

Reviews of Books and Teaching Materials

Statistical Consulting.

Javier CABRERA and Andrew MCDUGALL. New York: Springer-Verlag, 2002, xi + 390 pp., \$74.95 (H), ISBN: 0-387-98863-7.

Since a large fraction of statisticians use their trade by interacting with non-statisticians working on applied problems, this book should have broad appeal. The authors have developed a diverse and readable book that tackles many aspects of consulting in good proportion. It describes the consulting process and common statistical methodologies, strategies of effective communication skills, and provides a broad range of case studies as a means of gaining consulting-related experience. It illustrates what information a consultant needs to obtain from the client, and what motivates those needs.

Although there are other books that effectively tackle the individual aspects described above, this book seems to be the one most ideally suited to teaching a well-rounded statistics course at the undergraduate or graduate level. Derr (2000) did a more thorough job of dealing with all aspects of oral and written communication, since that is the primary focus of her book. Boen and Zahn (1982) discussed the business aspects of being a consultant more comprehensively. Spurrier (2000) and Chatfield (1988) gave a more complete overview of statistical methods and the problem-solving aspects of determining an appropriate method for a particular situation. Other sources for case studies, ranging from straightforward to quite advanced, are Holcomb (2002), Peters and Gray (1994), Spurrier (2000), and Andrews and Herzberg (1985). *Statistical Consulting* gives informative and self-contained discussions for the many aspects of consulting in balanced proportions that would make using the book for a textbook delightfully straightforward. The collection of case studies is diverse in disciplines considered and level of difficulty, and seems to focus on interesting problems that students will find highly motivating.

The book is divided into two major parts: The first part, called the Methodology of Statistical Consulting, discusses the communication and statistical skills that a statistical consultant needs to be successful. The second part consists of three collections of increasingly complex case studies. Details of some of the contents of individual chapters are given in the following.

Chapter 1 gives an overview of the scientific method and addresses where a statistician should actively seek to contribute. As well, it discusses broad categories of problems commonly found in five major areas where statisticians are typically employed. This chapter gives a context for the discussions in the remainder of the book and does a superior job of giving students an overview of their possible roles in a number of careers. Chapter 2 addresses oral and written communication skills commonly used by consultants.

Chapter 3 discusses a wide array of statistical issues and frequently occurring methodologies, with terminology that a moderately sophisticated client would be able to grasp. Statistical methods described range from the elementary to the quite sophisticated. In addition, a good list of textbook references are given to provide a starting point for learning some of the methods presented. Key terminology and ideas of each topic are presented. Given the richness of the list of topics presented, it is clearly beyond the scope of the book to give complete details about any individual topic. The style of presentation of topics is ideal for readers with some general statistical expertise but with no previous exposure to the particular methods. A nice comparison of the statistical software packages SAS and S-Plus is given, outlining some of their relative strengths and weaknesses.

The final chapter in part I of the book provides details of the entire consulting interaction between consultants and clients, from initial contact to final written report. Excellent information is given about the first meeting, administrative details, the nature of the correspondence between client and consultant, as well as the thought process of the consultants at various stages of the project. In our graduate-level course in statistical consulting, we frequently teach consulting methods to students who are waiting for their first real-world experience as a consultant, which occurs partway through the course. This chapter will be particularly beneficial to help students to develop more realistic expectations about the consulting process and their roles in it.

Part II of the book contains case studies grouped into categories by their level of complexity. Twelve studies with partial analyses are presented with four from each of three categories, including "simple" (contingency tables, surveys, analysis of variance, and observational study assessment), "more-complicated" (regression, logistic regression, response surface methods, and time series analysis) and "research-oriented" (mixed models and some multivariate methods). Sufficient details about the contexts of the problems are given so that the reader has a realistic sense of the importance of each project to the client, the complications faced in analyzing the data, and the requirements for summarizing the results. Unlike so many textbooks, the data in these case studies has the feel of "real data," with inconsistencies, missing values, and multiple responses. Careful tables at the start of each case study make it easy to get a quick sense of the statistical methods suggested for the analysis, and the nature of the data collected. In addition, there are eight additional partial case studies that present only the setting of the problem.

The appendixes contain helpful information grouped into several categories. The resources section has links to Web pages where one can obtain datasets for all of the detailed case studies and some of the partial case studies. A second section has basic help for both SAS and S-Plus. The authors have designed this section to get a beginning user started, but it would need to be supplemented with other readily available resources.

Overall, this book is a valuable resource for statistical consultants, both beginning and established. As an instructor of a statistical consulting class, this is a prime candidate for use as a stand-alone textbook in my course, since it contains a desirable balance of materials with statistical methodology, oral and written communication skills, and rich case studies. A sample course outline is given in the book to give suggestions about how to organize a single-semester 15-week project-oriented course around the book. It will make a solid long-term reference for students. Also, for instructors of more traditional senior undergraduate and junior graduate courses, it provides useful case studies to illustrate standard methods in realistic settings that can easily be implemented.

C. M. ANDERSON-COOK
Virginia Polytechnic and State University

REFERENCES

- Andrews, D. F., and Herzberg, A. M. (1985), *Data: A Collection of Problems from Many Fields for the Student and Research Worker*, New York: Springer-Verlag.
- Boen, J. R., and Zahn, D. A. (1982), *The Human Side of Statistical Consulting*, Belmont, CA: Lifetime Learning Publications.
- Chatfield, C. (1988), *Problem Solving: A Statistician's Guide*, New York: Chapman and Hall.
- Derr, J. (2000), *Statistical Consulting: A Guide to Effective Communication*, Pacific Grove, CA: Brooks/Cole.
- Holcomb, Z. C. (2002), *Interpreting Basic Statistics*, Los Angeles: Pyrczak.
- Peters, L. H., and Gray, J. B. (1994), *Business Cases in Statistical Decision Making*, Englewood Cliffs, NJ: Prentice Hall.
- Spurrier, J. D. (2000), *The Practice of Statistics: Putting the Pieces Together*, Pacific Grove, CA: Duxbury.

Calculated Bets: Computers, Gambling, and Mathematical Modeling to Win.

Steven SKIENA. New York: Cambridge University Press, 2001, xiii + 256 pp., \$17.95 (P), ISBN: 0-521-00962-6.

It is interesting to look back at your childhood and see how you got interested in your chosen career. When I was growing up, I was interested in sports data, especially the statistics printed on baseball cards, and in playing indoor sports games such as tabletop baseball. I am pretty sure that these early interests in data and probability models led me naturally to a career as a statistician and, in particular, to my current research interests in statistics in sports. *Calculated Bets*

describes the author's similar lifetime fascination with a game called jai alai and, in particular, his "obsession with predicting the results of jai alai matches in order to bet on them successfully."

What is jai alai? It is a sport, originally played in the Basque provinces of Spain and France, where opposing players or teams alternate flinging a ball against a wall and catching it until one player misses and loses the point. To me, the sport seems to be a hybrid of lacrosse and racquetball. In the United States, jai alai is currently being played in Florida, Connecticut, and Rhode Island, and there are seven major jai alai stadiums (or "frontons") in these states.

There are eight players in a typical jai alai match. A special scoring system called Spectacular Seven Scoring is used. At the beginning of the match, two players compete to win one point. The loser leaves the court and the winner plays against the next player in the queue. After the seventh contest, the number of points given for a win increases from one to two points. The play continues until one player accumulates seven points—he is the winner of the match.

The fascination in attending a jai alai match comes from the betting aspect of the game. One can place bets on individual players to win the match. Or, like a horse race, one can bet on a particular player to place or show, or you can bet on pairs of players to finish first or second, or bet on a triplet of players to finish first, second, and third. At a fronton, the fronton (manager) will keep a percentage of the amount of money bet and give out payoffs dependent on the amount of money bet on each possible outcome.

Skiena attended a jai alai match in Florida with his family at a young age and was immediately curious whether one could consistently win by placing the right bets. This curiosity led him to a long-term study of the game. This book describes in detail the construction and implementation of a mathematical model for jai alai betting.

An important characteristic of a jai alai match is the ordering of the eight players at the start. As a first step in this study, the author constructs a Monte Carlo simulation in chapter 3 to learn about the probabilities of different outcomes assuming that the players are equally skilled. But to get a better handle on the likely outcomes, one needs to relax the unrealistic assumption that the players have equal abilities. To estimate the different abilities of the jai alai players, the author downloads a large number of player win/loss statistics from the Internet. He then proposes a model for estimating the probability that player A will defeat player B, given that one knows the proportion of points won by each player.

The next stage of the analysis in chapter 6 looks carefully at the economic structure of jai alai. There is a discussion of pari-mutuel wagering, and there is a detailed analysis of the payoffs from a large number of matches. This enables one to predict the payoff given any possible outcome of the match. This chapter also discusses the impact of one's wager on the payoffs and investigates the bets that are typically made by the public.

In chapter 7, knowledge about the abilities of the jai alai players and the economic structure is combined to develop an efficient system for betting. Skiena develops a computer program to look at the scheduled players for the next day, estimate the probabilities that player A will defeat player B in each competition, and then perform a large simulation to determine the outcome of all possible bets that day—the bets with the highest expected payoffs would be listed. The book provides a detailed description of how the system actually worked for a four-month time period. In a concluding chapter, the author provides practical guidance for betting for the novice who plans to go to a jai alai match.

Although I do not have a particular interest in gambling at jai alai (or gambling in any other sport), I found the book to be a very enjoyable read. The book is written in a leisurely, relaxed style. The author doesn't assume any mathematical, statistical, or computer science knowledge of the reader, and gives plenty of background material on the subject. In many places, the author takes side roads to interesting topics outside of the main themes. In particular, the book discusses the basics of Monte Carlo simulation; generating pseudo-random numbers; the publish-or-perish atmosphere at most universities; parsing text to extract information; PERL and HTML languages on the Internet; why programmers hate Microsoft; why jai alai players are like baseball players; classes of smooth curves, such as polynomials, trigonometric functions, and exponential; problems with over-fitting data; what makes a computer program beautiful; and the differences between conventional and Internet gambling.

Could this book be used in a particular course? It is a clearly written case study on the use of mathematical modeling and Monte Carlo simulation to study an interesting real-world problem. So it may be useful as a text for a course in applied mathematics or simulation. The book does contain a nice list of possible projects in the areas of gambling, sports, "people prediction," and financial modeling where the methods described in the book may be useful. I

recommend this book for anyone who is curious about how mathematics and simulation can be used to learn how to win at gambling.

Jim ALBERT
Bowling Green State University

News and Numbers: A Guide to Reporting Statistical Claims and Controversies in Health and Related Fields, (2nd Ed.).

Victor COHN and Lewis COPE. Ames, IA: Iowa State University Press, 2001, 211 pp., \$24.95 (H), ISBN: 0-8138-1424-3.

If you are a researcher or statistician who has been interviewed by a science writer in the past 10 years, chances are that the interview questions were influenced by the material in the first edition of *News & Numbers* by Victor Cohn. Reviews of that edition can be found in *JASA* (Tanur 1990) and *Chance* (Fienberg 1989). Cohn was in the process of writing this second edition when he was diagnosed with a fatal illness. He spent his remaining months making sure the book would be completed, in part by inviting colleague Lewis Cope to join him as a co-author. Much of the material from the first edition remains, expanded to include more recent examples and topics of current interest.

The book is written for science writers, but there are many audiences, including statisticians, who could learn a great deal from reading it. There are three themes seamlessly interwoven throughout the book. First, there are tutorials on basic statistics, study designs, and polls. Most of the material is excellent, but there are some important mistakes (as noted below) and this is the weakest aspect of the book. Second, there is information about how medical statistics are collected, summarized, and reported, and most of these explanations are exceptionally clear. For instance, the authors describe the difference between the incidence rate and the prevalence rate of a disease in a population by writing, "if incidence is the entering class, prevalence is the whole school" (p. 92).

The third theme addresses the intricacies of good science journalism. For instance, a scant five pages into the book there are "four areas for improvement" for science writers, starting with "we sometimes overstate and oversimplify" (p. 5). There is an entire chapter (5) titled "Questions You Can Ask," and the first one is, "is the study large enough to pass statistical muster?" (p. 61). There is an appendix listing the eight questions discussed in that chapter, as well as "four key areas for assessing a poll" (p. 180), which are discussed in detail in chapter 10. Given the length of the book and the number of topics it covers, the authors include an impressive breadth of questions related to designing and interpreting studies. There is one common mistake made in the interpretation of statistical studies that I wish had received more attention, that of making a causal conclusion from an observational study. The topic is only briefly mentioned in the book, and in wording that only includes a subset of such studies: "[Observational] studies, without comparison groups, are uncertain when it comes to determining cause and effect" (p. 51). Perhaps the lack of attention to this issue in the first edition of *News & Numbers* explains why it is still a common source of mistakes among journalists!

An informal survey of statisticians I know reveals that few of them had heard of the first edition of this book, but the same cannot be said of science writers. When I asked two of my favorite science writers to comment on it, they had glowing praise for Cohn and the book. Carol Morton, science writer for Harvard Medical School and correspondent for the *Boston Globe*, commented that "the book is considered a must-have on any science writer's shelf. . . [Cohn] was considered the dean of American medical writing, defining a standard of excellence that has outlived him" (Morton 2002, personal comm.).

Deborah Blum, Pulitzer Prize winner and President-elect of the National Association of Science Writers, wrote: "Victor Cohn was a pioneer, in the best sense of the word, and his book *News & Numbers* was that rare thing: a groundbreaking publication. No other journalist understood so quickly and so well that good reporters must have at least a common-sense understanding of statistics. Cohn was a science writer first, and his book addresses the needs of the science-writing community. But the book's reach is grander—it remains one of the best tools for journalists of every specialty. And so he remains one of the most influential science writers of his time, and through *News & Numbers* he continues to educate and to improve the rest of us" (Blum 2002, personal comm.).

A list of some of the chapter headings and subheadings for this book will illustrate the breadth it manages to cover. The first five chapters and chapter 10 (on polls) generally cover statistical and reporting issues. Chapter 6 includes "Patients' Tests," "Drugs and Drug Trials," and "Animals as Stand-Ins for Us." Chapters 7 to 9 are titled "Vital Statistics: Measuring Life and Health," "Health Plans and Hospitals: Measuring Their Quality," and "Our Environment and Measuring All Risks." Chapter 11 is "Politics, Your Pocketbook, and Plenty More."