Developing Statistical Concepts for Engineering Students Using Computer Packages

Brian Phillips and Peter Jones - Melbourne, Australia

1. Introduction

The idea of using the computer as a teaching tool is not new. It has been used as an effective aid in mathematics lectures for a number of years (see, for example, the report produced by the Shell Centre for Mathematical Education, 1986). Hunka (1988) reports on a computer-assisted instructional course which has been used for fifteen years at the University of Manitoba for a first course in statistics for graduate students in educational psychology. While such a system has the advantage of allowing students with diverse backgrounds to progress through the course at their own pace, it is difficult to update and take advantage of modern developments such as new techniques and easy to use statistical packages. Another approach is described by Harrington (1988) who discusses the dilemma which statistics educators face in deciding where the balance between hand and computer calculations should lie in statistics courses. She advocates the use of a spreadsheet in which the students can enter data using "macros" as a means of achieving this balance. They can also investigate "what if" questions by altering data values and observing the effect on the statistics being studied.

Like the calculator, many mathematics and statistics educators see the computer basically as a computational tool. Joiner (1982) stated that the goal of a basic statistics course should be to help students understand as much as possible of the statistical analysis process. He listed a number of steps which are involved in the statistical process and explained how the use of a computer can help free the student's mind from the need to worry about calculation details and concentrate on the process. In recent years the computer has opened the way to readily handling large amounts of real-life data, giving students the opportunity to investigate many situations for which the calculations would have previously been too tedious.

Although the computer has proved most useful in facilitating virtually any form of statistical analyses, it has not necessarily greatly improved the understanding of statistical concepts for many students. In fact the easy access to computational facilities can too readily lead to inappropriate analysis being carried out. Further, with an increasing number of people needing to come to grips with at least the basic ideas of statistics, both in their formal studies and for everyday life, it is more important than ever to develop strategies to help get the statistical message to a broader audience. It is our belief that the computer has an important role to play in this endeavour.

While recently published introductory statistics texts often incorporate output from packages such as MINITAB, for example Weiss and Hassett (1987), Groeneveld (1988) and Scheaffer (1990), most of these texts tend to be of fairly traditional format mainly using the computer to illustrate the applications of statistics so taught. This approach certainly helps satisfy the Joiner criterion by allowing students to concentrate more on interpretation, rather than becoming lost in masses of computations. While this is a marked improvement on the older style of texts, in which key concepts could easily be lost in the computational formulae used, too rarely, in our view, is the computer used to full advantage as a teaching and learning tool. In fact we would say that the computational use of the computer is less important than its use in developing statistical concepts, especially at the early stages of a statistics course. In this paper we will outline a course we are developing which aims to implement these ideas and to indicate a few techniques we have used to help develop some important statistical notions. In doing so, we will focus on the teaching of statistics to engineering students.

2. A brief history

At Swinburne Institute of Technology, statistics for engineering students is taught as a component of the second year mathematics course. Up till three years ago it was taught in a fairly traditional way which included a reasonable amount of probability theory and some mathematical statistics. The main reference was Bajpai et al. (1978) which at the time was one of the few texts that focussed on data drawn from engineering applications. However it did not use any statistical packages to assist with the computations although it did provide some FORTRAN programs for this purpose. Too often calculations rather than concepts dominated statistics classes and assessment, even though calculators were extensively used. In 1986 the MINITAB package was introduced into the course but its role was primarily that of a computational tool rather than a teaching and learning aid. At the time one of the main drawbacks for us was that access to MINITAB was via the mainframe and at Swinburne this was not very user friendly. In fact using MINITAB in this unfriendly environment unfortunately had a somewhat negative effect on students' attitudes to both computers and statistics! The advent of student access to the PC version of MINITAB in the classroom overcame many of the problems we had experienced with using MINITAB in the past and offered us the chance of remodelling our approach to teaching statistics.

3. The remodelled course

The remodelled statistics course is of ten weeks' duration with three hours of instruction per week. Students attend a one-hour lecture given near the start of the week and a two-hour computer practical workshop session run later in the week. The aim of the lecture is to be motivational and to give an overview of the week's topic. It makes use of relevant practical exercises and discusses issues but usually does not involve detailed solutions of problems or extensive development of theory. A few exercises are set on each lecture topic; they sometimes involve theory questions and always interpretation questions.

The practical workshop is used to help develop and reinforce the statistical concepts introduced in the lectures, and to give students experience with handling data. With the time that is saved in computations students can concentrate on the concepts and the interpretation of the results. At the moment the laboratory being used contains 18 computers which can be used as stand alone computers, as terminals to a PC network, and as terminals to the Institute's IBM mainframe computer. The groups consist of 25 to 30 students so many of the terminals have to be shared. While this may not be seen as preferable in certain situations, it has proved to be useful in encouraging discussion of concepts and self-help with the computing aspects.

Each week the lecturing team prepares a running sheet on which the workshop is based. This is often very detailed giving the aim of the session, syntax instructions where necessary, spaces for results to be written in, graphs from the screen to be sketched in and conclusions/interpretations to be made. The session is taught as a lesson in which the workshop leader introduces the topic with extensively prepared overheads, has the students proceed with the first task, and allows time for most to finish. Then class results are recorded and relevant discussion occurs. The leader aims to keep any talking up front to a reasonable minimum, 5-10 minutes being normal at any one time. A datashow connected to a PC running MINITAB is used to guide students and serve as a focus for discussion of issues as they arise during the workshop. A roll is kept to keep track of attenders since the workshop is regarded as an integral part of the course.

4. Use of simulations to demonstrate statistical concepts

Our experience suggests that before the techniques are applied on real data the student needs to have developed an understanding of the underlying concepts. For many students a traditional theoretical approach appears to do little to develop these concepts.

One instructional strategy aimed at promoting the development of basic statistical concepts has been through the use of concrete models. For example, the use of coin tossing and drawing balls without replacement from a box can be used to introduce the binomial and hypergeometric distributions respectively, while the Central Limit Theorem can be established by sampling from a box of counters (Glencross, 1986). However, in the classroom situation, these approaches become time consuming, noisy, and it is ultimately boring to carry out large-scale experiments to try to make a point. The use of computer packages such as MINITAB, which have the facility to draw random samples from a variety of populations, greatly facilitates the demonstration of many statistical concepts.

Simulation plays a key role in the conceptual development in the workshop sessions. For example the basic ideas of hypothesis testing were developed through simulating the binomial distribution. Examples of these situations were drawn from traffic flow and acceptance sampling. Results from this type of simulation proved to be particularly useful in getting across the ideas of Type I and Type II errors, which are often difficult for students to fully appreciate.

For example, in one experiment the number of successes from ten Bernoulli trials were obtained for several values of p, 0.4, 0.5, 0.6, 0.7 etc. A large amount of data is quite quickly generated, and patterns emerge clearly from the grouped class data. The binomial distribution is then used to calculate the Type I errors for each value of p, which are compared with the observed values. Further, the Type II errors can be examined by determining the proportion of times when the null hypothesis, p = .5, was not rejected when it should have been for each value of p.

Another concept which is often difficult to get across is the fact that you should not accept a null hypothesis as being true, rather you either reject it or do not reject it. This is readily seen by examination of such a class table: just because the number of successes falls between 2 and 8 does not prove that p = .5!

Other concepts which were investigated using simulations included the distribution of the sample mean, starting out with samples from different distributions. The shape, mean and variance can readily be investigated for different size samples and the concept behind the Central Limit Theorem demonstrated.

Another use is in developing the concept of confidence intervals. Simulating samples from populations of known characteristics and obtaining the confidence interval for each, we can see the proportion of times that the population parameter of interest is included within the interval.

The sampling distributions of the t, F and chi-square distributions can be investigated by randomly sampling from a normal population and using the appropriate transformations of the variables to generate columns of these statistics which can then be graphed. Also the sampling distribution of the correlation coefficient can be examined by taking repeated samples from a bivariate population for which the population coefficient is known.

Students and staff can be surprised at the variability which results from such simulations from known distributions. This helps develop a feeling for what is a reasonable degree of variability, an important lesson for future engineers to be learnt from a statistics course.

5. Other aspects of the course

While simulations are very useful ways of generating large amounts of data quickly and helping develop statistical concepts, there is no real substitute for the students actually generating data themselves. With this in view the students carry out several projects involving data collection. One concerned collecting information on current used car prices for a model of their choice. The analysis included exploratory data techniques with the main point being to aim to build a model to predict price from the year. The other exercise involved an experimental design in which students were randomly assigned to one of six cells in a two factor factorial design to investigate the effect of "handedness" and practice on reaction time. These exercises proved to be very useful in both stimulating interest and for illustrating some of the problems associated with data collection and the importance of experimental design.

Videotapes were also used, particularly ones which showed the use of statistics in engineering situations.

At present, the course is built around the MINITAB package because it is particularly easy to use and extremely useful in developing computer concepts due to its ability to generate samples from a wide range of distributions. The text used is the *MINITAB Handbook* (Ryan et al., 1989) which provides a good source of examples based on datasets available with the package. Other software is used when appropriate such as SAS for Operations Research applications and SPSS (PC+ or X) when survey research questions are involved. We also plan to use spreadsheets to help develop understanding of formulae and for seeing the effects of varying parameters.

Assessment should reflect the style and flavour of the course being taught. We have attempted to do this with the use of some practical assignments and some tests. The tests include an open book computer-based test in which students are provided with a dataset on which they have to carry out a number of analyses and make interpretations. There is also a final examination to test understanding with standard type problems but minimal computation requirements. Often questions involve interpretation of computer output. Students are not expected to remember many statistics formulae and class attendance and participation is also considered when determining final grades.

The course is under continual revision to try and improve students' interest in and appreciation of statistics as applied to engineering. Some of the ideas we have for the future are greater cooperation with engineering departments and use of engineering applications including more quality control work, greater use of student generated data and student projects, and the use of spreadsheet packages to help answer "what if" questions. In terms of equipment, we are seeking improved computer facilities, possibly including a class set of laptops which can be used in any classroom, and greater use of videos to show practical applications. We would like to see students spending more time on tutorial work, and we are analysing the material included in the course to see what can be dropped so that we can get the most efficient use of the time available.

6. Conclusions and discussion

A survey of student reaction to this course revealed that a large majority of students preferred this more practical, hands-on approach, over 75% of the 72 students opting for this method over alternatives. In particular, all nine students repeating the course strongly favoured this approach. Considering the teething problems, this is quite encouraging.

The main area of dissatisfaction arose from giving too little time to tutorial activities. In the future we intend to use some of the practical time for tutorial work. There was also lack of time to do technique-oriented problems in class.

Some problems we found in running the labs were that the computer can be a distraction from class discussion; students must stop working at the computer while the leader is talking. Lateness to class needs to be discouraged; not only is the tardy student confused, but the class as a whole is disrupted. Problems in timetabling the computer laboratory were encountered. Complete trialling of the worksheet before each session is important to save hold-ups and loss of interest if something goes wrong. Appropriate technology needs to be available in the laboratory, for example two overhead projectors and screens, datashow, whiteboard, video facilities.

In summary, our view is that traditional statistics teaching for engineering students should be reviewed in the light of modern technology and the needs of the modern professional engineer. One approach has been outlined in this paper in which practical work and computer simulations are integrated into the statistics programme. Statistical packages such as MINITAB should not be seen merely as computational tools. Rather they can have an important teaching role since they can be used to supplement, and sometimes replace, experiments to help develop statistical concepts. Computer simulation, provided it is simple to use, gives one way to help gain these concepts in an efficient manner but where possible should be used in conjunction with physical models.

References

Bajpai, A C, Calus, I M and Fairley, J A (1978) *Statistical Methods for Engineers and Scientists*. Wiley, Chichester.

Bisgaard, S (1989) *Teaching Statistics to Engineers*. Presented at the American Statistical Association Annual Meeting, Washington, DC.

Glencross, M (1986) A practical approach to the Central Limit Theorem. In: R Davidson and J Swift (eds) *Proceedings of the Second International Conference on Teaching Statistics*, University of Victoria, 91-95.

Groeneveld, R R (1988) *Introductory Statistical Methods - An Integrated Approach Using MINITAB*. PWS-Kent, Boston.

Harrington, P (1988) Using microcomputer spreadsheet programs to teach statistical concepts. *Proceedings of the Annual Meeting of the American Educational Research Association*, New Orleans.

Hunka, S (1988) *Fifteen Years of Teaching Elementary Applied Statistics Using CAI*. Report No RIR-88-6, Faculty of Education, University of Alberta.

Joiner, B L (1982) The case for using computers in the teaching of statistics. In: D R Grey et al. (eds) *Proceedings of the First International Conference on Teaching Statistics*, University of Sheffield, 307-312.

Ryan, B, Joiner, B L and Ryan, B (1989) *MINITAB Handbook* (2nd ed). Duxbury Press, Boston.

Shell Centre for Mathematical Education (1986) Micros in the lecture theatre. *Proceedings of the Undergraduate Mathematics Teaching Conference*, University of Nottingham.

Scheaffer, R L and McClave, J T (1990) *Probability and Statistics for Engineers and Scientists*. PWS-Kent, Boston.

Weiss, N A and Hassett, M J (1987) *Introductory Statistics*. Addison-Wesley, Reading.