My Life as a Statistical Scientist

by Brian S. Yandell

Yandell

I was always interested in mathematics. For a long time, I wanted to be a mathematician, but a few chance experiences (and people) guided me toward statistics. When I was in junior high school, I did a summer intern program at the Lawrence Berkeley Laboratory that gave me my first exposure to computers. I loved it, and looked forward to a repeat experience the next summer. However, funding evaporated, and I had to look elsewhere. As it happened, my father asked a friend if he knew of any jobs for someone who liked math and was interested in computers. Dave Wood said, "As a matter of fact, I am looking for someone. Have him give me a call." Dave, a professor at UC Berkeley in forest entomology, had a project modeling how bark beetles infest pine trees. He assigned me to work with Bland Ewing. Bland exposed me to the interdisciplinary world of math, computers, and biology in 1970, when computers were just becoming available.

He also introduced me to mathematical problems inspired by biological investigation. We considered spatial patterns of beetles in a forest and how to estimate beetle abundance using field traps and model the dynamics between short-lived beetles and long-lived pine trees. I worked with Bland each summer while I attended Caltech to get my bachelor's degree in mathematics. During my junior year, I became disillusioned with math. Pure math seemed too rarefied. An older college friend, who was then in graduate school, had switched from math to medicine—should I make a similar switch? Talking with Bland led me to consider biostatistics, a degree that would let me blend my

developing interests in math, computers, and biology. I applied during my senior year and was accepted at UC Berkeley. I delayed a year, taking a Thomas J. Watson, Jr., traveling fellowship to study the interface of math and biology. I spent a half-year in Europe, visiting scientists and attending conferences. Then, by chance, I traveled to India, where I spent a half-year immersed in wildlife biology, walking the trails of sanctuaries with field ornithologist Paul Spitzer in search of rare birds, rhinos, and tigers.

Returning to the United States was a welcome shock. The intellectual stimulation of statistics and probability theory at UC Berkeley was exciting, giving me a new way to think mathematically. I renewed my association with Bland and the forest entomology team under Dave, but their funding had shifted and I could not see how their research would lead to a thesis for me. Much later, as I became full professor, Bland, Dave, and I reunited.

At UC Berkeley, I continued my interest in ecological questions, rather than public health—the focus of the biostatistics degree program. I took courses in ecology and made friends with biologists. In biostatistics, I gravitated toward professors Betty Scott and Jerzy Neyman, eventually working for both of them. Betty was involved in numerous applied projects with colleagues on campus and encouraged me to deepen my collaborative interests. Neyman was in his 80s at that time, but retained a humorous, inquiring mind. He taught using the Socratic Method, sending me and other students to the blackboard to figure out problems a step at a time.

During the summer of 1976, I had another opportunity to travel while a graduate student, this time to Costa Rica with the Organization for Tropical Studies. I became fascinated with plant-insect interactions and tried to design a field study as the basis of my research. I returned to Costa Rica in 1979 for a few months, after completing most of my graduate coursework, but I could not crystallize a coherent project. What was I to do? Once again, I was at a crossroads. I considered leaving graduate school and completed the statistics master's degree as insurance.

Neyman suggested I work on the serial sacrifice problem upon my return from Central America. From there, I became interested in survival analysis. While I enjoyed his inspiration, I gravitated toward Kjell Doksum for formal guidance on my dissertation. (Perhaps it was our mutual interest in soccer, or maybe it was my fascination with nonparametric statistics.) As my research finally coalesced into a thesis, Neyman became ill and died. Shortly thereafter, Jack Kiefer died. It seemed time to move on from student life.

I asked my biology friends about jobs that combined ecology and statistics. They chuckled, as there was little funding in ecology and evolution in 1981. One friend suggested I consider statistical agriculture. As I pored over job ads, I found exactly that combination at the University of Wisconsin-Madison. I landed that position and remain today. My formal assignment is joint faculty in statistics and horticulture, bridging quantitative and (mostly nonhuman) biological fields as part of the Biometry Program. Half my teaching assignment is one-on-one consulting with students, staff, and faculty about experimental design and data analysis issues. My job is to guide statistical thinking and build collaboration, rather than to conduct the mechanics of inference. The goal is to raise the bar of statistical acumen among the biological community, to make research more efficient and cost-effective, and to advance science.

Some of these "in-house" consulting projects are short, but others have developed into long-term collaborations. More importantly for me, they have guided my thinking about statistical research areas. Because I did not see survival studies, my research efforts in that arena dried up. Nonparametrics, while important and useful in applied problems, did not adapt readily in my hands to repeated measures. I did produce some (I think) nice methodological work in those areas, but I often felt split between my theoretical interests and the pragmatic problems in front of me.

At some point, I was asked to teach two courses: linear models and statistical consulting. I noticed our applied training did not seem to go deep enough into nested designs and issues of unbalanced data, so I used examples from joint research in the classroom to ground ideas, giving students "word problems" in the spirit of Betty to help them uncover the design issues and methods of analysis most appropriate to the task. I pulled these ideas and examples into a book, *Practical Data Analysis for Designed Experiments*, which was published about 10 years ago.

The statistical consulting course was inspired by George Box, who founded our department with collaboration in mind, and built by Brian Joiner, who went on to build his own successful private consulting firm, Joiner Associates. I like our course model, which is very structured. The first third is focused on many small, one-week assignments that require students to think hard about how to communicate data concisely using graphs and words. Assignments overlap, forcing students to manage their time through triage, cutting corners to get practical results that capture 90% of the story in 10% of the time, rather than aiming for the full, 'correct,' sophisticated analysis. I get students talking right away, first with each another in small groups and, later, in the larger group.

The second third, I assign students major projects. A scientist from campus or local industry or government comes to the class with a project, which I pre-screen to keep from being too easy, but tractable, with a week or two of student effort. Students get a description and data a week ahead, and then meet face to face for two sessions. It is up to them to ask questions, learn what they need to know, and write a 12-page report that the scientist can understand and use.

In the third portion of this course, students work on an individual major collaborative consulting project guided by me. Meanwhile, I bring in local professionals—including Box, Joiner, Kevin Little, and Miriam Goldberg—to share their excitement about being consultants.

The richest consulting experiences for me have been those that persisted for several years. These collaborations have resulted in joint publications and research grant support. In the past decade or so, my focus has been on statistical genetics—again guided by the scientists who have approached me with problems. These longer collaborations have forced me to delve deeper into the biological questions. My two most successful collaborations have been with Tom Osborn in agronomy, working on flowering time of *brassica* species (canola oil), and Alan Attie in biochemistry, working on diabetes and obesity.

Notice I describe these collaborations in terms of the focus of study, rather than the statistical methods. In fact, many statistical

methods have been important over the years. But, the most exciting parts for me often have been in developing or adapting statistical methodology to the research question at hand. I can do this only by spending a lot of time with scientists in their labs. I attend weekly lab meetings, and then meet with members of the lab individually about specific questions. I now regularly teach workshops to quantitative biologists on quantitative trait loci or gene mapping, which build directly on these collaborations. I am continually refining how I communicate ideas, trying to ground ideas more in relevant examples. Yet, my goal in these collaborations and workshops is not to just give 'the answer.' I want my colleagues to understand the thinking behind the results, so they can share the ideas with their colleagues. Thus, I use mathematics in carefully prescribed doses to frame the statistical concepts. Communicating across disciplines is hard and requires me to be on my toes continually.

Along the way, I have had a few opportunities to consult in industry. Once again, these have arisen through personal contacts. A horticulture colleague put me in touch with a seed company that needed statistical perspective on their breeding program. This experience was a mixed success for me. I learned something about industrial priorities, but I situated myself as a technician too early. That made it hard to suggest major changes in thinking about process later.

I did a brief public/private consultation with a dairy company. My main lesson there was to document my contributions along the way. I was not satisfied with the proposed experimental design; it was too unfocused and small to give reliable results. When it came to the difficult meeting in which I had to explain the inconclusive experiment, I was able to point to my early objection. The middle management person at the meeting was not happy, as his job was on the line, but I stuck to my story and walked away with a clear conscience.

The most successful and longest-running consultation I have had with industry involves my brother, Bruce. He is now vice president of Tragon Corporation, a company specializing in market research/product testing. Bruce approached his bosses at a critical time, saying, "I know someone with the skills we need but…he's my brother." They had no objection, and we have been working together on fractional factorial designs, response surfaces, and various analyses for years. I helped them build a software product that automates much of their routine and now am helping them think about how to outsource some of their data analysis efficiently to increase productivity. Often, we talk more about human resources than statistical issues. It often comes down to time, money, and the unique people at hand.

People remain central to my career as a statistician. As I stated above, personal contacts led to school and job choices and yielded the best collaborations. When I was editor of Amstat Online, the ASA's web site, my primary responsibility was building communication among ASA staff and members, rather than technical issues concerning the internet. Emotions play a big part in how we think and solve problems. More and more, I realize how important a supportive work environment is for statistical science. Lately, I have been involved in hiring new faculty that bridge quantitative and biological sciences and mentoring these faculty through tenure. My consulting experience, particularly the building of communication skills across disciplines, plays a central role in my thinking about how to guide the next generation of statistical scientists as they establish their own careers. ■

Recommended

Amstat News would like your participation in the creation of a new section called "Recommended." Included in this section will be a list of statistical or mathematical books, movies, web sites, or publications that you and other ASA members have found useful or entertaining. If there is a resource you would like to recommend to friends or colleagues, email its citation—with a short description of why you recommend it—to *amstat@amstat.org*. Recommendations are accepted from ASA members and Section and Chapter representatives only. Commercial enterprises should contact the ASA Advertising Department at *advertise@amstat.org*.

Summer Reading

The Lady Tasting Tea by David Salsburg

"Salsburg's book is the story of statistical theory in the 20th century, its time of triumph, and of the mathematical/scientific geniuses who made it happen."

– Bradley Efron

"A fascinating description of the kinds of people who interacted, collaborated, disagreed, and were brilliant in the development of statistics." – Barbara A. Bailar

Web Sites

Social Science Statistics Blog:
www.iq.harvard.edu/blog/ss.

A source of thought-provoking social commentary and conversations about statistical methods and analysis.

Television

Numb3rs–CBS

NUMB3RS is a drama about an FBI agent who recruits his mathematical genius brother to help solve a range of crimes in Los Angeles, California. The math used in each episode is real and accurate and based on actual FBI cases. Additionally, the National Council of Teachers of Mathematics and Texas Instruments worked with CBS to design a program based on NUMB3RS to help students (and their parents) realize how relevant math is to everyday activity and success.