Journal of Official Statistics, Vol. 18, No. 2, 2002, pp. 305–315

Book and Software Reviews

Books for review are to be sent to the Book Review Editor Jaki Stanley McCarthy, USDA/NASS, Research Division, Room 305, 3251 Old Lee Highway, Fairfax, VA 22030, U.S.A.

S. James Press and Judith M. Tanur. The Subjectivity of Scientists and the Bayesian Approach. New York: Wiley, 2001. ISBN 0-471-39685-0. 224 pp+appendix, references, index, 79.95 USD.

One of the things that are difficult to teach to methods and statistics students is the important role of subjective beliefs in science. The leading conception of the business of scientists is that they study facts, taking great care to be objective, and then cautiously come to their conclusions about these facts. The public idea of scientific research is still firmly grounded in 19th-century inductivism, a philosophic perspective that is nicely summed up by a quotation from Sherlock Holmes: ''It is a capital mistake to theorize before one has data'' (from Scandal in Bohemia, first published in 1892). It is precisely this view of science that Press and Tanur set out to eradicate. In their own words: ''We will show that the most famous scientists in history have all used their hunches, beliefs, intuition, and deep understanding of the processes they study, to one extent or another, to arrive at their conclusions'' (p. 1, Introduction).

The authors set the scene in the first chapter. Here, they explain that they have two goals in their book. First, they want to show that the notion of scientific objectivity is only partially true; science has both objective and subjective elements. Second, they want to show that modern Bayesian statistics is a valuable way to incorporate subjectivity in scientific research. Using various examples, they go on to discuss what they mean by subjectivity and objectivity, the evident diversity in scientific methodology, and the role of creativity and thought experiments in science. In addition, they discuss the practice of blinding the scientists against knowing which of the subjects are in the experimental group and which are in the control group. This practice is common in biomedical research, but is now also used in some physical experiments. This is an indirect admission of the existent risk of subjectivity even in a strict science such as particle physics.

Press and Tanur demonstrate the role of subjectivity in the scientific process by reviewing the work of a set of prominent scientists. In Chapter Two, they describe how they made their selection of prominent scientists. First, they selected five scientists because there were strong subjective elements in their work: Johannes Kepler, Gregor Mendel, Robert Millikan, Cyril Burt, and Margaret Mead. These are all discussed in Chapter Three. In addition, they take all the scientists from a well-known book (Meadows 1987) with scientific biographies of twelve scientists: Aristotle, Galileo, William Harvey, Isaac Newton, Antoine Lavoisier, Alexander von Humboldt, Michael Faraday, Charles Darwin, Louis Pasteur, Sigmund Freud, Marie Curie, and Albert Einstein. The lives and accomplishments of these twelve scientists, selected by someone else than the authors, are discussed in considerable detail in Chapter Four. Chapter Five, the last chapter, introduces Bayesian statistics, and argues that this is a fruitful approach to incorporate subjectivity in scientific research. Press and Tanur give several examples from different scientific fields to explain their position.

The stories Press and Tanur tell about their chosen scientists are interesting, and demonstrate clearly the strong role of subjective beliefs and hunches in science. This applies most to the five scientists chosen for their extreme subjectivity: Kepler, Mendel, Millikan, Burt, and Mead. Kepler had strong beliefs which inspired his cosmology that he massaged his data to fit his theory. Mendel and Millikan also massaged their data, and Burt has even been accused of fabricating his data. Mead appears to have let her subjective judgment steer the data collection to such an extent that one critic called her work so unscientific as to be ''not even wrong'' (p. 47). Yet some of these scientists, who were so convinced of their beliefs that their scientific practices bordered on scientific fraud, did discover important scientific laws. Kepler established the three laws of planetary motion, which are still known as ''Kepler's laws.'' Mendel's laws of heredity still stand, and Millikan's value of the electric charge of the electron was accurate enough, despite his practice of discarding observations that did not fit his theory. Apparently, strong subjectivity can go together with making major scientific discoveries. On the other hand Mead's interpretation of Samoan culture has been strongly criticized, and whether Burt's ideas about the hereditary basis of intelligence are correct is still unsettled. But the extent to which Burt and Mead let their subjectivity dominate their scientific practices has damaged their scientific reputation beyond repair. So, subjectivity in science is not always good.

The twelve other scientists discussed in Chapter Four are less extreme, but still show considerable subjectivity in their work. Together with the stories about the five extremely subjective scientists, these other stories present a strong case for the important role of subjective beliefs in science. In my view, Press and Tanur should have made a stronger distinction between the logic of scientific discovery and the logic of scientific proof. Some scientists were bright or lucky enough for their subjective beliefs to be right. So even if they massaged or misinterpreted their data to fit their beliefs, in the end they were absolved, because other scientists using rigorous methods proved they were right. In the logic of discovery, anything goes, including strong subjectivity. In the logic of scientific proof, there is considerably less freedom. The plea to recognize the important role for subjectivity in science is most relevant for making discoveries, not for proving hypotheses.

The chapter on Bayesian statistics is brief (25 pages) but manages to give a good description of the basic ideas and workings of Bayesian statistics. Press and Tanur demonstrate Bayes rule using the example of a medical diagnosis problem, where the diagnosis (posterior probability) is strongly influenced by the incidence of the disease (prior probability). They go on to discuss more complex examples, and argue that Bayesian statistics are a useful way to combine subjective beliefs (the prior probability) and empirical data into a more appropriate belief (the posterior probability).

Although I am convinced of the value of Bayesian statistics in scientific research, I am skeptical about their value in assimilating subjective beliefs of the kind described in this book. In Bayesian statistics, we are uncertain about the population value of a specific parameter (or set of parameters) in a statistical model. Bayesian statistics can then be used to combine our prior beliefs about this unknown value with empirical data, which then produces plausible values for this parameter. Press and Tanur argue (p. 217) that one could assign prior probabilities to the hypotheses that the theory is ''true'' or ''false'' and use the binomial distribution. For example, assume an experiment that can have two outcomes: a success or a failure. The probability of observing a success is 0.5 if the theory is true and 0.1 if the theory is false. Of course, if we have observed a success or a failure, we can now use Bayesian statistics to modify the prior probability that our theory is true. However, subjective beliefs like Kepler's belief that planets move in a certain way, or Einstein's belief that ''God does not throw dice'' (his main argument against the uncertainty principle) seem to fall into a different category. For instance, if Kepler's theory is true, the probability that planets move in ellipses is 1. If it is false, we have no idea what this probability is. The correct theory might also predict ellipses, or it might predict anything else. Philosophers of science like Kuhn and Lakatos have argued convincingly that the growth of scientific knowledge is a complex process. Kuhn has pointed out that competing theories may be incommensurable, which means that they are so different that they cannot be directly compared. Of course we can always choose to ignore such complications, somehow assign probability values to our theories and to experimental outcomes, and apply Bayesian statistics. But, somehow I doubt that any amount of Bayesian statistics would have settled the dispute between Albert Einstein and Niels Bohr as to whether the world is deterministic or probabilistic.

Still, it is fun to read all the stories about the selected scientists, and the importance of their subjective beliefs for their work is undeniable. It is good to realize that scientists are not cold-blooded logical machines, but human beings with all their follies and inconsistencies. In addition, the chapter on Bayesian statistics is a good and readable introduction. I just do not agree that Bayesian statistics solve the problem of subjectivity.

References

Conan Doyle, A. (1892). A Scandal in Bohemia. In The Adventures of Sherlock Holmes.
Meadows, J. (1987). The Great Scientists. New York: Oxford University Press.

Joop Hox
Department of Methodology and Statistics
Utrecht University
P.O.B. 80140
NL-3408 TC Utrecht
The Netherlands
Telephone: +31 30 2539236
Fax:+31 30 2535797
E-mail: j.hox@fss.uu.nl
http://www.fss.uu.nl/ms/jh

Michael Cowles. *Statistics in Psychology.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 2001, 235pp, ISBN 0-8058-3510-5, 24.95 USD.

Michael Cowles'' *Statistics in Psychology: An Historical Perspective* (2nd edition) ''presents an historical overview of the field – from its development to the present – at an accessible mathematical level'' (Quotes in this paragraph are taken from the back cover). ''Intended for advanced undergraduate and graduate students in psychology and other social sciences, this book will also be of interest to instructors and researchers interested in the origins of this omnipresent discipline.'' Through 15 mostly short chapters, the book ''provides insight into the disciplines of statistics and experimental design through the examination of the character of its founders and the nature of their views, which were sometimes personal and ideological, rather than objective and scientific.'' The value of such a treatment is presumed to be that it ''motivates further study by illustrating the human component of this field, adding dimension to an area that is typically very technical.''

As an applied statistician with a background in psychology, and psychological methodology and statistics in particular, and as someone who has spent over a decade interacting with thousands of students and practitioners in the social sciences concerning the use of statistical methods and software, I found the idea of a book with these goals very valuable, and the opportunity to review such a book truly exciting. Unfortunately, I fear that the effect of reading this book on many students (and instructors, for that matter) is less likely to be that of motivating further study than it is of serving as an excuse to continue denigrating the importance of understanding statistical methodology.

The introductory chapter, ''The Development of Statistics,'' sets the framework for the rest of the book. It situates the presumably relevant origins of statistical theory and methods for psychologists in the study of variation in natural populations, highlighting the centrality of the biometric and eugenic concerns of figures such as Francis Galton, Karl Pearson, and R. A. Fisher to the early development of the modern conception of statistics. The distinction (of no small importance to readers of the *Journal of Official Statistics*) is made between statistics as ''numerical descriptions of political and economic *states* (from which the word is derived)'' (p. 6), and ''. . . the process of reasoning about them'' (p. 7).

Brief discussions of probability, the normal distribution, and biometrics are followed by a final section entitled ''Statistical Criticism,'' in which the author states that one of the greatest dangers of using statistics in psychological measurement is a lack of awareness of the limitations of their use. Cowles notes that even experimental psychologists who see direct experimental control as the key to doing true science are often forced by the nature of their subject matter to substitute statistical for experimental control, and appropriately cautions that this substitution, while often inevitable, should not be done without considering its dangers.

The final paragraph of the chapter provides the fundamental assumption of the author's approach to the assessment of statistics in psychology: ''A balanced, but not uncritical, view of the utility of statistics can be arrived at from a consideration of the forces that shaped the discipline and an examination of its development'' (p. 20). The goal of the remainder of the book is to illustrate and defend this assertion.

Chapter 2, ''Science, Psychology, and Statistics,'' discusses some basic philosophical issues, and sums up the status of psychology as a science, with Cowles stating that many areas of psychology ''. . . are marked by undertones of ethics and ideology that the scientific purist would see as outside the notion of an autonomous science'' (p. 29). However, psychology is not alone in being affected by subjectivity in science, and this fact does not render the methods useless. Cowles states that even though the development of statistics may have been influenced by ideology, ''. . . its latter-day users do not have to subscribe to the particular views of the pioneers in order to appreciate its utility and apply it successfully'' (p. 29). This is one of the most important points in the book, and I fear that its importance will likely be lost on many readers amongst the numerous descriptions of the foibles and personal and ideological weaknesses of the founders of modern statistical methods.

No treatment of statistics in psychology would be complete without a discussion of measurement issues, which are the focus of Chapter 3. Some well taken cautions on the dangers of the worship of quantification, or the ''lust for measurement,'' are found here. Cowles states: ''To equate science with measurement is a mistake. Science is about systematic and controlled observations and the attempt to verify or falsify those observations. And if the prescription of science demanded that observations *must* be quantifiable, then the natural as well as the social sciences would be severely retarded'' (p. 37). These are indeed valuable insights for social scientists who are prone to falling into ''physics envy'' and placing far too much emphasis on the value of quantification.

The chapter ends with a discussion of errors in measurement and the limitations of measurement in a world where the Heisenberg Uncertainty Principle ultimately ''applies to *all* acts of measurement'' (p. 46). Though this claim may be a bit overstated, Cowles is definitely correct when he states that ''The world of psychological measurement is beset with system-disturbing features'' (p. 46).

After laying the philosophical groundwork in the first three chapters, Cowles proceeds in Chapter 4 to briefly discuss the origins of statistical information gathering as ''political arithmetic'' and vital statistics. Chapter 5 briefly discusses the origins of probability theory, the meaning of probability and some of the controversy surrounding this subject, and closes with a short section on the fundamental definitions of formal probability theory. Chapter 6 discusses binomial, Poisson, and normal distributions, and Chapter 7, entitled ''Practical Inference,'' goes into more detail on the controversies among statisticians over the meaning and use of probability, including Fisher's concept of fiducial probability and his battles against Bayesian approaches. The theme that runs throughout the book, that many of the pioneers of modern statistics were less than objectively scientific in their arguments and beliefs, is well illustrated by a quote from Maurice Kendall on the controversy between Fisher and those who refused to accept his notion of fiducial probability, to the effect that ''a man's attitude towards inference, like his attitude towards religion, is determined by his emotional make-up, not by reason or mathematics'' (p. 81).

Chapter 8 discusses sampling, estimation, and randomization. Chapter 9 covers χ^2, F, and t distributions, as well as the Central Limit Theorem. Chapter 10 covers Galton's discovery of the regression phenomenon, the invention of measures of correlation, and some of the controversies surrounding these measures. Chapter 11 discusses the early development of factor analysis methods and some of the controversies that have arisen surrounding the use of factor analytic techniques. Cowles closes the chapter by observing that factor analysis has been used for both theory confirmation and as an exploratory tool. He states: ‘‘Unless these two quite different views are recognized at the start of discussion the debate will take on an exasperating futility that stifles all progress’’ (p. 170).

Chapter 12 covers experimental design and the concepts of experimental and statistical control, the linear model and the design of experiments. Chapter 13 discusses Fisher’s analysis of variance, multiple comparison procedures, and the use of hypothesis tests vs confidence intervals, and closes with a note on one- vs two-tailed tests. Chapter 14 focuses on the increasing use of ANOVA methods in psychology and their coverage in textbooks. My own experience certainly validates the author’s assessment that, ‘‘in general, it is fair to say that many psychological researchers were not in tune with the statistical methods that were appearing. ‘‘Statistics’’ seems to have been seen as a necessary evil! Indeed, there is more than a hint of the same mindset in today’s texts’’ (p. 206).

The final chapter, entitled ‘‘The Statistical Hotpot,’’ discusses the disagreements and the enmity between Fisher and Neyman and Pearson, including sections entitled ‘‘Fisher versus Neyman and Pearson’’ and ‘‘Statistics and Invective.’’ Speaking of both sides, but particularly of Fisher, Cowles asserts that ‘‘Had the protagonists been more concerned with rational debate rather than heated argument, statistics would have had a quite different history’’ (p. 233).

Cowles appears to believe that the immaturity and pettiness of the pioneers of modern statistics set the stage for the failure of later practitioners in psychology and related fields to achieve sufficient understanding of the uses and limitations of statistical methods. One particular expression of this concerns what Cowles views as the unfortunate manner in which psychological statistics texts discuss the methods without reference to the people and problems that were involved in their genesis: ‘‘Failure to acknowledge the work of others, which was a characteristic of both Pearson and Fisher, and which, to some extent, arose out of both spite and arrogance, at least partly explains the anonymous presentation of statistical techniques that is to be found in the modern textbooks and commentaries’’ (p. 189).

While I find the behavior of Pearson and Fisher as regrettable as does Cowles, I am not at all certain that one need invoke this as an explanation for the relatively anonymous treatment of statistical methods in modern texts and commentaries. A simpler, and to me much more compelling, explanation is that statistical methods are grounded in mathematics. Most standard presentations of mathematical ideas and methods are concerned with the internal logic and structure of the methods themselves, or with assessing their usefulness in solving problems, rather than with discussing their development over time or the conditions and personalities involved in their genesis. Perusal of my own collection of statistics and mathematics texts indicates that mathematical subject matter in general simply seems to be treated rather anonymously. References to the inventors of methods appear to be less common in mathematics and mathematical statistics texts, and somewhat more common in the more applied statistics texts. Indeed, to the extent that a count of references or sources listed serves to assess the level of attention to the history of the methods, Fisher’s *Statistical Methods for Research Workers* (14th Edition) does as least as good a job as do most of my more modern texts. Given the fact that Fisher was writing while so much less literature was available, this admittedly imprecise assessment would seem to contradict Cowles'' interpretation.

Most of the problems I have with Cowles'' treatment of statistics in psychology concern what he did not write rather than what he did. Attempting to cover the history of statistics in psychology in 235 pages is more than a tall order. A number of discussions leave out important aspects of the issues, and fail to take the opportunity to clarify areas where a great deal of misunderstanding exists among psychologists and other social scientists.

An example is the section in Chapter 13 entitled ''A Note on ''One-Tail'' and ''Two-Tail'' Tests.'' Cowles could have done his readers an important service by making explicit the fact that the issue at stake is really not the number of tails but whether the alternative hypothesis under consideration is directional or nondirectional. Instead, he makes the following statement: ''Put baldly and simply, using a one-tail test means that the researcher is using only half the probability distribution, and it is inconceivable that this procedure would have been acceptable to any of the founding fathers'' (p. 204). In making this claim he fails to note that standard applications of both F- and χ^2-tests, while concerned with nondirectional alternative hypotheses, employ only one tail of the chosen probability distribution. Furthermore, the claim that use of a directional hypothesis would never have been acceptable to any of the founding fathers is plainly contradicted by the example on pages 96–97 of *Statistical Methods for Research Workers*, in which Fisher illustrates the use of his exact test for 2×2 tables and produces a directional and one-tailed probability value. (If memory serves me correctly, the tea-tasting example Fisher originally used to illustrate the exact test also involved use of a directional alternative hypothesis).

Another place where I have to take some issue with Cowles'' interpretations concerns the claim made in the discussion of multiple comparison methods in Chapter 13 that ''the automatic invoking, from the statistical packages, of any one of half a dozen procedures following an F-test has helped to promote the emphasis on the comparison of treatment means in psychological research'' (p. 196). I understand that the availability of multiple comparison methods in statistical software does make their use more common. I also understand that some people will be led to conclude that something is valid if it is included in software. Nonetheless, the fact is that multiple comparison procedures were invented and popularized long before the existence of statistical software packages. I also know from firsthand experience that social scientists have been taught to use such methods regardless of their implementation in software packages. I have spoken with numerous SPSS users who have requested them in contexts in which they are not available. These have actually been among the most common types of enhancement requests I have received.

One concern of Cowles'' that my experience supporting widely used statistical software allows me to validate is the widespread failure to understand the essential unity of the linear model. Unfortunately, Cowles'' own brief treatment of the use of expected mean squares and mixed models suffers from an apparently narrow understanding of the topic. Moreover, he fails to mention the fact that for over a decade the treatment of mixed models has received important methodological contributions from psychologists and other social scientists, generally under different names such as hierarchical or multilevel models.

This is an example of my biggest problem with the book: it is seriously incomplete in its coverage of the topic of statistics in psychology, in large part due to a failure to discuss the many areas in which psychologists have made important contributions to statistical methodology. Some of this appears to be due to a failure to really present a history of the topic that brings things up to the present, as it is purported to do (perhaps this is the work of editors or publicists rather than the author, the quote being taken from the book cover rather than the book text itself). For example, in addition to not covering the modern treatment of mixed/hierarchical/multilevel models, the treatment of factor analysis as a methodology in essence stops at least 15 years too soon in its failure to cover developments in confirmatory factor analysis and structural equation modeling. The quote earlier about the failure to recognize the difference between exploratory and confirmatory approaches to factor analytic methodology left me wondering whether the author was familiar with what has been perhaps the most widely attended to area in psychological statistics over the last decade and a half, namely structural equation modeling (SEM). The distinction between exploration and theory confirmation is commonly discussed in treatments of SEM, and psychologists would have to have been in virtual hiding to not have been repeatedly exposed to this.

Another example of a modern methodology that is old enough to be of historical interest and that has received important contributions from psychologists is meta-analysis. Meta-analysis is widely used in a variety of fields, particularly medicine and biostatistics, and the work of psychological pioneers in the area has been highly influential. Power analysis, an area where psychologists have been among the leading authorities, and which has become a standard part of research protocols in a variety of fields, receives little attention. Similarly, modern item response theory methods have been pioneered by researchers in psychology and education. Prior to that, important work in classical item analysis, reliability theory, and generalizability theory was produced by psychologists, sociologists, and educational theorists. The study of the reliability of measurements and ratings has been a major area of research and theoretical work among psychological statisticians for over half a century, but it receives almost no coverage in this book. The more modern generalizability theory and item response theory areas (which now have histories spanning several decades) are not even mentioned. Another area that has a history spanning several decades and that has seen major contributions from psychologists is that of multidimensional scaling (not to mention unidimensional deterministic scaling, which dates back even further). Even nonparametric statistics, which are commonly taught in psychological statistics courses, and which received important contributions from psychologists dating back a half century or so, are not discussed.

Perhaps the presentation of the book as an overview, which implies being reasonably comprehensive in coverage, is the work of an editor or a publicist rather than the author, but even if the intention is not to provide a reasonably comprehensive summary of the interaction between statistics and psychology, the failure to include discussions of just about any of the topics where statistics has been the recipient of major work done by psychologists makes this, in my opinion, a seriously flawed book, and will, I think, likely prevent it from having the desired effect on students of psychology. In my experience, many psychologists and psychology students do indeed see statistics as an evil, and one that is only necessary because authorities tell them that it is, rather than because they see its relevance to solution of problems in their chosen particular areas of study. Many psychologists have a great deal of trouble handling mathematical subject matter, and there is a very strong tendency to deny its relevance to what they wish to study. An exploration of some of the areas in which the problems of psychology and related subject matter fields have motivated developments in statistical theory and practice would seem to me to be a valuable way to demonstrate this relevance, but it is sorely lacking in this book.

The main theme of the book seems to be that the pioneers of modern statistical methods were often not rational and scientific in their reasoning and argumentation. The author seems to feel that psychology students who are intimidated by mathematics will be better disposed to critically evaluate statistical methods once they understand the fragility of figures such as Pearson and Fisher. I fear that the effect of the book is more likely to be to serve as an excuse for some of these students to dismiss the importance of statistical methods, rather than to study them more carefully and more critically. Furthermore, I fear that the truly important fact that the value of statistical methodology in psychological research does not depend on how rational or scientific were statistical pioneers, which is clearly stated by Cowles in the quote above from page 29, will be lost on most readers amongst the plethora of examples of unfortunate behavior by historical figures.

The author states in the preface to this 2nd edition that ''It is still my firm belief that a little more mathematical sophistication and just a little more historical knowledge would do a great deal for the way we carry on our research business in psychology.'' Though I share the general desires voiced here, I believe that a good deal more mathematical sophistication is required on the part of most psychological researchers, and I fear that this incomplete treatment of the history of statistics in psychology will not have the desired effect on many of its readers. While psychologists with substantial knowledge in statistical methodology will find a good deal of interesting information in the book, I cannot recommend its use as a sole or primary source of study of the history of statistics in psychology for students who are unable to put this information in perspective.

David P. Nichols
Principal Support Statistician and Manager of Statistical Support
SPSS Inc.
233 S. Wacker Drive
Suite 1100
Chicago, IL 60606
U.S.A.
Telephone: 312-651-3412
E-mail: nichols@spss.com

Rand R. Wilcox. *Fundamentals of Modern Statistical Methods*. New York: Springer-Verlag, 2001. ISBN 0-387-95157-1. 258 pp+refs and index, 49.95 USD.

Many conventional methods of statistical inference are based on unbiased estimates of the first and second moments of distributions. For example, standard confidence intervals and hypothesis tests use sample means and variances, and regression analyses use sample covariances. Inferences from such conventional methods are valid when the parametric assumptions underlying them hold true. When these assumptions are violated, in particular the normality assumption, conventional inferences may be inaccurate.

To avoid the potential breakdowns of conventional methods, one approach is to use robust methods, for example medians, trimmed means, and M-estimators. Although software for implementing these methods exists, robust methods are not widely used by applied researchers. Wilcox offers one explanation for this: applied researchers do not know enough about robust methods to feel comfortable using them. By writing this text, Wilcox seeks to fill this gap by introducing readers to robust methods.

The text is written in two parts. The first part discusses the inaccuracies of inferences based on conventional methods when the parametric assumptions underlying them are violated. The second part explains how robust methods can improve inferences when standard parametric assumptions are violated. The text focuses on conceptual explanations, examples, and simulation studies as opposed to mathematical proofs. The language is very clear, and the ideas are accessible to a wide audience. The text is pitched at applied researchers who understand conventional inference for means, correlation analysis, and regression analysis. It does not have any exercises. It does not consider robust inference in complex sampling designs.

A detailed summary of the chapters of the book follows. Chapter 1 contains an informative history of the development of the normal distribution. Chapter 2 reviews the definitions of weighted means, variances, medians, and regression lines. This chapter introduces a main point of the text: sample means, variances, and regression coefficients can be sensitive to outliers.

Chapter 3 reviews the normal distribution and the central limit theorem. In this chapter, Wilcox argues that it is difficult to determine when sample sizes are large enough for the central limit theorem to kick in. He uses simulation studies to show that light-tailed distributions typically require larger sample sizes. This chapter also discusses outlier detection methods. Wilcox shows that calling observations outliers when they are more than two estimated standard deviations from the sample mean can fail to identify outliers when data are not normally distributed.

Chapter 4 shows how inferences based on conventional, unbiased estimates of means and regression coefficients can be inaccurate when variances are not constant. The chapter focuses mainly on the negative effect of non-constant variance on confidence interval coverage.

Chapter 5 begins with a review of hypothesis testing, in particular the *t*-test for single means and differences in two means. Using simulation studies, Wilcox shows the consequences regarding Type I and Type II error rates of violations of the normality assumption.

Chapter 6 introduces methods of inference based on the bootstrap, such as the percentile bootstrap and percentile *t* bootstrap. Wilcox argues that there are situations for which bootstrap methods are more effective than conventional methods. He outlines how to apply the bootstrap for inferences about means, correlations, and regression coefficients.

Chapter 7, the last chapter in Part 1, presents what Wilcox calls the fundamental problem with conventional methods. When normality is assumed but not true, the population variance implied by the assumed normal distribution may be so large that it is impossible to achieve sufficient power with conventional methods. Addressing this problem is a main focus of Wilcox's discussion of robust methods in Part 2.

Chapter 8 introduces trimmed means and M-estimators. Wilcox's discussions are conceptual in nature; readers who seek the mathematical derivations of these methods can refer to the references at the end of the chapter.

Chapter 9 focuses mainly on how inferences can be obtained from trimmed means by using Winsorized variances. For M-estimators, Wilcox suggests a bootstrap approach to inference.

Chapter 10 discusses robust measures of association, including Winsorized correlations, Kendall's tau, and Spearman's rho. As Wilcox points out, these measures do not effectively deal with points that do not fit the overall trend of the data. In this chapter, Wilcox also discusses the use of smoothers.

Chapter 11 presents robust regression, including the Thiel-Sen estimator, M-estimators, L_1-estimators, and various regression methods based on trimming observations. The explanations focus on regression with a single predictor. Wilcox ultimately does not recommend any one particular method; instead, he recommends examining multiple methods.

Finally, Chapter 12 briefly describes nonparametric methods, including rank tests, permutation tests, and nonparametric regression.

The robust methods described by Wilcox are of primary interest to researchers who work with small sample sizes, for example in randomized experiments. In official statistics, there is less need for these methods. In design-based inference for population means, parametric assumptions typically are not made. And, since most national surveys have sufficiently large sample sizes, central limit theorems often apply. There are settings in which sample sizes are small and central limit theorems do not apply, for example describing characteristics of sub-populations. In these settings, I believe parametric methods that borrow strength across small areas, like hierarchical Bayesian models, are likely to provide bigger gains in accuracy than separate robust methods.

There are two main themes raised by Wilcox that deserve close attention from official statisticians. First, as Wilcox emphasizes, outliers can greatly affect estimates of means and variances. This is also true in survey sampling, particularly when the survey weights of outliers are large. Second, inferences can be inaccurate when assumptions are violated. It is therefore crucial in every data set to check methods'' assumptions before applying them, such as by examining graphical displays of data. For official statisticians, these graphical explorations are also relevant, even when using design-based estimation. As examples, identifications of differences across sub-populations can lead to improved post-stratification, and identifications of strong linear relationships can lead to more efficient regression estimators.

To summarize, Wilcox's book is a very clear introduction to robust methods. For those official statisticians seeking to expand their general knowledge of statistics, it is an informative read.

Jerome Reiter
Institute of Statistics and Decision Sciences
Duke University
Durham, NC 27708-0251
U.S.A.
Telephone: 919-684-4210
Email: jreiter@stat.duke.edu