Chapter 9

History and Teaching Statistics

Deborah Nolan

When Terry Speed arrived in Berkeley in the 1980s, I too was a new arrival. He was coming to Berkeley as a senior hire and I as a junior. It was through our connection to David Pollard that we discovered our mutual interest in teaching statistics. We first collaborated on a small project to introduce computing into the advanced undergraduate theoretical statistics course. The computing exercises we developed were aimed at students uncovering, through simulation studies, some of the rules of thumb that a practicing statistician regularly uses. This was not as successful as we had hoped because our students didn't see any reason to care about the simulation results. We had fallen into the trap Terry warns against in Speed [10]: teaching pseudo-applied statistics with context-free numbers. Subsequent attempts led us to connect the work to real applications and then to the template described in Nolan and Speed [8] and used in *Stat Labs: Mathematical Statistics through Applications* [9]. It would seem that this template should have been an immediate and obvious result of Speed [10]. It wasn't. While Terry modestly claims to be "no exception – for allowing ourselves to forget the fundamental importance of the interplay of questions, answers and statistics", I dare conjecture that one of his goals in the project was for me to gain experience through trial and error in developing an effective approach to teaching statistics.

Speed [10] successfully argues that the "whole point of statistics lies in the interplay between context and statistics." Others share this viewpoint as noted in the quotes included in the article from James, Cox, Dawid, and Tukey. However, Terry takes this assertion into the education arena and compels us to reflect this important thesis in our teaching. Following Speed [10], others have made similar arguments to change statistics education. According to Cobb and Moore [3], "The focus on variability naturally gives statistics a particular content that sets it apart from mathematics itself and from other mathematical sciences, but there is more than just content that distinguishes statistical thinking from mathematics. Statistics requires a different kind of thinking, because data are not just numbers, they are numbers

D. Nolan
Department of Statistics, University of California, Berkeley
e-mail: nolan@stat.berkeley.edu

with a context.” Higgins [6] and Nicholls [7] echo these statements; e.g., Higgins claims that “for the past 40 years, statistics has been doing a great job of training theoretical statisticians, but we have a more data based society and it is crucial that we identify changes to course content and delivery that need to occur.” Similarly, Wild and Pfannkuch [11] note “the biggest holes in our educational fabric, limiting the ability of graduates to apply statistics, occur where methodology meets context (i.e. the real world).”

One teaching strategy offered in Speed [10] is to meet people with data by, for example, pairing the statistics teacher with a teacher in an empirical field of inquiry or pairing statistics students with students who have subject matter knowledge. Anecdotal evidence of the success of this approach appears in Field et al. [5]. There we learn of the preparation of Betty Allan, Mildred Barnard, and Helen Turner for successful careers in biometrics at CSIR in the 1930s. All three women spent significant time in Rothamsted Station where they learned statistics by designing and carrying out experiments under the guidance of Fisher, Wishart, and Yates.

Terry raised and answered in his 1986 paper two common objections to working with real problems: that these problems are too complex and the data too large to be practical in the classroom and that only the most advanced students who have a sufficiently large set of tools can successfully attack real problems. Today, we face a new version of these same concerns. Data are now free and ubiquitous. People with all sorts of backgrounds have ready access to data. This data explosion is an enormous opportunity for us to make better, more informed decisions. However, this opportunity presents challenges because people expect to be able to interact with data in new ways and the role of the statistician is changing.

As I reflected on Terry’s call to change how we teach statistics, it was at first disconcerting to see that we are still asking statistics educators to consider this issue. Cobb [2] explains that “What we teach was developed a little at a time, for reasons that had a lot to do with the need to use available theory to handle problems that were essentially computational.” Efron [4] describes the mathematical statistics course as “caught in a time warp” that “does not attempt to teach what we do and certainly not why we do it.” Brown and Kass [1] examine statistics graduate training and warn us to break away from the view of the statistician’s role as “short-term consultant” because that model “relegates the statistician to a subsidiary position, and suggests that applied statistics consists of handling well-formulated questions, so as to match an accepted method to nearly any kind of data.” I have since realized that we must periodically revisit this question of how best to teach statistics and that is precisely the point. We are not aiming at a fixed target that once arrived at we will have accomplished our goal.

Efron [4] suggests starting over by imagining “a universe where computing preceded mathematics in the development of statistics” and advocates “starting more muscularly without worrying about logical order of presentation” and focusing instead on the basic kinds of reasoning and explanations that can be arrived at through randomization-based inference. Cobb [2] further develops this notion, explaining how randomization-based inference “makes a direct connection between data production and the logic of inference that deserves to be at the core of every intro-

ductory course.” Cobb further posits that “Technology allows us to do more with less: more ideas, less technique. We need to recognize that the computer revolution in statistics education is far from over.” Brown and Kass [1] advocate taking a “less restrictive view of what constitutes statistical training.” They see a blurring of the distinction between people with data and people with statistical expertise and state “some of the most innovative and important new techniques in data analysis have come from researchers who would not identify themselves as statisticians.” Brown and Kass recommend we minimize prerequisites to research, require real-world problem solving in our courses, and embrace a deeper commitment to cross-disciplinary training. Efron [4], Cobb [2], and Brown and Kass [1] advocate twenty-first century changes to statistics education that echo Terry’s call to include the value of statistics in our training programs.

Terry Speed’s advice from twenty-five years ago remains extremely relevant today as computational and data challenges continue to evolve and shape our field.

References

- [1] E. Brown and R. Kass. What is statistics? *Am. Stat.*, 63:105–110, 2009.
- [2] G. Cobb. The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education*, 1:1–15, 2007.
- [3] G. Cobb and D. S. Moore. Mathematics, statistics, and teaching. *Am. Math. Mon.*, 104:801–823, 1997.
- [4] B. Efron. Is the math stat course obsolete? In *Panel Discussion at the Joint Statistics Meeting, San Francisco, CA*, 2003. www.amstat.org/sections/educ/MathStatObsolete.pdf.
- [5] J. B. F. Field, F. E. Speed, T. P. Speed, and J. M. Williams. Biometrics in the CSIR: 1930–1940. *Aust. J. Stat.*, 30B(1):54–76, 1988.
- [6] J. J. Higgins. Nonmathematical statistics: A new direction for the undergraduate discipline. *Am. Stat.*, 53:1–6, 1999.
- [7] D. F. Nicholls. Future directions for the teaching and training of statistics at the tertiary level. *Int. Stat. Rev.*, 69:11–15, 2001.
- [8] D. Nolan and T. P. Speed. Teaching statistics: Theory through applications. *Am. Stat.*, 53(4):370–375, 1999.
- [9] D. Nolan and T. P. Speed. *Stat Labs: Mathematical Statistics Through Applications*. Springer Texts in Statistics. Springer-Verlag, New York, 2000.
- [10] T. P. Speed. Questions, answers and statistics. In R. Davidson and J. Swift, editors, *Proceedings: The Second International Conference on Teaching Statistics*, pages 18–28, Victoria, BC, Canada, 1986. University of Victoria.
- [11] C. Wild and M. Pfannkuch. Statistical thinking in empirical enquiry. *Int. Stat. Rev.*, 67:223–248, 1999.

QUESTIONS, ANSWERS AND STATISTICS

Terry Speed
CSIRO Division of Mathematics and Statistics
Canberra, Australia

A major point, on which I cannot yet hope for universal agreement, is that our focus must be on questions, not models. . . . Models can – and will – get us in deep troubles if we expect them to tell us what the unique proper questions are.

J.W. Tukey (1977)

1. Introduction

In my view the value of statistics, by which I mean both data and the techniques we use to analyse data, stems from its use in helping us to give answers of a special type to more or less well defined questions. This is hardly a radical view, and not one with which many would disagree violently, yet I believe that much of the teaching of statistics and not a little statistical practice goes on as if something quite different was the value of statistics. Just what the other thing is I find a little hard to say, but it seems to be something like this: to summarise, display and otherwise analyse data, or to construct, fit, test and evaluate models for data, presumably in the belief that if this is done well, all (answerable) questions involving the data can then be answered. Whether this is a fair statement or not, it is certainly true that statistics and other graduates who find themselves working with statistics in government or semi-government agencies, business or industry, in areas such as health, education, welfare, economics, science and technology, are usually called upon to answer questions, not to analyse or model data, although of course the latter will in general be part of their approach to providing the answers. The interplay between questions, answers and statistics seems to me to be something which should interest teachers of statistics, for if students have a good appreciation of this interplay, they will have learned some statistical thinking, not just some statistical methods. Furthermore, I believe that a good understanding of this interplay can help resolve many of the difficulties commonly encountered in making inferences from data.

My primary aim in this paper is quite simple. I would like to encourage you to seek out or attempt to discern the main question of interest associated with any given set of data, expressing this question in the (usually non-statistical) terminology of the subject area from whence the data came, before you even think of analysing or modelling the data. Having done this, I would also like to encourage you to view analyses, models etc. simply as means towards the end of providing an answer to the question, where again the answer should be expressed in the terminology of the subject area, although there will always be the associated statement of uncertainty which characterises statistical answers. Finally, and regrettably this last point is by no means superfluous, I would then encourage you to ask your-

self whether the answer you gave really did answer the question originally posed, and not some other question.

A secondary aim, which I cannot hope to achieve in the time permitted to me, would be to show you how many common difficulties experienced in attempting to draw inferences from data can be resolved by carefully framing the question of interest and the form of answer sought. A few remarks on this aspect are made in Section 6 below.

2. Why speak on this topic?

Over the years I have had many experiences which have lead me to think that the interplay between questions, answers and statistics is worthy of consideration. Let me briefly mention four, each of a different type.

The first experience is a common one for me. Someone is describing an application of statistics in some area, say biology. The speaker usually begins with an outline of the background science and goes on to give an often detailed description of the data and how they were collected. This part is new and interesting to any statisticians listening, most of whom will be unfamiliar with that particular part of biology. Sometimes the biologist who collected the data is present and contributes to the explanation, but at a certain stage the statistician starts to explain what she/he did with the data, how they were "analysed". By now the biologist is quiet, deferring to the statistician on all matters statistical, and terms like main effects, regression lines, homoscedacity, interactions, and covariates fly around the room. Sooner or later I find myself thinking "Here are the answers, but what was the question?" All too frequently in such presentations neither the statistician nor the biologist has posed the main question of biological interest in non-statistical terms, that is, in terms which are independent of analyses or models which may or may not be appropriate for the data, and I can certainly remember occasions when the analysis presented was seen to be inappropriate once the forgotten question was formulated. Of course many scientific questions can be translated into statements about parameters in a statistical model, so that I am not condemning all instances of the above practice.

A similar sort of experience is surely familiar to all who have helped people with their statistical problems. This time a scientist, say a psychologist, comes to me with a set of data and one or more questions. She/he knows some statistics, or at least some of the jargon. After being briefed on the background psychology and the mode of collection of the data I usually say something like "What questions do you want to answer with these data?", implicitly meaning "What psychological questions . . . ?" Not infrequently the answer comes back "Is the difference between such and such significant?" meaning, of course, statistically significant. [In my perversity I often think to myself: "Well, you should know; it's your data and you are the psychologist!"] Another similar query might concern interactions, or regression coefficients of covariates etc. What this has in common with the previous example is the unwillingness or inability of the psychologist to state her/his questions of interest in nonstatistical terms. We should all be familiar with the idea that scientific (e.g. psychological) significance and

statistical significance are not necessarily the same thing, but how many of us keep in mind the fact that the latter involves an analysis or a statistical model, and that there may be as many answers to this question as there are analyses or models? Surely much of the blame for such thinking rests with us, the teachers of statistics, who never fail to popularize the rigid formalism of Neyman-Pearson testing theory.

My third type of experience concerns recent graduates in statistics, students I and my colleagues have taught and whom we believe should be able to operate independently as statisticians. Many of these graduates go into jobs in big public enterprises: railways, agriculture bureaux, mining companies, government departments and so on, and a few – far too many for comfort – get in touch with us when they meet a difficulty in their new job. It is not the fact that they get in touch which is discomfiting, but the questions they ask! For we then learn how little they have grasped. They have questions in abundance, often important policy questions, access to lots of data, or at least the possibility of collecting any data that they deem necessary, but they are quite unsure how to proceed, how to answer the questions. Out there in the world there are "populations" of real trains, field plots, cubic metres of ore or people, and even the simplest question relating to a mean or a proportion or a sample size can be forbidding. Perhaps they should standardize something to compare it with something else, perhaps include the variability of one factor when analysing another, or something else again, all things which we feel that a graduate of our course should be able to cope with unaided. But how well did we train them for this experience?

Finally, and briefly, let me castigate my professional colleagues – and myself, since I am no exception – for allowing ourselves to forget the fundamental importance of the interplay of questions, answers and statistics, for in so many of our professional interactions we act as if it is irrelevant. How many times have we presented new statistical techniques to one another, illustrated on sets of "real" data, drawing conclusions about those data concerning questions no one ever asked, or is ever likely to ask? And how often do we derive statistical models or demonstrate properties of models which are unrelated to any set of data collected so far, and certainly not to any questions from a substantive field of human endeavour. We are, so we tell ourselves, simply adding to the stock of statistical methods and models, for possible later use. Is it any wonder that we or our co-workers then find ourselves using these models and methods in practice, regardless of whether or not they help us to answer the main questions of interest. For a discussion of some closely related issues of great relevance to teachers of statistics, see the two excellent articles Preece (1982, 1986).

3. Why this audience?

I don't think I will be very wide off the mark if I assume that most of you – at least the active teachers of statistics amongst you – have come from a background of mathematics rather than statistics, and that few of you have actually been statisticians before you started teaching the subject. I would further guess that many of you still teach mathematics, and perhaps at the

school level, statistics within a mathematics curriculum. It is on this assumption that I have chosen to focus on non-mathematical aspects of our subject, ones with which I feel you will generally be less familiar. As I said in the introduction, I hope that my talk will encourage you to give more attention to the non-mathematical aspects of statistics in your teaching, in particular to spend more time considering real questions of interest with real sets of data.

It is a curious thing that interest in the teaching of statistics in schools, colleges and universities has sprung up worldwide as an extension of mathematics teaching, because I certainly feel that the practice of statistics is no closer to mathematics than cooking is to chemistry. Both mathematics and chemistry are reasonably precise subjects in their own ways, and in general what goes on in them both is repeatable; perhaps they are true sciences. On the other hand, statistics and cooking are as much arts as they are science, although both have strong links to their corresponding science: mathematics in the case of statistics, and chemistry in the case of cooking. Who would recommend that a chemistry teacher with no cooking experience be appointed as cooking teacher as well? If I can convey to you some of the enjoyment and intellectual challenge that lies in my particular variety of cooking, and encourage you to try it yourself, I will have succeeded in my aims.

4. Two further examples

In this necessarily too brief section I offer two more concrete illustrations of interplay of the questions, answer and statistics. The first one is a very simple paraphrase of Neyman's classic illustration of hypothesis testing involving X-ray screening for tuberculosis, and I refer you to Neyman (1950, Section 5.2.1) for a fuller background and further details.

You have a single X-ray examination and, after the photograph has been read by the radiologist, you are given a clean bill of health, that is, you are told that there is no indication that you are affected by tuberculosis. With Neyman we will assume that previous experience has led to

$$\text{pr}(\text{clean bill} \mid \text{no TB}) = 0.99$$

$$\text{pr}(\text{clean bill} \mid \text{TB}) = 0.40$$

You now ask the radiologist "What are the chances that I have TB?" She says "I can't answer that question but I can say this: Of the people with TB who are examined in this way, 60% are correctly identified as having TB, and of" You interrupt her. "Doctor, I know the procedure is imperfect, but you have just examined my X-ray . . . What are the chances that I have TB?"

If your radiologist is sufficiently flexible and well informed, she will answer "Well, that depends on the prevalence of TB in your population, that is, on the proportion of people affected by TB in the (a?) population from which you may be regarded as a typical individual". Indeed a simple application of Bayes' theorem yields:

$$\begin{aligned}\text{pr}(\text{TB} | \text{clean bill}) &= \frac{\text{pr}(\text{clean bill} | \text{TB})\text{pr}(\text{TB})}{\text{pr}(\text{clean bill})} \\ &= \frac{0.40\text{pr}(\text{TB})}{0.40\text{pr}(\text{TB}) + 0.99[1-\text{pr}(\text{TB})]}\end{aligned}$$

At last you see how to get an answer to your question. It may not be easy to obtain a value for $\text{pr}(\text{TB})$: your smoking habits, the location of your home, your occupation, your ancestry . . . may all play a part in defining "your population", but this is what is needed to answer the question and it is far better to recognise this than to fob you off with the answer to another question not of interest to you.

If this example smacks of Bayesian statistics it is not entirely accidental, for there are many occasions where the Bayesian view (which is certainly not necessary in this example) helps answer the question of interest, whereas classical statistics refuses, frequently answering another, unasked, question instead. For a more complex, explicitly Bayesian example, see the very fine paper Smith and West (1983) concerning the monitoring of renal transplants.

My second example concerns the determination of the age of dingos, Australia's wild native dogs. A statistician was given a large body of data relating the age of a number of dingos to a set of physical measurements including head length. The data concerned both males and females, a number of breeds and animals from a number of locations, but for this discussion we will restrict ourselves to a single combination of sex, breed and location. The question, or at least the task, to be addressed was the following: produce an age calibration curve for dingos based upon the most suitable physical measurement, that is, produce a curve so that the age of a dingo may be predicted by reading off the curve at the value of the physical measurement. This curve was for use in the field and it was taken for granted that an estimate of the precision of any age so predicted would also be obtained.

It was found that a curve of the general form $h = a + b[1 - \exp(-ct)]$, where h and t are head length and age, respectively, and a , b and c are parameters of the curve, fitted the data from each dingo extremely well over the range of ages used. This was an exercise in non-linear regression with which the statistician took great care, special concern being given to the different possible parametrizations of the curve, the convergence of the numerical algorithm used, the residuals about the fitted line and to the validity of the resulting confidence intervals for a , b and c . The parameters estimated for different dingos naturally differed, although, not surprisingly, the values of a (head length at birth) showed less variation than those of b (ultimate head length $-a$) and c (a growth rate parameter).

All this seems fine, and you might wonder why I am mentioning this example at all in the present context. My answer is as follows. The statistician in question knew, or knew where to find, lots of information about the fitting of individual growth curves, and so he focussed on this aspect of the problem. To answer the original question, however, his attention

should have been pointed in quite a different direction, towards: the calculation of a population or group growth curve for the calibration procedure; features of the sample of dingos measured that may affect the use of their measurements as a basis for the prediction of the age of a new dingo; properties of the parameters which are relevant to this question; and, finally, towards obtaining a realistic assessment of the prediction error inherent in the use of the curve in the field.

In summary, he was willing and able to spend a lot of time on the individual animals' curves; he was less willing and less able to focus on the issues demanded by the question, those concerning population parameters, population variability, problems of selection, unrepresentativeness, and other issues including the use of normal theory, with real but not very well defined populations.

5. What is the problem?

Let me oversimplify and put my message like this. In the beginning we taught mathematics and called it statistics; much of this was probability, a quite different subject. Then, with the help of computers, we started to teach data analysis and statistical modelling; this was fine apart from one feature: it was largely context-free. The real interest (for others and many statisticians), the important difficulties and the whole point of statistics lies in the interplay between the context and the statistics, that is, in the interplay between the items of my title.

Let me offer a few similar views. A.T. James (1977, p. 157) said in the discussion of a paper on statistical inference:

The determination of what information in the data is relevant can only be made by a precise formulation of the question which the inference is designed to answer. . . . If one wants statistical methods to prove reliable when important practical issues are at stage, the question which the inference is to answer should be formulated in relation to these issues.

Cox (1984, p. 309) makes the following characteristically brief contribution to our discussion:

It is trite that specification of the purpose of a statistical analysis is important.

Dawid (1986) is even more to the point:

Fitting models is one thing; interpreting and using them is another, . . . If the model is correct and we know the parameters, how ought we to compare [schools]? . . . There is in fact no unique answer; it all depends on our purpose. . . . there remains a strong need for a careful prestatistical analysis of just what is required: following which it may well be found that it is conceptually impossible to estimate it!

Tukey and Mosteller (1977, p. 268) offer seven purposes of regression, or, as I would paraphrase it, seven types of questions which regression analysis may help answer. Summarized, these seven purposes are:

1. to get a summary;
2. to set aside the effect of a variable;
3. as a contribution to an attempt at causal analysis;
4. to measure the size of an effect;
5. to try to discover a mathematical or empirical law;
6. for prediction;
7. to get a variable out of the way.

Similarly, Tukey (1980, pp. 10-11) gives the following six aims of time series analysis;

1. Discovery of phenomena.
2. "Modelling".
3. Preparation for further inquiry.
4. Reaching conclusions.
5. Assessment of predictability.
6. Description of variability.

Similar numbers of aims, purposes, or types of questions could be given for the analysis of variance, the analysis of contingency tables, multivariate analysis, sampling and most other major areas of statistics. Yet how often do our students meet these techniques in context with even one of these aims, much less the full range? And how else are they going to learn to cope with the special difficulties which arise when questions are asked of them in context whose answers require statistics? This is the problem.

6. Some General Comments

In this section I will mention a few difficulties which I believe can be resolved in a given case when the relation between the questions asked, the form of the answers desired and the statistical analysis to be conducted are carefully considered. A full discussion of any one of the difficulties is out of the question, and even if that had been given, there would probably remain an element of controversy, something which would be out of place in a talk like this. The section closes with some further general comments about questions.

Some elementary difficulties which I think arise include

- What is the population?
- When are population characteristics (e.g. proportions) relevant?

- What is the "correct" variance to attach to a mean or proportion?
- When should we standardize (for comparison)?

I have found that the relations between statistical models and analyses on the one hand, and populations and samples on the other, with parameters playing a role in both, are something which puzzle many students of our subject. The former play a big role in standard statistics courses whereas the latter are prominent in applications. Just how they connect is not a trivial matter.

A few somewhat more advanced difficulties include

- Which regression: y on x , x on y or some other?
- When should we use correlation and when regression analysis?
- When can/should we adjust y for x ?
- Which error terms do we compare (in anova)?
- Should we regard a given effect as fixed or random?
- Which classifications (of a multiway table) correspond to factors and which to responses?

More subtle difficulties are associated with general questions such as

- Should we do a joint, marginal or conditional analysis?

I believe that in all of the above cases the difficulties arise because insufficient attention has been given to the nonstatistical context in which the discussion is taking place, and that when the question of interest is clarified and the form of answer sought understood, the difficulty either disappears completely or is readily resolved. Of course doing so takes some experience. Note that many of the difficulties listed involve, implicitly or explicitly, the notion of conditioning, or its less probabilistic forms, standardizing or adjusting. Just what we regard as being "held fixed" and what we "average over" in any given context is crucial, and here our questions and answers determine everything. The simplest form of this issue is usually: "Are we interested in just these units (the ones we have seen), or in some population of units from which these may be regarded as a (random?) sample, or both?" Models are no help here.

A simple but easy to forget aspect of the use of a statistical method is that not all questions which could be asked and answered by that method, are necessarily appropriate in a particular context. Lord's paradox, see Cox & McCullagh (1982) and references therein, provides a good example here.

7. What can/should be done?

It hardly needs saying that the best way to promote interest in the interplay between questions, answers and statistics is to put trainee statisticians into situations where they are required to provide answers to clearly stated questions on the basis of real data sets. Note that this can be a very different thing from "illustrating" a statistical technique on a set of data. In particular, much more background to the data is usually required, and this is rarely available in data sets presented in statistics texts. Indeed technical journals are now so tight with their space that it is rare to find full data sets published together with analyses and conclusions in scientific articles. This means that the best sources of suitable material of the kind being discussed, that is, of questions and data, are often one's colleagues or clients: teachers and researchers in other disciplines who make use of experimental or observational data in their work. Seeking out such material can be a way of forging links with the users of statistics and of course sandwich courses are designed with this general aim in mind.

One practice which I believe is valuable is the conduct of regular practical statistics sessions where students are asked to help answer specific questions on the basis of sets of data supplied together with background material. This is much more like the situation they will meet after their training is over. Two objections which are often expressed to me when I recommend this approach are (i) Surely it is unrealistic, except with the most advanced students, for unless they have learned a wide range of techniques, they will not be able to begin attacking "real" problems with any likelihood of success?; and (ii) Surely it is unrealistic, because real problems are so complex and real data sets so large, or even ill-defined, that nothing like what happens in practice can be presented in the classroom?

Both these objections have some validity, but let me make a few observations concerning them. Firstly, it is not necessarily a bad thing for a student (or anyone!) to attempt to answer a particular question (solve a particular problem) without knowing of the tools or techniques that may have been developed to answer just that type of question (or problem). This goes on all the time in the real world: parts of the wheel are rediscovered time and time again, and locomotion is even found to be possible without the wheel! And of course there is very seldom a single "correct" way to answer a question; an approach using less knowledge of techniques may well be better than one which uses greater knowledge. In the hands of a good teacher, such experiences can provide valuable object lessons, and, at the very least, valuable motivation for techniques not yet learned. Surely nothing could be more satisfying than hearing a student say: "What I need (to answer this question) is a way of doing such and such, under the following circumstances (e.g. errors in this variable, that factor misclassified, these observations missing or censored, that parameter chosen in a particular way, etc.)? Group discussions, where ideas are shared and knowledge pooled, are also most appropriate for this sort of work, and most enjoyable. The teacher can then play a subsidiary role, at times focussing the discussion back on the questions, perhaps at other times supplying a sought-for technique.

It would seem to me that this is just the sort of statistics which should be taught in secondary schools, not the watered-down and frequently sterile mathematical material which is often found at that level.

The second objection, that real problems are often very complex and rarely amenable to the sort of trimming that would be necessary before they could be used in a classroom, is harder to dismiss. It is certainly true that many (most?) problems are like this, but surely this highlights even more the difference between "illustrative" data sets, taken out of context, with no realistic questions or idea what would be satisfactory answers, and what we expect students to be able to cope with upon graduation. There is certainly a big gap here – between "pseudo-applied" statistics involving context-free sets of numbers, to illustrate arithmetic, and fully-fledged "warts and all" consulting problems – and I can only state that in my experience it is possible to find problem data sets which can be presented in the way I am suggesting. It certainly takes a little effort to find such material, particularly if you are not in the habit of meeting people with data and statistical problems. But as teachers of the subject, that is not such an unreasonable thing for me to expect of you is it?

A teaching strategy which could provide a means of putting these ideas into practice might be the following: pair yourself (the statistics teacher) with a teacher in an empirical field of enquiry, e.g. biology, agriculture or medicine, and also pair your statistics students with students in the corresponding class, requiring them to work together on a practical project which will enrich their understanding of both disciplines, and how statistics helps to answer questions. Many variants on this suggestion could be devised; the important thing is try something along these lines. Statistics students must meet more than mathematics and sets of numbers in their training, and it is the teachers of statistics who must arrange for this to happen.

Acknowledgement

I am very grateful for the discussions and comments on this topic offered to me by my CSIRO colleagues Peter Diggle, Geoff Eagleson and Emlyn Williams.

8. References

- Cox, D.R. (1984). Present position and potential developments: Some personal views. Design of experiments and regression. J.R. Statist. Soc. Ser. A., 147, 306-315.
- Cox D.R. & McCullagh P. (1982). Some aspects of the analysis of covariance. Biometrics 38, 541-553.
- Dawid, A.P. (1986). Contribution to the Discussion of: "Statistical modelling issues in school effectiveness studies" by M. Aitkin and N. Longford. J. Roy. Statist. Soc. Ser. A, 149, 1-43.

- James, A.T. (1977). Contribution to the Discussion of: "On resolving the controversy in statistical inference" by G.N. Wilkinson. J. Roy. Statist. Soc. Ser. B, 39, 157.
- Mosteller, Frederick & Tukey, John W. (1977). Data analysis and regression. Sydney: Addison-Wesley Publishing Company.
- Neyman, J. (1950). First Course in Probability and Statistics. New York: Henry Holt & Company.
- Preece, D.A. (1982). t is for trouble (and textbooks): a critique of some examples of the paired samples t-test. The Statistician, 31, 169-195.
- Preece, D.A. (1986). Illustrative examples: illustrative of what? The Statistician, 35, 33-44.
- Smith, A.F.M. & West, M. (1983). Monitoring renal transplants: An application of the multiprocess Kalman filter. Biometrics, 39, 867-878.
- Tukey, J.W. (1977). Contribution to the Discussions of "A reformulation of linear models" by J.A. Nelder. J. Roy. Statist. Soc. Ser. A, 140, 72.
- Tukey, John W. (1980). Can we predict where "Time Series" should go next? Directions in Time Series Eds D.K. Brillinger & G.C. Tiao. IMS. pp.1-31.

Teacher's Corner

Teaching Statistics Theory Through Applications

D. NOLAN and T. P. SPEED

This article presents a model for developing case studies, or labs, for use in undergraduate mathematical statistics courses. The model proposed here is to design labs that are more in-depth than most examples in statistical texts by providing rich background material, investigations and analyses in the context of a scientific problem, and detailed theoretical development within the lab. An important goal of this approach is to encourage and develop statistical thinking. It is also advocated that the labs be made the centerpiece of the theoretical course. As a result, the curriculum, lectures, and assignments are significantly restructured. For example, the course work includes written assignments based on open-ended data analyses, and the lectures include group work and discussions of the case-studies.

KEY WORDS: Instruction; Quantitative literacy; Statistical thinking.

1. INTRODUCTION

Over the past 10 years we have developed a course that teaches undergraduate upper-level mathematical statistics through the use of in-depth case studies, which we call labs. It is through the labs that the theory of statistics is introduced, which leads to an integration of statistical theory and practice in a way not commonly encountered in an undergraduate course. The labs raise scientific questions that are interesting in their own right, and they contain datasets for use in addressing these questions. The context of the scientific question is the starting point for developing statistical theory.

This article presents a model that we have found to be successful for designing these labs. The model calls for the lab to be a substantial exercise with nontrivial solutions that leave room for different analyses, and for it to be a central part of the course. The lab should offer motivation and a framework for studying theoretical statistics, and it should give students experience with how statistics can be used to answer scientific questions. An important goal of this

approach is to encourage and develop statistical thinking while imparting knowledge in mathematical statistics.

It often can be difficult for students to bring the mathematical statistics learned in the classroom to an independent project in school or on the job. It can also be difficult for students to make the transition from reading and understanding a critical review of a statistical analysis to successfully working on their own problem. In our attempt to address these issues, we propose here to mix theory with substantial case studies. In addition to increasing student understanding of statistical theory, through the labs students develop their quantitative reasoning and problem-solving skills in a broad, multidisciplinary setting. They become practiced in communicating their ideas orally and in writing, and they become versed in the use of statistical software.

2. BACKGROUND

The labs were designed to accompany a typical theoretical statistics course for juniors and seniors. We have used these labs for two such courses—one for mathematics and statistics majors and one for engineering and computer science majors. Both courses require students to have studied calculus for two years and probability for one semester. There is no statistics prerequisite for the course.

The material covered in the course is outlined in the following; it is divided into five broad categories. In a typical semester, roughly one week is spent on the first topic, summary statistics, and about three weeks are spent on each of the remaining four topics: sampling, estimation and testing, regression and simple linear least squares, and analysis of variance and multiple linear least squares.

1. Summarizing data

- a. *Univariate summaries*: histogram, box plots, bar charts, and quantile plots; measures of location and spread; normal approximation, skewness, kurtosis.
- b. *Bivariate summaries*: cross-tabulations, quantile-quantile plots, and scatterplots.

2. Sampling

- a. *Simple random*: sample mean, percent, and total; expectation, variance, and estimator for variance; normal approximation and confidence intervals.
- b. *Stratified*: expectation, variance; proportional and optimal allocation; bootstrap.

3. Parameter estimation, goodness-of-fit, and hypothesis testing.

D. Nolan is Professor, and T. P. Speed is Professor, Department of Statistics, University of California, 367 Evans Hall, Berkeley CA 94720-3860 (Email: nolan@stat.berkeley.edu). This work has been partially supported by NSF grant 9720490. The authors thank the reviewers for their comments which led to an improved and more informative article.

- a. *Parameter estimation*: method of moments; maximum likelihood; mean square error, asymptotic bias, and variance; information and Cramer-Rao inequality.
 - b. *Goodness-of-fit and testing*: chi-square tests; two-sample tests; test statistic, null distribution, and Type I and II errors.
4. Regression and the simple linear model.
- a. *Regression*: correlation; least squares; residuals, confidence intervals; bivariate normal.
 - b. *Simple linear least squares*: linear model, RMS error, transformations, polynomial regression, *t*-tests and *F*-tests.
5. Extensions to linear models
- a. *Analysis of variance*: one-way and two-way classification, additive model, *F*-test; least squares with indicators; sum of squares decomposition.
 - b. *Multiple linear regression*: geometry of least squares; diagnostics, model checking.

We started to develop labs for these courses in 1989. At that time, each lab consisted of a list of exercises designed to illustrate a particular statistical concept and to be applied to the data provided. These original labs were unsatisfactory for many reasons. They lacked context; the presentation was too much like an instruction manual; and they were separate from the development of the course material. Students viewed the labs as drudge work; they did not see them as important or relevant to the course. In response to these shortcomings, we began to expand the role of the labs. We kept those labs for which we had interesting scientific problems to accompany them, and we added to them detailed descriptions of the data, background material to put the problem in context, and suggestions for how to proceed in the analysis. In a further evolution, we incorporated the statistical theory into the lab. This was an important step in the lab development, because it made the labs the centerpiece of the theoretical course, not a secondary accompaniment.

As a result of this fundamental restructure of the labs, several changes took place in the course. The format of lectures changed. More time was spent on determining how to answer general scientific questions with statistical analyses and on deriving a statistical method from its application to a specific problem. Less time was spent covering many small examples constructed to illustrate a single statistical technique. Because the application serves as the motivation for working out the theoretical material, we still cover all the material traditionally covered in the course. Also, the course work changed; now lab assignments required written reports with accompanying graphs and charts, and exams contained questions on theoretical topics related to the labs. Finally, the material distributed to the students for each lab grew from a 2-page list of instructions to a 10- to 20-page in-depth case study. More details are provided in Section 4 on how the labs are integrated into the course.

3. A LAB MODEL

Over time, we have settled on a model for the organization of a lab. Each lab is divided into five main parts: an introduction, data description, background material, investigations, and theory. Sometimes we include an extension section for a more advanced analysis and related theoretical material.

We describe the content and important features of each section in the following. The descriptions are accompanied by examples taken from a sample lab, one that introduces parameter estimation and hypothesis testing by asking students to search for unusual clusters of patterns in the DNA of a herpes virus. Despite the technical nature of the topic, this lab is one of the most popular among the students. The students derive hypothesis tests that study features of the DNA against a model for random scatter based on the homogeneous Poisson process.

3.1 Introduction

Here a clear scientific question is stated, and the motivation for answering it is given. The question is presented in the context of the scientific problem, and not as a request to perform a particular statistical method. Typically the introduction ends with the question to be addressed by the data described in the following section. We do not address questions suggested by the data, instead we consider questions raised by researchers then use statistics to seek answers.

The introduction often begins with a current newspaper article on a related subject to convey the relevance of the question at hand. The DNA lab includes an excerpt from "Feat is Milestone in Study of Evolution" by Nicholas Wade, *New York Times* (08/01/95). The article describes the sequencing of a bacterium's DNA. It includes a quote from Watson, one of the codiscoverers of the structure of DNA, explaining the importance of the event.

Life is a mystery, ineffable, unfathomable, the last thing on earth that might seem susceptible to exact description. Yet now, for the first time, a free-living organism has been precisely defined by the chemical identification of its complete genetic blueprint.

The creature is just a humble bacterium known as *Hemophilus influenzae*, but it nonetheless possesses all the tools and tricks required for independent existence. For the first time, biologists can begin to see the entire parts list, as it were, of what a living cell needs to grow, survive and reproduce itself. . . .

Following the article, students are provided a brief introduction to the problem: how to complete the "parts list" for the human cytomegalovirus (CMV) by locating regions of DNA that may contain information on how the virus reproduces. To further motivate them, they are told that CMV is a herpes virus which is potentially life threatening for people with suppressed or deficient immune systems, and to develop strategies for combating the virus, scientists study the way in which it replicates. To study how the virus reproduces itself scientists need to find the site on the DNA that contains the instructions for starting the process of replication, called the origin of replication. DNA can be thought of as a long coded message made from the four-letter alphabet A, C, G, and T, and for other herpes viruses, the origin of replication is specially marked with a

Table 1. Locations for the First and Last 30 Palindromes in the Cytomegalovirus DNA

177	1321	1433	1477	3248	3255
3286	7263	9023	9084	9333	10884
11754	12863	14263	14719	16013	16425
16752	16812	18009	19176	19325	19415
20030	20832	22027	22739	22910	23241
...	...	204548	205503	206000	207527
207788	207898	208572	209876	210469	215802
216190	216292	216539	217076	220549	221527
221949	222159	222573	222819	223001	223544
224994	225812	226936	227238	227249	227316
228424	228953				

Source: Chee et al. (1990)

pattern of letters in their DNA code. Students are told that they are to identify any unusual patterns in the 229,354-letter-long CMV DNA. The search is restricted to one type of pattern—clusters of complementary palindromes. In the DNA alphabet, A is complementary to T, C is complementary to G, and a complementary palindrome is a sequence of letters that reads in reverse as the complement of the forward sequence—for example, GGGCATGCC.

3.2 Data Description

Documentation for the data collected to address the question is provided. It includes a detailed description of the study protocol, as appropriate. The data for the DNA lab are quite simple to describe; the data are the locations along the CMV DNA of any palindrome longer than 10 letters. See Table 1 for a list of the first 30 locations; the entire dataset is available on the Web site www.stat.berkeley.edu/users/nolan/statlabs. Chee et al. (1990) published the DNA sequence of CMV, and Leung, Blaisdell, Burge, and Karlin (1991) implemented a hash-coding computer program to screen the CMV DNA in search of patterns. Altogether, they found 296 palindromes that were at least 10 letters long. Palindromes shorter than 10 letters were ignored as they can occur too frequently by chance. The longest ones found were 18 letters long.

3.3 Background

Material to put the problem in context is provided in this section. Information is gathered from a variety of sources, and it is presented in nontechnical language. The idea is to present a picture of the field of interest that is accessible to a broad college audience.

For example, in the DNA lab, the material is at the level of the *Cartoon Guide to Genetics* (Gonick and Wheelis 1991). The background information provided includes: a summary of the history of the discovery of DNA; a brief introduction to the structure of DNA; the definition of a complementary palindrome in the DNA language; a description of how a virus' DNA replicates; information on the prevalence and symptoms of the herpes *Cytomegalovirus*; and mention of the burgeoning science of genomics, the study of living things through their full DNA sequences.

Teacher's Corner

3.4 Investigations

Suggestions for answering the question posed in the introduction are provided next. These suggestions are written in the context of the problem, using very little statistical terminology. The ideas behind the suggestions vary in difficulty, and are grouped to enable the assignment of subsets of investigations. Also included are suggestions on how to write up the results—for example, as an article for a widely read magazine; as a memo to the head of a research group; or as a pamphlet for consumers.

For the DNA lab, the students already have had experience with histograms and quantile plots, and it is expected that they will derive diagnostic tools and statistical tests for comparing the DNA data against a benchmark homogeneous Poisson process. Following is an excerpt:

How do we find clusters of palindromes?

How do we determine whether a cluster is just a chance occurrence, or a potential replication site?

To begin, pursue the point of view that structure in the data is indicated by departures from a uniform scatter of palindromes across the DNA. Of course, a random uniform scatter does not mean that the palindromes will be equally spaced as are milestones on a freeway. There will be some gaps on the DNA where no palindromes occur, and there will be some clumping together of palindromes. To look for structure, examine the locations of the palindromes, the spacings between palindromes, and the counts of palindromes in nonoverlapping regions of the DNA.

Simulation—One starting place might be to first see how random scatter looks by using the computer to simulate it. The computer can simulate 296 palindrome sites chosen at random along a DNA sequence of 229,354 letters, using a pseudo-random number generator.

Locations and spacings—Use graphical methods to examine the spacings between palindromes that are one apart, two apart, and so on. Compare what you find for the CMV DNA to what you would expect to see in a random scatter. Consider graphical techniques for examining the locations of the palindromes.

Counts—Use graphical displays and more formal statistical tests to investigate the counts of palindromes in various regions of the DNA. Split the DNA into nonoverlapping regions of equal length to compare the number of palindromes in an interval to the number that you would expect from uniform random scatter. The counts for shorter regions will be more variable than those for longer regions. Also consider classifying the regions according to their number of counts.

The biggest cluster—Does the interval with the greatest number of palindromes indicate a potential origin of replication?

How would you advise a biologist who is about to start experimentally searching for the origin of replication? Write your recommendations in the form of a memo to the head biologist of a research team of which you are a member.

3.5 Theory

The theoretical development of the statistical concepts and methodology appear after the problem is introduced, at the end of the lab handout. The material includes information on the general topics described in Section 2 of this article, as well as topics more specific to the individual lab, such as goodness-of-fit tests for the Poisson distribution and parameter estimation for the log-normal distribution.

In the DNA lab, the homogeneous Poisson process on a line is introduced. We discuss how the Poisson process seems a reasonable model for describing the scatter of palindromes on a sequence of DNA. That is, the CMV DNA has 229,354 base pairs and 296 complementary palindromes, and given the large number of base pairs, we can think of the location of a palindrome as a point on a line segment. If the palindromes are scattered randomly across the DNA, then the number of palindromes found in any small region should be independent of the number in another nonoverlapping region, and the chance that one small region has a palindrome in it should be the same for all regions of the same length.

In addressing the main question of the lab, the analysis centers around issues of how to test the fit of the model. To help students in this task, we explore various properties of the Poisson process—for example, the uniform distribution of the locations of hits, the exponential and gamma distributions for gaps between hits, and the Poisson distribution for counts of hits in intervals. Quantile plots and simulations are used as diagnostic tools. Hypothesis testing is introduced through the chi-square goodness-of-fit test for count data. We also derive a test statistic based on the maximum number of hits (palindromes) in a set of nonoverlapping intervals (segments of DNA). To perform the tests, we need to estimate the intensity parameter, and we derive method of moments and maximum likelihood estimators for the parameter. The students also need to address practical issues of how to select interval lengths, what to do if a cluster is split across intervals, and how to reconcile various results.

4. IMPLEMENTATION

We have incorporated the labs into two theoretical statistics courses—one for statistics and mathematics majors and one for engineering students. The typical enrollment is 20 students and 40 students, respectively. The classes meet three hours a week with the instructor and one or two hours a week, respectively, with a teaching assistant. (We have also used versions of these labs with simpler investigations and theoretical material in a course for social and life science majors).

We have created more than a dozen labs for these two courses, but we use only about eight in a semester. The labs are chosen by the instructor according to the topic (theoretical or practical) and the background of the students. These eight labs are divided between those labs that are discussed primarily in lecture and those that require students to do ex-

tensive analyses outside of class and to write short papers containing their observations and solutions.

Typically we ask students to write reports for four labs, about one for each of the five main topics in the course (see Section 2). Table 2 provides two example sets of assignments. Option A could be used in a class for students in the social and life sciences, and option B could be used for engineering students. In addition to learning the core material, many labs also include special topics, which are listed in the table. In the course, we cover all of the material that we formerly covered in the theoretical course (see Section 2) plus a few of these additional topics.

4.1 Lecture

Although both types of labs feature prominently in the course, they are handled somewhat differently in class. For both types of labs we found it optimal when lectures include: discussion of the background to a particular problem, where students who have taken courses in related fields can bring their own expertise to the discussion; and motivation of the theoretical material through discussion of how to address a problem from a lab. In addition, for labs on which students are to write reports, we hold regular question-and-answer periods where students raise concerns about their work. For the other labs, we have students work in groups as described in the following.

Roughly about one class period in three is spent on these types of activities. The remainder of time is spent in a more traditional presentation of theoretical results (with reference to the current lab).

4.2 Group Work

To facilitate group work in class, we supply handouts to the students. The handouts have one of two formats. In one case, they are abbreviated lists of investigations, and students are asked to come up with a plan of attack for addressing the questions, such as a way in which to summarize and present the data. Sometimes we prepare in advance a slew of possible summaries, and we sketch results on the blackboard as students discuss their plans. In the other case, students are given a set of charts and graphs and they are asked to further summarize and interpret the output in order to answer questions from the investigation. Groups write their solutions on the blackboard, and the instructor leads the class in a discussion of the analysis. It would be ideal to lead these discussions in a multimedia classroom where output can be projected onto a screen, or where students can be seated at display stations.

An alternative format for group work that we have tried, has students sign up for a lab in groups of three to five (one of the nonreport labs). Each group meets with the instructor outside of class for assistance in their investigations. They prepare a presentation of their results for the class with their own handouts and materials, and these presentations are synchronized with the curriculum and scheduled throughout the semester.

This second type of group work has the advantage over the former in that students receive individual attention from the instructor, and they have the opportunity to work on

Table 2. Sample Sets of Lab Assignments

Lab	Option		Topics	Additional material
	A	B		
Maternal smoking and infant health I		×	summary statistics	
Student use of video games	×		simple random sampling	ratio estimator
Household radon levels		×	stratified sampling	parametric bootstrap
Patterns in DNA		×	estimation and testing	Poisson process
Crab molting	×		regression	prediction
Hispanic voting behavior	×		regression	weighted regression
Snow gauge calibration		×	simple linear model	inverse regression
Maternal smoking and infant health II	×		multiple linear regression	indicator variables
Transgenic mice with down syndrome	×		analysis of variance	unbalanced design
Designing a paper helicopter		×	analysis of variance	2 × 2 factorial

NOTE: The data and investigations for these labs are available on the Web site: www.stat.berkeley.edu/users/nolan/statlabs.

the lab in greater detail than if it were only discussed in class. However, this approach exposes students to a fewer number of labs. As for the first approach, students seem to really enjoy working on problems in class where they receive immediate feedback from the instructor and see the variety of ideas their classmates have come up with.

4.3 Reports

Students work outside of class on a lab for two to three weeks. Their time is spent analyzing data and preparing reports on their findings. The datasets are complex, the analysis is open ended, and they must synthesize their findings coherently on paper. The students find this work very challenging, and we sometimes allow them to work in groups of two or three on their assignments.

It can be difficult to grade the lab reports because the investigations allow students to be creative in their solutions to the problem. We usually break down the score into four equal parts: one part for composition and presentation, including statistically sound statements; one part for a basic minimum set of analyses; one for relevant, readable, and understandable graphs and tables; and finally we make a list of several advanced analyses and look for a subset of these in the write-up. Sometimes we also request an appendix to the report for technical material.

The lab reports typically constitute 40% to 50% of the course grade, the remaining being grades for homework and exams. We also think it is important to include questions related to the labs on exams, because it maintains consistency with the approach we have taken in teaching theoretical statistics through applications.

4.4 An Alternative

M. Hahn at Tuft University teaches an undergraduate course similar to ours. The course requires three semesters of calculus and one semester of probability, and there are 15 to 20 students in the class, including statistics majors, mathematics majors, and graduate students in economics, biology, and engineering. The class meets 3 hours a week with the instructor, and there is an optional 1.5-hour problem-and-computer session led by a teaching assistant.

Last semester, she adopted some of our labs for her course, which she ran somewhat differently than discussed so far. The students worked on three labs—maternal smok-

ing and infant health, video game usage, and patterns in DNA—plus a final project of their own choosing. Together, project and lab reports were worth about one-third of the course grade.

The course was run in a seminar style. Students were responsible for reading the background materials on their own. Class time had plenty of discussion and brainstorming on how to solve the problems in the three labs, and instruction focused on the nature of the statistical problems needed to be solved to motivate the theoretical material to be covered. Assignments were not set in advance, but made on a daily or weekly basis. This allowed the course to be somewhat flexible in meeting demands raised by the students. As a result, all the material traditionally included in the course was still covered, new mini-topics were interjected as the need arose, and students felt that class time was essential.

The inclusion of the research project has an advantage over sole reliance on the labs we have prepared in that students tend to have greater enthusiasm for projects of their own choosing, which can boost retention. Students also can present their projects in class and their projects can be the source for theoretical exam questions. Projects can be problematic, however, for it can be difficult to ensure that all students wind up with successful, challenging projects.

4.5 Computing

Bringing the computer into the theoretical course enables us to go far beyond the traditionally small, artificial examples found in textbooks. But care is taken to keep the demands made upon the student at an appropriate level. For example, the teaching assistant is mainly responsible for providing assistance in the weekly section meeting on how to use the statistical software. Often the section meets in a laboratory room, where students double up at workstations to work on the assignment and the teaching assistant provides advice as needed. Other times, the assistant distributes handouts with sample code. In a course without a teaching assistant, handouts and office hours may be able to meet the students' needs, or a mini-course on the software could be organized.

5. EVALUATION

The student feedback to our labs has been very positive. In anonymous end-of-term evaluations, students report that

the labs helped them better understand statistical theory. They liked the practicality and concreteness of the course and the relevance to the real world. However, not all of the response to the laboratory exercises has been favorable. Most negative reactions center around computer anxiety, and students have requested more help with using software packages. We are considering offering a short course in the use of statistical software, for most students have no prior experience with such software.

In rating the overall effectiveness of the course, the median score was 6 with a mean of 5.6 on a scale of 1 (poor) to 7 (excellent), with 143 students from 5 classes reporting. The departmental average for courses at this level is 5.2.

We have also been pleased with feedback from students after the course. Many students report they have used their lab reports as writing samples for job applications, and that interviewers have focused on one of the lab projects in job interviews.

Hahn also met with success by including labs in her course. She found the students were motivated to learn the theoretical material necessary to understand and complete the lab assignments, and that they approached the topics with vigor since they saw them as relevant. She reports that

The students were extremely enthusiastic about the labs. They took them as a challenge and liked the fact that they were not cookbookish. The labs generated more discussion among the students than anything I have used in the past.

The change in the course also seemed to generate an atypical interest among students in continuing their study of statistics. Students requested advanced statistics courses to be offered, they asked for advice and assistance on studying statistical applications to engineering and biology, and several decided to apply for actuarial jobs.

Although some of our evidence is anecdotal, the success of the labs is encouraging.

6. CONCLUSION

Recently Cobb and Moore (1997) called for the design of a better one-semester statistics course for mathematics majors that both strengthens their mathematical skills and integrates data analysis into the curriculum. Others have called for similar courses (Foster and Smith 1969; Hogg et al. 1985; Kempthorne 1980; Moore and Roberts 1989; Mosteller 1988; Petrucci, Nandram, and Chen 1995; Whitney and Urquhart 1990) and many consider training in statistical thinking important (e.g., Daisley 1979; Joiner 1989; Schuyten 1991; Riffenburgh 1995; and Nash and

Quon 1996). But it is a challenge to bring statistical thinking into the advanced undergraduate mathematical statistics course. We advocate that we are better able to, as Mosteller puts it, affect broadly the minds and lives of our students if we include in the curriculum case studies that are derived from problems with important scientific questions that contain rich background information, and are integrated with development of the theoretical material.

[Received June 1998. Revised March 1999.]

REFERENCES

- Chee, M. S., Bankier, A. T., Beck, S., Bohni, R., Brown, C. M., Cerny, R., Hosnell, T., Hutchinson III, C. A., Kourzarides, T., Martignetti, J. A., Preddie, E., Satchwell, S. C., Tomlinson, P., Weston, P. M., and Borell, B. G. (1990), "Analysis of the Protein Coding Content of Human Cytomegalovirus Strain ad169," *Current Topics in Microbiology and Immunology*, 154, 126-169.
- Cobb, G. W., and Moore, D. S. (1997), "Mathematics, Statistics, and Teaching," *American Mathematical Monthly*, 104, 801-823.
- Daisley, P. (1979), "Statistical Thinking Rather Than Statistical Methods," *The Statistician*, 28, 231-239.
- Foster, F. G., and Smith, T. M. F. (1969), "The Computer as an Aid in Teaching," *Applied Statistics*, 18, 264-269.
- Gonick, L., and Wheelis, M. (1991), *The Cartoon Guide to Genetics*, New York: Harper Perennial.
- Hogg, R. V. (1985), "Statistical Education for Engineers: An Initial Task Force Report," *The American Statistician*, 39, 21-24.
- Joiner, B. L. (1989), "Statistical Thinking: What to Teach and What Not to Teach Managers," *ASA Proceedings*, 150, Alexandria, VA: American Statistical Association, 448-461.
- Kempthorne, O. (1980), "The Teaching of Statistics: Content Versus Form," *The American Statistician*, 34, 17-21.
- Leung, M. Y., Blaisdell, B. E., Burge, C., and Karlin, S. (1991), "An Efficient Algorithm for Identifying Matches With Errors in Multiple Long Molecular Sequences," *Journal of Molecular Biology*, 221, 1367-1378.
- Moore, T. L., and Roberts, R. A. (1989), "Statistics at Liberal Arts Colleges," *The American Statistician*, 43, 80-85.
- Mosteller, F. (1988), "Broadening the Scope of Statistics and Statistical Education," *The American Statistician*, 42, 93-99.
- Nash, J. C., and Quon, T. K. (1996), "Issues in Teaching Statistical Thinking With Spreadsheets," *Journal of Statistics Education* (online at <http://www.amstat.org/publications/jse/>), 4, 1.
- Petrucci, J. D., Nandram, B., and Chen, M.-H. (1995), "Implementation of a Modular Laboratory and Project-Based Statistics Curriculum," in *ASA Proceedings of the Section on Statistical Education*, Alexandria, VA: American Statistical Association, pp. 165-170.
- Riffenburgh, R. H. (1995), "Infusing Statistical Thinking into Clinical Practice," in *ASA Proceedings of the Section on Statistical Education*, Alexandria, VA: American Statistical Association, pp. 5-8.
- Schuyten, G. (1991), "Statistical Thinking in Psychology and Education," in *Proceedings of the International Conference on Teaching Statistics*, pp. 486-489.
- Whitney, R. E., and Urquhart, N. S. (1990), "Microcomputers in the Mathematical Sciences: Effects on Courses, Students, and Instructors," *Academic Computing*, 4, 14.