

CASE STUDIES IN THE MATHEMATICAL STATISTICS COURSE ®

Deborah Nolan
The University of California
USA

We have developed a model for teaching mathematical statistics through detailed case studies. We use these case studies to bridge the gap between statistical theory and practice, and to help students develop an understanding of the basic ideas in mathematical statistics. We also use them to motivate students to explore the concepts of statistics. Although we strongly advocate teaching mathematical statistics through case studies, there are many challenges that arise from this approach. In this paper, we describe how we incorporate case studies in the course, outline the challenges that we face in adopting this approach, and discuss our efforts to overcome these challenges.

INTRODUCTION

Cobb and Moore (1997) call for the design of a better undergraduate mathematical statistics course that both strengthens students' mathematical skills and integrates data analysis into the curriculum. Others have called for similar courses (Foster & Smith, 1969, Hogg et al., 1985, Kempthorne, 1980, Moore & Roberts, 1989, Mosteller, 1988, Petrucci et al., 1995, Whitney & Urquhart, 1990). However, it is a challenge to bring data analysis skills into the mathematical statistics course. We advocate that we are better able to achieve this integration by including case studies in the curriculum.

Nolan and Speed (1999) have developed a course that teaches mathematical statistics through in-depth case studies. Our approach integrates statistical theory and practice in a way not commonly found in an undergraduate course in mathematical statistics. Each case study centers on a scientific question; it contains a dataset to address the question, and we develop statistical theory in order to answer this question. There are three salient aspects to our case studies approach:

- The problem central to the case is introduced first, and background information on the problem and a description of data collected to address the problem are provided before any relevant statistical theory is discussed.
- The solution to the problem raised in the case study is not provided to the students. In fact, there are many possible solutions, which use many different types of analyses.
- The student plays the role of a consultant, analyst, government official, textbook author, etc. in developing and presenting the solution to the problem.

We have found that there are many advantages to incorporating case studies in this way in the advanced mathematical statistics course. For example, when case studies are central to the course, we find that the cases motivate students to learn the theory, and they provide a meaningful framework for the theoretical material. Solving a case gives the student experience with how statistics can be used to answer scientific questions, and it helps the student develop skills in statistical thinking. The student also becomes practised in communicating his or her ideas in writing, and he or she becomes versed in the use of statistical software. However, the instructor faces many challenges when incorporating case studies in the mathematical statistics course, challenges which are different from the familiar ones faced in teaching mathematical statistics in a more traditional style. These challenges include finding ways to:

1. format class meetings to integrate theory and practice,
2. foster productive group work,
3. help students develop effective writing and data presentation skills,
4. assist students in data analysis without turning assignments into cookbook analyses,
5. connect the data analysis project to the theoretical material,
6. teach statistical software rapidly,
7. give students fair and informative evaluations of their reports and analyses.

In this paper, we describe our model for teaching mathematical statistics through case studies, and we provide examples of how we address each of these seven challenges.

THE CASE STUDY APPROACH

Using case studies we (Nolan & Speed, 2000) blend mathematical statistics with modern statistical practice. Each topic of study is centered on a case study, and it is via the case study that the mathematical statistics is introduced. The presentation of a case is divided into five main parts: introduction, data description, background material, investigations, and theory. In the introduction, a clear scientific question is stated, and 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 operation. We also provide students with documentation for the data collected to address the question, including a detailed description of the study protocol. In addition, students read background material to put the problem in context. This information is gathered from a variety of sources, and presented in nontechnical language. With the background material, we attempt to create a detailed picture of the field of application that is understandable to a broad college audience.

We make suggestions to the students on how to begin to answer the scientific question. These suggestions use very little statistical terminology. They provide a way to consider what is needed in a statistical approach to a problem, and so help students learn to apply statistical methods to different contexts and data. Students use statistical software to analyze the data. These data are often complex, and the analysis is open-ended. We require students to write a coherent synthesis of their findings. We have them play the role of a statistician or scientist when writing their reports. For example, the student may be asked to write an article for a widely read science magazine, a memo to the head of a research group, or a pamphlet for consumers. In addition, we often require a technical appendix for their "boss" the instructor. Essentially, the process of understanding the question, analyzing data, and summarizing findings in a report mimics what we do as data analysts, statisticians, and scientists.

The theoretical development of the statistical concepts and methodology required to address the question are introduced as the students begin their investigations. This approach to teaching mathematical statistics provides students with a reason to learn statistical theory because it is needed to solve a real problem, and it also generates an interest in learning the material because the problems have depth and merit. Students gain practical experience, problem-solving skills, and the ability to apply what they learn in the theoretical course to real applications. We view the divide between applied and theoretical statistics as unnecessary and unnatural, and think the integration of the two using case studies makes for a more compelling and complete mathematical statistics course.

CHALLENGES

1. Classroom Activities

The enrollments in our classes range from 20 to 60 students, and class meets three hours a week with the instructor and one hour a week with a teaching assistant. We break from the traditional lecture style to teach this course, because we find that an interactive teaching approach with in-class activities helps students develop their critical thinking skills and it helps prepare them for analyzing data on their own. In smaller classes, we run the class in a seminar style with time spent brainstorming on how to solve the problems in the investigations. Solving these problems leads to new statistical methods, and class time then switches over to lectures when we cover the mathematical side to these new methods. In the larger classes, we rely on group work in class to facilitate the discussion and analysis.

We analyze some case studies primarily through in-class activities, and others are mainly out-of-class group work. For the in-class cases, we supply handouts to the students to get them thinking about the problem presented in the case. These handouts typically take one of two forms. One type gives an abbreviated list of investigations, and we ask the students to work in pairs to come up with a plan of attack for how to begin to address the questions on the list. After they present their plan, we provide results from analyses we have prepared in advance in anticipation of their suggestions. Other times, we give students a set of charts and graphs and ask them to

further summarize and interpret the output in order to answer questions from the investigations. Groups present their solutions on the blackboard, and we lead the class in a discussion of their findings including how to interpret the analyses and how to address any shortcomings.

We also dedicate class time to a discussion of the background to a particular problem, where students who have taken courses in a field related to the application (e.g. biology, engineering, and sociology) bring their expertise to the discussion. In addition, when students are working on cases outside of class, we hold regular question and answer periods in class where students raise questions that they have come across in their investigation of the data. 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.

2. Fostering Productive Group Work

At the beginning of the course, we schedule several in-class activities such as those described in the previous section for students to meet each other and interact. These activities help the class establish as a group, and they help students choose partners for their work outside the classroom. As the semester progresses, we reduce the frequency of these prepared class time activities, and rely more on student initiated discussion.

One format for group work that we have had some success with has groups of students work on different cases, and lead a class discussion on their findings. Each group prepares a presentation of their results, including handouts for the class. These presentations are synchronized with the curriculum and scheduled throughout the semester. Students are given leeway in designing their presentation; some lead discussions modeled after our in-class activities, and others use role-playing as a research team to present their new research, or they hold a debate where they present conflicting statistical evidence to make their arguments. To help students prepare their presentations, we schedule two appointments with each group, one to discuss their investigations and one to review their handouts and plan the class presentation. In this type of group work, we expect a more thorough level of analysis, and a more in-depth report than for the other written assignments.

3. Encouraging Student Writing

It is important for statisticians to have good writing skills, and writing is an important means for helping students develop their critical thinking skills. We find the writing process to be an effective learning process where students piece their data analyses into a coherent story and develop arguments in support of their thesis. The importance that we place on writing skills is reflected in the grading of the reports: 25% of the grade for a writing assignment is based on the report's organization, clarity of argument, and correctness of the statistical statements.

We find it especially effective to require students to write for a real-life setting that fits the case study. Overall, the students enjoy such concrete assignments, and we find it helps them better focus on the problem when they have an audience, albeit imagined, other than the instructor. We dedicate class time to a discussion of how to prepare the written report. We emphasize the importance of making clear, careful arguments, and we encourage students to revise and proof read their manuscripts. Communicating ideas through charts and graphs is an particularly important part of a statistician's communication skills, and we dedicate one to two lectures on this topic, where we carefully review the many examples of how to display data badly found in Wainer (1984). Some of our favorite writing resources are Gopen and Swan's (1990) article on how to write clear scientific arguments, Tollefson's (1988) booklet containing examples of common grammatical errors, and Wainer's (1984) examples of poor data displays. We have prepared a synopsis of these writing tips (Nolan & Speed, Appendix A, 2000) for our students to read before their first assignment. In addition, Davis (1993) provides many good ideas on how to help students write better.

4. Avoiding Cookbook Analyses

When we first started to develop these data analysis and computer assignments, we were disappointed with the results. This was in part because the assignments were too prescribed. They were one-page instruction sheets with sample code, and the students followed the instruction

sheets without exploring or developing their own analysis of the data. To remedy this, we dramatically changed the format of the assignment, which led to the case studies approach described in this paper. In particular, the specific list of instructions for data analysis and simulation were replaced by suggestions for investigating the main question posed in the introduction of the case. For example, students are asked to "compare two groups," "assess the importance of the differences found," "pursue the point of view" exposed by an expert, or "develop a procedure for" making predictions. These sorts of investigations are written in the context of the case, and the student must figure out which statistical techniques are appropriate for accomplishing the task.

5. Connecting Data Analysis to Theory

We find it important to have the students connect the theory taught in the classroom to their work with data done outside of class. We try to achieve this connection between statistical theory and practice using simulation. That is, we have our students use the computer to explore the ideas in mathematical statistics as part of their analysis of a case. For example, we teach our students that the average from a simple random sample has an approximate normal distribution when the sample size is large in absolute terms and small relative to the population size. They invariably ask us how large is large and how small is small, and our answer is that it depends on the population. This is not a very helpful answer when faced with actual data.

To help address this problem, we demonstrate how to employ the bootstrap to get a picture of the sampling distribution of an estimator, and we expect a student to use this simulation technique to conduct his or her own study of the statistical properties of an estimator used in the student's analyses. The results from the simulation study are then included in a technical appendix to the report. We are continuing our efforts in this area and exploring how to develop simulation studies and interactive activities that use multimedia. These efforts are described in Buttrey et al. (2001).

6. A Crash Course in Software

We have had success in using the S-plus and R software to analyze the case studies. The syntax of both R and S-plus is based on the S-language. In addition, the R software is free, which makes it particularly attractive for students (lib.stat.cmu.edu/R).

Bringing the computer into the theoretical course enables us to go far beyond the small, artificial examples found in traditional textbooks. However, we have found that we need to be careful to keep the programming demands made on the students at an appropriate level. To help this situation, we often dedicate the one-hour spent with the teaching assistant each week to providing assistance on how to use the statistical software. The teaching assistant meets the students in a computer laboratory, where students double up at workstations to work on the assignment and the teaching assistant provides advice as needed. We also build a Frequently Asked Questions (FAQ) web page for each assignment. The page contains sample code and answers to questions students have asked in office hours, class, and section. Some of these FAQs are available at the website www.stat.berkeley.edu/users/statlabs/. In addition, we provide the code for the graphs and plots in Nolan and Speed (2000) to serve as code examples for the students.

7. Fair and Informative Grading

It can be difficult to grade students' written work because the investigations allow students freedom in finding a solution to the problem. To stress the importance of the case studies, we allocate 40-50% of the total grade to the written reports. We generally base our grading of a report on four aspects: composition, basic analyses, graphs and tables, and advanced analyses. Each of these four aspects receives equal weight in our grading scheme.

For composition, we look for an organized presentation, persuasive argument, descriptions made in the student's own words (not in the words of a reference), and statistically sound statements. As for the basic analyses, we list and expect to find in each report a minimum set of standard analyses. For the advanced analyses, we list several more complex analyses that would be appropriate for addressing the problem, and we look for a subset of these in the

student's write-up. The final 25% of the grade is based on relevant, readable, and understandable graphs and tables. We do not provide our list of acceptable analyses in advance because we want the student to determine what is relevant to his or her argument, and we want to be flexible in crediting a student who has come up with something not on our list.

We provide feedback to the students via the four-part grade and through our written comments on each paper. In addition, we handout a "solution" to the assignment where we cut and paste snippets from several reports to provide multiple examples of good introductions, analyses, and arguments. Finally, we also place restrictions on the length of a report, and the number of charts and graphs that may be included in it. For instance, short reports are up to two single spaced pages with at most 4 plots and tables, and long reports are typically a maximum of four pages plus 6-8 plots and an additional two pages for a technical appendix. We also include questions related to the cases on the course exams, which maintains consistency with our approach to integrating theory and applications.

CONCLUSIONS

The student feedback to our case study approach has been very positive. In anonymous end-of-term evaluations, students report that the cases helped them better understand statistical theory. They liked the practicality of the course and the relevance to the real world. However, not all responses have been favorable. Most negative reactions center around computer anxiety, and students have requested more help with using software packages. We have also been pleased with feedback from students after the course. Many students report they have used their reports as writing samples for job applications, and that interviewers have focussed the interview on a discussion of the student's analysis of a particular case. Although some of our evidence is anecdotal, the success of our efforts to incorporate case studies in the mathematical statistics course is encouraging.

REFERENCES

- Buttrey, S., Nolan, D., & Temple Lang, D. (2001). Computing in the Mathematical Statistics Course. To appear in *Proceedings of the Joint Statistical Meetings '01*.
- Cobb, G.W. & Moore, D.S. (1997). Mathematics, statistics, and teaching. *American Math Monthly*, 104, 801-823.
- Davis, B.G. (1993). *Tools for Teaching*. San Francisco: Jossey-Bass.
- Foster, F.G., & Smith, T.M.F. (1969). The computer as an aid in teaching. *Appl. Statistics*, 18, 264-269.
- Gopen, G.D., & Swan, J.A. (1990). The Science of Scientific Writing. *American Scientist*, 78, 550-558.
- Hogg, R.V., et al (1985). Statistical education for engineers: an initial task force report. *American Statistician*, 39, 21-24.
- Kemphorne, O. (1980). The teaching of statistics: content versus form. *American Statistician*, 34, 17-21.
- Moore, T.L., & Roberts, R.A. (1989). Statistics at liberal arts colleges. *American Statistician*, 43, 80-85.
- Mosteller, F. (1988). Broadening the scope of statistics and statistical education. *American Statistician*, 42, 93-99.
- Nolan, D., & Speed, T.P. (1999). Teaching statistics theory through applications. *American Statistician*, 53, 370-375.
- Nolan, D., & Speed, T.P. (2000). *Stat Labs: Mathematical statistics through applications*. New York: Springer-Verlag.
- Petrucelli, J.D., Nandram, B., & Chen, M-H. (1995). Implementation of a modular laboratory and project-based statistics curriculum. In *ASA Proceedings of the Section on Statistical Education*, 165-170.
- Tollefson, S. (1988). *Encouraging student writing*. University of California.
- Wainer, H. (1984). How to display data badly. *American Statistician*, 38, 137-147.
- Whitney, R.E., & Urquhart, N.S. (1990). Microcomputers in the Mathematical Sciences: Effects on courses, students, and instructors. *Academic Computing*, 4, 14.