins of procedural misconceptions*. Cambridge, MA: MIT Press.

Velleman, P. F., & Hoaglin, D. C. (1992). Data analysis. In D. C. Hoaglin (Ed.), *Perspectives on contemporary statistics* (pp. 19-39). Washington, D.C.: Mathematical Association of America.

Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher, 21,* 14-23.

Wainer, H. T., & Thissen, D. (1981). Graphical data analysis. *Annual Review of Psychology, 32*, 191-241.

Yu, C. H., & Behrens, J. T. (1995). Identification of misconceptions concerning statistical power with dynamic graphics as a remedial tool. *Proceedings of the American Statistical Association Convention*. Alexandria, VA: American Statistical Association.