RECRUITING AND TRAINING UNDERGRADUATES THROUGH STATISTICAL CONSULTING

DONALD L. BENTLEY
POMONA COLLEGE

TECHNICAL REPORT NO. 91-006

SLAW is supported by a grant from the Exxon Corporation
Statistics in the Liberal Arts Workshop (SLAW) is a group of educators concerned with the teaching of statistics. The workshop was initially funded by the Sloan Foundation. Continuing support has been provided by the Exxon Corporation.

Participants in the summer of 1991 are:

Donald L. Bentley  
Pomona College

George W. Cobb  
Mt. Holyoke College

Janice Gifford  
Mt. Holyoke College

Katherine T. Halvorsen  
Smith College

Homer T. Hayslett, Jr.  
Colby College

Gudmund Iversen  
Swarthmore College

Robin H. Lock  
St. Lawrence University

David S. Moore  
Purdue University

Thomas L. Moore  
Grinnell College

Norean Radke-Sharpe  
Bowdoin College

Rosemary Roberts  
Bowdoin College

Dexter Whittinghill III  
Colby College

Jeffrey Witmer  
Oberlin College
I: INTRODUCTION

There is the story of the famous mathematician who was teaching a course to undergraduate students. After he had completed the explanation of a particular result a student expressed confusion as to the line of reasoning which had been presented. The mathematician, with thick Eastern European accent, repeated the explanation word for word, and definitively stated "There, you understand now! We will go on."

Those of us who devote our professional efforts to the training of undergraduate students recognize that a major skill we must develop is the ability to explain the same concept in a variety of ways. What appears as a clear argument to one student may well be confusing to another. Extending this notion, if different students require different explanations of the material, why should it not be the same with the motivations? The aim of this paper is to address that question as it applies to the motivation and training of undergraduate students for careers in the field of statistics.
II: BACKGROUND

Pomona College is a selective, small liberal arts college in Southern California which has, over the past twenty five years, been successful in the production of statisticians. Each year during this period the college has graduated less than three hundred students, yet has averaged greater than one doctorate per year in Statistics or the related fields of Biostatistics and Epidemiology from among her alumni. A similarly large percentage of graduates have pursued graduate work in these fields and terminated at the Master's degree.

The undergraduate statistics curriculum at Pomona is similar to that of many colleges and universities within the United States. Several departments outside the sciences, including economics, sociology and psychology, provide their own introductory statistics courses for their majors. These courses do not have a calculus prerequisite. In addition, each year the Mathematics Department offers a one semester "service course", with a univariate calculus prerequisite. This course is directed toward the physical and biological science majors.

The majority of the Statistics curriculum within the Mathematics Department is directed at the Mathematics major. The first course is the traditional one year Probability, Statistical Inference sequence which has as prerequisite a course which includes an introduction to vector spaces and linear
transformations. The prerequisite represents the level of mathematical maturity we expect from the students rather than the specific mathematical material required for the course. Approximately seventy five percent of the Mathematics majors will take the Probability course during their junior or senior year, and the vast majority of those will continue on with the Statistical Inference course.

Two other courses are then available for students with an interest in Statistics. The first of these is a course titled "Advanced Topics in Statistics". The purpose of this course is to introduce those students planning to enter graduate work in the field to topics they are likely to come across later in their education. Included are introductions to Decision Theory, Nonparametric Inference, and a smorgasbord of other topics as fits the interests of the students and the instructor. At the end of the semester each student is required to make both a written and an oral presentation on a topic of his or her own interest. Choices have included, to name a few, Bootstrapping, Logistic Regression, and Darwin's influence on the development of statistics. This course is offered every year.

The second course can be taken by students who have completed a one semester service course in statistics. This course is an introduction to Experimental Design, and has been offered once every other year. What is perhaps evident from the above
descriptions is that the college does not have a data oriented course available for the students. A common suggestion (complaint) from our alumni is that we need to offer a course in which the student can gain experience working with data. For this, among other reasons, the Design course is being dropped, and in its place will be a data oriented course which will be offered at least once each year. The format and evaluation of this course will be the topic of another paper.

III: STATISTICAL CONSULTING

A unique component of the Statistics program at Pomona College which might well attribute to the success in recruiting students to continue in the field is the exposure the undergraduate students receive in the application of statistics to biomedical problems through consulting experiences with that industry. That the involvement has played a role in their decision to enter the field is continually supported by comments from former students. To understand the way in which the consulting experience has contributed to the recruitment of students at Pomona, it is worth while considering the different paths students follow to reach their initial exposure to the discipline.

A first would be the applied science student. This is the student who has been considering a major in the physical, biological, or social sciences. Such a student might well have
taken several semesters of calculus (either in college or, increasingly, in high school). At Pomona, over three fourths of all freshmen take a mathematics course in the first semester. These students have been exposed to data in laboratories connected with courses in their applied area, and are first exposed to statistics in connection with analyses of these data. They become excited about the way statistical methodology can be used to answer scientific questions. Further, they feel comfortable with their quantitative abilities, and find applied statistics a satisfying combination of their interests. This combination of factors naturally leads them to take more statistics, and consequently more mathematics.

A second route which students might take which would lead to a career in statistics would be through the service course. Students in this track usually have an interest in applied science, but are more generalists with regard to that interest. In many instances, these students tend to have an aversion to the laboratories required in many of these courses, and are searching for a science related major which does not require that form of exercise. Statistics offers a satisfying alternative.

The above two groups might well characterize the undergraduate students concentrating in statistics at the vast majority of institutions. But by far the greatest source of statistics concentrators at Pomona have come from the Mathematics major.
These students typically enter the college with the intent of concentrating in mathematics, and spend their first two years taking the courses required for such a major. This includes multivariable calculus followed by a year sequence studying linear algebra and differential equations. Students who are tending toward some area of applied mathematics are encouraged to take the Probability, Statistical Inference sequence during their Junior year (I feel it absolutely necessary that this sequence be offered every year - a necessity difficult to provide at many small undergraduate institutions). Since the course is primarily intended for Mathematics majors, the formal description and syllabus describe a fairly mathematically rigorous course. Mention of applications is kept to a minimum. This is the point at which statistical consulting enters the scene.

From the beginning of my tenure at Pomona I have been actively involved with consulting in the biomedical industry in Southern California. The proximity of the college to numerous smaller companies provides a variety of opportunities for such activities. These projects are incorporated into the statistics education in two ways. First, undergraduate students are hired to help with the required analyses. In this way students can experience the excitement that we professional statisticians feel as we attack and solve the methodological problems required for legitimate analysis of a data set, and participate in the interaction that takes place between statistician and
experimenter. The student consultants obtain a real ownership in the project, for it is one which no predecessor has worked on. There are other educational benefits which are experienced by the students who work on such projects. The need for accuracy in their work becomes emphasized, for a numerical error on an exam or homework assignment is usually forgiven (at least in part) but cannot be tolerated on a consulting report. The students find that they need to keep asking questions until they fully understand the project, and then be able to explain the results clearly to someone not necessarily trained in statistical methodology. The latter forces the students to a deeper understanding of that methodology than is usually acquired in the classroom.

The consulting projects have impact on students other than those directly involved in the analyses. In accepting a project I require that "Corporate Confidentiality" extend to the classroom. In other words, I reserve the right to talk about the projects to statistics classes. In this way the current projects become part of the discussion in class. The whole class can sense the enthusiasm I have for participation in original research, and they freely enter into discussion about the projects. Homework assignments or exam questions can be centered around issues involved in such projects, and the students then gain a greater sense of the importance of statistical methodology. In other words, the students experience statistics in real world application.
For many students, the application of their mathematical talents to real world problems is an exciting combination. Mathematics (statistics) becomes more than a mental exercise and grows into an area in which they can excel while at the same time make contribution to society. First hand experience to job opportunities, meaningful applications, and challenging quantitative exercise sway them to enter graduate study in statistics.

IV: SUMMARY

The field of statistics requires people with varying talents and interests. While many of these people will enter the field because of experiences they have in the traditional course structure, there are potential students who need different motivation. Included in this group are those who pursue a mathematical curriculum, but become excited with statistics when they recognize that it presents a meaningful area for application of their mathematical talents. An effective way of demonstrating this application is by involving these students in problems arising in statistical consulting.