Behavior Research Methods, Instruments, & Computers
1989, 21 (2), 96-98

SESSION II

TOOLS FOR TEACHING AND RESEARCH IN METHODOLOGY, STATISTICS, AND DECISION ANALYSIS

Drake R. Bradley, *Presider*
Bates College

Using the microcomputer as a visual aid in the statistics classroom

RICHARD L. ROGERS
Georgia Southern College, Statesboro, Georgia

A microcomputer running a spreadsheet program makes an excellent visual aid for demonstrating statistical concepts and procedures. Some advantages of this setup over traditional visual aids are described. Several statistical demonstrations that I have found to work well are presented.

Surveys of instructional computing in psychology (Butler & Kring, 1984; Castellan, 1982) indicate that computers are most often used in statistics courses, most commonly for statistical computations, of course. However, there is another excellent use for microcomputers in statistics classes. A microcomputer with a suitable display device running a spreadsheet program can be a splendid visual aid for demonstrating statistical concepts and procedures.

A spreadsheet program is a general-purpose tool for microcomputers that is designed for the manipulation of numbers. It is essentially a table of rows and columns that define cells into which words, numbers, or formulas may be entered. The formulas may reference—that is, get numbers from—other cells in the spreadsheet. As the numbers in the referenced cells are changed, the value calculated by the referencing formula automatically changes. For example, one can put the formula to calculate the mean of a column of numbers in the cell at the bottom of the column. As one changes any of the data values, the new mean will be displayed immediately. This is a simple example, but such programs can carry out very complex calculations with many arithmetic, logical, statistical, and financial functions. The best spreadsheet programs can also display graphical representations of the data. These displays will also change automatically to reflect changes in the data.

Correspondence may be addressed to Richard L. Rogers, Department of Psychology, Georgia Southern College, Statesboro, GA 30460-8041.

From the preceding example, one can see how a spreadsheet can be used to show the effect of extreme data values on the arithmetic mean. If the numbers are ordered in the column (which the program can also do), then the median can easily be identified and shown not to be affected by the extreme values. This point can still be made with numbers on a chalkboard too, but spreadsheets are quicker and more impressive to students. In fact, the author of the original microcomputer spreadsheet program thought of it as an electronic chalkboard.

Another demonstration concerning the arithmetic mean involves the fact that the sum of deviations about the mean is equal to zero. A second column containing the deviations of the numbers in the first column from the mean can be set up quickly and easily. The formula "X − mean X" can be entered into the first cell of the second column and then with one command replicated into the rest of the column. The sum of this second column will, of course, be zero. As one changes any value in the first column, students can see that the mean and all of the deviations change but that the sum of the deviations remains zero. This provides an interesting, and for some students more effective, alternative to the algebraic proof of this concept.

Compared to traditional visual aids, such as the chalkboard, transparencies, or slides, the spreadsheet presents a "live," dynamic demonstration in which students can see changes take place. One way to take advantage of this is to build some of the simpler demonstrations in class, such as the one described above concerning the sum of deviations from the mean. The presentation of all the steps

Copyright 1989 Psychonomic Society, Inc.

required to arrive at the sum of deviations should promote a better understanding of this concept than just presenting the final table of deviations and their sum. This is even more likely to be true when one extends the demonstration to the sum of squared deviations and measures of variability.

The graphic capabilities of the spreadsheet are useful for demonstrating any concept or procedure that has a graphic interpretation. With a grouped frequency distribution in the spreadsheet and a line graph of it defined, one can switch between the display of the data and the graph with one keystroke, so that it is easy to alter the data, changing the skewness or kurtosis, and then show the resulting change in the graph of the frequency distribution. The relationship between the shape of a bivariate scatterplot and the strength of the relationship between the two variables can be demonstrated in a similar way. The data and the calculations for the correlation coefficient are entered in the spreadsheet. As the data are changed, the value of the correlation coefficient changes and the scatterplot changes. For this one, I have several different bivariate distributions already stored in the spreadsheet, representing $r = 1.00$, $r = -1.00$, $r = .00$, and $r = .50$. To display the graphs of these different distributions is quicker than to change the original distribution to the desired values. The easiest way to prepare bivariate distributions with a desired degree of correlation is to prepare a bivariate distribution of z scores, adjusting them to get the desired correlation, and then to convert the z scores to raw scores on any chosen scale. It is best to use a data set small enough so that all of the values will fit into one screen; otherwise one will have to scroll the screen around to see all of the data.

My entire classroom presentation on the concept of correlation and the calculation of Pearson's product-moment correlation coefficient is done with the spreadsheet. First I show the students the bivariate distributions of raw scores described above and ask them to describe relationships or patterns they see in the data. They have no difficulty for $r = 1.00$ and $r = -1.00$, but the difficulty they do have with the $r = .50$ and $r = .00$ data sets points out the need for statistical devices for describing such relationships. I then show the scatterplots for the data sets, relating the patterns to the degrees of relationship. The next step is to convert the raw scores to z scores, a task quickly accomplished with the spreadsheet. The relationships are easier to see with z scores than with raw scores. Finally, I have the spreadsheet compute a column of z-score crossproducts and the mean z-score crossproduct, which is, of course, the Pearson's correlation coefficient. The columns of raw scores, z scores, et cetera, and their accompanying graphs are not included in the present article, however, for they are no different than what one can find in most statistics textbooks. The value of presenting the material in the way described is that the students can see every step and calculation in the development of the tables and graphs. The spreadsheet is so quick and convenient to use, however, that the entire presentation described in this paragraph can be done in one class period.

The same data sets may be used to show linear regression calculations and the corresponding lines of regression. I have found this demonstration to be helpful in explaining that there are two lines of regression, X on Y and Y on X. I plot standardized values of Y' opposite X and of X' opposite Y and show that the lines converge as the relationship between the two variables gets stronger. The effect of outliers on the correlation and the regression statistics can also be demonstrated here.

The spreadsheet can be used to generate sampling distributions too. First, a column of values of the desired statistic is entered. Then the formula that generates the corresponding value for the ordinate of the graph is entered next to the first value of the statistic. This formula, which can be very complex, needs to be entered only once. The replication or copy function in the spreadsheet enters the appropriate formula for all the remaining values of the statistic. This makes it relatively easy to generate graphs of the standard normal distribution or sampling distributions of the mean for various sample sizes.

I have used two spreadsheet demonstrations in teaching analysis of variance (ANOVA). The first shows the relationship of the linear structural model of the ANOVA design to the observed data and final variance estimates as described by Bolton (1975). Each effect in the model is estimated by a deviation score, for example:

linear model for one-way ANOVA: $X-\mu = \alpha+\epsilon$

estimates of effects: $X_{ij}-\bar{X} = (\bar{X}_j-\bar{X}) + (\bar{X}_{ij}-\bar{X}_j)$

Every data value in the design is put in a column of the spreadsheet, and the deviation scores in the model are computed in succeeding columns. The deviations are squared and summed, showing the partitioning of the total sum of squares into its components. These sums of squares are divided by their appropriate degrees of freedom, to arrive at the final variance estimates. Rather than generate this in front of the class, I sometimes simply print and distribute copies.

The other ANOVA demonstration is designed to help present the concept of interaction in a factorial ANOVA design. The data, a summary table of group means, and a source table are put in the spreadsheet, and a line-graph plot of the group means is defined. Some of the data (or just the group means) can be changed to produce different kinds of interactions or no interaction. The associated graph changes to show parallel lines for no interaction or nonparallel lines for an interaction. This is a good place to point out that the presence of an interaction cannot be determined from any particular pattern of main effects. The meaning, as represented in group means and graphs, of catalytic, antagonistic, and terminative interactions can be readily demonstrated.

For many instructors, the major obstacle to the effective presentation of such demonstrations is the lack of a suitable means of displaying the computer images to the class. One needs certain hardware that is not usually part of the standard microcomputer system. A graphics video adapter in the computer is necessary to display the graphs. Obviously something other than the ordinary 12- or 13-in. computer monitor is needed for the whole class to see the display. A 25-in. color monitor can be very satisfactory for small classes of up to 10-12 students. For larger classes, some kind of computer projection system is desirable. The new transparent LCD panels (e.g., Kodak Datashow), which are used with a standard overhead projector, have brought computer projection technology into a more reasonable price range. A review of these screens has appeared recently (Sherer, 1988).

REFERENCES

Bolton, B. (1975). Teaching ANOVA models via miniature numerical examples. *Improving College & University Teaching*, **23**, 241-244.

Butler, D. L., & Kring, A. M. (1984). Survey on present and potential instructional use of computers in psychology. *Behavior Research Methods, Instruments, & Computers*, **16**, 180-182.

Castellan, N. J., Jr. (1982). Computers in psychology: A survey of instructional applications. *Behavior Research Methods & Instrumentation*, **14**, 198-202.

Sherer, P. M. (1988, October 3). LCD projection panels are replacing awkward and expensive video systems. *PC Week*, pp. 49-54.