Consulting : An Aid in Recruiting Statistics Students

Donald L Bentley - Claremont, California, USA

1. Introduction

The small four year liberal arts colleges in the United States have traditionally produced a disproportionately large percentage of graduates who have continued their educations beyond the baccalaureate to earn doctorate degrees in the sciences (Stanitski et al., 1986; Fuller). Unfortunately, this phenomenon does not occur in the field of statistics. This paper discusses the causes for the low production of statistics graduates from these small colleges, and attempts to gain insight into approaches which might improve the recruiting of undergraduate students from all colleges and universities into graduate programmes in statistics and related fields.

2. The liberal arts college

To understand the reasons for the successful yield of scientists by a small liberal arts college we must first recognise the educational environment which it provides. To begin, the primary purpose of the institution is the education of its students. While research activity is (or should be) present to a limited degree, it is secondary to that primary purpose. The greatest part of the college's resources is spent on teaching. And the most precious resources are the faculty and their time.

In contrast, at the research university most faculty have limited teaching loads. It is common for undergraduate courses to be taught by graduate students, with only a limited number taught by tenured faculty. Further, when courses are taught by permanent faculty, graduate assistants are assigned to absorb the responsibilities of the course other than the main lectures, thus limiting the opportunity for interaction between students and faculty. In particular, laboratories are conducted by graduate students, whereas these are run by tenured faculty at the small liberal arts college.

A second important and distinguishing attribute of the liberal arts college is the lack of graduate students. If a faculty member in the sciences requires a student assistant to help with his or her research, there is no alternative but to employ an undergraduate. In contrast, the faculty member at the university can use the advanced training of the graduate students. Such students might make major contributions to research projects in which undergraduates would be excluded from participation due to their limited background and experience. At the least, the graduate student requires less faculty time in supervision.

These two differences, limited research requirements on faculty, and use of undergraduates in research, go a long way to explain the success of small liberal arts colleges in the production of scientists. Faculty do participate in research, but the nature of the research can be limited to the level at which the undergraduate student can participate. And since there are no graduate students in competition for the research slots, undergraduates participate. Hence, the environment permits greater opportunities for the exposure of undergraduates to the thrill of scientific exploration.

3. The statistics environment

We statisticians are extremely dependent upon other areas of science. The sustenance of our professional lives comes from the data generated by scientists in other fields. There are relatively few statisticians who do not need this outside stimulation to remain professionally motivated. If each of us were to describe the events which led us to enter the field of statistics, the vast majority would respond with reflections centered upon the excitement of applying our tools to "real world data".

Working with data is an important educational exercise when trying to motivate students in statistics (Moore and Roberts, 1989). Recognition of this fact is demonstrated by the change in emphases that is occurring in many introductory texts. The current trend is toward "data-driven" curricula. But we must inquire as to which aspects of data analyses can potentially motivate students. Perhaps we can get some clues from our colleagues in the laboratory sciences.

There are at least three types of exercises involving data which should be available for students to work with at the introductory level. The first two types are becoming fairly common, but the third is not. The first is the standard fictional data set. The data set is small, and the numbers are made up so that the calculations and results are simple. Such problems are important in allowing the student to confirm an understanding of the formulae for various statistical procedures, but do nothing to demonstrate the true value of the statistical technique to scientific investigation. Exercises based on such data might be considered comparable to the exercises that students in the laboratory sciences deal with on homework sets or examinations.

A second type of exercise, which has only recently received much emphasis in our introductory courses, is that which is based on real data. These data are frequently extracted from the literature. With the computer it is reasonable for students to perform the analyses required of such data sets, provided they are properly selected. This is particularly true where the data are provided on a computer system so the student is only required to perform the analysis, and not enter and verify the raw data. A library of such data sets is starting to develop. Sudents who perform analyses on such real data sets

will gain a feeling for the role that statistics plays in the larger scientific community. Such an experience is analogous to that which students in other areas of science gain from the structured laboratory portion of their classes. The exercise is with the real thing, but it is an exercise that has been performed by generations of past students.

There is a third situation in which students in the laboratory sciences gain motivation in their discipline. This is the point at which the small liberal arts colleges currently have the advantage in the recruitment race. And that is the opportunity provided the undergraduates for involvement in original research. The analogous experience is lacking in statistical education. Working with real, published data sets as described above does not provide this experience. With the types of data sets and exercises discussed above, there is no sense of originality associated with the effort. The data have been manipulated by many students and professional statisticians before. The results of the analyses are well established. There is no sense of *ownership* of the analyses on the part of the student.

Students who have a potential interest in statistics as a career should be given the opportunity to become meaningfully involved in the analysis of data from original scientific investigations. Such an exposure allows them to experience the thrill of original research, and to legitimately "take ownership" of the data. They would encounter the excitement of being the first to know the results of the statistical analyses, knowing that these data had never been analysed before. Such experience is the analog of the research opportunities provided undergraduate students in the laboratory sciences. And such experience will recruit undergraduates to careers in statistics.

4. Statistical consulting

Statistical consulting provides the undergraduate statistics student with the analogous experience to undergraduate research in the laboratory sciences. However, the opportunities for such an experience are extremely limited on most liberal arts college campuses. First, the number of scientific projects requiring statistical analyses is limited by the small size of the faculty, and then further by the reduced percentage of faculty involved in research. Additionally, while most of these projects are scientifically less elaborate than would be found at a research university, the level of statistical sophistication required to properly analyse the data does not necessarily correlate with the scientific sophistication of the project. Hence, a further reduction in the number of statistical projects which are available for undergraduate participation. Finally, this lower level of scientific sophistication affects the perceived value of the research and leads to a reduced enthusiasm for participation in the projects.

Pomona College is a small liberal arts college in the greater Los Angeles area which is successful in encouraging students to enter the field of statistics. Over the last quarter century the college has graduated only 300 students a year, yet has averaged greater than one doctorate a year earned by its alumni in statistics and the related fields of biostatistics and epidemiology. A main ingredient in this recruiting success has been the exposure of its students to applications of statistics which have been provided through consulting experiences in the biomedical industry. This feeling is continually supported by comments from former students.

A list of projects with which students have been involved include: (a) soaking and cleaning solutions for hard contact lenses; (b) sterilising solutions for soft contact lenses; (c) design of sun screens; (d) proving efficacy of intraocular lenses; (e) evaluation of a retroprofusion process used with angioplastic surgery; and (f) investigation of the use of ambulatory tocodynamometry in high risk pregnancies.

These industry generated projects provide a wealth of opportunities for students to become involved in the analyses of meaningful data. The necessity for accuracy is clearly defined, an element frequently lacking in the classroom experience. The important role of statistics in the research is also made very clear to the student, both by the financial implications of the project to the corporation, and the biological implications to health care in general. In other words, when evaluating a new biomedical product the consequences of both the Type I and Type II errors gain real meaning.

There are numerous benefits derived from participation in industrial consulting. Not only is there benefit to the students who are assigned to work on the actual data, but the benefits carry over into the classroom. The faculty member overseeing the project can include examples from the project for classroom lectures, exercise sets, and examinations. Other students will gain an interest in the work and will keep questioning the progress of the project. The opportunity to present statistics in action motivates students to consider the field as a vocation.

5. Summary

Traditionally, the small liberal arts colleges have produced a disproportionately large percentage of the leaders in the laboratory sciences. A major factor behind this phenomena might be the opportunity that is provided the undergraduate student to participate in original research. Statistics educators can take a lesson from this success, and work at providing an analogous experience for those students with an interest in a career in statistics. Such experience can be provided by involvement in statistical consulting on scientific research which is *perceived* by the student to be important. At Pomona College, a prominent source of such research projects is the biomedical industry. Other undergraduate colleges should seek comparable projects drawn from industries located within a convenient geographical proximity.

References

Stanitski, C, Frankfort, F and Muir, M (1986) Science-active liberal arts colleges and the future of basic science. *Change* **18(6)**, 52-53.

Fuller, C *An Analysis of Leading Undergraduate Sources of PhD's, Adjusted for Institutional Size. Revised and Expanded.* A report prepared for the Great Lakes Colleges Association. ERIC Accession Number ED273187.

Moore, T and Roberts, R (1989) Statistics at liberal arts colleges. *The American Statistician* **43(2)**, 80-85.