From **Brunelli, Lina & Cicchitelli, Giuseppe (editors). Proceedings of the First Scientific Meeting (of the IASE).** Università di Perugia (Italy), 1994. Pages 191-197.

USING COMPUTERS IN TEACHING STATISTICS IS IT WORTHWHILE?

Andrej Blejec
Institute of Biology, University of Ljubljana
Karlovska 19 POB 141, 61001 Ljubljana, Slovenia

1. Introduction

Computers were invented and constructed to compute. Since statistical analyses are computing intensive it is natural that computers are widely used in statistical research and applications. Statistical applications, such as census with its problems of sorting, counting and tabulating, were among the motives for constructing the ancestors of modern computers.

It was always clear that students have to be taught how to use computers, because they will use them in their later careers. With the development of modern technology, computers evolved from large mainframes to personal computers, available for individual use. Availability of personal computers changed the way of computer usage and allowed computers to be incorporated into the teaching process in various disciplines.

In particular, can computers be useful in teaching of statistics? If yes, to what extent? What changes in the teaching process are needed if we want to apply computers efficiently? There are many other questions related to the usage of computers in teaching. I would like to present some of my views about computers in teaching of statistics.

Computers can be used in teaching in many ways. Some are of a practical nature, speeding up routine actions such as calculating or sorting. Some help students to understand, and teachers to explain, certain difficult topics. Computers can be used as electronic textbooks. They can be incorporated into practical student work as well as into lectures.

2. Computation

One aspect of the use of computers in teaching of statistics is the computer as an aid to calculation. The benefit of different types of calculators in the past, and computers at the present time, is enormous. Computers take the burden of computing from users, leaving them time for final decisions based on statistical calculations. I wrote users, not

students, because the burden of computing certainly falls on users of statistics. But what about students? Is it wise to take *all* calculations from students in the learning phase? I don't think it is. Some personal experience and involvement in the calculation process can help to better understand the formulas and statistical methods considered. Such calculations *by hand* should certainly be limited to small datasets, which are quite common in teaching.

Moderate or large datasets, more realistic for statistical decision-making, cannot be analysed without computers. In the past, the calculation efforts were too large and reasonably large data sets were seldom analysed by students. What software should now be used: statistical packages, spreadsheets or both? To be on the safe side it is wise to answer-both. Let us first consider spreadsheets.

There are some advantages in using *spreadsheets* in teaching statistics. Using computers, people are sometimes confused by the fact that data are hidden in a mystery called "file", where they are kept invisible. In spreadsheets, data are kept in tabular form, similar to the hand-written tables used for data recording data on paper. Formulas, necessary for calculations, are entered step by step into individual cells, showing the intermediate results of calculations. Many statistical functions are implemented in modern spreadsheets, from average values to inverse probability functions, regression analysis and hypothesis testing. For teaching purposes, the sum of a range of cells is the crucial one, and always present. It can be clearly shown that average value, variance and many others are indeed functions, defined on a sample (or realisation of a random variable, process...). Entering statistical formulas in fragments, showing the steps of the calculation scheme on actual data, can aid understanding of statistical formulas. The automatic recalculation facility can be used to experiment with "what-if" questions, varying some data values and observing changes in final or intermediate results.

Spreadsheets, even if they are flexible and straightforward, are not the final answer. Some methods are too complicated to be efficiently coded in spreadsheets. In such cases, *statistical packages* can be used. The range of statistical packages is very wide. All traditional statistical packages, available for mainframes in the past, are now available on personal computers and workstations. There are many new packages available, heavily supported with graphics and attractive user interfaces. They range from simple and small to sophisticated and huge. It is not easy to decide which package to use in teaching, nor is my aim to recommend any specific package.

First of all, we must decide whether we want to simply teach statistical concepts or, at the same time to train students to use a certain

statistical package as potential future users. In the latter case, the choice is influenced by availability of the package for future users, implementation of specialized methods important for the field of research, and so on.

If we want to select a statistical package for teaching, other aspects should be considered. Most statistical packages are powerful enough - if not too powerful - for teaching, because for teaching only a limited range of the available methods will be used. The selection should be based on ease of use, clarity of presentation of results and flexibility in data manipulation.

In the past, the burden of calculations was exchanged for the burden of complicated computer usage. Modern trends in computer usage are headed to "point-and-click" methods. It is certainly a great step towards "easy-to-use" but is also a step to "easy-to-forget". Maybe I am too traditional (or should I say old fashioned?) but I think that certain mixture of clicking and typing should be used. It is fine if a menu-driven or windows-like shell leads the student over pits and traps of computer usage, taking him to the core of selected application. But it is better to type in commands for statistical procedures since one has to think what one will type. I believe that things are remembered better if they are written down or typed in the computer than if they are just pointed or clicked on the screen.

In the area of presentation of results, two aspects should be considered: printed results and graphical presentation. Printed results are traditionally organised in lists and tables, but it is a nice feature if a log file for inspection of previous results is available. In some cases, results are kept in a specific object format that can be viewed at any time and even analysed by another method. Sometimes, results are presented in spreadsheet-like log file, parts of which can be copied to the data editor and reanalysed.

Graphics depend greatly upon hardware capabilities and will be discussed in the next section. In regard to data manipulation, many statistical packages incorporate spreadsheet-like data editors. The ability to apply mathematical and statistical functions to the data in a spreadsheet is a fine feature of such editors.

3. Statistical graphics

Graphical presentation of statistical data and results of statistical calculations is a very important part of statistical practice. Computers can greatly help the teaching of appropriate graphical presentation.

In the past, graphical capabilities of computers were limited to low-resolution printed bar charts, histograms or scatterplots. Modern

spreadsheets and statistical packages are well equipped with a wide range of possible types of plots. A dataset can easily be graphed in many different ways and their appropriateness can be discussed in class. Such discussions are hardly possible without computers, because the effort for producing a lot of different graphs - some of which may be inappropriate - by hand is too large.

On the other hand, it is advisable to train students to be able to sketch graphs without the help of computer graphics. Only the usual types of graphs are implemented in computer programs, especially in spreadsheets, where more emphasis is given to presentation graphics than to analytical graphs. Some skills to construct uncommon or new graphs by hand should be developed.

4. Simulation

Statistical reasoning, based on concepts of probability, variation and uncertainty has to be supported by intuition or experience with data in different situations. Using computers one can repeat the same analysis on different, simulated, sets of data. Such simulated samples, taken from known distributions enable us to compare the results of an analysis of a sample with the known, exact state of the original distribution. Such repeated analysis can be conducted many times, showing to students the variety of possible results, giving them the opportunity to feel the probability of the expected outcome. In that way students can see the power and limits of applicability of statistical methods.

To support such simulations, generators of random numbers from different distributions should be implemented in spreadsheets and statistical packages. While some statistical analyses and concepts can be shown by simulated data with spreadsheets and general statistical packages, some should be shown by special programs, designed for statistical simulations.

Some very important statistical concepts, such as confidence intervals, sampling distribution or standard error, are difficult to understand for many non-mathematically inclined students. Specialized graphically supported *statistical experimental/simulation systems* (SE/SS) should be developed to show such concepts in an intuitive way. They should show not only standard results, but also the backstage performance of the statistical method. In the examples mentioned above, using repeated sampling from a simulated population, one can plot relevant sample indicators (sample itself, mean value, confidence interval...) for each sample and the sampling distribution at the same time. In that way, spectators can see a variety of

samples and compare their properties with the known state of the population. It is important to show to students how representative or unrepresentative samples of the same size can be.

The purpose of SE/SS is to demonstrate difficult statistical concepts graphically. For non-mathematically inclined students formulas and proofs are not always enough. In some cases attractive, animated graphical simulation can help students to understand the underlying principles.

In that way, the effect of different sampling schemes and techniques can be shown. Among others, concepts of significance in hypothesis testing, problems of outliers in regression analysis and some aspects of analysis of variance are interesting for simulated presentation. Computer supported SE/SS can be incorporated into individual learning programmes and lab work, but they can also serve as illustrative tools in lectures.

5. Electronic textbooks

Electronic textbooks or programmed learning sequences are not limited to the field of statistics. A general trend is to implement them as a sort of self-explaining tool, suitable for individual learning at one's own pace. As in other disciplines, they are useful in statistics too. They provide an excellent mixture of classical textbook and self assessment. Compared to classical textbooks, they have the advantage of animated examples and graphs. Graphically supported simulations can be incorporated and explained by the surrounding text. Problems and questions are used as an instant test of understanding, leading the student to the next topic or guiding him back to reread some topics in the case of failure. Testing is not a one-way process and can serve as a guide for the author, indicating weak points in explanatory text in case of frequent student failures. A textbook can be improved with feedback between the author and the student.

Electronic textbooks have, at the moment, two great disadvantages compared to classical textbooks: they are rare and they cannot be studied elsewhere. The first problem can be solved by lecturers of statistics and the second by new technology. But even so, they are not practical enough to push paper based textbooks from student's hands.

I have to mention *multimedia*, a very promising field of application of computer technology to education. The multimedia approach is closely interconnected with electronic textbooks, unfortunately putting even greater demands to computer equipment, limiting the availability of such systems. While anyone can take a paper textbook to any place in the world, and be able to study it as well as anyone else, electronic textbooks are restricted in use.

6. Teaching with computers

Using computers influences the way of teaching. Some skills, traditionally taught, may not be needed in the future. Some calculation schemes, used in the past to ease the calculation process, are no longer used. Should they be abolished? Let me present some examples.

Calculation of the mean by the method of deviations from an arbitrary value is no longer needed if calculation is performed with computers. But it illustrates the concept of data transformation and connection between the statistics of the original and transformed data. One student asked me, why are we learning formulas for calculations of the arithmetic mean and variance from grouped data, if statistical packages calculate statistics from raw data. Maybe the computational reason vanished but as a concept they are still important.

Should we teach reading of values from probability the density tables, knowing that statistical packages report "exact" probabilities and significance levels? As it teaches the concept of a table and shows the connection with probability density functions, yes. And we can even do it better than in the past, if computer graphics are used to show what we actually read from tables.

We can compare such dilemmas with the role of the logarithm function in the past and at the present time. The logarithm was very important, and according to some indications, was invented to simplify multiplication to addition. It is nowadays no longer needed for that purpose, but is still very important as a concept of certain relations in mathematics and nature.

So we don't have less to teach, but I hope that we can teach better with computers. They can be used in individual lab work of students-mostly for calculation and graphing purposes. For me, more promising and interesting is the incorporation of computers, especially the demonstrations of simulations, into lectures. They should be used as animated transparencies for overhead projections. As is true with transparencies, we should pay attention to visibility of details. Characters should be of appropriate size to be read from back row of the classroom. Resolution for graphics should be reduced, plotting lines as bars, several pixels wide. Colours on the screen should be carefully selected if monochrome projection is used. Projectors should be luminous enough to present clear picture. If these recommendations are not met, the goal of the presentation will not be fulfilled, regardless of a carefully prepared spreadsheet calculation or excellent simulation.

7. Conclusion

Computers can be incorporated into the teaching of certain statistical topics.

Using computers, one can spare a lot of time otherwise devoted to calculations. We can show more extensive examples in lectures. Students can analyse larger samples, plot more graphs and get more experience with data. Different approaches to the same problem can be tested, graphs of different kind can be easily presented. As the burden of calculation or graphing can be left – to a certain extent – to computers, more time is left for the discussion and explanation of statistical concepts. Understanding of different statistical concepts can be enhanced by the use of computerized statistical experiments and simulations.

Computers can help us to teach statistics on some levels. Using computers in lectures is an additional effort for teachers. Computers should be used, if they are available, for tasks that cannot be efficiently shown on the blackboard.

Availability of computers is not a question to be neglected. In many schools around the world they are simply absent. Where they are available, they are frequently unable to run modern software. It is wise to keep in mind that teaching software must be able to run even on simple computers.

Computers, however capable they might be, will not push teachers from their desks in the near future. If a good teacher is a flame, boiling the statistics teaching stew, computers are spices: the stew is better if they are used with care, but is likely to be tasty without them too.