What is Basic Statistics? Lessons from a Canadian-Indonesian Project

K Larry Weldon and P Tham - Burnaby, Canada

1. Introduction

If one were to judge by the topics covered in textbooks for introductory statistics courses, it would appear that nothing dramatic has happened in the foundations of the discipline over the last 30 years (Cobb, 1987). This is the same time interval when there has been an explosion of data-based research in both the natural and social sciences, and computers have changed the way this research is done, including statistical theory and method. Perhaps the claim by many science majors that statistics is not important to them is justified in the sense that the statistics they have learned is irrelevant to modern science. There is little doubt that statistics has a bad reputation among students: many enrol in statistics courses only because it is a requirement of their degree.

The reaction of teachers of the subject to these negative attitudes is often to assume that the unpopularity of their subject can be explained by its intrinsic difficulty and abstract orientation. The fact that mathematics majors tend to find statistics courses boring (unless they are offered in the style of a pure mathematics course) suggests that the difficulty and abstraction are not the essential problem. An alternative explanation is that the subject of statistics has been developed in ever more isolation from the applied science that gave birth to the discipline (see Mosteller, 1988). One thing is clear - the status of teaching statistics in the more developed countries is not so successful that one would want to reproduce this approach in the less developed countries (see Kempthorne, 1980; Minton, 1983; Shoemaker et al., 1986; and Hotelling, 1949 (reprinted 1988)).

This consideration of the question "What is basic statistics?" was initiated by an examination of training needs for statistics faculty at certain target universities in eastern Indonesia. The situation in eastern Indonesia is one of severe resource limitations (computers are available to faculty only), a poorly-trained cadre of teachers of statistics, a lack of scientific research activity at the universities, and a tradition of passive learning. In some ways, the special context of teaching statistics in eastern Indonesia has the

potential for encouraging a backward step in statistical pedagogy: with no student access to computers, and little real research taking place in the science disciplines, and with an emphasis on passive learning and avoidance of confrontation, the path of least resistance is to teach a course in the mathematics of statistics, in the mode common in North America in the 1960s. Since this tradition in North America is now considered to be part of the problem rather than of a solution, it was necessary to overcome this natural inclination in eastern Indonesia.

2. Changes in statistics instruction implied by the advent of computers

Although it is trite to recount the revolutionary impact on the discipline of statistics of computers, of the information explosion, and of the quantisation of academic research in both the sciences and the social sciences, it is nevertheless true that university instruction in statistics has been very slow to adapt to these changes.

For example, one would expect a modern curriculum to have a much greater emphasis on the modelling process, computer-intensive methods, non-parametric estimation, and graphics and related pattern-recognition skills. Perhaps too the special problems of handling very lage data sets, quality control of data and data screening, and data collection methods, require increased emphasis. And a more concerted effort is needed to explain the relevance of statistics to students of the natural and social sciences in terms free of statistical jargon.

Some statistics instructors in the developed world are making progress towards these goals (see, for example, Urquhart (1986), Chatfield and Schimek (1987)), but for many these goals are not clearly seen. For example, most instructors acknowledge the importance of including computer-based calculations in their course - but not everyone has changed their course content to reflect the impact of the computer. It is now possible to teach data display, data modelling, multiple regression, the analysis of factorial designs, simulation, and the bootstrap, in the very first course in statistics (see the attempt in Weldon (1986)) - whereas sophisticated ideas like the calculus of probabilities, transformations of random variables, formal confidence intervals, and tests of hypotheses, should perhaps be left for a future course. Use of a computer in a course is a step in the right direction, but it is perhaps less important than the modification of content to reflect the impact of computers on statistics.

"Basic statistics" in an academic context should include those topics that are most generally useful for future application. Surely such topics deserve to be called "basic theory". A student who understands this "basic theory" will be well on the way to knowing how to use it. But this basic theory is not entirely mathematical. In fact, many calculation ideas which were the focus of old-fashioned statistics courses are now simply quaint relics of no current relevance. (Choice of cell size for histograms, sum of squares formulas suitable for hand calculation, fixed-alpha tests, approximations to "intractable" probabilities, etc.)

We must consider what kind of introduction to the subject will best induce those who complete the introductory courses to successfully incorporate the key concepts of statistics into their professional lives. To expect scientists and others to see "science" in the mathematical abstractions of statistical techniques is expecting too much (see Julian, 1987). The science is actually part of the basic theory, and should be included in basic

statistics courses, even at the expense of some mathematical abstraction.

Statistics courses are often divided into "service" courses and "mainstream" courses. The argument above should indicate that the service courses are the mainstream courses. It is hard to imagine that a student could get an adequate training in statistics without having been exposed to a fairly heavy dose of applied problems in science or social science. It is equally true that a heavy dose of applied problems in science or social science would not necessarily entail any training in statistical theory. This is why teachers of statistics must introduce the theory of statistics in the context of applied problems (see Mosteller, 1980; Preece, 1986).

Before computers, basic statistics courses were of two types: one type was a purely mathematical course with little emphasis on the utility of the techniques, and the other was a concentration on calculation procedures with little emphasis on the theory. But as computers gradually solved the problem of getting the calculations done quickly, the mathematical courses should have included more applications and the calculation courses should have found more time for theory. But prejudices about "statistical theory" being "mathematical statistics" seem often to have thwarted both these potential improvements. Instead, the theory-oriented courses became entirely mathematical and the calculation-oriented courses incorporated a bewildering armoury of obscure procedures guaranteed to mesmerise any thinking person.

Attempts at a more reasonable incorporation of the computer into statistics instruction involve a substitution for the emphasis on calculation methods inherited from the old days with a new emphasis on the theory of statistics: this new theory involves scientific considerations which include mathematical abstractions but also a wide range of non-mathematical issues (see Weldon and Buenaventura, 1988).

In the bygone years when multiple regression with more than one or two independent variables presented a serious computational challenge, the methodologists of the time could be forgiven for not considering "influential" data values or the non-normality of errors. When a nonparametric density estimate would require a whole career to complete, a limitation of consideration to a small class of parametric forms was reasonable. But with modern computational facilities, a new perspective has evolved, so that these important modelling questions can now be addressed feasibly, and even routinely.

3. Use of computers without computer labs

Students in eastern Indonesia universities do not have access to computers. For a modern course in statistics, we must rely on an alternative strategy: have statistics instructors use computers and do scientific "quasi-research", and involve students in these activities as much as possible. This proposal is explained below.

Instructors at the target universities have recently been given access to computing equipment. This equipment is quite new, and the software provided is modern and adequate. The problem that remains is the lack of training of the instructors in either statistics or computing. But a package like MINITAB can be learned quickly by such novices, and an eager instructor would soon be able to do some computing for his statistics class. There is some early indication that the novelty and practical usefulness of the computer will provide some new motivation for the modernisation of

computer instruction. There are several activities that this might involve:

(i) Demonstrations to students with computer-generated simulations, either with hard copy, a live screen, or a video film of a computer session (see Saunders, 1986).

(ii) Consultant model: Students undertake simple data collection studies and use the instructor as a consultant and computer-assistant.

(iii) Instructor quasi-research: The instructor can initiate a research project and involve students in it. Students could each be asked to provide full details for data collection, description, analysis, or conclusions, having been provided with the appropriate input information. The fact that the instructor may not be in a position to suggest a genuine research problem in which publishable new findings are uncovered is not important - since teaching is the aim here, the main requirement is that the instructor and the students be interested in the project.

(iv) Re-analysis projects: Undergraduate theses are required of most students at the target universities. These theses sometimes involve the collection and analysis of data. A possible project for students is a critique of the statistical methods, and again the instructor can provide the computing facility if the student wishes to try a variation of the method in a thesis.

These are four fairly obvious ways to build in some contact with computing and scientific research problems to a statistics course. Some additional ideas for implementing these suggestions are given in Tanner (1985).

4. Lessons for teaching statistics in the more developed countries

The wider context of the Indonesian Project was the improvement of the quality of instruction in the basic sciences at the target universities. This science-oriented nature of the upgrading in statistics has encouraged consideration of the interaction of statistics with the sciences. One positive feature of the Indonesian context is a new pressure on university faculty to do research - many faculty look upon statistics as a key to the world of research. They reason that if they can become knowledgeable in statistics, they are more likely to win research contracts or be included in the contracts obtained by others. We make some use of this angle in the developed countries to promote statistical education. Our case will be much stronger if we integrate the theoretical aspects of scientific work into our statistics courses.

In considering how to overcome the lack of student access to computers, a number of strategies for instructor computer use in teaching statistics have been suggested. Some of these strategies will be effective even if students do have access to computers.

The current review of the problems with statistics instruction in a lesser developed country is like a look back thirty years or so at the situation in the more developed countries. If hindsight is clearer than foresight, we should at least be able to help the less developed countries to side-step problems we have met. Moreover, by looking back in this way, we may see where we have inadvertently missed the early promise of modern developments such as computerisation. Have we in the developed

countries adapted appropriately to the new environment for teaching statistics?

References

Chatfield, C and Schimek, M G (1987) An example of model formulation using IDA. *The Statistician* **36**, 357-363.

Cobb, G W (1987) Introductory textbooks : a framework for evaluation. *JASA* **82**, 321-339.

Hotelling, H (1988) The place of statistics in a university. Reprinted from a 1949 paper. *Statistical Science* **3**, 72-83. See also current comments on this paper by Moore, Bradley, Hogg and Hotelling Jr.

Julian, K (1987) Statisticians - keep it simple. *The Statistician* **36**, 487-492.

Kempthorne, O (1980) The teaching of statistics : content versus form. *The American Statistician* **34**, 17-21.

Minton, P D (1983) The visibility of statistics as a discipline. *The American Statistician* **37**, 284-289.

Mosteller, F (1980) Classroom and platform performance. *The American Statistician* **34**, 11-16.

Mosteller, F (1988) Broadening the scope of statistics and statistical education. *The American Statistician* **42**, 93-99.

Preece, D A (1986) Illustrative examples : illustrative of what? *The Statistician* **35**, 33-44.

Saunders, D J (1986) Computer graphics and animations for teaching probability and statistics. *International Journal of Mathematical Education for Science and Technology* **17**, 561-568.

Shoemaker, H H Jr et al. (1986) Academic versus applied training in statistics : approaches used by the US Bureau of the Census. *Proceedings of the International Conference on Teaching Statistics II.* University of Victoria, Victoria, BC Canada, 249-253.

Tanner, M A (1985) The use of investigations in the introductory statistics course. *The American Statistician* **39**, 306-310.

Urquhart, N S (1986) Use of computers in teaching statistics : some current and projected uses of computer technology in the teaching of statistics. *Proceedings of the International Conference on Teaching Statistics II.* University of Victoria, Victoria, BC Canada, 184-189.

Weldon, K L (1986) *Statistics : A Conceptual Approach.* Prentice-Hall.

Weldon, K L and Buenaventura, M (1988) *The Foundations of Applicable Statistics.* Research Report No 88-06, Department of Mathematics and Statistics, Simon Fraser University, Burnaby, BC Canada V5A 1S6.