USE OF COMPUTERS IN TEACHING STATISTICS: SOME CURRENT AND PROJECTED USES OF COMPUTER TECHNOLOGY IN THE TEACHING OF STATISTICS II

N. Scott Urquhart
Department of Experimental Statistics
New Mexico State University

1. Introduction

This introduction expands on the instructional environment in applied statistics to provide the setting for the five illustrations of computer enriched instruction. The first two illustrations were discussed in more detail by Southward, Urquhart, and Ortiz (1983). The third illustration was discussed in more detail, but at an earlier stage of its development, by Urquhart (1985).

Students in an introductory statistics course exhibit several forms of heterogeneity which should influence how statistics is taught. They come from a wide variety of disciplines: from accounting to zoology; from agriculture to engineering technology; from computer science to police science. They differ greatly in previous experience: The student who has gathered and tried to understand data has an entirely different perspective from one who has few real world experiences with data. Some, but not all, of this diversity can be removed by creating different sections for students with different backgrounds, but the advantage of extensive sectioning is limited by scheduling and funding.

In spite of requirements and remedial courses, many students still are afraid of symbols, do not understand what a formula tells them, and often cannot see past a particular application to the general concept it is intended to illustrate. Thus, diversity in mathematical insight provides another major source of heterogenity with which we must deal.

The appearance of comprehensive statistical computing packages has led to a popular but rarely stated myth: If students can learn to use a statistical computing package well, they do not need to take a statistics course. Nothing could be farther from the truth. To effectively use a statistical computing package, the user must select appropriate procedures, options, and perhaps the linkage of procedures, resulting in thousands of possible configurations of analyses. After a particular analysis is done, the user must interpret a stack of output in the context of the original problem. Informed decision making in this computational context requires a good understanding of the theory – NOT necessarily the mathematics – behind the various procedures. This presents a real instructional challenge when students arrive with limited overall mathematics preparation and seemingly are more interested in immediate applications than in concepts.

2. Computer Generated Assignments

The first illustration concerns assigned lab or homework exercises. Students often are concerned about how a topic relates to their discipline. Their learning is enhanced when some of their exercises have contexts which are specifically associated with their respective disciplines. Further value accrues if each student's data is unique. The latter characteristic raises the level of interaction among students from "What answer did you get?" to "How did you approach that problem?" This shifts discussion among students from numbers to concepts.

We have developed and used a computer system for creating such problems and for preparing personalized assignments for each student. Necessarily, accompanying answer sheets also must be produced. Once problems are developed and the instructor makes decisions about what subject matter areas – students from related majors – get what problems, an assignment can be set up, executed, and sent to a high speed printer in less than ten minutes. The development of individual problems takes from one to several hours; but, once done, they can be improved and reused. Students provide a good source of relevant problems and are good critics of the wording of specific questions.

Completed problems are stored in four parallel computer files: one each for the problem text, the specification for the generation of the random data to be embedded in the text, the answer text, and the way to compute the answers to appear in the answer text. The problem text carries substantial identification and keys which characterize features of the problem, such as, its subject matter area and its statistical type. The wording of stored problems has improved in response to student suggestions.

Another file contains information about a specific class, including a class list which gives each student's name, major, and broad subject matter area This file also contains information about each assignment which is to be passed out. A class assignment contains header information appropriate for that assignment and a table which relates the student's subject matter area and the assignment's problem number to the identifier of a stored problem. For example, in assignment four, assignment problem one could be stored problem 72 for all students. Assignment problem two might use stored problems 78, 79 and 102; 78 might be used for social science and business, 79 for agriculture and biology, and 102 for the remaining areas. There could be more problems in the assignment, if appropriate.

When an assignment is run, the controlling computer program requests the class identifier and assignment number. From this it retrieves the class list and the specifications of the assignment. It then loops through the class list placing each student's name in the appropriate place in the assignment header. Using the student's subject matter information, it retrieves the appropriate stored problems, generates that student's random data, embeds the data in the text, and otherwise prepares that student's assignment for printing. After the assignments for all of the students are completed, the accompanying answer sheets are prepared in an analogous manner.

3. Decision Making in an Advanced Regression Course

The second illustration was designed to help students learn to make decisions within the context of regression analysis. To make an analysis appropriate for a particular situation, the user must select among many alternative analyses.

We have developed a regression analysis package which is oriented toward instruction rather than production. It forces the user to make many decisions which might be made for him in more production-oriented environments. Further, so that a student is not overwhelmed by the number of analysis alternatives – many not yet discussed in class – the instructor can release alternatives as students become ready for them.

The system has three distinct parts: preliminaries, analysis, and termination. The preliminaries begin by identifying the student, providing an opportunity for information about the interaction, and ends with current messages from the instructor. The body of the analysis has two major parts: the development and identification of the data set to be analyzed and the actual analysis. During the termination phase, the user can choose to have the session log printed at the central computing facility; otherwise, it disappears.

This system uses menus, probes for yes/no responses, and probes for numeric or literal (character) responses to communicate with the user. Menus are used to present major alternatives. Specific questions elicit specific information; this system allows the imposition of various conditions on the user's response. User responses are evaluated for agreement with the type and range appropriate for the response. When the computer requests a piece of information, the prompt is taken from a file. The prompts come from any of three parallel files. This allows the designer to have very informative prompts in one file with equivalent, but brief, prompts in another file. A specific response from the user can change the file providing the prompts at any time. This allows the user to get brief prompts in familiar parts of the system but more lengthy ones in new parts. Users will encounter this situation because the instructor can prevent any menu entry from appearing until an appropriate time, such as, after the topic associated with the menu entry has been discussed. This locking feature has another characteristic: A menu alternative may be prevented from appearing until the user has completed other, possibly prerequisite, menu entries. For example, a user cannot proceed to residual analysis until a regression analysis has been executed.

The data entry menu allows a user to retrieve his/her own data set previously generated by the system described in the preceding section, enter data directly, make a variety of data transformations, display any currently defined data set, identify a data set for subsequent analysis, or proceed to analysis. Two of these options currently invoke other menus to accomplish their tasks.

The data analysis menu presents several preanalysis options for examining a data set or defining an analysis model, an analysis, and several postanalysis alternatives. Plotting of any currently defined aspect of the data set always is available. A user can return to the data entry menu to terminate the analysis menu. A user may use any available menu alternative or sequence of alternatives as frequently as desired; this is true for both the data entry and analysis menus, provided the prerequisite conditions have been met.

This system has evolved in response to student suggestions and criticisms over 15 years. When other instructors have used this system, they have made good suggestions for improvements.

4. Computer Drawn Transparencies

The third illustration concerns the development and use of overhead transparencies. These instructional tools solve some problems but pose others. Instructors who use an overhead projector often use handwritten transparencies due to the various "costs" of developing nicer ones. Furthermore, students often feel, legitimately, that instructors go through transparencies too fast. In the past, we partially solved these problems by placing copies of handwritten transparencies on reserve in the library. Many students reproduce these but complain, again legitimately, about the cost. An instructor rarely is willing to invest much more than this in handwritten transparencies.

We have overcome most of these problems by developing appropriate computer software. A microcomputer can drive an inexpensive pen plotter to make transparencies. By implementing appropriate translation tables, the plotter can do special things, such as, change pen color, make subscripts and superscripts, or draw objects of special meaning in a discipline. This includes formulas and statistical symbols. Complex formulas and graphs can be coded once and later be simply inserted into a transparency's text. Once set up, the micro can have the plotter make a transparency in five to ten minutes. This goes a long way toward solving the problem of making decent transparencies quickly and in a timely manner. It also allows simple editing and re-creation of ones which do not quite accomplish their objectives.

But what about the students? The micro can have the plotter reduce the transparency to half size in both height and width, so that four fit on one side of a piece of paper and four more on the other side. Most lectures require no more than eight transparencies; a copy of all eight can be given to each student inexpensively, on a single sheet of paper. A set of eight transparencies which accompanied the oral presentation of this paper is available from the author.

5. Computer Graphics for Demonstrating Statistical Ideas

Dynamic computer graphics can effectively illustrate ideas which are difficult to understand from word descriptions or single figures. For example, many students have difficulty associating distributions of random variables with underlying population structures. We have developed a computer graphic demonstration of this relation using the bionomial probability distribution. The computer draws an appropriately labeled binomial distribution on the screen, say with n = 20 and P(success) = 0.10. Then it steps through the series of distributions with P(success) = 0.11, 0.12, . . . where this quantity is displayed in a "thermometer" above the distributions. As the distributions move to the right, the thermometer fills, thus forming the basis for a discussion of the relation of interest. Students have responded positively to this demonstration.

The central limit theorem submits to equally good visualization through dynamic computer graphics. The graph of a parent distribution can be placed in a corner of the screen while a sequence of distributions of means of samples of size n appear in the center of the screen. Experimentation with normal, uniform, and v-shaped distributions show that this application is quite feasible; but we have not yet fully implemented it.

Another planned demonstration will show how power increases as the underlying population deviates from the hypothesized value. This demonstration will have a split screen, part relating to the null and alternative distributions, which a student can pull apart, and on which the power will show as a shaded area. The other split of the screen will contain the associated graph of power.

6. Lecture Graphic Support

The transparencies provide static instructional support usable in a lecture setting. The computer graphics described in the last section appear on the screen of a micro, so they are not very usable in a lecture setting.

Video projectors exist. They accept the same signal as a monitor but project the image toward a screen, rather than on the back side of a video tube.

The coding for the transparencies could be translated for screen display rather than for drawing on a plotter. The graphics demonstrations could then be integrated with text and formulas that now go on transparencies. Prepared text could be presented part of a page at a time, as transparencies are. The student copies still could be made in advance. The integrated text and graphics could be stepped through using controls much like those for a slide projector. In fact, the instructor could enter text at a micro keyboard during lecture in responding to student questions. If implemented, such a system would be a step toward the development of dynamic instructional material which students could use anywhere, and could be encoded on videotape or compact videodisks.

7. Concluding Comments

Does it take much time? Is it worth it? Emphatically, YES to both questions.

Appropriate technology, appropriately adapted to your context, can greatly improve your communication as a teacher. It may open avenues so you can reach students you previously lost. It may provide totally new ways to explain a topic. You may see the topic in a wholly new way. You can save your successes, improve on your acceptable presentations, and discard your failures. Do not plan to always succeed. If your channels of communication with students are open, you will know when you have succeeded, and when you have failed.

Success will nearly always require planning what you are going to try to achieve. The instructional objectives should remain as the guiding force. For almost all of us, the computer should be viewed as a tool, not as the object of interest.

8. References

Southward, G.M., N.S. Urquhart, and M. Ortiz (1983). Computer enriched instruction of intermediate level statistical methods. Proceedings of the Section on Statistical Education, American Statistical Association, 6-9.

Urquhart, N.S. (1985). Some current and projected uses of computer technology in the teaching of statistics. Proceedings of Teaching to Potential, New Mexico State University, Las Cruces, 422-440.