

# | Data Science and Its Applications

-La science des données et ses applications-

Edited by

Yves Escoufier

Bernard Fichet

Edwin Diday

Ludovic Lebart

Chikio Hayashi

Noboru Ohsumi

Yasumasa Baba



ACADEMIC PRESS, INC.

*Harcourt Brace & Company*

Tokyo San Diego New York Berkeley

Boston London Sydney Toronto

## ACADEMIC PRESS RAPID MANUSCRIPT REPRODUCTION

COPYRIGHT © 1995 by Harcourt Brace Japan

ALL RIGHTS RESERVED.

NO PART OF THIS PUBLICATION MAY BE REPRODUCED OR  
TRANSMITTED IN ANY FORM OR BY ANY MEANS, ELECTRONIC  
OR MECHANICAL, INCLUDING PHOTOCOPY, RECORDING, OR ANY  
INFORMATION STORAGE AND RETRIEVAL SYSTEM, WITHOUT  
PERMISSION IN WRITING FROM THE PUBLISHER.

ACADEMIC PRESS/HARCOURT BRACE JAPAN, INC.

Ichibancho Central Bldg., 22-1 Ichibancho, Chiyoda-ku, Tokyo 102

ISBN : 0-12-241770-4

Printed in Japan

95 96 97 98    9 8 7 6 5 4 3 2 1

# Preface

In March 1987 the first Japanese-French Scientific Seminar was held in Tokyo. The proceedings were published in book form by Academic Press, Inc. (1988) under the title “Recent Developments in Clustering and Data Analysis,” edited by C. Hayashi, E. Diday, M. Jambu, and N. Ohsumi. The seminar resulted from an ongoing exchange between French and Japanese specialists in what is known as data analysis, *analyse des données* in French, and *deta kaiseki* in Japanese. One might wonder at the common interest of groups so greatly separated by distance.

What follows is an interesting subject for anyone curious about the evolution of scientific thought. Since the early sixties, Japanese and French researchers traced parallel paths, without mutual consultation, aware of the widening gulf between mathematical statistics and the practical use of computers that were becoming more and more commonplace. Both communities were dismayed by the existing methods of mathematical statistics - far too rigid a framework for the realities of modern experimental data collecting. Indeed either the hypothesis of mathematical statistics was fully disengaged, or contemporary data analysis presented a poor reflection of its prototype. Whichever, the conscientious researcher was left dissatisfied: Either he had applied a wrongly adapted method for the data available, or had sacrificed his work by using data that failed to adequately express the original substance.

Meanwhile, the same individuals realized the vast potential of rapidly expanding computer technology as a means to process enormous amounts of data. Experimentation produced good results with simple mathematical manipulations of large groups (organizational layout, graphic display, computing characteristics). It showed the need to develop theoretical methods embracing new advances in computer processing. Not only the French and Japanese arrived at this conclusion; J. Tukey’s decidedly is the best example of a similar approach in the U. S. Interestingly, both French and the Japanese had something in common: Being minority scientific communities, they sought to express their individuality in spite of the American giant. Beyond that, their overall outlook was so similar that, minus prior consultation, both C. Hayashi of Japan and J.-P. Benzécri of France, the two generally regarded

as the most exemplary researchers of their respective countries, placed *qualitative data analysis* at the center of their work. Converging thoughts led to mutual interest, fed by international meetings and exchange programs.

The first Japanese-French Scientific Seminar in Tokyo, held during 1987, finally made this common viewpoint official, thus allowing for an increase number of researches plus an enhanced reciprocal flow of ideas. Should the parties concerned gather again to assess the latest developments? All who took part in the Tokyo conference favored another meeting. Accordingly it convened during 1992 in Montpellier and the outcome of that highly successful seminar -this book- clearly showed that the decision to assemble was most worthy.

Since 1987 computer science has steadily provided scientists with increasingly powerful means to automatically collect and store data. Organizational concepts have been developed that allow for prompt retrieval of stored data either partly or in mass. Static or dynamic graphics, with or without color, are efficient and esthetic. Meanwhile, computing power has been enhanced tenfold. New computer potential multiplied the number of possible tools for *data analysis*. Unless clearly based on a consistent set of concepts, newly developed tools cannot experience their full potential in data processing and interpretation. All of this by no means waives overall reflection on the course these general concepts will take, but rather demands it. Bearing this in mind is how the present volume should be used. The authors propose ways to formalize steps in *data analysis*. They have studied their properties, drawing from them methods of analysis. They then propose software with which to implement the methods.

Such an approach gives birth to a new science with data at its core. Its nature, numerical, qualitative or symbolic, determines the type of operations possible with them. Their origin, whether exhaustive collection or sample, conditions the objective expected in their analysis. It seems justified to coin the term *data science* for this particular activity. Of course this construction must rest on solid ground. Several texts presented herewith would find a place in publications on multidimensional statistical analysis, exploratory data analysis, or clustering. Bringing these texts together in this book expresses the common conviction of its authors: To take data as a starting point provides a complementary vision of theory and practice, and avoids creating an unfortunate gap between these two steps both of which essential in any scientific process.

The Editors