

Behaviormetrics:
Quantitative Approaches to Human Behavior 8

Shizuhiko Nishisato
Eric J. Beh
Rosaria Lombardo
Jose G. Clavel

Modern Quantification Theory

Joint Graphical Display, Biplots, and
Alternatives

 Springer

Behaviormetrics: Quantitative Approaches to Human Behavior

Volume 8

Series Editor

Akinori Okada, Professor Emeritus, Rikkyo University, Tokyo, Japan

This series covers in their entirety the elements of behaviormetrics, a term that encompasses all quantitative approaches of research to disclose and understand human behavior in the broadest sense. The term includes the concept, theory, model, algorithm, method, and application of quantitative approaches from theoretical or conceptual studies to empirical or practical application studies to comprehend human behavior. The Behaviormetrics series deals with a wide range of topics of data analysis and of developing new models, algorithms, and methods to analyze these data.

The characteristics featured in the series have four aspects. The first is the variety of the methods utilized in data analysis and a newly developed method that includes not only standard or general statistical methods or psychometric methods traditionally used in data analysis, but also includes cluster analysis, multidimensional scaling, machine learning, corresponding analysis, biplot, network analysis and graph theory, conjoint measurement, biclustering, visualization, and data and web mining. The second aspect is the variety of types of data including ranking, categorical, preference, functional, angle, contextual, nominal, multi-mode multi-way, contextual, continuous, discrete, high-dimensional, and sparse data. The third comprises the varied procedures by which the data are collected: by survey, experiment, sensor devices, and purchase records, and other means. The fourth aspect of the Behaviormetrics series is the diversity of fields from which the data are derived, including marketing and consumer behavior, sociology, psychology, education, archaeology, medicine, economics, political and policy science, cognitive science, public administration, pharmacy, engineering, urban planning, agriculture and forestry science, and brain science.

In essence, the purpose of this series is to describe the new horizons opening up in behaviormetrics — approaches to understanding and disclosing human behaviors both in the analyses of diverse data by a wide range of methods and in the development of new methods to analyze these data.

Editor in Chief

Akinori Okada (Rikkyo University)

Managing Editors

Daniel Baier (University of Bayreuth)

Giuseppe Bove (Roma Tre University)

Takahiro Hoshino (Keio University)

More information about this series at <http://www.springer.com/series/16001>

Shizuhiko Nishisato · Eric J. Beh ·
Rosaria Lombardo · Jose G. Clavel

Modern Quantification Theory

Joint Graphical Display, Biplots,
and Alternatives

Shizuhiko Nishisato
University of Toronto
Toronto, ON, Canada

Rosaria Lombardo
Department of Economics
University of Campania Luigi Vanvitelli
Capua, Caserta, Italy

Eric J. Beh
School of Mathematical
and Physical Science
University of Newcastle
Newcastle, NSW, Australia

Jose G. Clavel
Department of Quantitative Methods
Universidad de Murcia
Murcia, Spain

ISSN 2524-4027 ISSN 2524-4035 (electronic)
Behaviormetrics: Quantitative Approaches to Human Behavior
ISBN 978-981-16-2469-8 ISBN 978-981-16-2470-4 (eBook)
<https://doi.org/10.1007/978-981-16-2470-4>

© Springer Nature Singapore Pte Ltd. 2021

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Foreword

Quantification theory has a long history and its popularity has reached every corner of the world during the past 100 years. It has drawn us into the fascinating world of data analysis and as such has played a key role as an attractive research tool for diverse areas of scientific disciplines.

During the first half of the last century, the foundation of quantification theory was firmly established by such researchers as Pearson, Gleason, Ramensky, Richardson, Kuder, Hirschfeld, Fisher, Guttman, and Maung; in the latter half of the last century, it flourished into routine methodology through concerted efforts by groups of researchers led by Hayashi (Japan), Benzécri (France), de Leeuw (the Netherland), Nishisato (Canada) and the trio of Young, de Leeuw, and Takane. There were of course a countless number of other outstanding individual researchers as well.

I am greatly honoured to write the foreword for this book. As a long-time researcher at the Institute of Statistical Mathematics (ISM), I am privileged to have worked with the late Dr. Chikio Hayashi who is well known for his theory of quantification and many other researchers at ISM. During my career, I spent several months as a visiting scholar at the University of Toronto with Prof. Nishisato and in return, I hosted him as a foreign visiting professor at ISM. When I heard about his dual scaling, what came to my mind was: when we expand real space to complex space, thus enriching our field of mathematical exploration, dual scaling must have the same effect of expanding our scope through re-directing our attention from simple space to dual space.

I met Prof. José Garcia Clavel at a conference and learned that he too had spent his sabbatical year in Toronto with Prof. Nishisato. I finally met Prof. Eric J. Beh and Prof. Rosaria Lombardo at the 2017 IFCS conference in Tokyo, whose fame I had known through their highly acclaimed book on correspondence analysis from Wiley.

By then, I was supporting Nishisato for his battle over the joint graphical display and learned that the three co-authors of the current book were also behind him. Nishisato reminisces in Chap. 1 the memorable CGS scaling debate at the 1989

IFCS meeting in Charlottesville, Virginia. I feel proud to say that I was there, and it was the first time I met him.

The current book offers an interesting mixture of two groups of researchers, Nishisato-Clavel and Beh-Lombardo. I heard that Nishisato had met Clavel at a conference in Barcelona where he was writing his Ph.D. thesis under the supervision of Michael J. Greenacre. As my generation knows, Nishisato is one of the last persons from the old school of traditional psychometrics, while Clavel is a contemporary researcher, gifted with modern technology. These two researchers with very different backgrounds have been working in unison for the last 20 years or so.

Beh and Lombardo are statisticians by training who have been exceptionally productive over the past two decades. It is my belief that they will play the role of the leaders in the development of quantification theory to the next level of advancement.

The four authors from Canada, Australia, Italy, and Spain have combined their unique talents in writing this book. The book itself is a collection of essays and technical writings on quantification theory. As a seasoned researcher, Nishisato offers his personal reminiscence over the history of quantification theory, where we see rare personal observations of the past researchers and their work. He then directs his focus on the joint graphical display of quantification results, a topic that has been ignored for decades; see his meticulous effort to solve the problem. Clavel and Nishisato present cluster analysis (hierarchical, partitioning, bi-clustering) as an alternative to the joint graphical display, where their main task is to explore groupings of row and column variables in dual space. Then, Beh and Lombardo present biplots as yet another alternative to the joint graphical display and expand their expert writings of other important topics of quantification theory.

This book represents a unique collaboration of two groups of researchers with different backgrounds, diverse viewpoints, and superb presentations. Their collaboration is very successful and the book demonstrates their own unique talents. As a whole, this is a very informative and uniquely helpful book as a technical guide for modern quantification theory. I would strongly recommend this book to many researchers in diverse disciplines.

Tokyo, Japan
October 2020

Yasumasa Baba

Preface

This book is a product of the collaborations by researchers from four countries with different backgrounds. The first contact was made when Beh and Lombardo published a highly innovative book, entitled *Correspondence Analysis: Theory, Practice and New Strategy* (Wiley, 2014), and Nishisato reviewed it (*Psychometrika*, 2016). With this background, Beh, Lombardo, Clavel (Nishisato's collaborator for some 25 years), and Nishisato proposed a session and presented papers at the IFCS (International Federation of Classification Societies) meeting in Tokyo in 2017. By then, through a number of correspondence, our close friendship was firmly forged.

The idea of writing a book together emerged through unfortunate and fortunate events. Beh was a reviewer of Nishisato's paper on joint graphical display which was, according to Beh, "controversial" but warranted a broader discussion because the issue had been largely ignored for decades. He recommended the paper for publication since he did not see strictly speaking anything wrong with it, the paper was well written and well argued. In spite of Beh's strongly positive review, the paper was unconditionally rejected as "fundamentally wrong". In the meantime, Nishisato successfully tested his solution to the long plagued problem of joint graphical display in quantification theory, the topic of his paper, at a conference. What happened then was sheer luck; although the review process of the aforementioned journal was strictly double blind, Beh could tell the identity of the author from the writing style and contacted Nishisato with strong encouragement. From these unfortunate and fortunate events, an idea of writing a book emerged, and two pairs of collaborators (Lombardo and Beh; Clavel and Nishisato) finally reached a decision to put our different ideas together into a book. After the 2017 IFCS meeting in Tokyo, we received encouragement from Akinori Okada (Series editor), and we finalized our decision to publish a book as a joint work of the Nishisato-Clavel team and the Beh-Lombardo team.

So, this is a product of our forged friendship and the book is by no means a unified product, for Beh/Lombardo and Nishisato/Clavel represent two different schools of thought. Beh and Lombardo are frontier researchers in statistics and their work is highly technical, while Nishisato and Clavel are more practice-oriented. In

spite of our different backgrounds, we have come together to highlight the pros and cons of different ways of thinking about the same problem.

We do not describe anything that is strictly new, but rather discuss various issues in essay and technical form from both sides of the fence. Due to the different flavours of the two partnerships, you will see distinct differences in how we have described the topics, while using the same notation.

Part I consists of six chapters. Chapters 1–5 are based on Nishisato’s reminiscence on his endeavour over half a century of research career with a particular perennial problem of joint graphical display. Clavel and Nishisato will discuss cluster analysis as an alternative to joint graphical display in Chap. 6.

Part II consists of five chapters edited by Beh and Lombardo. Chapter 7 provides a brief outline of the inner workings of reciprocal averaging and its role in correspondence analysis, while some previously unseen variants of reciprocal averaging are proposed in Chap. 8. Chapter 9 provides a brief historical introduction to biplots. Further discussions on biplots are presented in Chap. 10, although its focus is based on ordered categorical variables and multi-way data quantification. Finally, Chap. 11 explores some new ideas to deal with over-dispersed categorical data and its visualization.

Quantification theory as known by many aliases will continue to evolve and will capture the hearts of many researchers. This is a book written with the collaboration of four international researchers with different backgrounds and viewpoints. We hope that you will find this book as a useful addition to your bookshelf.

Toronto, Canada
Newcastle, Australia
Capua, Italy
Murcia, Spain
November 2020

Shizuhiko Nishisato
Eric J. Beh
Rosaria Lombardo
Jose G. Clavel

Acknowledgements

First, we would like to thank Prof. Akinori Okada, the editor of the Springer *Behaviormetrics* series, for accepting our proposal for the current book and constantly encouraging us, and Mr. Yutaka Hirachi of Springer Japan for his guidance. Our special thanks go to Prof. Yasumasa Baba for his kind Foreword, which makes all of us feel uplifted to the first-class leading researchers! We now would like to take this opportunity to express our appreciation to those who always offered us their helping hands with understanding, watchful eyes, and affections.

Jose G. Clavel

There are many combinations of fantastic events that have made my contributions to this book possible. Some of them I know, but others have happened without my awareness. I would mention only three of them.

It is evident that I owe greatly to the co-directors of my Ph.D. degree: Dr. Joaquín Aranda (Universidad de Murcia, Spain) and Dr. Michael J. Greenacre (Universitat Pompeu Fabra, Spain). Dr. Aranda, my mentor, was the Head of the Department of Quantitative Methods when I began my teaching career at Universidad de Murcia. In my view, he always had confidence in my possibilities and has offered me his clear guidance not only for my teaching but also for my life. The second mentor, Michael J. Greenacre, is a renaissance man of our times, who opened my mind to what a university professor should be and he generously shared not only his knowledge with me, but also his friends at Sant Fruitosós de Bages and researchers all over the world.

The second important event was, of course, a casual meeting with Nishi in the hall of Universitat Pompeu Fabra, Barcelona, in July 1993. We were then attending the European Meeting of the Psychometric Society. I cannot remember why, but I was alone for lunch and decided to invite him to a truly Spanish meal—he clearly looked like a Japanese professor being alone. We had lunch together, but I did not know who he was. We continued our lunch together during the conference. Only

later a friend of mine told me that he was Nishisato of dual scaling. This accidental meeting was the beginning of our cooperative work and that simple Spanish luncheon was the event which made a decisive influence on my life; it brought me later to Toronto, and more specifically, to Saint Georges Road, and then Old Mill Road, where I found a home away from home; thanks Lorraine.

Those matters of enormous importance, however, are nothing in comparison with the magnitude of compassion involved in the next recipients of my heart-felt appreciation: my dearest parents who, through their love, joy, and generosity, have given me the best family in the world; Dodo, Manolo, Mariuca, Beatriz, Elena, Joaquín, Javier, and Ciuca, my siblings, with whom I share this unimaginably great luck and fortune.

Rosaria Lombardo

I would like to thank my sisters Giovanna and Savina and my brother Federico who, differently from me, are *artists* and often inspired me to do my work *creatively*, especially when visualizing data. I would also like to thank Eric for having involved me in this project, but especially for his genuine friendship, and Nishi and José for their tenacity, patience, and understanding over the last years.

Eric J. Beh

Over the past few years, I've had the pleasure of working with some wonderful people. The last few years have also been extremely difficult for me, with various health issues getting in the way. So I'd like to acknowledge the collegiality and friendship I've received from Pieter Kroonenberg who I deem to be a legitimate *quasi-Aussie*; thanks mate. Of course, I am extremely grateful for the patience and understanding I have received from Nishi and José over the years and, especially, Rosaria, whose friendship I've always valued.

There are others that I would also like to acknowledge who have helped me greatly over a much longer period of time and probably haven't been told often enough. My wife, Rosey, and son, Alex, know how much they mean to me but I'd like to give a special shout out to my Mum (Donella), my sister (Emma), her wife (Suzi), and their son and my (favourite, and only) nephew Oli. I love you all and thank you for everything.

Shizuhiko Nishisato

Since my first publication in 1960, it is already 60 years. I am indebted to a countless number of people, in particular, my Japanese mentors Masanao Toda and Yoshio Sugiyama at Hokkaido University, my American mentors R. Darrell Bock and Lyle V. Jones at the University of North Carolina, and my Canadian colleague Ross E. Traub at the University of Toronto. I was lucky to have had international contacts from the early days, such as Gaul, Bock, Mucha, Ihm, and I. Böckenholt in Germany; Lebart, Morineau, Tennenhaus, Saporta, and Le Roux in France; Mirkin and Adamov in USSR; Gower and Goldstein in Britain; de Leeuw, Heiser, Meulman, van der Heijden, and Kroonenberg in the Netherlands; Lauro and Coppi in Italy; and too many names to list here in the USA, Japan, and Canada. I owe greatly to all members of the Psychometric Society since 1961. Born in Japan, I spent my student days in both Japan and the USA, and professional life in Canada. For my truly fulfilled life, I do not have sufficient words to thank my wife Lorraine (née Ford), son Ira, his wife Samantha Dugas, and dearest grandson Lincoln Dugas-Nishisato, as well as my sister Michiko Soma and brother Akihiko Nishisato in Japan. Finally, my sincere and heart-felt thanks to the three wonderful co-authors Pepe, Rosaria, and Eric, who helped me to fulfill my life-long wish.

The most important step for the publication of the book was provided by the editorial staff of Springer Singapore, and we would like to express our sincere appreciation to all the editorial staff for their meticulous editing and helpful advice. Through their devotion, we can now see this beautiful book.

Contents

Part I Joint Graphical Display

| | | |
|----------|---|----|
| 1 | Personal Reflections | 5 |
| 1.1 | Early Days | 6 |
| 1.2 | Internationalization | 9 |
| 1.3 | Books in French, Japanese and English | 10 |
| 1.4 | Names for Quantification Theory | 11 |
| 1.5 | Two Books with Different Orientations | 13 |
| 1.6 | Joint Graphical Display | 14 |
| 1.7 | A Promise to J. Douglas Carroll | 15 |
| 1.8 | From Dismay to Encouragement | 16 |
| | References | 19 |
| 2 | Mathematical Preliminaries | 27 |
| 2.1 | Graphs with Orthogonal Coordinates | 28 |
| 2.1.1 | Linear Combination of Variables | 28 |
| 2.1.2 | Principal Axes | 30 |
| 2.2 | Correlation and Orthogonal Axes | 31 |
| 2.3 | Standardized Versus Non-standardized PCA | 32 |
| 2.4 | Principal Versus Standard Coordinates | 33 |
| | References | 36 |
| 3 | Bi-modal Quantification and Graphs | 37 |
| 3.1 | Likert Scale | 38 |
| 3.1.1 | Its Ubiquitous Misuse | 38 |
| 3.1.2 | Validity Check | 39 |
| 3.2 | Quantification Theory | 44 |
| 3.2.1 | Quantification by Reciprocal Averaging | 44 |
| 3.2.2 | Simultaneous Linear Regressions | 47 |
| 3.3 | Bi-linear Decomposition | 47 |
| 3.3.1 | Key Statistic: Singular Values | 49 |

| | | |
|----------|---|------------|
| 3.4 | Bi-modal Quantification and Space | 50 |
| 3.5 | Step-by-Step Numerical Illustrations | 50 |
| 3.5.1 | Basic Quantification Analysis | 51 |
| 3.6 | Our Focal Points | 55 |
| 3.6.1 | What Does Total Information Mean? | 55 |
| 3.6.2 | What is Joint Graphical Display | 58 |
| 3.7 | Currently Popular Methods for Graphical Display | 58 |
| 3.7.1 | French Plot or Symmetric Scaling | 58 |
| 3.7.2 | Non-symmetric Scaling (Asymmetric Scaling) | 59 |
| 3.7.3 | Comparisons | 59 |
| 3.7.4 | Rational 2-D Symmetric Plot | 61 |
| 3.7.5 | CGS Scaling | 62 |
| 3.8 | Joint Graphs and Contingency Tables | 64 |
| 3.8.1 | A Theorem on Distance and Dimensionality | 66 |
| | References | 66 |
| 4 | Data Formats and Geometry | 69 |
| 4.1 | Contingency Table in Different Formats | 70 |
| 4.2 | Algebraic Differences of Distinct Formats | 71 |
| 4.3 | CGS Scaling: Incomplete Theory | 74 |
| 4.4 | More Information on Structure of Data | 74 |
| | References | 77 |
| 5 | Coordinates for Joint Graphs | 79 |
| 5.1 | Coordinates for Rows and Columns | 79 |
| 5.2 | One-Component Case | 80 |
| 5.3 | Theory of Space Partitions | 84 |
| 5.4 | Two-Component Case | 87 |
| 5.5 | Three-Component Case | 93 |
| 5.6 | Wisdom of French Plot | 97 |
| 5.7 | General Case | 98 |
| 5.8 | Further Considerations | 102 |
| 5.8.1 | Graphical Approach and Further Problems | 104 |
| 5.8.2 | Within-Set Distance in Dual Space | 105 |
| | References | 106 |
| 6 | Clustering as an Alternative | 107 |
| 6.1 | Decomposition of Input Data | 108 |
| 6.1.1 | Rorschach Data | 108 |
| 6.1.2 | Barley Data | 108 |
| 6.2 | Partitions of Super-Distance Matrix | 111 |
| 6.3 | Outlines of Cluster Analysis | 117 |
| 6.3.1 | Universal Transform for Clustering (UTC) | 117 |
| 6.4 | Clustering of Super-Distance Matrix | 119 |
| 6.4.1 | Hierarchical Cluster Analysis: Rorschach Data | 119 |

- 6.4.2 Hierarchical Cluster Analysis: Barley Data 120
- 6.4.3 Partitioning Cluster Analysis: Rorschach Data 120
- 6.4.4 Partitioning Cluster Analysis: Barley Data 123
- 6.5 Cluster Analysis of Between-Set Relations 124
 - 6.5.1 Hierarchical Cluster Analysis of Rorschach Data (UTC) 124
 - 6.5.2 Hierarchical Cluster Analysis of Barley Data (UTC) 125
 - 6.5.3 Partitioning Cluster Analysis: Rorschach Data and Barley Data (UTC) 125
 - 6.5.4 Effects of Constant Q for UTC on Cluster Formation 126
- 6.6 Overlapping Versus Non-overlapping Clusters 126
- 6.7 Discussion and Conclusion 128
- 6.8 Final Comments on Part 1 128
- References 130

Part II Scoring Strategies and the Graphical Display

- 7 Scoring and Profiles 135**
 - 7.1 Introduction 135
 - 7.2 Profiles 136
 - 7.3 The Method Reciprocal Averaging 137
 - 7.3.1 An Overview 137
 - 7.3.2 Profiles 138
 - 7.3.3 The Iterative Approach 139
 - 7.3.4 The Role of Eigendecomposition 143
 - 7.3.5 The Role of Singular Value Decomposition 145
 - 7.3.6 Models of Correlation and Association 146
 - 7.4 Canonical Correlation Analysis 147
 - 7.4.1 An Overview 147
 - 7.4.2 The Method 148
 - 7.5 Example 150
 - 7.5.1 One-Dimensional Solution via Reciprocal Averaging 150
 - 7.5.2 K -Dimensional Solution via SVD 154
 - 7.5.3 On Reconstituting the Cell Frequencies 155
 - 7.6 Final Remarks 156
 - References 156
- 8 Some Generalizations of Reciprocal Averaging 159**
 - 8.1 Introduction 159
 - 8.2 Method of Reciprocal Medians (MRM) 159
 - 8.3 Reciprocal Geometric Averaging (RGA) 161
 - 8.3.1 RGA of the First Kind (RGA1) 161

| | | |
|-----------|---|------------|
| 8.3.2 | RGA of the Second Kind (RGA2) | 162 |
| 8.3.3 | RGA of the Third Kind (RGA3) | 163 |
| 8.4 | Reciprocal Harmonic Averaging (RHA) | 164 |
| 8.5 | Final Remarks | 165 |
| | References | 166 |
| 9 | History of the Biplot | 167 |
| 9.1 | Introduction | 167 |
| 9.2 | Biplot Construction | 170 |
| 9.3 | Biplot for Principal Component Analysis | 174 |
| 9.4 | Final Remarks | 177 |
| | References | 177 |
| 10 | Biplots for Variants of Correspondence Analysis | 181 |
| 10.1 | Introduction | 181 |
| 10.2 | Biplots for Simple Correspondence Analysis—The Symmetric Case | 183 |
| 10.3 | Biplots for Simple Correspondence Analysis—The Asymmetric Case | 185 |
| 10.4 | Ordered Simple Correspondence Analysis | 186 |
| 10.4.1 | An Overview | 186 |
| 10.4.2 | Biplots for Ordered Simple Correspondence Analysis | 189 |
| 10.4.3 | The Biplot and a Re-Examination of Table 3.1 | 192 |
| 10.5 | The Biplot for Multi-Way Correspondence Analysis | 199 |
| 10.5.1 | An Overview | 199 |
| 10.5.2 | TUCKER3 Decomposition | 199 |
| 10.6 | The Interactive Biplot | 203 |
| 10.6.1 | The Biplot and Three-Way Correspondence Analysis | 205 |
| 10.6.2 | Size and Nature of the Dependence | 206 |
| 10.6.3 | The Interactive Biplot | 207 |
| 10.7 | Final Remarks | 209 |
| | References | 210 |
| 11 | On the Analysis of Over-Dispersed Categorical Data | 215 |
| 11.1 | Introduction | 215 |
| 11.2 | Generalized Pearson Residual | 216 |
| 11.3 | Special Cases | 217 |
| 11.3.1 | Generalized Poisson Distribution | 217 |
| 11.3.2 | Negative Binomial Distribution | 219 |
| 11.3.3 | Conway-Maxwell Poisson Distribution | 219 |
| 11.4 | Over-Dispersion, the Biplot and a Re-Examination of Table 3.5 | 220 |

| | | |
|--------|------------------------------------|-----|
| 11.5 | Stabilizing the Variance | 227 |
| 11.5.1 | The Adjusted Standardized Residual | 227 |
| 11.5.2 | The Freeman-Tukey Residual | 227 |
| 11.6 | Final Remarks | 229 |
| | References | 230 |

Part I

Joint Graphical Display

Preface

Over the past century and a half, quantification theory (QT) has been presented in many papers and books in diverse languages. For the current stage of its developments, please refer to Beh & Lombardo (2014), which is an excellent compendium of our current knowledge—other reference books in English, French, and Japanese will be presented later.

In the quantification of a two-way table of data (*e.g.*, contingency tables), we use singular value decomposition of the data matrix, and as such Torgerson (1958) called our quantification procedure *principal component analysis of categorical data*. The traditional principal component analysis (PCA) (Pearson, 1901; Hotelling, 1933), however, is different from quantification theory in two aspects, namely (1) continuous data for PCA and categorical data for QT and (2) primarily uni-modal analysis for PCA and bi-modal analysis for QT. These differences have led to respective courses of development.

As for the traditional uni-modal analysis of PCA, a typical data set may be patients-by-medical measurements (*e.g.*, blood pressure, heart rate, body temperature), and the object of the analysis is to find multidimensional relations of these medical statistics, where patients are considered a random sample, hence no direct interest in analyzing individual patients. The main task lies in finding multidimensional coordinates of these medical measures. This is a straightforward mathematical problem and there is no theoretical problem in finding the Euclidean coordinates for these variables.

In contrast, the bi-modal analysis of QT deals with such a data set as collected from different age groups of people on their most preferred life styles out of say ten choices. In this case, the main object of analysis is to find multidimensional relations between two sets of variables (age groups and life styles), and we must find multidimensional coordinates for the two sets of variables. It is almost certain that age groups and life styles are correlated to some degrees. This correlation makes the graphical display of QT results much more complicated than that of PCA, for we must face at least two questions: (1) how many dimensions are needed to describe

the complete relations between two sets of correlated variables, and (2) how we can find the Euclidean coordinates of two sets of correlated variables in common space.

Thus, one-mode and two-mode tasks lead to distinct tasks. Note that we have one set of variables for one-mode analysis and two sets of correlated variables for two-mode analysis. One set of multidimensional coordinates of medical statistics is the direct output of PCA, while we have two sets of multidimensional coordinates, which is the reason why we talk about *joint* graphical display for QT. It can easily be inferred that the two-mode QT output requires a space of larger dimensionality than the one-mode PCA output. How to expand the space for QT appropriately, however, is not a simple matter, and this is known as *the perennial problem of joint graphical display*.

In typical courses of introductory statistics, we learn a geometric interpretation of *correlation* between two variables: the two variables are expressed as two vectors and the correlation (Pearson, 1904) is defined as the cosine of the angle between the vectors. Then, assuming that the correlation between the rows and the columns is not 1, a single component of QT requires two-dimensional space. Our problem then is how to double the space for the QT outcomes. Keep in mind that we are interested in placing both rows and columns of the contingency table in *common space*.

In QT, we must find coordinates to accommodate two sets of correlated variables in expanded space. In the history of QT, however, this task of space expansion has been intentionally or non-intentionally avoided or even ignored for the sole reason of practical graphing. Unbelievable as it may sound, this avoidance of space expansion for QT has become the main-stream of QT and it has dominated the QT literature and history. The main aim of Part I is to revisit the currently popular compromised graphical procedure and replace it with a logically correct method for joint graphical display.

French plot is currently the most widely used graph. This name became popular in the 1970s and 1980s. It is also referred to as *symmetric scaling*. We can define symmetric scaling as a joint graph of principal coordinates of rows and principal coordinates of columns, assuming that the *row-column correlation is perfect*. The problem is that this assumption of perfect correlation does not generally exist nor is interesting at all. Thus, the method is one of the simplified and logically compromised graphical methods. It ignores the role of correlation between rows and columns in joint graphical display. Instead, it is a graphical method that overlays the graph for row variables over that for column variables. This is currently the most widely used practice and is logically not correct.

After a few attempts to rectify this time-honoured practice, most researchers have so far ignored the problem and have used it as a standard procedure and have moved away from the inherent problem. Thus, the problem of joint graphical display has been ignored for decades. For some researchers, it remained as a perennial problem.

In the current book, we will re-introduce Nishisato's 1980 work to show how to solve this perennial problem. But, why did the solution to this graphical problem remain unknown to most researchers for so many years? Sadly or tragically, there was a good reason for it as we will see it later.

Once we derive exact Euclidean coordinates for rows and columns in common space, we encounter another problem, that is, how to graph a configuration in multidimensional space. This is one major problem that must await future investigation and implementation. An immediate alternative is to rely on *dimensionless analysis*, of which one popular method is cluster analysis. Thus, we will consider cluster analysis as an alternative to joint graphical display. We will examine the merits and demerits of the two approaches, joint graphical display and cluster analysis.

In Part 1, Nishisato will present his view on the traditional quantification theory, starting with early days, then some background information for solving the perennial problem of joint graphical display, and a solution in terms of his theory of doubled common space in the first five chapters. Then, Clavel will join him to discuss cluster analysis as an alternative to graphical display in Chap. 6. This final chapter of Part 1 will contain the two authors' final words on Part 1.

Notes on Shizuhiko Nishisato

Born in 1935 in Sapporo, Japan. After BA and MA in experimental psychology from Hokkaido University, Japan, I went to the University of North Carolina at Chapel Hill (UNC) as a Fulbright student and obtained Ph.D. under a joint programme of psychometrics and mathematics. At UNC, I was taught by excellent professors (e.g., former Presidents of the Psychometric Society [Bock, Jones, Kaiser, Adkins Woods] and Hotelling), had great fellow students (e.g., Rapoport, Messick, Mukherjee, Wiesen, Smith, Das Gupta, Abbe (née Niehl), Gordon, Zyzanski, Cole (née Stooksberry), Kahn, and Norvick). I served as the only subject for Prof. T. G. Thurstone's project (the late L. L. Thurstone's wife) for an English proficiency test for foreign students, each session was always followed by tea and cookies with her at the Thurstone Psychometric Laboratory.

Since 1961, the Psychometric Society became my home ground for the next 60 years, for its annual meetings became my personal arena to forge a wide acquaintance with key researchers (e.g., Gulikesen, Horst, Guttman, Torgerson, Green, Coombs, Harman, Kruskal, J. D. Carroll, Luce, Jöreskog, Cliff, Bentler, Bloxom, Ramsay, Arabie, Hubert, Ackerman, Molenaar, Fischer, Young, de Leeuw, Takane, Heiser, Meulman, van der Linden, Cudeck, U. Böckenholt, Thiesen, and others).

After one year of post-doctoral work at McGill University, I was recruited by Ross Traub to the Ontario Institute for Studies in Education, the University of Toronto (OISE/UT) in 1967. The Department of Measurement and Evaluation at OISE/UT soon became one of the centres of psychometrics in North America.

I coined the name **Dual Scaling** in 1976, retired in 2000 as Professor Emeritus, the University of Toronto, and currently live in Toronto with my wife Lorraine. I Served as President of the Psychometric Society, Editor of *Psychometrika*, Fellow of the American Statistical Association, Fellow of the Japanese Classification Society, Year 2000 Distinguished Alumnus of the UNC Psychology Alumni Association,

and President of the Metropolitan Toronto Japanese Family Services, and served 20+ years on the editorial board of Springer Verlag for the German Classification Society Data Analysis Series.

Shizuhiko Nishisato
University of Toronto, Canada.

Notes on Jose G. Clavel

Like most of the economists of my time, I finished my degree without knowing a word about the analysis of categorical data. Of course, we were taught in the first week of Statistics I that there were attributes that could be summarized in pie charts and so on, but after that, my classes moved into the quantitative world for the subsequent topics of Econometrics.

Thus, like Vasco Núñez de Balboa, I also saw my Pacific Ocean, when I was asked to write a thesis on multidimensional scaling (MDS), from which I moved years later to Correspondence Analysis, Classification and Regression Trees, Dual Scaling, and so on. In those days, more and more friends started pouring into my office with all types of data (not only economics and business data, but also stress data of race horses, non-cognitive skills data, LaLiga, and lately, covid-19 medical data). My policy of open doors, despite its interference of my work, has kept my mind fresh. It is like that "stay hungry and stay foolish" recommended by Steve Jobs.

In addition to those job requirements and my own curiosity, I have had the fortune to be in this age of computers. Starting with Framework and its spreadsheets, I followed the rainbow through Lotus, Excel, S-plus, Eviews, MATLAB, SPSS, Stata, R, and finally (so far) Python. This voyage had an important stop in Goregaon at the *Indira Gandhi Institute of Development Research*, Mumbai, India, where, during my sabbatical year, under the supervision of Dr. Dilip Nachane, a wise econometrician and friend, I learned how to program and write my own codes.

Looking back, I am sure that all these facts have contributed to my teaching life: my first-year statistics students will hopefully benefit from my background, as I take advantage of their desire to know more. I hope that the readers of this book will see that the categorical data have much more information inside than a simple pie chart.

José Joaquín García Clavel
Universidad de Murcia, España.

Chapter 1

Personal Reflections



Over half a century of his research career, Nishisato has observed the historical developments of quantification theory. He himself was involved in the heated arguments over the problem of joint graphical display of quantified results, the problem which lasted until recently when he solved it. So, please do allow him to reflect on his personal involvements in the controversy over the problem together with his overview of the early history of quantification theory.

It was the summer of 1960, in Tokyo, Japan, when Prof. Masanao Toda introduced Nishisato to Dr. Chikio Hayashi, one of the early pioneers of quantification theory. In 1961, with Hayashi's 1950 paper in his briefcase, Nishisato arrived as a Fulbright student at Chapel Hill, North Carolina, USA, and started his graduate work at the Psychometric Laboratory of the University of North Carolina. His supervisor was Prof. R. Darrell Bock, another pioneer of quantification theory, who was then teaching optimal scaling. Only after a while, Nishisato realized and understood that *Hayashi's theory of quantification* (Hayashi 1950) was essentially the same as *Bock's optimal scaling* (Bock 1960).

Later he learned that there were many other aliases for quantification theory. Outstanding among them was *French Analyse des Correspondances* (correspondence analysis), for it was unique with a strong emphasis on joint graphical display. To interpret the outcome of quantification, the French group used graphical display extensively, resulting in what we now know as *French plot*, or more neutrally *symmetric scaling* or *correspondence plot*. As mentioned earlier, this graphical method has a serious problem.

It has become a routine tool for visual display of quantification results. Because of its wide use, many researchers today are not even aware that this routine tool has had a long history of fierce debates, pro's and con's. We should note that some toughest critics of French plot would even denounce the use of French plot, but they were a minority and were ignored completely.

What is the problem with the French plot? The answer is that it does not provide an exact configuration, but is only a practical approximation to the true configuration. More concretely, the French plot employs half the space required for a complete description of the data configuration. In other words, a four-dimensional configuration of data is depicted in a two-dimensional graph. This is so because the French plot represents rows and columns of the contingency table as if the row-column correlation is perfect, that is, 1. But, we should realize that if the correlation is 1, we do not need a graphical display.

Critical views against any popular method are typically sidelined or suppressed, often into oblivion. This is the reason why Nishisato has chosen his own personal reminiscence as a useful vehicle to overview a historical background surrounding the joint graphical problem—his 1980 description of the total information in the contingency table has been more or less ignored, and when he tried to revisit the relevant theory in the book, a well-respected journal rejected his criticism on the current joint graphical display as “fundamentally wrong” for no obvious reasons.

We will see later that all the necessary information for finding multidimensional coordinates for correlated sets of variables was thoroughly discussed in his 1980 book. To understand the nature of his struggles, let us first go back to the early days of quantification theory, and then look at his solution to the perennial problem of joint graphical display in Chap. 5.

1.1 Early Days

In early 1960s, joint graphical display in data analysis was promoted by French researchers and the graphical display was uplifted to the level that it became almost synonymous to French *Analyse des Correspondances* (Benzécri et al. 1973). Since ideas of graphical display go back almost to the beginning of quantification theory, let us look at its early history first.

The birth of quantification theory goes as far back as the early years of the twentieth century (see Nishisato 2007a). First, we see an international group of ecologists interested in optimal mapping of two sets of variables (e.g., plants and environments) using gradient methods (e.g., Gleason, Lenoble, Ramensky), then we see similar developments advanced by famous social and statistical scientists in the 1930s and the 1940s (e.g., Edgerton and Kolbe, Ellenberg, Fisher, Guttman, Hirschfeld, Horst, Johnson, Maung, Mosier, Richardson and Kuder, Whitakker, Wilks), and then further developments in the 1950s and 1960s (e.g., Hayashi, Benzécri, R.D. Bock, Baker, Bourroche, Carroll, Escofier, Lebart, Lord, McDonald, Slater). These researchers established solid foundations for further developments of quantification theory. After 1980s, the number of publications on quantification theory increased substantially.

At the 1976 Annual Meeting of the Psychometric Society, Jan de Leeuw of the Netherlands, identified the following four groups of researchers as distinct promoters of quantification theory:

- **Japanese Quantification Theory Group:** Starting with Hayashi's papers in 1950 and 1952 in English, a large number of papers and several books were published, mostly in Japanese. The Institute of Statistical Mathematics (ISM), in Tokyo, played an important role in disseminating Hayashi's theory of quantification with Hiroshi Akuto's classification paradigm. In addition to those ISM researchers, there were also other outstanding researchers at several universities and research institutes throughout the country.
- **French Correspondence Analysis Group:** They published many relevant papers and books in French in the 1970s and 1980s, the best known being the two volumes by Benzécri and his many collaborators. Another outstanding French contribution is the journal *Les Cahiers de l'Analyse des Données*, devoted mostly to correspondence analysis. In terms of the number of active researchers, including those outside the Benzécri school, France was outstanding out of the four groups listed here.
- **Dutch Homogeneity Analysis Group:** Starting with de Leeuw's doctoral thesis at the University of Leiden, the Netherlands, in 1973, many young Dutch students published their doctoral theses on quantification theory and related topics in English from DSWO Press. Their contributions to the field were outstanding, and the University of Leiden was one of the centres of quantification theory, together with the above two groups. The other institutions in the Netherlands have also produced many outstanding researchers in psychometrics and related disciplines.
- **Canadian Group of Optimal/Dual Scaling:** For the purpose of dissemination to a large number of researchers in North America and abroad, Nishisato published five reports on Bock's optimal scaling and its generalizations (Nishisato 1972, 1973, 1976, 1979; Nishisato and Leong 1975), which culminated into his 1980 book. Since 1970, Nishisato and his students at the Ontario Institute for Studies in Education of the University of Toronto started presenting their studies on optimal/dual scaling and its generalizations to other types of categorical data at international conferences. Those days, Toronto was one of the centres of psychometrics in North America, with Ross E. Traub (Princeton University) in classic and modern test theory, Roderick P. McDonald (University of Queensland, Australia) in factor analysis and covariance structure analysis, Shizuhiko Nishisato (University of North Carolina) in optimal (dual) scaling, multidimensional scaling and other psychometric methods, and Raghu P. Bhargava (Stanford University) in multivariate analysis of discrete and continuous variables. The solid graduate programme was unfortunately terminated around 2000, and no more students were trained since then.

Out of these groups, the most colourful and vibrant was the French group, the main promoter of joint graphical display as its flagship. This is important to note because they had laid a solid mathematical foundation for quantification theory, together with a practical method for summarizing the quantification results in graphs.

In 2017, Nishisato chaired the Awards Committee for the conference of the International Federations of Classification Societies (IFCS), in Tokyo, and observed a dramatic change in submitted papers such that the reviews of relevant studies went

only as far back as 10 years, as opposed to typically 30 to 100 years some 50 years ago. In old days, it was very important to identify the first author who developed a particular procedure. What a surprise for an old-timer! A proper review of the relevant studies used to be essential, and the paper with a scant review of relevant studies was typically rejected! In this modern time when any information is instantly available through the internet, old studies perhaps do not matter as much as half a century ago.

Here strictly for old-timers' sake, let us list those researchers one could regularly see at conferences in 1960s–1980s, namely those involved in quantification theory and related areas. As you will see below, the old academia was quite vibrant, and those were the days when international travels were extremely difficult, due to cost and visa restrictions. Hopefully, some readers would recognize their predecessors or mentors on this list. The names below are in the alphabetical order and represent only a sample of active researchers

- *Japan*: Adachi, Akiyama, Akuto, Aoyama, Asano, Baba, Haga, C. Hayashi, F. Hayashi, Higuchi, Inukai, Ishizuka, Iwatsubo, Kamisasa, Katahira, Kobayashi, Kodake, Komazawa, Kyogoku, Maeda, Maruyama, Miyahara, Miyano, Mizuno, Morimoto, Murakami, Nakamura, Nojima, Ogawa, Ohsumi, Okamoto, Otsu, Saito, Sakamoto, Shiba, Sugiyama, Takakura, Takeuchi, Tanaka, Tarumi, Tsuchiya, Tsujitani, Yamada, Yanai, Yoshino, Yoshizawa.
- *France*: Benzécri, Besse, Bouroche, Caussinus, Cazes, Choulakian, d'Aubigny, Daudin, Deville, Diday, Escofier-Cordier, Escoufier, Fenélon, Fichet, Foucart, Jambu, Kazmierczak, Lebart, Leclerc, Lerman, Le Calve, Le Roux, Marco-torichino, Morineau, Nakache, Pagés, Rouanet, Roux, Saporta, Schektman, Tabard, Tenenhaus, Tomassone, Trecourt, Vasserot (Note: Tenenhaus, Rouanet and Le Roux visited Nishisato in Toronto).
- *The Netherlands*: de Leeuw, Heiser, Israëls, Kiers, Kroonenberg, Meulman, Sikkell, Stoop, ten Berg, ter Braak, van der Burg, van der Heijden, van Rijckevorsel, van Schuur. Imagine that these young researchers were already in the frontiers of research!
- *USA, Canada, Australia*: Abbey, Arabie, Arri, Austin, Baker, Bechtel, Bentler, Bloxom, R.D. Bock, W. Böckenholt, Bradley, Bradu, Brown, J.D. Carroll, Chang, Chase, Cliff, Clogg, Coons, Cronbach, Curtis, Dale, de Sarbo, Edgerton, Evans, Fienberg, Franke, Gabriel, Gauch, B.F. Green, P.E. Green, Guttman, Hartley (Hirschfeld), Helmes, Horst, Hotelling, Hubert, Jackson, Jones, Katti, Kessell, Kolbe, Kruskal, Lawrence, Leong, Lord, McDonald, McKeon, Moore, Nishisato, Noma, Norvik, Odoroff, Olkin, Orłóci, Peet, Perreault, Prentice, Ramsay, Rao, Schönemann, Sheu, Singer, Sokal, Spence, Takane, Torgerson, Torii, Tucker, Wang, Wentworth, Whittaker, Young.
- *Britain*: Burt, Cox, Critchley, Digby, Everitt, Goldstein, Gower, Hand, Healy, Hill, Kendall, Krzanowski, Slater, Stuart.
- *Germany*: Baier, H. H. Bock, I. Böckenholt, Decker, Gaul, Ihm, Mucha, Pfeifer, Schader.
- *Italy*: Bove, Coppi, D'Ambra, Decarli, Lauro.