Teaching Statistical Inference within Applied Statistics

Roger Mead and John Whitehead - Reading, England

1. Introduction

The study of statistical inference is a component of almost all statistics degree courses. In an applied statistics course, with the objective of training future professional statisticians, it remains proper to acquaint students with the principles on which their discipline is based. We have rejected formal mathematical presentations of inference which involve a rigorous sequence of definitions and algebraic examples. In our hands at least, such an approach fails to excite the practically-minded student uninterested in mathematics itself, while failing to convey *statistical* insight to those who can master the algebraic manipulations.

This paper is based on our attempts to develop an inference course at Reading in which students learn about concepts primarily through project work. The course is preceded by courses on:

(i) probability and random variables;
(ii) descriptive statistics;
(iii) distribution theory;
(iv) initial ideas of regression and linear models;
(v) statistical computing.

The distribution theory course is taught in a traditional mathematical manner and the students tend not to enjoy it and consequently approach our course with (at best) misgivings. The statistical computing course introduces students to INSTAT, a PC based statistical package which has been developed in Reading, and is rather like MINITAB. Students continue to use INSTAT within our inference course.

2. Course structure

The course is built around a series of six problems. Each is used to motivate a different topic of statistical inference, showing how practical problems create the need to develop theory. The six topics included are:

(i) order statistics;
(ii) estimation theory;
(iii) likelihood;
(iv) confidence intervals;
(v) hypothesis testing;
(vi) robustness.

Each topic is introduced by means of the problem. A summary of the essential theory for the topic is provided in distributed notes, making reference to a standard textbook (Beaumont, 1980), and amplified in one or two lecture hours per topic. Other class contact hours are spent in guided computer practicals and in tutorials in which students attempt theoretical problems and exercises under supervision. Students produce a report on their investigation of the problem, which will usually include the development of theory and the use of computer simulation. It is stressed to students that problems have been oversimplified in order to make them approachable by second year undergraduates and in the short time available.

The assessment of the course is based on the students' reports and on a three-hour examination paper. In preparation for the latter, students are set additional exercises on each topic.

3. Computers and simulation

Computers are familiar to our students both for their power in calculation and their capacity in storage and manipulation of databases. They should also be an important part of statistical education because they can be used to deepen understanding of numerically-based concepts.

The use of computers in teaching statistics has been discussed in Mead and Stern (1972) and in previous ICOTS Proceedings. Much of the previous ICOTS discussion has concentrated on the use of the computer for illustration. The speed of the computer allows numerical investigation of sampling distributions. The computer can also be used to provide scenarios, through simulation, of practical problems for which students are required to design and analyse experiments or surveys and to produce answers to the practical problems (Mead and Freeman, 1970; Pollock et al., 1979).

Computers for rapid and complex calculation and for numerical illustration are very useful aids to teaching. However the real opportunities they offer will become apparent as we allow ourselves to rethink our syllabi and educational approaches in the context of the permanent availability of computers in education.

In our course we use the simulation capacity of the computer extensively and it is therefore appropriate to consider the role of simulation in statistical theory itself. Three situations in which simulation is used can be distinguished as follows.

1 **Exact Theory Available**	2 **Approximate Theory Available**	3 **No Theory Available**
Theory must be presented (within the mathematical capacity of the students).	Approximate asymptotic results can be derived	Theory is either unknown or requires more computing than the simulation.
Simulation can be effective to illustrate and make the theory intuitive.	Simulation enables the assessment of reliability of the approximate theory for specific purposes	Simulation provides the only available information on the properties of the assumed models.
THE EXACT THEORY IS DEFINITIVE.	THE SIMULATION RESULTS ARE DEFINITIVE.	ONLY SIMULATION RESULTS ARE AVAILABLE.

4. Topics and problems

4.1 *Accident data (estimation)*

Students are provided with an INSTAT worksheet containing 65 observations. These are the numbers of accidents attended to in a small casualty department on each of 65 consecutive days. The problem is to estimate the probability p of an accident-free day. Three estimators are suggested:

$$p_1 = D/65, \quad p_2 = \exp(-T/65), \quad p_3 = (64/65)^T,$$

where D = number of accident-free days and T = total number of accidents. The data give the values $p_1 = 0.154$, $p_2 = 0.092$ and $p_3 = 0.090$. Which is the most appropriate?

Exploration of the data shows that a Poisson model fits the daily accident counts. The accident rate λ is estimated by $\hat{\lambda} = T/65 = 2.385$. Students then simulate 25 samples, each of 65 Poisson observations with parameter 2.385. The corresponding 25 values of p_1, p_2 and p_3 are derived. Histograms and summary statistics are prepared.

Students usually claim that p_2 and p_3 are better estimators than p_1. Discussion reveals that closeness of the mean to the true answer ($p = \exp(-\lambda) = 0.092$) and the concentration of the distribution about that value are the criteria which influence them. They have discovered unbiasedness and relative efficiency!

That guided and practical start now leads to a more mathematical investigation of these estimators. In two formal lectures the concepts of sufficiency, unbiasedness, relative and absolute efficiency, the Cramer-Rao lower bound and Rao-Blackwellisation are introduced. In answering the problem, students establish which estimators are unbiased, and which is more efficient than the others. Usually they need help in seeing that p_3 is the minimum variance unbiased estimate, and that p_2 and p_3 depend for their advantages on the Poisson assumption, whereas p_1 does not. However, the introductory computing exercise, and the concrete situation of this example, do help the concepts of estimation to be meaningful in a way which lectures alone fail to do.

4.2 *Disease incidence (likelihood)*

Data are provided on the number of cases of a disease in a sample of households for household sizes between 2 and 11. Data are available only for those households with at least one case of the disease and the zero-truncated Binomial distribution (probability of case = p) is a sensible simple model for the number of cases per household. Students are provided with the likelihood function, maximum likelihood estimating equations, and asymptotic variance (σ^2) for the zero-truncated Binomial distribution.

They are asked to investigate the likelihood function, with particular reference to shape, maximum and change with asymptotic standard deviation, for each household size and for the data combined over household sizes. It is also suggested that they simulate additional data and examine the equivalent behaviour of the likelihood for the simulated data.

Students have to work out how to construct an INSTAT program to find the ML estimator when the ML equations cannot be solved directly. Those who immediately type in the sample values for each household soon find that they need only the sufficient statistic, Σr, for each household size.

Examining the change in the log likelihood function as p is changed in steps of σ leads to discussion of how consistent the change of log likelihood should be for different household sizes. The variation of $\hat{p}$ (between 0.19 and 0.47) for different household sizes produces varying responses from students, some assuming intuitively that such a range is large, others that it is small, and yet others that it is about right relative to the values of σ^2.

This project provides the students with a "feel" for likelihood information and persuades most of them that they need to check the algebraic derivation of the theoretical results they are given. In the discussions the students themselves suggest the questions about the interpretation of estimates which lead naturally to the next two topics (confidence intervals and hypothesis testing).

4.3 *Other topics and problems*

The Order Statistics project is based on the idea of machines which include five identical "special connectors". Information on failure times of a sample of individual connectors is provided and students are asked to investigate the distribution of times to breakdown for a machine assuming either that the machine will operate only if all five connectors remain good, or alternatively, if only four remain good.

The project for Asymptotic Confidence Intervals is based on initial data of numbers of defects in lenses, one sample being of 150 values with a mean defect number about 1.5 and the other sample of 20 values with a mean defect number over 2.0. The data do not include any zeros, because only defective lenses are collected, and so the relevant distribution is the zero-truncated Poisson. Students are asked to calculate 90% confidence intervals for the true defect rate in all lenses and to then investigate the validity of the confidence level for these estimated intervals.

The Hypothesis Testing problem is entitled "The Giddiness Trial" and concerns the evaluation of a new drug in a clinical trial. The observed variable is the number of attacks suffered by the patient in the month following the administration of the drug. Students determine their hypothesis testing procedure for specified conditions, using power curves. They then work in pairs, each generating data for parameter values unknown to the partner who makes the test decision.

The last part of the course addresses robustness. This topic is neglected in many conventional textbooks and courses on inference, but we feel that it is an essential part of an inference course within applied statistics. Through theory and simulation, students investigate the properties of "optimal" procedures when the model assumptions under which they have been derived are violated. The principal problem is a reconsideration of the accident data (Section 4.1), with an exploration of the properties of the estimators p_1, p_2, p_3 when the 65 accident counts are not Poisson or not independent or not of constant rate.

Students do find this section of the course to be the most difficult. However, for those who understand it, the strengths and limitations of parametric inference are clarified. One important message which is conveyed is that for samples of reasonable size model violations which do not manifest themselves in simple goodness-of-fit checks seldom seriously compromise the model-based inference.

5. Discussion

As statisticians, we should at this point produce data demonstrating the (significant?) benefits of our approach when compared to a control group of students educated traditionally. We have no such data, nor any claims of a dramatic improvement of examination or assessment results. However, we do believe that our computer-project based statistical inference course has been more effective in persuading students to think about, and ask questions about the concepts that we have been trying to teach. This conclusion could be expected of lecturers developing and enthusiastically implementing a different approach to a subject. We are certain that students who put the necessary effort into the course derive much more enjoyment, as well as knowledge, from the experiential form of learning. Of course it is not possible for students simply to sit in lectures hoping that the ideas will seep in; but is that a loss?

In our form of course, the lecturer has less direct control over the direction in which students' learning proceeds. In general this seems to us to be no bad thing. Lecturers are not always the best judges of the order in which ideas should be assimilated and students may be more receptive to concepts introduced in response to their own questioning. For example, sufficiency is discussed in the estimation topic. However, in the likelihood topic students discover for themselves the clear dependence of the maximum likelihood estimate on only the sum of the set of values. The time they have spent laboriously simulating individual values is then both irritating and, later, enlightening. For some lecturers the serendipity factor in learning may be frustrating, but we should not undervalue the benefits of learning, disorganisedly, by experience.

We are also prompted by what we regard as the success of our approach to ask whether this approach to teaching could be applied to other topics. Developments in teaching the design of experiments and sample surveys have already been referred to. We believe that there are possible developments in courses on distribution theory, where there is traditionally a substantial amount of mathematical theory. Such theory might be better relegated to support notes for projects based on Poisson processes, derived variates, independence and joint distributions.

It is arguable that the syllabus content of our inference course, and of many others, is the result of the relative chronology of the development of classical statistical inference and modern computing. If modern computing facilities had been available eighty years ago then resampling techniques, such as jack-knifing, boot-strapping and randomisation tests, might occupy a more central position in statistical inference than they do today. And if these methods could have been developed earlier, at the expense of some of the more mathematical distribution theory, perhaps the statistical inference we teach should represent the body of basic knowledge which would have arisen. On the other hand, students' findings on robustness tend to increase faith in traditional parametric inference.

We hope that our approach will generate discussion and some reassessment of present teaching methods. If the teaching of any one of us ever becomes routine and lacking in originality then we shall have forfeited any right to expect our students' interest.

Acknowledgement

We have benefitted from discussions with various colleagues at Reading, but wish particularly to acknowledge how much we have learnt from our students.

References

Beaumont, G P (1980) *Intermediate Mathematical Statistics.* Chapman and Hall.

Mead, R and Freeman, K H (1973) An experiment game. *Applied Statistics* **22**, 1-6.

Mead, R and Stern, R D (1973) The use of a computer in the teaching of statistics. *JRSS (A)*, 191-225.

Pollock, K H, Mead, R and Ross-Parker, H M (1979). A sequence of games useful in teaching experimental design to agriculture students. *The American Statistician* **33**, 20-76.

INSTAT Manual.

MINITAB Manual.