DOCUMENTATION

Review of books

(In this section we publish reviews of the books from which we receive a copy in our library)

Compositional Data Analysis in Practice, by Michael Greenacre. CRC Press. Taylor & Francis Group. Boca Raton (FL, USA), 2018. XIV+121 pages.-ISBN 978-1-138-31661-4 (Hardback), ISBN 978-1-138-31661-0 (Paperback).

Compositional Data Analysis (CODA) is a statistical line of research that has recently emerged. In general, compositions are non-negative data with values summing to a constant, usually 1 or 100. Such a definition includes fat and oil fatty acid compositions, which are typically expressed as a percentage of the whole peak areas referred to mass.

Karl Pearson was the first to advise that the relationships between proportions of a whole could lead to spurious correlations and wrong conclusions. After a period of confirmation of this possibility, the publication of the book "The Statistical Analysis of Compositional Data", written by John Aitchison in 1986, was the first compendium on a positive approach to compositional data analysis, with the microcomputer statistical package CODA developed by the same author being an essential complement for the implementation of the special statistical methods required at that time.

Many statisticians have cultivated this special field of statistics from then and contributed to its surprising development. As an example of the interest generated, it suffices to mention the several books and statistical programs progressively published during the last few decades. However, the use of the new tools is still rather restricted to statisticians and the theoretical progress has not been accompanied by a parallel application by users from other fields. A major inconvenience for spreading CODA has been the lack of flexibility of researchers to change their minds from the mathematical principles learnt at school and university to those new concepts such as perturbation, powering, etc. But another important cause could also be the absence of publications focused on the basic principles of the new statistics in a simple language, able to be understood by scientists from other disciplines, and, simultaneously, to promote its application to other fields and situations.

The recently published book "Compositional Data Analysis in Practice" possibly covers this niche. The publication avoids cumbersome theoretical digressions and only presents to the reader the essential basic concepts for the application of CODA, using ratios and logratios that retain most of the original data structure and, subsequently, may lead to proper conclusions. After the preface, where the author explains his conversion to the CODA *religion*, the following chapters deal with the definition of compositional data, their characteristics, geometry, and visualisation. Then, the different logratio transformations (additive, centred and isometric) are explained, paying particular emphasis to the comparison of logratios and their interpretations. Next, the properties and distributions of logratios are commented as well as the procedures available for testing their normality. After that, a series of practical chapters follows, successively explaining regression models involving compositional data, dimension reduction, which is a critical issue when handling numerous variables, clustering and methods for solving the problems caused by the presence of zeros. The subsequent chapter deals with variable selection and the introduction of several alternatives such as stepwise variable selection, parsimonious selection, and the use of amalgamations. Finally, a chapter is devoted to a complete case study, using a dataset from the fatty acid compositions of four species of amphipods. Supplementary concise information on the theory of CODA is supplied in the ten sections of Appendix A, while a selected bibliography is offered as Appendix B. Several examples, including most of the methods and practical exercises of the book, are described in Appendix C, devoted to computations of CODA. The book ends with two

final Appendices: a glossary of terms in alphabetic order and an Epilogue, in which the author comments his ideas regarding the practice of CODA which, according to his point of view, should be directed to simplify the analysis as much as possible while producing interpretable results. The use of logratios may achieve both objectives, although the procedure will not represent a completely orthodox approach. The name of the R program developed for applying the CODA according to these principles cannot be more explicit, easyCODA, in agreement with the philosophy behind its development.

Therefore, readers will not find in this book a complete compendium of the CODA theory. On the contrary, they will be faced with a practical approach to CODA, requiring only a few elementary concepts. The simplification of the analysis and the straightforward interpretability of results is, clearly, one of the primary values of the publication. In addition, the emphasis on the general application of weights in the calculus of most of the operations and methodologies used throughout the book deserves a special mention. In fact, its general application in most of the calculus explained throughout the publication (together with easyCODA) is a novel contribution of "Compositional Data Analysis in Practice" to CODA which, as demonstrated, leads not only to more interpretable results but, in parallel, to more reliable conclusions. Finally, the application of Procrustes analysis, to measure the degree of similarity between data structures, and permutation testing, to assess statistical significance, are also outstanding contributions.

However, one would also expect, given the supposed elemental knowledge of practitioners on CODA, a more detailed explanation and, mainly, a fuller description in some sections on the way the results are obtained: for example, in Chapter 6 (Dimension reduction using logratio analysis) or Chapter 9 (Simplifying the task: variable selection). Also, a coordination between the case study (Chapter 10) and the CODA computations included as appendix would also be welcome.

In any case, the book represents an exciting contribution which introduces a simple and reliable alternative approach, although not fully orthodox, to CODA research. Besides, the accompanying R program, after fixing the few initial difficulties, will also contribute to a wider spreading of the proposed methodology. Altogether, the book and the easyCODA R package may represent a promising instrument for introducing CODA in the fat and oils field, where fatty acid compositions have been treated until now exclusively by classical multivariate techniques without considering their compositional structure.

Predicting the future is risky, but the book may represent an essential instrument for CODA spreading since it represents just what many practitioners were expecting to initiate their experience in this promising new statistical field of compositional data analysis.

A. Garrido Fernández