Reprinted from JOURNAL OF MATHEMATICAL PSYCHOLOGY
All Rigts Reserved by Academic Press, New York and London

Vol. 32, No. 2, June 1988
Printed in Belgium

Book Review

Statistics: Art or Craft?

WILLIAM A. GALE (Ed.). **Artificial Intelligence and Statistics.** Reading, MA: Addison–Wesley, 1986. pp. xiv + 418. $39.95.

Reviewed by YOSHIO TAKANE

The author, William A. Gale, is a member of the Technical Staff at AT & T Bell Laboratories. He obtained his Ph.D. in Physics at Rice University in 1969, and his current primary research interest is the application of artificial intelligence in statistics. He has published *Life in the Universe*, Westview Press, Boulder, 1979, and *Inflation*, Oelgeschlager Gunnhaim, Cambridge, 1981.

The reviewer, Y. Takane, is a Professor of Psychology at McGill University. He was President of the Psychometric Society in 1986–87, and some of his research interests concern stimulus recognition models, multidimensional scaling, and statistical expert systems.

> By 1995 or so, the largest single driving force in guiding general work on data analysis and statistics is going to be the needs that have to be met to understand and improve data-analytic expert systems, ...

The above statement was excerpted from the last chapter (Chapter 18) of *Artificial Intelligence and Statistics.* It is rather compelling, because it was made by John Tukey, a modern day "prophet" in statistics, to whom we owe much of the current proliferation of research in robust techniques, exploratory data analysis, and interactive computer graphics for statistical data analysis (SDA). I like the statement. It is a bit exaggerated, but I interpret it as his usual first-rate rhetoric to emphasize the importance of nonmathematical aspects of SDA. I like it, because I share his strong conviction that more work needs to be done on strategic aspects of SDA and their implementation in the form of statistical expert systems (Takane, 1986).

A statistical expert system is a computer implementation of statistical expertise in such a form that "it can provide intelligent statistical advice, make intelligent decisions, and explain its lines of reasoning for such decisions." Gale and Pregibon (1984; Pregibon & Gale, 1984) have already taken a big step toward this goal. They have developed a computer program for simple regression analysis, called REX (regression expert), which checks assumptions underlying regression analysis, suggests possible corrective actions, carries out necessary computations, and writes

Requests for reprints should be sent to Dr. Yoshio Takane, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, PQ, H3A 1B1, Canada.

0022-2496/88 $3.00

a summary report on major results of the analysis. Gale (Chapter 10 of *Artificial Intelligence and Statistics*) is now working on a meta expert system, called STUDENT, which learns strategies for SDA through examples worked out by human expert statisticians, and builds an expert system (like REX for regression analysis) in other analysis domains. This latter attempt is, of course, much more challenging than merely developing domain specific statistical expert systems.

Both REX and STUDENT extensively use artificial intelligence (AI) techniques. One active area of AI research concerns acquisition and representation of knowledge in a computer in such a way that it can be efficiently retrieved, modified, and manipulated to arrive at some form of conclusion. New concepts and programming styles have been invented to suit this purpose. It was quite natural then that Gale chose to utilize the AI methodology in representing and manipulating strategic knowledge in statistics. This was the point of initial contact between AI and statistics, which later culminated in the workshop on AI and statistics at AT & T Bell Labs and the book entitled *Artificial Intelligence and Statistics* based on the workshop.

For those who, like myself, missed the workshop, while having been interested in the topic, *Artificial Intelligence and Statistics* is indeed a welcome volume. In this volume Gale reviews what has been done, what is being done, and what is going to be done in the future in the development of statistical expert systems, and demonstrates how AI concepts are useful (Chapters 8 and 10). Pregibon presents a concise, yet vivid, guide to statistical strategy in the light of his experience in building REX (Chapter 17). The book also contains chapters presenting other people's views and approaches to statistical strategy and statistical expert systems. Among them I found Chapter 9 by Ellman, Chapter 15 by Oldford and Peters, and Chapter 16 by Hand particularly insightful. Ellman discusses a way of representing computational knowledge particularly suitable for explanation. Oldford and Peters present an "operational level" view of data analysis, using the collinearity detection problem as an example. Hand, on the other hand, views data analysis as consisting of several stages (1. formulation of aims, 2. translation into formal terms, 3. numerical processing, and 4. interpretation). Actual data analysis takes place by cycling through these stages several times to reach a final conclusion.

As indicated in Chapter 1 of the book, Gale has considerably widened his initial contact between AI and statistics. He also includes aspects of interaction between these two areas other than the statistical expert systems. Again I found three chapters of this nature particularly entertaining. Chapter 2 by Spiegelhalter addresses the problem of uncertainty propagation in expert systems. He argues quite convincingly that nothing other than probability (e.g., belief function, fuzzy set) is really necessary, and advocates the subjective Bayesian approach to deal with the problem. (There is some follow-up of this in DeGroot, 1987.) In Chapter 4 Fisher and Langley present a survey of AI approaches to cluster analysis, called conceptual clustering, which is radically distinct from the numerical taxonomy that we are so accustomed to. In Chapter 13 Butler and Corter discuss possible applications of psychometric methods to knowledge acquisition. As a psychometrician I was par-

ticularly intrigued by the novel use that they are proposing of the methods with which I am already familiar.

As is usual with a collection of chapters written by different authors, some portions of the book are extremely readable, while others are not. In general I found it more difficult to understand the chapters written by AI people (with the exception of Chapters 2, 4, and 13). This, however, may merely reflect my state of knowledge in AI. On the other hand, it may be more deeply rooted to a general difficulty in communicating AI products. The non-algorithmic nature of the AI programs makes it difficult to understand what they really do. It also makes evaluation of the programs extremely difficult.

The book is the first of its kind. (See also Haux, 1986, which is, as far as I know, the second of its kind.) The book demonstrates that although the useful interplay between AI and statistics has just begun, it is already producing interesting results. However, many challenging tasks (e.g., standardization of statistical knowledge, standardization of statistical consulting processes) remain to be done in order to go beyond REX and STUDENT, i.e., to build more sophisticated statistical expert systems fully equipped with the sort of capabilities described by Hand (1984, 1985). A contributor to the book (Huber in Chapter 12) thinks that the task is formidable if not impossible. I remain curious, not as a bystander, but as a committed researcher in this area, as to what follow the lead of *Artificial Intelligence and Statistics* in the next ten years. The only way to prove that statistics is not an art but a craft is to demonstrate the feasibility of the statistical expert systems.

References

De Groot, M. H. (Ed.). (1987). *Statistical Science*, **2** (1), 3–44.

Gale, W. A. & Pregibon, D. (1984). Constructing an expert system for data analysis by working examples. *Compstat 1984* (Proceedings in computational statistics) (pp. 227–236). Wien: Physica-Verlag.

Hand, D. J. (1984). Statistical expert systems: Design. *The Statistician*, **33**, 351–369.

Hand, D. J. (1985), Statistical expert systems: Necessary attributes. *Journal of Applied Statistics*, **12**, 19–27.

Haux, R. (Ed.). (1986). *Expert systems in statistics*. Stuttgart: Gustav Fischer Verlag.

Pregibon, D., & Gale, W. A. (1984). REX: An expert system for regression analysis. *Compstat 1984* (Proceedings in computational statistics) (pp. 242–248). Wien: Physica-Verlag.

Takane, (1986). Development of cognitive statistics and statistical expert systems. Invited Address. In *Proceedings of the 14th annual meeting of the Behaviormetric Society* (pp. 47–48); *Japanese Journal of Behaviormetrics*, **14**, 38–40.

Printed by Catherine Press, Ltd., Tempelhof 41, B-8000 Brugge, Belgium