



**Electronic Journal of Applied Statistical Analysis  
EJASA, Electron. J. App. Stat. Anal.**

<http://siba-ese.unisalento.it/index.php/ejasa/index>

e-ISSN: 2070-5948

DOI: 10.1285/i20705948v12n2p552

**A genealogy of correspondence analysis: part 2 –  
the variants**

By Beh, Lombardo

Published: 14 October 2019

This work is copyrighted by Università del Salento, and is licensed under a Creative Commons Attribution - Non commerciale - Non opere derivate 3.0 Italia License.

For more information see:

<http://creativecommons.org/licenses/by-nc-nd/3.0/it/>

# A genealogy of correspondence analysis: part 2 – the variants

Eric J. Beh<sup>\*a</sup> and Rosaria Lombardo<sup>b</sup>

<sup>a</sup>*University of Newcastle, School of Mathematical and Physical Sciences , Callaghan NSW  
2308, Newcastle, Australia*

<sup>b</sup>*University of Campania, Economics Department, Gran Priorato di Malta, 81043 Capua (CE),  
Italy*

Published: 14 October 2019

In 2012, a comprehensive historical and genealogical discussion of correspondence analysis was published in *Australian and New Zealand Journal of Statistics*. That genealogy consisted of more than 270 key books and articles and focused on an historical development of the correspondence analysis, a statistical tool which provides the analyst with a visual inspection of the association between two or more categorical variables. In this new genealogy, we provide a brief overview of over 30 variants of correspondence analysis that now exist outside of the traditional approaches used to analyse the association between two or more categorical variables. It comprises of a bibliography of a more than 300 books and articles that were not included in the 2012 bibliography and highlights the growth in the development of correspondence analysis across all areas of research.

**keywords:** Correspondence analysis, dual scaling, optimal scaling, reciprocal averaging, R, history of statistics.

## 1 Introduction

Much has been said in the past on the development of correspondence analysis, most of which has simply treated it as being merely an “exploratory” technique for visualising the association between categorical variables. Despite this, there is an extensive amount in the literature which has highlighted the myriad of technical issues and have

---

\*Corresponding author: [eric.beh@newcastle.edu.au](mailto:eric.beh@newcastle.edu.au)

compared correspondence analysis with other aspects of statistics including Bayesian analysis (Braga et al., 2005; De Tibeiro and Murdoch, 2010) and the numerous categorical models that are available, including log-linear models, bilinear models, association models and correlation models; see, for example Van Der Heijden and De Leeuw (1985); Van Der Heijden et al. (1989). Irrespective of the varied perception of the need for, and utility, of correspondence analysis, it remains a valuable analytical procedure for understanding the (sometimes) complex association patterns that exist in categorical data. The varied history of its development is equally interesting and has been extensively documented by (Greenacre, 1989; Gifi, 1990; Nishisato, 2007; Kroonenberg, 2008; Beh and Lombardo, 2014). Bibliographies have also appeared that provide a list of the key papers concerned with the various issues of correspondence analysis. For example

- [1] NISHISATO, S. (1986). Quantification of Categorical Data: A Bibliography 1975 – 1986. Toronto, Canada: MicroStats.
- [2] BIRKS, H. J. B., PEGLAR, S. M. & AUSTIN, H. A. (1996). An annotated bibliography of canonical correspondence analysis and related constrained ordination methods 1986–1993. *Abstracta Botanica*, 20, 17 – 36.
- [3] BEH, E. J. (2004). A Bibliography of the Theory and Application of Correspondence Analysis, Vol. III – By Year. Available through the author’s ResearchGate site.
- [4] BEH, E. J. & LOMBARDO, R. (2012). A genealogy of correspondence analysis. *The Australian and New Zealand Journal of Statistics*, 54, 137 – 168.

provide a list of references that apply correspondence analysis or extend its theoretical development.

In addition to these bibliographies and the ever-growing number of papers that have been published using correspondence analysis since 2012, there has been an increasing number of texts beyond those described in that [4] describes. New books, or new editions of existing books, have recently been published that provide a mix of theoretical, practical and computational insight into the development of correspondence analysis. Some of the recent additions include those of

- [5] BEH, E.J. & LOMBARDO, R. (2014), *Correspondence Analysis: Theory, Practice and New Strategies*. Chichester: Wiley.
- [6] BLASIUS, J. & GREENACRE, M. (eds) (2014). *Visualization and Verbalization of Data*. Boca Raton: CRC Press.
- [7] GREENACRE, M. (2017). *Correspondence Analysis in Practice* (3rd ed). Boca Raton: CRC Press.
- [8] HUSSON, F., LÊ, S. & PAGÈS, J. (2017). *Exploratory Multivariate Analysis by Example Using R* (2nd ed). Boca Raton: CRC Press.

- [9] HJELLBREKKE, J. (2018). *Multiple Correspondence Analysis for the Social Sciences*. Routledge.

A Japanese translation of the six-chapter short book

- [10] CLAUSEN, S. E. (1988). *Applied Correspondence Analysis: An Introduction*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-121. Thousand Oaks, CA: Sage.

by Kazuo Fujimoto (Sakushin Gakuin University) was published in 2015 and includes not just the translation of the book but also six additional chapters that highlight the use of the FactoMineR package in R to perform correspondence analysis; see

- [11] HUSSON, F., JOSSE, J., LE, S. & MAZET, J. (2018). FactoMineR: Multivariate exploratory data analysis and data mining. R version 1.41, <https://cran.r-project.org/web/packages/FactoMineR/index.html>

and <http://factominer.free.fr/> for details on this package. The choice of this R package was made since it is amenable to Japanese characters. The six extra chapters also cover topics outside of its original source including the analysis of three-way data, multiple response tables and the link between correspondence analysis and log-linear models. An overview of the translation was presented at the 2015 conference “Correspondence Analysis and Related METHODS” (CARME2015) held in Naples, Italy. Another short 39-page text, aimed at archaeological researchers, is

- [12] SIEGMUND, F. (2015). *How to Perform a Correspondence Analysis: A Short Guide to Archaeological Practice*. CreateSpace Independent Publishing Platform.

Unlike the many new editions to the correspondence analysis literature, this contribution describes its implementation using PAST, not R. PAST – an abbreviation of PAleontological Statistics – was developed by Øyvind Hammer, David Harper and Paul Ryan and can be freely downloaded from <http://folk.uio.no/ohammer/past>. A comprehensive description of the programs functionality can be found by referring to its 277-page manual. However, a more concise description can be found in

- [13] HAMMER, Ø., HARPER, D. A. T. & RYAN, P. D. (2001). PAST: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, 4(1), 9 pages.

New texts have appeared that dedicate substantial chapters, or sections, to discussing various aspects of correspondence analysis. Some excellent inclusions are

- [14] FRIENDLY, M. & MEYER, D. (2016). *Discrete Data Analysis with R: Visualization and Modeling Techniques for Categorical and Count Data*. Boca Raton: CRC Press.
- [15] KATERI, M. (2014). *Contingency Table Analysis: Methods and Implementation using R*. New York: Birkhäuser.

- [16] GOWER, J., LUBBE, S. & LE ROUX, N. (2011). *Understanding Biplots*. Chichester: Wiley.
- [17] GREENACRE, M. & PRIMICERIO, R. (2013). *Multivariate Analysis of Ecological Data*. Bilbao: Fundación BBVA.

This paper complements the genealogy of 4 which discussed much of the early theoretical groundwork of correspondence analysis focusing on key measures of association; some of which are accredited to the most imminent and influential statisticians of the early 20th century, including Pearson, Fisher, Yates and Yule. Much of this groundwork focused on finding the “best” scaling of the categories while maximising the association between the variables. In fact, on page 371 of his essay on scaling,

- [18] HORST, P. (1935). Measuring complex attitudes. *The Journal of Social Psychology*, 6, 369 – 375.

says

One of the fundamental requisites of any scaling procedure is that a distribution of scores calculated . . . should exhibit a significant amount of variability. . . . In general, then, the most unsatisfactory scaling procedure conceivable is one which results in identical scores for all persons measured

Early rigorous mathematical derivations for determining the “best” scaling was undertaken by

- [19] HOTELLING, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417 – 440 & 498 – 520.
- [20] THURSTON, L. L. (1933). *The Theory of Multiple Factors*. Ann Arbor, Michigan: Edwards.

In fact, [18, p. 372] also commented that Thurstone “derived a set of equations very similar to those of Hotelling and for the same purpose”. So, it was from this ground work that scoring methods for categorical variables began to blossom. One of the most influential proponents of this area of statistical research at the time, and one that was made independently of Hotelling and Thurston’s contributions, was

- [21] FISHER, R. A. (1940). The precision of discriminant functions. *Annals of Eugenics*, 10, 422 – 429.

Since these early days, exploring the association between categorical variables, whether it be through purely numerical means or more graphical approaches, has blossomed and correspondence analysis lies at the heart of the visual tools for doing so. This has resulted in various synonyms being used for the term “correspondence analysis” (including dual scaling, optimal scaling, reciprocal averaging, homogeneity analysis and correspondence

factor analysis). In this paper we explore the genealogy of correspondence analysis by providing a brief overview of the many variations that now exist in the literature. All these variations have one common thread – scaling the categories according to some criteria applied to the scales and/or the association.

## 2 Scaling Categorical Data

Before we discuss the various adaptations to the classical approach to correspondence analysis we need to keep in mind that, at its heart, is the scaling of categorical data. Such scaling involves identifying scores (on a continuous scale) that capture the variation between a set of variables as well as the association between them. There are many different ways in which categorical data can be scaled, some of them are known under different names but are effectively lead to the same solution. This is an important issue in correspondence analysis since, once a scaling has been made on a variable, a visual interpretation of the association between the variables can be made. How these scales (also called scores) is one of the core foundations of correspondence analysis.

Like many others, [4, 5] describe the origins of correspondence analysis and noted that its name is derived from the English translation of *analyse des correspondance* adopted by Jean-Paul Benzécri. However, with the technique having its origins lying in various places around the globe, it is also known under different names which stem from different ways of viewing the data but lead to the same solution. There are many ways of determining the scales for the categories of a two-way or multi-way contingency table which lead to scaling obtained from performing correspondence analysis. Nishisato (2007) and Tenenhaus and Young (1985) provide excellent technical summaries of the most popular approaches; Tenenhaus and Young (1985) demonstrated the equivalency of these techniques for multiple variables by showing that virtually all scaling approaches lead to the singular value decomposition of some (common) matrix of association. From this decomposition, the singular vectors serve as the scales (or scores) of interest.

We shall not provide an account of the technical features shown in Tenenhaus and Young (1985). Instead we provide a brief overview of some of the key contributions. In Section 3, we expand upon our scaling discussion and describe the evolution of the graphical features which have made correspondence analysis a very versatile tool for the data analyst wishing to explore the association between their categorical variables.

### 2.1 Optimal scaling

While much of the ground work to scaling categorical data had been undertaken early in the 20th century, one such approach referred to as optimal scaling was discussed by

- [22] BOCK, R. D. (1956). The selection of judges for preference testing. *Psychometrika*, 21, 349 – 366.

and is so named since the categories are scaled by optimising (maximising) the association between the variables. This method of scaling was later elaborated upon by the same author in

- [23] BOCK, R. D. (1960). *Methods and Applications of Optimal Scaling*. The University of North Carolina Psychometric Laboratory and Memorandum, No 25, Chapel Hill, NC.

Further developments under the guise of optimal scaling were made by

- [24] NISHISATO, S. (1978). Optimal scaling of paired comparison and rank order data: An alternative to Guttman's formulation. *Psychometrika*, 43, 263 – 271.
- [25] NISHISATO, S. & INUKAI, Y. (1972). Partially optimal scaling of items with ordered categories. *Japanese Psychological Research*, 14, 109 – 119.

These references focus on the scaling of nominal categorical variables but many including

- [26] BRADLEY, R. A., KATTI, S. K. & COONS, I. J. (1962). Optimal scaling for ordered categories. *Psychometrika*, 27, 355 – 37.
- [27] SNELL, E. J. (1964). A scaling procedure for ordered categorical data. *Biometrics*, 20, 592 – 607.
- [28] NISHISATO, S. & INUKAI, Y. (1972). Partially optimal scaling of items with ordered categories. *Japanese Psychological Research*, 14, 109 – 119.
- [29] NISHISATO, S. & ARRI, P. S. (1975). Nonlinear programming approach to optimal scaling of partially ordered categories. *Psychometrika*, 40, 525 – 548.
- [30] TANAKA, Y. (1979). Optimal scaling for arbitrarily ordered categories. *Annals of the Institute of Statistical Mathematics*, 31, 115 – 124.

developed some of the first “optimal scaling” strategies for ordered categorical data.

## 2.2 Reciprocal averaging

The introduction to correspondence analysis of [5] is from a perspective whereby the scores for the row and column categories are found by reciprocating (moving between) two linear combinations of (weighted) averaged scales. Such an approach is therefore referred to as “reciprocal averaging” and its origins were discussed by

- [31] RICHARDSON, M. & KUDER, G. F. (1933). Making a rating scale that measures. *Personnel Journal*, 12, 36 – 40.

with the term “method of reciprocal averaging”, or more simply “reciprocal averaging”, adopted “for convenience” by [18, p. 370.]

These authors [18, 31] did not discuss this procedure from a mathematical perspective, instead discussing it from a psychometric point of view. Interestingly Horst appears

to have not been aware of Richardson and Kuder's work, instead reflecting upon the contributions of Hotelling and Thurstone (as described above). The key statistical link between reciprocal averaging, correspondence analysis and the English literature was that of

- [32] HILL, M. O. (1973). Reciprocal averaging: An eigenvector method of ordination. *Journal of Ecology*, 61, 237 – 251.
- [33] HILL, M. O. (1974). Correspondence analysis: A neglected multivariate methods. *Applied Statistics*, 23, 340 – 354.
- [34] GAUCH JR, H. G., WHITTAKER, R. H. & WENTWORTH, T. R. (1977). A comparative study of reciprocal averaging and other ordination techniques. *Journal of Ecology*, 65, 157 – 174.

with others, such as

- [35] DE LEEUW, J., YOUNG, F. W. & TAKANE, Y. (1976). Additive structure in qualitative data: an alternating least squares method with optimal scaling features. *Psychometrika*, 41, 471–503.
- [36] BEKKER, P. & DE LEEUW, J. (1988). Relations between variants of non-linear principal component analysis. In *Component and Correspondence Analysis*, eds J. van Rijkevorsel, J. de Leeuw, pp. 1 – 31, Chichester: Wiley.

contributing to this area of research.

Hill's approach is based on Whittaker (1967) gradient analysis of categories and identifies scores for each category of two categorical variables, while maximising the correlation between them. Given the biological emphasis of Hill's articles, further publications on reciprocal averaging that appear in the biological area of research include those of

- [37] FASHAM, M. J. R. (1977). A comparison of nonmetric multidimensional scaling, principal components and reciprocal averaging for ordination of simulated coenoclines and coenoplanes. *Ecology*, 58, 551 – 561.
- [38] CULP, J. M. & DAVIS, R. W. (1980). Reciprocal averaging and polar ordination as techniques for analysing lotic macroinvertebrate communities. *Canadian Journal of Fisheries and Aquatic Sciences*, 37, 1358 – 1364.
- [39] HACKER, R. B. (1983). Use of reciprocal averaging ordination for the study of range condition gradients. *Journal of Range Management*, 36, 25 – 30.
- [40] EZCURRA, E. (1987). A comparison of reciprocal averaging and non-centred principal components analysis. *Vegetatio*, 71, 41 – 47.

Further mathematical developments of reciprocal averaging, in line with the dual scaling approach to scaling, were made by

- [41] NISHISATO, S. (1984). Dual scaling by reciprocal medians. Proceedings of the 32nd Scientific Conference of the Italian Statistical Society (Estratto dagli Atti della XXXII Riunione Scientifica), Sorrento, Italy, pp. 141 – 147.
- [42] SACHS, J. (1994). Robust dual scaling with Tukey’s biweight. *Applied Psychological Measurement*, 18, 301 – 309.

### 2.3 Dual Scaling

Another early form of scaling from which the solution to correspondence analysis rests is that of “dual scaling”. Essentially, this involves simultaneously scaling two spaces (a row space and column space) yielding a technique that determines simultaneously scales two variables. This approach was first thoroughly discussed in

- [43] NISHISATO, S. (1980). *Analysis of Categorical Data: Dual Scaling and its Applications*. Toronto: University of Toronto Press.

and was also extensively discussed in two further books of his:

- [44] NISHISATO, S. (1994). *Elements of Dual Scaling: An Introduction to Practical Data Analysis*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- [45] NISHISATO, S. (2014). *Multidimensional Nonlinear Descriptive Analysis*. London: Chapman & Hall/CRC.

While the mathematical description of dual scaling leads to the same solution that is obtained for scoring categories using reciprocal averaging or any other synonymous technique, where dual scaling differs from correspondence analysis (and its related methods) is that there is no visual component to the analysis. While this philosophy may be at odds with the traditional (and French influences of) correspondence analysis approach, [44] dedicates a whole chapter (Chapter 14) to providing rationales as to why a visual display is not helpful. To help illustrate this point, one may refer to

- [46] NISHISATO, S. (1988). Assessing quality of joint graphical display in correspondence analysis and dual scaling. In *Data Analysis and Informatics V*, eds Y. Diday, Y. Escoufier, L. Lebart, J. Pagès, Y. Schktman, Y., R. Tomassone, pp. 409 – 416, Amsterdam: North-Holland.

for a measure of “badness” of a graphical display. Nishisato [45, p. 61] explains that his influence for not adopting a visual inspection of the association is due to a letter he had received from Michael Browne in June 1976. Browne also drew to Nishisato’s attention that optimal scaling was mathematically equivalent to the French approach of correspondence analysis and said

Browne’s view that the mathematical justification of plotting rows and columns together was not sound

and that this

deeply affected Nishisato . . . who consequently refrained from using graphical displays.

Other issues raised by Nishisato, and his colleague José Clavel, on the visual display from a correspondence analysis can be found by referring to

- [47] NISHISATO, S. (1995). Graphical representation of quantified categorical data: its inherent problems. *Journal of Statistical Planning and Inference*, 43, 121 – 132.
- [48] NISHISATO, S. & CLAVEL, J. G. (2003). A note on between-set distances in dual scaling and correspondence analysis. *Behaviormetrika*, 30, 87 – 98.
- [49] NISHISATO, S. & CLAVEL, J. G. (2010). Total information analysis: Comprehensive dual scaling. *Behaviormetrika*, 37, 15 – 32.

Further insights into dual scaling and its links to correspondence analysis can be found by referring to the extensive list of publications of Nishisato and his colleagues, as well as, for example,

- [50] BECHTEL, G. G. (1971). A dual scaling analysis for paired compositions. *Psychometrika*, 36, 135 – 154.
- [51] NISHISATO, S. (1984). Forced classification: a simple application of a quantification method. *Psychometrika*, 49, 25 – 36.
- [52] NISHISATO, S. & GAUL, W. (1988). Marketing data analysis by dual scaling. *International Journal of Research in Marketing*, 5, 151 – 170.
- [53] NISHISATO, S. (1995). Optimization and data structure: Seven faces of dual scaling. *Annals of Operations Research*, 55, 345 – 359.
- [54] NISHISATO, S. (1996). Gleaning in the field of dual scaling. *Psychometrika*, 61, 559 – 599.
- [55] NISHISATO, S. & BABA, Y. (1999). On contingency, projection and forced classification of dual scaling. *Behaviormetrika*, 26, 207 – 219.
- [56] VAN DE VELDEN, M. (2000). Dual scaling and correspondence analysis of rank order data. In *Innovations in Multivariate Statistical Analysis*, eds R. D. H. Heijmans, D. S. G. Pollock, A. Satorra, pp. 87 – 99, Boston: Springer.
- [57] TORRES, A. & GREENACRE, M. (2002). Dual scaling and correspondence analysis of preference paired comparisons and ratings. *International Journal of Research in Marketing*, 19, 401 – 405.

- [58] MUCHA, H.-J., BARTEL, H.-G. & DOLATA, J. (2014). Dual scaling classification and its application in archaeometry. In *Data Analysis, Machine Learning and Knowledge Discovery*, eds M. Spiliopoulou, L. Schmidt-Thieme, L. Janning, L., pp. 105 – 113, Cham: Springer.
- [59] SCHOONEES, P. C., VAN DE VELDEN, M. & GROENEN, P. J. F. (2015). Constrained dual scaling for detecting response styles in categorical data. *Psychometrika*, 80, 968 – 994.

For an excellent overview of these, and other, techniques we direct the reader to

- [60] TANAKA, Y. (1979). Review of the methods of quantification. *Environmental Health Perspectives*, 32, 113 – 123.
- [61] TENENHAUS, M. & YOUNG, F. W. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika*, 50, 91 – 119.

## 2.4 Pre-Visualisation

Before the impact of Benzécri's contributions, much of the literature concerned with the scoring of categorical variables was inherently one-dimensional, including those of reciprocal averaging and dual scaling. In fact, (Greenacre, 1984 p. 11) says of dual scaling that it

. . . is concerned with deriving numerical scores for categories with certain properties, a method pioneered by Guttman. No geometry of such scores is mentioned or intended in this framework and the results are not reported in the form of graphical displays. Neither is the framework specifically multidimensional, but rather a sequence of one-dimensional frameworks. This distinction is an important one and should also be mentioned in the case of reciprocal averaging . . . By contrast, correspondence analysis, as we know it, derives a set of multidimensional “scores” with a well-defined and intentional geometric interpretation

So, with all the groundwork of modern day correspondence analysis laid down in the 1950's and 1960's how did the term come about? None of the initial techniques developed were close in name to correspondence analysis. As described in [4] and elsewhere, the term “correspondence analysis” was not used until at least 1973 and is an English translation of the title to the two-volume work by Jean-Paul Benzécri's *L'Analyse des Données* where the geometric features of the analysis were discussed. Thus, Benzécri is often considered to be the “father” of correspondence analysis. It is his influence that has greatly impacted on how correspondence analysis has flourished – in particular, by

those in France, The Netherlands and Italy. However, the very essence of what correspondence analysis “is” can also be traced back to the Japanese school of data analysts lead by Chikio Hayashi with the 1952 paper

- [62] HAYASHI, C. (1952). On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics*, 3, 69 – 98.

An account of the influence of Hayashi in this area of statistics, and others, can be found by referring to

- [63] OHSUMI, N. (2004). Memories of Chikio Hayashi and his great achievement. *Electronic Journal for History of Probability and Statistics*, 4(2), 4 pages.

A great example of Hayashi’s influence in the development of correspondence analysis include that of

- [64] LEBART, L. & SAPORTA, G. (2014). Historical elements of correspondence analysis and multiple correspondence analysis. In *Visualization and Verbalization of Data*, eds. J. Blasius, M. Greenacre, pp. 31 – 44, Boca Raton: CRC Press.

who note that

- [65] HAYASHI, C. (1956). Theory and examples of quantification (II). *Proceedings of the Institute of Statistical Mathematics*, 4(2), 19 – 30.

was probably the first real application of multiple correspondence analysis. See also

- [66] MURTAGH, F. (2014). History of cluster analysis, In *Visualization and Verbalization of Data*, eds. J. Blasius, M. Greenacre, pp. 117 – 133, Boca Raton: CRC Press.

for a description of the common threads shared between the contributions of Hayashi and Benzécri to data analysis. Following the death of Prof Hayashi in 2002, Noboru Ohsumi reflected in the December 2002 issue of the *International Federation of Classification Societies (IFCS) newsletter* that when Benzécri and Hayashi met for the first time in 1979, Prof Benzécri said

I was surprised that, of all places, it was the Far East where there was a researcher who had come up with the same ideas as me, and quite clearly earlier than I had

with Hayashi stating

Professor Hayashi insisted that although there were similarities in equations and formulation, the basic ideas and philosophies behind were different

Hayashi's influence in statistics is recognised with the Chikio Hayashi Awards (CHA) which are awarded by the IFCS for the most promising early career researchers aged 30 - 35 years who make contributions to classification and data analysis.

## 2.5 A Glimpse into Data Visualisation through the Biplot

Once scaling of the variables has been undertaken, the matrices of these scales can be used to visualise the association between the variables. There are a variety of different ways in which a visual summary can be obtained. Traditionally, the row and column scales are weighted to yield what is commonly referred to as *principal coordinates* and the joint visual representation of these coordinates is called a *correspondence plot* – there is plenty in the literature on the definition and application of these coordinates and how to visualise the association using correspondence analysis; see, for example (Greenacre, 1984, pp. 88 - 89) and [5, Section 4.5.2]. An alternative method of visualising the association between categorical variables using correspondence analysis is through the *biplot*. While the biplot were first described by Gabriel (1971), it may now be considered the most utilised new feature of correspondence analysis for at least the last decade. This, in part, can be accredited (in our view) to the books of Greenacre (2010) and [16]. The following examples provide a glimpse into the different types of biplots that are now available, and their use in correspondence analysis and methods related to it:

- [67] GABRIEL, K. R. (1981). Biplot display of multivariate matrices for inspection of data and diagnosis. In *Interpreting Multivariate Data*, ed. V. Barnett, pp. 147 – 173, Chichester: Wiley.
- [68] BARR, B. S. & KANTOR, G. D. I. (1987). The weighted covariance biplot – an application. *South African Statistical Journal*, 21, 155 – 171.
- [69] GOWER, J. C. & HARDING, S. A. (1988). Nonlinear biplots. *Biometrika*, 75, 445 – 455.
- [70] GOWER, J. C. (1990). Three-dimensional biplots. *Biometrika*, 77, 773 – 785.
- [71] GABRIEL, K. R. & ORODOFF, C. L. (1990). Biplots in biomedical research. *Statistics in Medicine*, 9, 469 – 485.
- [72] GOWER, J. C. (1992). Generalised biplots. *Biometrika*, 79, 475 – 493.
- [73] DAIGLE, G. & RIVEST, L.-P. (1992). A robust biplot. *The Canadian Journal of Statistics*, 20, 241 – 255.
- [74] GOWER, J. C. (1993). Recent advances in biplot methodology. In *Multivariate Analysis: Future Directions 2*, eds. C. M. Cuadras, C. R. Rao, pp. 295 – 325, Amsterdam: North-Holland.

- [75] GREENACRE, M. J. (1993). Biplots in correspondence analysis. *Journal of Applied Statistics*, 20, 251 – 269.
- [76] GRASSI, M. & VISENTIN, S. (1994). Correspondence analysis applied to grouped cohort data. *Statistics in Medicine*, 13, 2407 – 2425.
- [77] GOWER, J. C. & HAND, D. J. (1996). *Biplots*. London: Chapman & Hall.
- [78] KROONENBERG, P. M. (1997). Introduction to biplots for  $G \times E$  tables. Research Report n.51, Centre for Statistics, The University of Queensland, Brisbane, Australia.
- [79] GABRIEL, K. R. (2002). Goodness of fit of biplots and correspondence analysis. *Biometrika*, 89, 423 – 436.
- [80] AITCHISON, J. & GREENACRE, M. J. (2002). Biplots in compositional data. *Applied Statistics*, 51, 375 – 392.
- [81] LA GRANGE, A., LE ROUX, N. & GARDNER-LUBBE, S. (2009). BiplotGUI: Interactive biplots in R. *Journal of Statistical Software*, 30(12), 37 pages.
- [82] GOWER, J. C., GROENEN, P. J. F. & VAN DE VELDEN, M. (2010). Area biplots. *Journal of Computational and Graphical Statistics*, 19, 46 – 61.
- [83] GREENACRE, M. J. (2013). Contribution biplots. *Journal of Computational and Graphical Statistics*, 22, 107 – 122.
- [84] GROENEN, P. J. F., LE ROUX, N. J. & GARDNER-LUBBE, S. (2015). Spline-based nonlinear biplots. *Advances in Data Analysis and Classification*, 9, 219 – 238.
- [85] GARDNER-LUBBE, S. (2016). A triplot for multiclass classification visualisation. *Computational Statistics and Data Analysis*, 94, 20 – 32.
- [86] GREENACRE, M. J. & GROENEN, P. J. F. (2016). Weighted Euclidean biplots. *Journal of Classification*, 33, 442 – 459.
- [87] GOWER, J. C., LE ROUX, N. J. & GARDNER-LUBBE, S. (2016). Biplots: qualitative data. *WIREs Computational Statistics*, 8, 82 – 111.
- [88] KIM, S.-K., MCKAY, D., TAYLOR, S., TOLIN, D. F., OLATUNJI, B. O, TIMPANO, K. R. & ABRAMOWITZ, J. S. (2016). The structure of obsessive compulsive symptoms and beliefs: A correspondence and biplot analysis. *Journal of Anxiety Disorders*, 38, 79-87.
- [89] ALBERS, C. J. & GOWER, J. C. (2017) Visualising interactions in bi and triaditive models for three-way tables. *Chemometrics Intelligent Laboratory System*, 167, 238 – 247.

### 3 The Correspondence Analysis Family Tree

Now that we have provided some insight into the early development of correspondence analysis and its visual links with the biplot, we now turn our attention to describing some of the various procedures that fall within the family of correspondence analysis techniques. Others have also provided a list of variations to correspondence analysis that have been proposed in the past. For example, Nishisato, 2007 (Chapter 3) provides an excellent overview of what he refers to as MUNDA (multidimensional nonlinear descriptive analysis) techniques. In his chapter, Nishisato also includes a “plethora of aliases” that list 52 techniques which include the biplot, the generalised biplot, reciprocal averaging, optimal scaling, dual scaling, correspondence analysis and three-way correspondence analysis. Beh and Lombardo [5, Section 1.6.3] list of 33 members of this family tree and, based on new contributions described at the CARME conference held in Naples, Italy, in 2015 this particular tree can now add another four members to its branches. Here we expand upon this list by providing a brief description of these techniques with reference to more than 200 articles. Some of these references, especially those that propose the earlier variations of correspondence analysis, are also referred to in [4]. Some of these techniques are extremely well established in the correspondence analysis literature, some have yet to gain any form of attention from the statistical and allied communities.

#### 3.1 Foucart’s Correspondence Analysis

One of the earliest variations of correspondence analysis, and one that is not well known, is *Foucart’s correspondence analysis*. Not much appears in the English literature on this early variant although it originally appeared in

[90] FOUCART, T. (1978). Sur les suites de tableaux de contingence indexés par le temps. *Statistique et Analyse des Données*, 3, 67 – 85.

[91] FOUCART, T. (1984). *Analyse Factorielle de Tableaux Multiples*. Paris: Masson.

and was focused on the social sciences. This member of the family involves the simultaneous correspondence analysis of multiple contingency tables where the row and column variables of each table are the same. Foucart’s correspondence analysis is performed by first calculating the matrix of cell proportions for each table then finding the “mean” contingency table, also referred to as a reference matrix – classical correspondence analysis is then performed on this reference matrix. Thankfully, the `foucart` function in the R package `ade4` provides the analyst with a means of performing this variant of correspondence analysis; see [143] for more on this function. Recent applications of Foucart’s correspondence analysis in the ecologically related disciplines can be found by referring to

[92] PAILLISSON, J.-M., REEBER, S. & MARION, L. (2002). Bird assemblages as bio-indicators of water regime management and hunting disturbance in natural wet grasslands. *Biological Conservation*, 106, 115 – 127.

- [93] PAVOINE, S., BLONDEL, J., BAGUETTE, M. & CHESSEL, D. (2007). A new technique for ordering asymmetrical three-dimensional data sets in ecology. *Ecology*, 88, 512 – 523.
- [94] BEAUCHARD, O., JACOBS, S., YSEBAERT, T. & MEIRE, P. (2013). Avian response to tidal freshwater habitat creation by controlled reduced tide system. *Estuarine, Coastal and Shelf Science*, 131, 12 – 23.
- [95] MANTE, C., BERNARD, G., BONHOMME, P. & NERINI, D. (2013). Application of ordinal correspondence analysis for submerged aquatic vegetation monitoring. *Journal of Applied Statistics*, 40, 1619 – 1638.
- [96] DOLÉDEC, S., TILBIAN, J & BONADA, N. (2017). Temporal variability in taxonomic and trait compositions of invertebrate assemblages in two climatic regions with contrasting flow regimes. *Science of the Total Environment*, 599-600, 1912 – 1921.

### 3.2 Canonical Correspondence Analysis

Another member of the correspondence analysis family that developed early as a result of its exposure to those within the ecological fields was that of *canonical correspondence analysis* (commonly abbreviated to CCA). The original technical development of CCA was described by

- [97] TER BRAAK, C. J. F. (1985). Correspondence analysis of incidence and abundance data: Properties in terms of a unimodal response model. *Biometrics*, 41, 859 – 873.

who was motivated by analysing presence/absence (or 1/0) data given some environmental factors that are taken into consideration. While classical correspondence analysis generally assesses the departure from complete independence of the cell frequencies in a contingency table, CCA assumes that the expected cell frequencies are modelled by a “Gaussian” log-linear model in terms of the row and column scores (to be estimated) thereby incorporating external, or ancillary, information about the categories that standard correspondence analysis techniques do not consider. By incorporating this feature, CCA has grown in popularity in the ecology. Further discussions, and applications, in this area have been described in, for example,

- [98] TER BRAAK, C. J. F. (1986). Canonical correspondence analysis: A new eigenvector technique for multivariate gradient analysis. *Ecology*, 67, 1167 – 1179.
- [99] PALMER, M. W. (1993). Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology*, 74, 2215 – 2230.
- [100] TER BRAAK, C. J. F. & VERDONSCHOT, P. F. M. (1995). Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquatic Sciences*, 57, 255 – 289.

- [101] YEE, T. W. (2004). A new technique for maximum likelihood canonical Gaussian ordination. *Ecological Monographs* 74, 685 – 701.
- [102] ZHU, M., HASTIE, T. J. & WALTHER, G. (2005). Constrained ordination analysis with flexible response functions. *Ecological Modelling* 187, 524 – 536.
- [103] MENDES, S., FERNÁNDEZ-GOMEZ, M. J., PEREIRA, M. J., AZEITEIRO, U. M. & GALINDO-VILLARDÓN, G. (2012). An empirical comparison of canonical correspondence analysis and STATICO in the identification of spatio-temporal ecological relationships. *Journal of Applied Statistics* 39, 979 – 994.

For an excellent discussion of CCA, and of correspondence analysis in general in ecology, refer to

- [104] TER BRAAK, C. J. F. (2014). History of canonical correspondence analysis, In *Visualization and Verbalization of Data*, eds J. Blasius, M. Greenacre, pp. 61 – 75, Boca Raton: CRC Press.

In R, canonical correspondence analysis can be performed the package `anacor`

- [105] DE LEEUW, J. & MAIR, P. (2009). Simple and canonical correspondence analysis using the R package `anacor`. *Journal of Statistical Software*, 31(5), 18 pages.
- [106] MAIR, P. & DE LEEUW, J. (2017). `anacor`: Simple and canonical correspondence analysis. R version 1.1-3, <https://cran.r-project.org/web/packages/anacor/index.html>

The package `vegan` also performs this variant of correspondence analysis. See

- [107] OKSANEN, J., GUILLAUME BLANCHET, F., FRIENDLY, M., KINDT, R., LEGENDRE, P., MCGLINN, D., MINCHIN, P. R., O'HARA, R. B., SIMPSON, G. L., SOLYMOS, P., STEVENS, M. H. H., SZOECS, E. & WAGNER, H. (2018). `vegan`: Community ecology package. R version 2.5-2, <https://cran.r-project.org/web/packages/vegan/index.html>

or Jari Oksanen's website <http://vegan.r-forge.r-project.org/> for more details and tutorials on how to use the package and a positive review of the package by

- [108] DIXON, P. (2003). `vegan`, a package of R functions for the community ecology. *Journal of Vegetation Science*, 14, 927 – 930.

### 3.3 Detrended Correspondence Analysis

One of the features of performing the classical approach to correspondence analysis is that sometimes the configuration of points that one obtains from visualising the association has what is referred to as the “arch effect”, also known as the “horseshoe effect”

(Kendall, 1971). This is where the configuration of points has a pattern that looks like an arch, or horseshoe. Some have felt that this is a problem and so consider an approach aimed at removing the arched “trend”, or to “detrend” the arch. Hence a variation of correspondence analysis that “detrends” the arch/horseshoe effect is referred to as *detrended correspondence analysis* (commonly abbreviated to DCA). One approach was proposed by

- [109] HILL, M. O. & GAUCH JR, H. G. (1980). Detrended correspondence analysis: an improved ordination technique. *Vegetatio*, 42, 47 – 58.

who proposed dividing the first axis of the graphical display into segments and centring the configuration of points according to the mean of each segment, thereby removing the “trend”. There has been much speculation about whether such detrending is necessary (some see it merely as an “artefact” of the analysis) and so its usage has been long debated. Regardless of where one stands on the issue, DCA continues to be extensively used. See, for example,

- [110] GAUCH, JR, H. G., WHITTAKER, R. H. & SINGER, S. B. (1981). A comparative study of nonmetric ordinations. *Journal of Ecology*, 69, 135 – 152.
- [111] MOHLER, C. L. (1981). Effects of sample distribution along gradients of eigenvector ordination. *Vegetatio*, 45, 141 – 145.
- [112] WILSON, M. V. (1981). A statistical test of the accuracy and consistency of ordinations. *Ecology*, 62, 8 – 12.
- [113] CARLETON, T. J. (1984). Residual ordination analysis: a method for exploring vegetation—environment relationships. *Ecology*, 65, 469 – 477.
- [114] PIELOU, E. C. (1984). *The Interpretation of Ecological Data*. New York: Wiley.
- [115] DJINDIJAN, F. (1985). Seriation and toposeriation by correspondence analysis. In *To Pattern the Past: Proceedings of the Symposium on Mathematical Methods in Archaeology*, eds. A. Voorrips, S. H. Loving, pp. 119 – 135, PACT.
- [116] WARTENBERG, D., FERSON, S., & ROHLF, F. J. (1987). Putting things in order: a critique of detrended correspondence analysis. *The American Naturalist*, 129, 434 – 448.
- [117] PEET, R. K., KNOX, R. G., CASE, J. S., & ALLEN, R. (1988). Putting things in order: the advantages of detrended correspondence analysis. *The American Naturalist*, 131, 924 – 934.
- [118] KNOX, R. G. (1989). Effects of detrending and rescaling on correspondence analysis: solution stability and accuracy. *Vegetatio*, 83, 129 – 136.

- [119] OKSANSEN, J. & MINCHIN, P. R. (1997). Instability of ordination results under changes in input data order: explanation and remedies. *Journal of Vegetation Science*, 8, 447 – 454.
- [120] SMITH, K. Y. & NEIMAN, F. D. (2007). Frequency seriation, correspondence analysis, and Woodland period ceramic assemblage variation in the Deep South. *Southeastern Archaeology*, 26, 47 – 72.

In R, detrended correspondence analysis can be performed using the `decorana` function in Jari Oksanen's `vegan` package. For details, see

- [121] OKSANEN, J. (2018). `vegan`: An introduction to ordination. R version 2.5-2. <https://cran.r-project.org/web/packages/vegan/index.html>

### 3.4 Non-Symmetrical Correspondence Analysis

While some variations of the classical approach to correspondence analysis were spawned because of the necessity of complexities arising from the data being analysed, one member of the family arose out of the necessity to generalise the type of association being considered. While, correspondence analysis generally considers categorical variables to be symmetrically associated (where, for two variables, neither variable is treated as a response variable to the other) and involves the partition of Pearson's chi-squared statistic, there are practical situations where the structure of two categorical variables may naturally be treated such that one variable is the predictor variable and the other is treated as the response variable. Such an association is commonly described as being asymmetric and the variant of correspondence analysis that accommodates this variation is called *non-symmetrical correspondence analysis* (commonly abbreviated to NSCA). Originally developed by

- [122] LAURO, N. C. & D'AMBRA, L. (1984). L'Analyse non symmetrique des correspondances. In *Data Analysis and Informatics III*, ed. E. Diday, pp. 433 – 446, Amsterdam: North-Holland.
- [123] D'AMBRA, L. & LAURO, N. C. (1989). Non-symmetrical correspondence analysis for three-way contingency table. In *Multiway Data Analysis*, eds R. Coppi, S. Bolasco), pp. 301 – 315, Amsterdam: North-Holland.

this variant is carried out by partitioning the Goodman-Kruskal tau index (Goodman and Kruskal, 1954) and its development and application has grown steadily over the past 30 years. This growth has been especially due to the ongoing developments by Italian statisticians, and their international teams, who have focused on examining the asymmetric association of variables in two-way and higher-way contingency tables. One may refer to, for example, the following contributions that describe the theoretical and practical development of NSCA:

- [124] SICILIANO, R. (1990). Asymptotic distribution of eigenvalues and statistical tests in nonsymmetric correspondence analysis. *Statistica Applicata*, 2, 259 – 276.

- [125] BALBI, S. (1992). On stability in nonsymmetrical correspondence analysis using bootstrap. *Statistica Applicata*, 4, 543 – 552.
- [126] VERDE, R. (1992). Nonsymmetrical correspondence analysis: a nonlinear approach. *Statistica Applicata*, 4, 453 – 463.
- [127] SICILIANO, R. MOOIJAART, A. & VAN DER HEIJDEN, P. G. M. (1993). A probabilistic model for non-symmetric correspondence analysis and prediction in contingency tables. *Journal of the Italian Statistical Society*, 2, 85 – 106.
- [128] BALBI, S. (1994). Influence and stability in non symmetrical correspondence analysis. *Metron*, 52, 111 – 128.
- [129] LOMBARDO, R., CARLIER, A. & D'AMBRA, L. (1996). Non-symmetric correspondence analysis for three-way contingency tables. *Methodologica*, 4, 59 – 80.
- [130] BALBI, S. (1998). Graphical displays in nonsymmetrical correspondence analysis. In *Visualization of Categorical Data*, eds J. Blasius, M. Greenacre, pp. 297 – 309, London: Academic Press.
- [131] GIMARET-CARPENTIER, C., CHESSEL, D. & PASCAL, J.-P. (1998). Non-symmetric correspondence analysis: An alternative for species occurrences data. *Plant Ecology*, 138, 97 – 112.
- [132] KROONENBERG, P. M. & LOMBARDO, R. (1998). Non-symmetric correspondence analysis: a tutorial. *Kwantitatieve Methoden*, 58, 57 – 83.
- [133] KROONENBERG, P. M. & LOMBARDO, R. (1999). Non-symmetric correspondence analysis: A tool for analysing contingency tables with a dependence structure. *Multivariate Behavioral Research*, 34, 367 – 396.
- [134] INVERARDI, P. L. N. & TAUFER, E. (1999). Interpreting asymmetrical display in correspondence analysis and in non-symmetric correspondence analysis. *Statistica Applicata*, 11, 49 – 58.
- [135] LOMBARDO R., KROONENBERG P. & D'AMBRA L. (2000). Non-symmetric correspondence analysis: a simple tool in market share distribution. *Journal of the Italian Statistical Society*, 9, 107 – 126.
- [136] KROONENBERG, P. M. (2002). Analyzing dependence in large contingency tables: nonsymmetric correspondence analysis and regression with optimal scaling. In *Measurement and Multivariate Analysis*, eds S. Nishisato, Y. Baba, H. Bozdogan, K. Kanefuji, pp. 87 – 96, Tokyo: Springer-Verlag.
- [137] TAKANE, Y. & JUNG, S. (2009). Tests of ignoring and eliminating in nonsymmetric correspondence analysis. *Advances in Data Analysis and Classification*, 3, 315 – 340.

- [138] TAKANE, Y. & JUNG, S. (2009). Regularized nonsymmetric correspondence analysis. *Computational Statistics and Data Analysis*, 53, 3159 – 3170.
- [139] BEH, E. J. & D’AMBRA, L. (2010). Non-symmetrical correspondence analysis with concatenation and linear constraints. *The Australian & New Zealand Journal of Statistics*, 52, 27 – 44.
- [140] BEH, E. J., LOMBARDO, R. & SIMONETTI, S. (2011). A European perception of food using two methods of correspondence analysis. *Food Quality & Preference*, 22, 226 – 231.
- [141] LOMBARDO, R. & RINGROSE, T. (2012). Bootstrap confidence regions in non-symmetrical correspondence analysis. *Electronic Journal of Applied Statistical Analysis*, 5(3), 413–417.

Non-symmetrical correspondence analysis may be performed using the R packages `ade4` and `CAvariants`. For details on these two packages, see

- [142] LOMBARDO, R. & BEH, E. J. (2016). Variants of simple correspondence analysis. *The R Journal*, 8/2, 167 – 184.
- [143] LOMBARDO, R. & BEH, E. J. (2017). `CAvariants`: Correspondence analysis variants. R version 3.4, <https://cran.r-project.org/web/packages/CAvariants/index.html>
- [144] DRAY, S., DUFOUR, A.-B. & THIOLOUSE, J. (2018). `ade4`: Analysis of ecological data: Exploratory and Euclidean methods in environmental sciences. R version 1.7-11, <https://cran.r-project.org/web/packages/ade4/index.html>

### 3.5 Partial Correspondence Analysis

Suppose we consider the case where, for classical correspondence analysis, we have two categorical variables, A and B. There may be an ancillary variable, C, that is related to both A and B, such as the gender of a person, their age group, or region in which they live. To gain an understanding of the association between A and B, the ancillary variable can influence this association. Therefore, the aim of *partial correspondence analysis* is to explore the association between the variables of interest by eliminating the impact of the ancillary variable. Such a variation of correspondence analysis was proposed by

- [145] YANAI, H. (1986). Some generalizations of correspondence analysis in terms of projection operators. In *Data Analysis and Informatics IV*, eds E. Diday, Y. Escofier, L. Lebart, J. Pagès, Y. Schectman, R. Tomassone, pp. 193 – 207, Amsterdam: North-Holland.
- [146] YANAI, H. (1988). Partial correspondence analysis and its properties. In *Recent Developments in Clustering and Data Analysis*, eds C. Hayashi, M. Jambu, E. Diday, N. Ohsumi, pp. 259 – 266, London: Academic Press.

and it has since been developed in the context of other variants of correspondence analysis. For example, one may consider the *partial canonical correspondence analysis* technique of

- [147] TER BRAAK C. J. F. (1987). Unimodal models to relate species to environment. Agricultural Mathematics Groups. Wageningen, the Netherlands: Box, Box 100, NL-6700 AC. 83 – 89.
- [148] TER BRAAK, C. J. F. (1988). Partial canonical correspondence analysis. In Classification and Related Methods of Data Analysis, ed. H. H. Bock, pp. 551 – 558, Amsterdam: North-Holland.

while

- [149] YANAI, H. & MAEDA, T. (2000). Partial multiple correspondence analysis. In Measurement and Multivariate Analysis, eds S. Nishisato, Y. Baba, H. Bozdogan, K. Kanefugi, pp. 110 – 113, Tokyo: Springer-Verlag.

extended the idea of partial correspondence analysis to the analysis of multiple categorical variables; such an extension is referred to as *partial multiple correspondence analysis*. Here, the aim is to eliminate the impact of a single ancillary variable when visually studying the association between multiple categorical variables. Similarly, when there exists an asymmetric association between two categorical variables, the impact of an ancillary variable may be eliminated through *partial non-symmetrical correspondence analysis* described by

- [150] LAURO, C. & BALBI, S. (1999). The analysis of structured qualitative data. Applied Stochastic Models and Data Analysis, 15, 1 – 27.

Applications of these partial versions of correspondence analysis include, for example,

- [151] BORCARD, D. & LEGENDRE, P. (1994). Environmental control and spatial structure in ecological communities: an example using oribatid mites (Acari, Oribatei). Environmental and Ecological Statistics, 1, 37 – 61.
- [152] BERSIER, L.-F. & MEYER, D. R. (1995). Relationships between bird assemblages, vegetation structure, and floristic composition of mosaic patches in riparian forests. Revue D'Écologie (La Terre et la Vie), 50, 15 – 33.
- [153] CARBONELL, A., PALMER, M., ABELLO, P., TORRES, P., ALEMANY, R. & DE SOLA, L. G. (2003). Mesoscale geographical patterns in the distribution of pandalid shrimps *Plesionika* ssp. in the Western Mediterranean. Marine Ecology Progress Series, 247, 151 – 158.
- [154] PAVOINE, S., BLONDEL, J., DUFOUR, A. B., GASC, A. & BONSALE, M. B. (2013). A new technique for analysing interacting factors affecting biodiversity patterns: Crossed-DPCoA. PLOS ONE, 8(1), e54530 (12 pages).

- [155] PENNINO, M. G., ROTTA, A., PIERCE, G. J. & BELLIDO, J. M. (2015). Interaction between bottlenose dolphin (*Tursiops truncatus*) and trammel nets in the Archipelago de La Maddalena, Italy. *Hydrobiologia*, 747, 69 – 82.

### 3.6 Residual Correspondence Analysis

Broadly speaking, the heart of correspondence analysis lies in assessing the difference between the observed cell frequencies from what is expected if the variables are independent. Therefore, many variations of correspondence analysis use Pearson's chi-squared statistic as the preferred measure of association between the variables. However, models of a more general nature than independence can be considered. For example, one may estimate the expected cell frequencies using a log-linear model, association model, correspondence model or correlation model. Thus, correspondence analysis can be modified to assess the departure of the observed cell frequencies from its expected cell frequency under the chosen model. That is, an analysis of such residuals using correspondence analysis provides a graphical goodness-of-fit procedure and is referred to as *residual correspondence analysis*. One may refer to, for example, the following papers for a discussion of various technical and practical aspects concerned with this member of the correspondence analysis family:

- [156] ESCOFIER, B. (1983). Analyse de la différence entre deux mesures définies sur le produit de deux mêmes ensembles. *Les Cahiers de l'Analyse des Données*, 8, 325 – 329.
- [157] VAN DER HEIJDEN, P. G. M. & DE LEEUW, J. (1985). Correspondence analysis used complementary to loglinear analysis. *Psychometrika*, 50, 429 – 447.
- [158] NOVAK, T. P. & HOFFMAN, D. L. (1990). Residual scaling: an alternative to correspondence analysis for the graphical representation of residuals from log-linear models. *Multivariate Behavioral Research*, 25, 351 – 370.
- [159] ROBERTS, JR. J. M. (1996). Alternative approaches to correspondence analysis of sociomatrices. *Journal of Mathematical Sociology*, 21, 359 – 368.

### 3.7 Decentred Correspondence Analysis

Another variation of correspondence analysis that was developed, and applied, within the ecological disciplines is *decentred correspondence analysis*. For some phenomenon it may be well understood how a particular categorical variable is observed in a population. Therefore, rather than considering the observed marginal proportions from the sample being analysed, this variant of correspondence analysis substitutes these observed proportions with weights (summing to 1) that reflect the understood phenomenon. Therefore, decentred correspondence analysis “decentres” the analysis for the variable where the observed marginal proportions are replaced. The technique was originally proposed in the study of fish communities in the upper Rhône River of France by

- [160] DOLÉDEC, S., CHESSEL, D. & OLIVIER, J. (1995). L'analyse des correspondances décentrée: application aux peuplements ichtyologiques du haut-rhône. Bulletin Français de la Pêche et de la Pisciculture, 336, 29-40.

with further technical and practical issues explored in

- [161] TILARD, E., FAYE, B., DESOUTTER, D., CHESSEL, D. & THIOULOUSE, J. (1997). Decentred correspondence analysis: A method for analysing a multivariate table of proportions. *Epidémiologie et Santé Animale*, 31 – 32, 3 pages.
- [162] KLINGEMAN, P. C., BRAVARD, J.-P., GIULIANI, Y., OLIVIER, J.-M. & PAUTOU, G. (1998), Hydropower reach by-passing and dewatering impacts in gravel-bed rivers. In *Gravel Bed Rivers in the Environment*, eds, P. C. Klingeman, R. Beschta, P. Komar, J. Bradley, pp. 313 – 344, Water Resources Publications, Littleton, Co.
- [163] FRUGET, J.-F., CENTOFANTI, M. & OLIVIER, J.-M. (1998). The fish fauna of the Doubs river prior to completion of the Rhine-Rhône connection. *Environmental Management*, 22, 129 – 144.
- [164] FRUGET, J.-F., CENTOFANTI, M., DESSAIX, J., OLIVIER, J.-M., DRUART, J. C. & MARTINEZ, P.-J. (2001). Temporal and spatial dynamics in large rivers: example of a long-term monitoring of the middle Rhône River. *International Journal of Limnology*, 37, 237 – 251.
- [165] CHESSEL, D., DUFOUR, A. B. & THIOULOUSE (2004). The ade4 package – I: One-table methods. *R News*, 4/1, 5 – 10.
- [166] PETTORELLI, N., GAILLARD, J.-M., YOCCOZ, N. G., DUNCAN, P., MAILLARD, D., DELORME, D., VAN LAERE, G. & TOÏGO, C. (2005). The response of fawn survival to changes in habitat quality varies according to cohort quality and spatial scale. *Journal of Animal Ecology*, 74, 972 – 981.

### 3.8 Joint Correspondence Analysis

The most common way of analysing more than two categorical variables is to transform the multi-way contingency table so that it is of a two-way form. The two most popular ways of doing this is to convert the multi-way table into an *indicator matrix* (of 0's and 1's for presence/absence of an observation in a particular category) or into its Burt matrix (a super-diagonal matrix comprising of two-way sub-matrices for each pair of variables and, along the diagonal of the Burt matrix, diagonal matrices summarising the marginal information of each variable). It must be pointed out that there exists an alternative approach to *multiple correspondence analysis* that may be used to visually analyse the association between the variables of a *multi-way contingency tables*. Such an approach is referred to as multi-way correspondence analysis and the interested reader is directed to, for example, Kroonenberg (2008) and [5, Chapters 10 & 11] for more details on this variant. There are key differences that make multiple correspondence

analysis and multi-way correspondence analysis distinctive. One such difference is that, unlike multiple correspondence analysis, multi-way correspondence analysis treats the data in its “cubed” form. That is, multi-way correspondence analysis considers not only bi- variate associations structures that may be present in the contingency table, it also considers higher order (tri- and multi- variate) association structures among variables, since it does not consider the classical singular value decomposition of two-way matrix, but its multi-way generalisation (Tucker, 1966).

One of the drawbacks of performing multiple correspondence analysis using the Burt table is that the quantity used to assess the association is inflated due to the presence of the matrices along the diagonal of the Burt matrix. To account for this

- [167] GREENACRE, M. J. (1988). Correspondence analysis of multivariate categorical data by weighted least squares. *Biometrika*, 75, 457 – 467.

proposed an iterative way of removing these diagonal matrices from the analysis thereby stabilising the association. Such an approach is referred to as *joint correspondence analysis* (or, simply, JCA). Additional drawbacks of performing multiple correspondence analysis were also described by

- [168] GREENACRE, M. J. (1990). Some limitations of multiple correspondence analysis. *Computational Statistics Quarterly*, 3, 249 – 256.

JCA was further described, and developed, by

- [169] BOIK, R. J. (1996). An efficient algorithm for joint correspondence analysis. *Psychometrika*, 61, 255 – 269.

- [170] TATENENI, K. & BROWNE, M. (2000). A noniterative method of joint correspondence analysis. *Psychometrika*, 65, 157 – 165.

- [171] GREENACRE, M. & BLASIUS, J. (eds.) (2006). *Multiple Correspondence Analysis and Related Methods*. Boca Raton: Chapman & Hall/CRC.

- [172] GREENACRE, M. (2010). Correspondence analysis. *WIREs Computational Statistics*, 2, 613 – 619.

- [173] CAMIZ, S. & GOMES G. C. (2013). Joint correspondence analysis versus multiple correspondence analysis: A solution to an undetected problem. In *Classification and Data Mining*, eds A. Giusti, G. Ritter, M. Vichi, pp. 11 – 18, Heidelberg: Springer.

- [174] CAMIZ, S. & GOMES, G. C. (2016). Alternative methods to multiple correspondence analysis in reconstructing the relevant information in a Burt’s table. *Pesquisa Operacional*, 36, 23 – 44.

JCA can be performed using the code outlined in the appendix of [171]:

- [175] NENADIĆ, O. & GREENACRE, M. (2006). Computation of multiple correspondence analysis, with code in R. In *Multiple Correspondence Analysis and Related Methods*, eds M. Greenacre, J. Blasius, pp. 523 – 551, Boca Raton: Chapman & Hall/CRC.

This variant can also be performed using the `mjca` function in (Nenadić and Greenacre, 2007) R package `ca`. For an update on the details on this package, see

- [176] GREENACRE, M. & NENADIĆ, O. (2017). `ca`: simple, multiple and joint correspondence analysis variants. R version 0.71. <https://cran.r-project.org/web/packages/ca/index.html>

### 3.9 Fuzzy Correspondence Analysis

As we described in Section 4.8, one way of performing a correspondence analysis of multiple categorical variables is to construct an indicator matrix with 0/1 elements. A value of 1 is in place for an individual that exhibits a categorical characteristic while a zero is assigned if this characteristic is not observed. However, sometimes it is not always clear whether a characteristic is observed or not and so fractional, or fuzzy coding, is performed. That is, rather than a 0 or 1 for the categories (so that for each variable, the total score is 1), a score of 0.8 for one category and 0.2 for another may be assigned. Such a strategy is very helpful for intervalised data where there is uncertainty as to which interval an observation belongs. While this is a broad view of the variant, performing correspondence analysis in this way is referred to as *fuzzy correspondence analysis* and is a special case of *multiple correspondence analysis*. There are several ways in which fuzzy correspondence analysis may be applied and one may refer to the following literature for examples of the application of this variant:

- [177] VAN RIJCKEVORSEL, J. (1987). *The Application of Horseshoes and Fuzzy Coding in Multiple Correspondence Analysis*. Leiden DSWO Press.
- [178] VAN RIJCKEVORSEL, J. (1988) Fuzzy coding and B-splines. In *Component and Correspondence Analysis*, eds J. L. A. van Rijckevorsel, J. de Leeuw, pp. 33 – 54, Chichester: Wiley.
- [179] MARTIN, J.-P. (1988). On probability coding. In *Component and Correspondence Analysis*, eds J. L. A. van Rijckevorsel, J. de Leeuw, pp. 103 – 114, Chichester: Wiley.
- [180] CHEVENET, F., DOLÉDEC, S. & CHESSEL, D. (1994). A fuzzy coding approach for the analysis of long-term ecological data. *Freshwater Biology*, 31, 295 – 309.
- [181] LOSLEVER, P. & BOUILLAND, S. (1999). Marriage of fuzzy sets and multiple correspondence analysis: Examples with subjective interval data and biomedical signals. *Fuzzy Sets and Systems*, 107, 255 – 275.

- [182] THEODOROU, Y. & ALEVIZOS, P. (2006). The fuzzy eigenvalue problem of fuzzy correspondence analysis. *Journal of Interdisciplinary Mathematics*, 9, 115 – 137.
- [183] THEODOROU, Y., DROSSOS, C. & ALEVIZOS, P. (2007). Correspondence analysis with fuzzy data: The fuzzy eigenvalue problem. *Fuzzy Sets and Systems*, 158, 704 – 721.
- [184] ALEVIZOS, P. D., THEODOROU, Y. & VRAHATIS, M. N. (2017). Correspondence analysis with grey data: The grey eigenvalue problem. *The Journal of Grey System*, 29, 92 – 105.
- [185] SERRA, S. R. Q., GRAÇA, M. A. S., DOLÉDEC, S. & FEIO, M. J. (2017). Discriminating permanent from temporary rivers with traits of chironomid genera. *International Journal of Limnology*, 53, 161 – 174.

Fuzzy correspondence analysis can be performed using the `dudi.fca` function in the R package `ade4`; see [143] for more details on this package.

### 3.10 Internal Correspondence Analysis

A variant of the classical correspondence analysis approach referred to as *internal correspondence analysis* was proposed by

- [186] CAZES, P., CHESSEL, D. & DOLEDUC, S. (1988). L'Analyse des correspondances internes dun tableau partitionné: son usage en hydrobiologie. *Review de Statistique Appliquée*, 36, 39 – 54.

This approach considers the case where, for an  $I \times J$  contingency table, the  $I$  rows are partitioned into  $L$  groups. Similarly, the  $J$  columns are partitioned into  $K$  groups. Thus, we have  $LK$  sub-tables each of size  $L_m \times K_n$  for  $m = 1, 2, \dots, L$  and  $n = 1, 2, \dots, K$ . For further developments, and applications, of how these sub-tables can be analysed using internal correspondence analysis refer to

- [187] CAZES, P. & MOREAU, J. (1991). Analysis of a contingency table in which the rows and the columns have a graph structure. In *Symbolic-Numeric Data Analysis and Learning*, eds E. Diday, Y. Lechevallier, pp. 271 – 280, New York: Nova Science Publishers.
- [188] LI, Y. X., YISHI, L. & ZHANPO, L. (1999). Internal correspondence analysis of gravimetric and magnetic data for the northern part of south–north tectonic zone. *Seismology and Geology*, 21, 370 – 376.
- [189] CAZES, P. & MOREAU, J. (2000). Analyse des correspondances d'un tableau de contingence don't les lignes et les colonnes sont munies d'une structure de graphes bistochastique. In *L'Analyse des Correspondances et les Techniques Connexes: Approches Nouvelles pour l'Analyse Statistique des Données*, eds J. Moreau, P. A. Doudin, P. Cazes, pp. 87 – 103, Berlin: Springer-Verlag.

- [190] LALOË, D., MOAZAMI-GOUDARZI, K., & CHESSEL, D. (2002). Contribution of individual markers to the analysis of the relationships among breeds by correspondence analysis. In Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, August 19 – 23, 2002, Communication 26 – 06, Montpellier, France.
- [191] LOBRY, J. R. & CHESSEL, D. (2003). Internal correspondence analysis of codon and amino-acid usage in thermophilic bacteria. *Journal of Applied Genetics*, 44(2), 235 – 261.
- [192] CHARIF, D., THIOULOUSE, J., LOBRY, J. R. & PERRIÈRE, G. (2005). Online synonymous codon usage analyses with the `ade4` and `seqinR` packages. *Bioinformatics*, 21, 545 – 547.
- [193] PAGÉS, J. & BÉCUE-BERTAUT, M. (2006). Multiple factor analysis for contingency tables. In *Multiple Correspondence Analysis and Related Methods*, eds M. Greenacre, J. Blasius, pp. 299 – 326, Boca Raton: Chapman & Hall/CRC.
- [194] SAMON, M., LOBRY, J. R. & DURET, L. (2006). No evidence for tissue specific adaptation of synonymous codon usage in humans. *Molecular Biology and Evolution*, 23, 523 – 529.
- [195] PARDO, C. E., BECUE-BÉRTAUT, M & ORTIZ, J. (2013). Correspondence analysis of contingency tables with subpartitions on rows and columns. *Revista Colombiana de Estadística*, 36, 115 – 144.
- [196] PAYNE, B. L. (2018). Differences in the protein evolutionary rates of Arabidopsis species and codon usage biases in the tissues of *Drosophila melanogaster*. Master's thesis, University of Nevada, USA.

Internal correspondence analysis can be performed using the `witwit.coa` function in the R package `ade4`; see [144] for more details on this package.

### 3.11 Grade Correspondence Analysis

Over the past 20 years a little known, and understood, variation of correspondence analysis has appeared, predominantly in the Polish literature. Originally proposed by

- [197] CIOK, A., KOWALCZYK, E., PLESZCYŃSKA, E. & SZCZESNY, W. (1995). Algorithms of grade correspondence cluster analysis. *The Collected Papers of Theoretical and Applied Computer Science*, 7, 5 – 22.

their approach, referred to as *grade correspondence analysis*, maximises the association between the variables using Spearman's correlation coefficient and not Pearson's statistic as is typically done. Therefore, rather than considering singular vectors from the SVD of a matrix (typically whose elements are standardised residuals of the cell proportions under the Poisson distribution) grade correspondence analysis considers instead “grade

regression variables”. The technique is performed by calculating these variables and iteratively permuting the rows and columns until the elements of Spearman’s correlation is maximised. Further technical and practical descriptions of this variant can be found in

- [198] CIOK, A. (1998). Discretization as a tool in cluster analysis. In *Advances in Data Science and Classification*, eds. A. Rizzi, M. Vichi, H. H. Bock, pp. 349 – 354, Berlin: Springer-Verlag.
- [199] MATYJA, O. & SZCZESNY, W. (2000). Visualization in prediction based on grade correspondence analysis. In *Proceedings of the IIS’2000 Symposium on Intelligent Information Systems*, eds M. A. Kłopotek, M. Michalewicz, S. T. Wierchon, pp. 289 – 301, Berlin: Physica-Verlag.
- [200] CIOK, A. (2001). Significance testing in the grade correspondence analysis. In *Proceedings of the International Symposium on “Intelligent Information Systems X”*, eds M. A. Kłopotek, M. Michalewicz, S. T. Wierchon, pp. 67 – 74, Berlin: Physica-Verlag.
- [201] SZCZESNY, W. & MATYJA, O. (2001). Using grade correspondence analysis to merge populations and detect latent orders. In *Proceedings of the International Symposium on “Intelligent Information Systems X”*, eds M. A. Kłopotek, M. Michalewicz, S. T. Wierchon, pp. 111 – 120, Berlin: Physica-Verlag.
- [202] CIOK, A. (2002). Grade correspondence-cluster analysis applied to separate components of reversely regular mixtures. In *Classification, Clustering and Data Analysis: Recent Advances and Applications*, eds K. Jajuga, A. Sokolowski, H. H. Bock, pp. 211 – 218, Berlin: Springer-Verlag.
- [203] SZCZESNY, W. (2002). Grade correspondence analysis applied to contingency tables and questionnaire data. *Intelligent Data Analysis*, 6, 17 – 51.
- [204] KOWALCZYK, T., PLESZCYŃSKA, E. & RULAND, F. (eds) (2004). *Grade Models and Methods for Data Analysis*. Berlin: Springer-Verlag.
- [205] GRZEGOREK, M. (2007). Homogeneity of pixel neighborhoods in gray levels images investigated by the grade correspondence analysis. In *Computer Recognition Systems 2: Advances in Soft Computing*, eds M. Kurzynski, E. Puchala, M. Wozniak, A. Zolnierok, pp. 76 – 83, Berlin: Springer-Verlag.
- [206] GAJOWNICZEK, K. & ZĄBKOWSKI, T. (2015). Data mining techniques for detecting household characteristics based on smart meter data. *Energies*, 8, 7407 – 7427.
- [207] WNOROWSKI, A., BILEK, M., STAWARCZYK, K., GOSTKOWSKI, M., OLSZEWSKI, M. WÓJCIAK-KOSIOR, M. & SOWA, I. (2017). Metabolic activity of tree saps of different origin towards cultured human cells in the light of grade

correspondence analysis and multiple regression modeling. *Acta Societatis Botanicorum Poloniae*, 86(2), 1 – 10.

### 3.12 Weber Correspondence Analysis

The traditional approach to scaling categorical data, and hence performing correspondence analysis, involves finding row and column scores so that the weighted sum-of-squares of the Euclidean distance between these points in a correspondence plot is minimised. Rather than considering the weighted sum-of-squares of these distances, one can consider instead their weighted *sum* as discussed by

- [208] HEISER, W. (1987). Correspondence analysis with least absolute residuals. *Computational Statistics and Data Analysis*, 5, 337 – 356.

which he described it as the reciprocal location problem; the term reciprocal is used to reflect the interrelationship between the row and column points (often described through transition formula), while location is used to reflect the characteristics of some monotonically increasing function of the Euclidean distances. This latter term is consistent with the 1909 Weber problem (see Weber, 1929, for an English translation). More on this variant to correspondence analysis can be found by referring to

- [209] DE LEEUW, J. & MICHAÏLIDIS, G. (2004). Weber correspondence analysis: the one-dimensional case. *Journal of Computational and Graphical Statistics*, 13, 946–953.

which the authors refer to it as *Weber correspondence analysis*.

### 3.13 Taxicab Correspondence Analysis

To perform the dimension reduction required in correspondence analysis, typically singular value decomposition (SVD) is the preferred option. An alternative method of decomposition is to use taxicab singular value decomposition (TSVD) which involves minimising the Manhattan, or equivalently  $L_1$  norm, distance between two points rather than their Euclidean distance. Its implementation to correspondence analysis was first made by

- [210] CHOULAKIAN, V. (2006). Taxicab correspondence analysis. *Psychometrika*, 71, 333 – 345.

who termed it *taxicab correspondence analysis*. Refinements of this variant of correspondence analysis include those of

- [211] CHOULAKIAN, V. (2008). Taxicab correspondence analysis of contingency tables with one heavyweight column. *Psychometrika*, 73 309 – 319.

- [212] CHOULAKIAN, V. & DE TIBEIRO J. (2013). Graph partitioning by correspondence analysis and taxicab correspondence analysis. *Journal of Classification*, 30, 397 – 427.
- [213] CHOULAKIAN, V. (2014). Taxicab correspondence analysis of ratings and rankings. *Journal de la Société Française de Statistique*, 155(4), 1 – 23.
- [214] CAMIZ, S. & GOMES, G. C. (2016). A comparison between simple and taxicab correspondence analysis and examples. In *Simpósio Brasileiro de Pesquisa Operacional*, pp. 1041 – 1050.

The development of taxicab correspondence analysis has seen it evolve to other variants. For example, *taxicab non-symmetrical correspondence analysis* has been discussed in

- [215] SIMONETTI, B. (2008). Taxicab non symmetrical correspondence analysis. In *MTISD2008: Methods, Models and Information Technologies for Decision Support Systems*, eds. E. Ciavolino, L. D'Ambra, M. Squillante, G. Ghiani, pp. 257 – 260, University of Salento, Italy.
- [216] CHOULAKIAN, V. & DE TIBEIRO, J. (2013). Graph partitioning by correspondence analysis and taxicab correspondence analysis. *Journal of Classification*, 30, 397 – 427.
- [217] CHOULAKIAN, V., SIMONETTI, B. & GIA, T. (2014). Some new aspects of taxicab correspondence analysis. *Statistical Methods and Applications*, 23, 401 – 415.

while *multiple taxicab correspondence analysis* was described by

- [218] CHOULAKIAN, V. (2006). Multiple taxicab correspondence analysis. In *COMPSTAT2006: Proceedings in Computational Statistics*, eds A. Rizzi, M. Vichi, pp. 557 – 564, Heidelberg: Physica-Verlag.
- [219] CHOULAKIAN, V. (2008). Multiple taxicab correspondence analysis. *Advances in Data Analysis and Classification*, 2, 177–206.
- [220] CHOULAKIAN, V., ALLARD, J. & SIMONETTI, B. (2013). Multiple taxicab correspondence analysis of a survey related to health services. *Journal of Data Science*, 11, 205 – 229.

Some applications of these variants include

- [221] CHOULAKIAN, V., KASPARIAN, S., MIYAKE, M., AKAMA, H., MAKOSHI, N. & NAKAGAWA, M. (2006). A statistical analysis of the synoptic gospels. In *JADT 2006: 8es Journées internationales d'Analyse statistique des Données Textuelles*, eds J.-M. Vipey, C. Condé, pp. 281 – 288, Besançon, Presses Universitaires de Franche-Comté.

- [222] SIMONETTI, B. & LUCADAMO, A. (2013). Taxicab non symmetrical correspondence analysis for the evaluation of the passenger satisfaction. In *Advanced Dynamic Modeling of Economic and Social Systems*, eds A. N. Proto, M. Squillante, J. Kacprzyk, pp. 175 – 184. Berlin: Springer-Verlag.
- [223] MALLET GAUTHIER, S. & CHOULAKIAN, V. (2015). Taxicab correspondence analysis of abundance data in archaeology: Three case studies revisited. *Acheologia e Calcolatori*, 26, 77 – 94.

Taxicab correspondence analysis can be performed using the `tca` function in the R package `TaxicabCA`. For details, see

- [224] ALLARD, J. & CHOULAKIAN, V. (2018). `TaxicabCA`: Taxicab correspondence analysis. R version 0.1.0, <https://cran.r-project.org/web/packages/TaxicabCA/index.html>

### 3.14 Focused Correspondence Analysis

*Focused correspondence analysis* is a variation of correspondence analysis that appears very obscure and so is worth a brief mention. A description of the technique is outlined in

- [225] ROSARIO, G. E., RUNDENSTEINER, E. A., BROWN, D. C. & WARD, M. O. (2003). Mapping nominal values to numbers for effective visualization. In *Proceedings of the IEEE Symposium on Information Visualization (INFVIS2003)*, pp. 113 – 120, Seattle, WA.

Several presentations can be found on the internet that provide an overview of this approach. For example, one may refer to the presentation from the INFOVIS2003 conference:

- [226] WARD, M. O. (2003). Mapping nominal values to numbers for effective visualization. <https://slideplayer.com/slide/4659413/>, (last accessed August 21, 2018).

In the notes to these slides, it is mentioned that focused correspondence analysis is performed by analysing one variable at a time and “using only the variables that are most associated with the target variable”. Mention is even made to *cascaded focused correspondence analysis* where “the results of the initial focused correspondence analysis can be used in succeeding analyses”

### 3.15 Over-dispersion in Correspondence Analysis

The traditional approach to correspondence analysis involves quantifying the association using Pearson’s chi-squared statistic. Partitioning this statistic is how a visual interpretation of the association is then produced. However, using this statistic assumes that the cell frequencies of the contingency table being analysed are Poisson random variables so that the expectation and variance are equivalent. Agresti (2002) and Haberman (1973)

demonstrate that the variance is always at least as big as the expectation – in which case the data in the contingency table exhibit over-dispersion. There are several strategies that may be considered to overcome this problem including that of

- [227] BEH, E. J. (2012). Simple correspondence analysis using adjusted residuals. *Journal of Statistical Planning and Inference*, 142, 965–973.

Another strategy is to make use of variance stabilising transformations. One such transformation was proposed by Freeman and Tukey (1950) and the resulting variant of correspondence analysis is called “*Freeman-Tukey*” *correspondence analysis* (or, simply, FTCA). While [5, Section 9.3] provided a very brief (skeletal) introduction of this variant,

- [228] BEH, E. J., LOMBARDO, R. & ALBERTI, G. (2018). Correspondence analysis and the Freeman-Tukey statistic: A study of archaeological data. *Computational Statistics and Data Analysis*, 128, 73 – 86.

described in far greater detail (including its links with association and correspondence models) the variant and explored its application to large, and sparse archaeological data. While such an approach deals with the instability in the variance, it also has the advantage that two row (or column) points in a low-dimensional space are separated by the Hellinger distance (rather than the Euclidean distance), a feature advocated in the correspondence analysis literature by

- [229] DOMENGENES, D. & VOLLE, M. (1979). Analyse factorielle sphérique: une exploration. *Annales de l’Inséé*, 35, 3–84.
- [230] RAO, C. R. (1995). A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance. *Qüestiió*, 19, 23 – 63.

### 3.16 Inverse Correspondence Analysis

Another interesting variation of correspondence analysis was proposed by

- [231] GROENEN, P.J.F. & VAN DE VELDEN, M. (2004). Inverse correspondence analysis. *Linear Algebra and its Applications*, 388, 221 – 238.
- [232] VAN DE VELDEN, M., VAN DEN HEUVEL, W., GALY, H. & GROENEN, P. J. F. (2019). Retrieving a contingency table from a correspondence analysis solution. (submitted).

As their title suggests, their approach is referred to as *inverse correspondence analysis*. They considered the problem of identifying the set of possible contingency tables that can be obtained from a low-dimensional plot. Thus, their variation involves identifying all possible contingency tables from a plot, rather than the more classical position of obtaining a unique plot from a given contingency table.

### 3.17 Parametric Correspondence Analysis

A generalised approach to correspondence analysis, called *parametric correspondence analysis*, was proposed by

- [233] CUADRAS, C. M. & CUADRAS, D. (2006). A parametric approach to correspondence analysis. *Linear Algebra and Its Applications*, 417, 64 – 74.

Under the classical approach to correspondence analysis, the squared distance between two row (or column) points in a correspondence plot is Euclidean. However there have been proponents for the use of the Hellinger distance [228, 229, 230]. So, to accommodate the use of Hellinger distances into the classical approach, a generalisation made by [233] was proposed by introducing a parameter whose choice of value would determine the nature of the distance being depicted.

### 3.18 Subset Correspondence Analysis

Another variant of correspondence analysis is that of *subset correspondence analysis*, proposed by

- [234] GREENACRE, M. & PARDO, R. (2006). Subset correspondence analysis: Visualizing relationships among a selected set of response categories from a questionnaire survey. *Sociological Methods & Research*, 35, 193 – 218.

This approach is applicable where there are multiple categorical variables and the analyst is interested in focusing on a selection, or “subset”, of them. Further details, and applications, on this variant can be found in

- [235] GREENACRE, M. & PARDO, R. (2006). Multiple correspondence analysis of subsets of response categories. In *Multiple Correspondence Analysis and Related Methods*, eds. M. Greenacre, J. Blasius, pp. 197 – 217, London: Chapman & Hall/CRC.
- [236] GREENACRE, M. (2009). Power transformations in correspondence analysis. *Computational Statistics and Data Analysis*, 53, 3107 – 3116.
- [237] GREENACRE, M. & LEWI, P. (2009). Distributional equivalence and subcompositional coherence in the analysis of compositional data, contingency tables and ratioscale measurements. *Journal of Classification*, 26, 29 – 54.
- [238] GREENACRE, M. (2010). Logratio analysis is a limiting case of correspondence analysis. *Mathematical Geosciences*, 42, 129–134.
- [238] HENDRY, G., NORTH, D., ZEWOTIR, T. & NAIDOO, R. N. (2014). The application of subset correspondence analysis to address the problem of missing data in a study on asthma severity in childhood. *Statistics in Medicine*, 33, 3882 – 3893.

- [240] DRAMALIDIS, A. & MARKOS, A. (2016). Subset multiple correspondence analysis as a tool for visualizing affiliation networks. *Journal of Data Analysis and Information Processing*, 4, 81 – 89.
- [241] HENDRY, G. M., ZEWOTIR, T., NAIDOO, R. N. & NORTH, D. (2017). The effect of the mechanism and amount of missingness on subset correspondence analysis. *Communications in Statistics (Simulation and Computation)*, 46, 7100 – 7115.
- [242] HENDRY, G. M., ZEWOTIR, T., NAIDOO, R. N. & NORTH, D. (2017). A comparative study of multiple imputation and subset correspondence analysis in dealing with missing data. *South African Statistical Journal*, 51, 183 – 193.
- [243] KORNELIUSSEN, T. & GREENACRE, M. (2018). Information sources used by European tourists: A cross-national study. *Journal of Travel Research*, 57, 193 – 205.

Subset correspondence analysis can be performed using the R package `ca`; see [175].

### 3.19 Symbolic Correspondence Analysis

Another more recent, and obscure, member of the family of correspondence analysis techniques is *symbolic correspondence analysis*. This variant was originally presented at CARME2007 in Rotterdam, the Netherlands by Oldemar Rodríguez (University of Costa Rica)

- [244] RODRÍGUEZ, O. (2007). Correspondence analysis for symbolic multi-valued variables. 12 pages, (last accessed August 21, 2018). [www.oldemarrodriguez.com/yahoo\\_site\\_admin/assets/docs / SymCA\\_CARME2007.229151706.pdf](http://www.oldemarrodriguez.com/yahoo_site_admin/assets/docs/SymCA_CARME2007.229151706.pdf),

and was later discussed by Rodríguez at the XVIII SIMMAC (Simposio Internacional de Métodos Matemáticos Aplicados a las Ciencias) hosted by the University of Costa Rica in 2012; the presentation can be accessed at the URL <http://www.youblisher.com/p/264843-Correspondence-Analysis-for-Symbolic-Multi-Valued-Variables/>. The developments made in symbolic correspondence analysis were also elaborated upon by

- [245] TAKAGI, I. & YADOHISA, H. (2011). Correspondence analysis for symbolic contingency tables based on interval algebra. *Procedia Computer Science*, 6, 352 – 357.

Broadly speaking, this variant is performed on a contingency table where, instead of analysing cell frequencies, intervals of possible values for each cell are considered. Note that such a table is not strictly a contingency table but was referred to as an interval contingency table by [244] and [245], while [245] referred to this variant of correspondence analysis as *interval algebraic correspondence analysis*. See also

- [246] KAO, C.-H., NAKANO, J., SHIEH, S.-H., TIEN, Y.-J., WU, H.-M., YANG, C.-K. & CHEN, C.-H. (2014). Exploratory data analysis of interval-valued symbolic data with matrix visualization. *Computational Statistics and Data Analysis*, 79, 14 – 29.

for more details on this variant. Symbolic correspondence can be performed using the `sym.cfa` function in the R package `RSDA`:

- [247] DE DIOS MURILLO, J., RODRÍGUEZ, O. & VILLALOBOS, J. (2012). `RSDA`: An R package for symbolic data analysis, <http://www.youblisher.com/p/264840-RSDA-An-R-package-for-Symbolic-Data-Analysis/>, last accessed August 21, 2018.
- [248] RODRIGUEZ, O. (2018). `RSDA`: R to symbolic data analysis. R version 2.0.5 <https://cran.r-project.org/web/packages/RSDA/index.html>

### 3.20 Discriminant Correspondence Analysis

While the aim of correspondence analysis is to visualise the association between categorical variables, a variant called *discriminant correspondence analysis* aims to cluster the categories into pre-defined groups. These groups may be identified by the label of the variable and the groups consist of those categories that define the variable. A technical, and practical, examination of this variant can be found by referring to

- [249] PERRIÈRE, G., LOBRY, J. R. & THIOULOUSE, J. (1996). Correspondence discriminant analysis: a multivariate method for comparing classes of protein and nucleic acids sequences. *Computer Applications in the Biosciences*, 12, 519 – 524.
- [250] ABDI, H. (2007). Discriminant correspondence analysis. In *Encyclopedia of Measurement and Statistics*, ed. N. J. Salkind, pp. 270 – 275, Thousand Oaks, CA: Sage.
- [251] WANG, J. Y., BAKHADIROV, K., ABDI, H., DEVOUS SR, M. D., MARQUEZ DE LA PLATA, C. D., MORE, C., MADDEN, C. J. & DIAZ-ARRASTIA, R. (2011). Longitudinal changes of structural connectivity in traumatic axonal injury. *Neurology*, 77, 818 – 826.
- [252] BEATON, D., CHIN FATT, C. R. & ABDI, H. (2014). An ExPosition of multivariate analysis with the singular decomposition in R. *Computational Statistics and Data Analysis*, 72, 176 – 189.

Once a correspondence analysis is performed on the groups, discriminant correspondence analysis treats the grouped categories as supplementary points. Discriminant correspondence analysis can be performed using the R function `tepDICA` in the `TExPosition` function. It can also be performed using `fast.dca` in the `svs` package. For details, see

- [253] BEATON, D., CHIN-FATT, C. R. & ABDI, H. (2013). ExPosition: Exploratory analysis with the singular value decomposition. R version 2.8.19, <https://cran.r-project.org/web/packages/ExPosition/index.html>

- [254] BEATON, D., RIECK, J., CHIN-FATT, C. R. & ABDI, H. (2013). TExPosition: Two-table ExPosition. R version 2.6.10, <https://cran.r-project.org/web/packages/TExPosition/index.html>

### 3.21 Regularised Multiple Correspondence Analysis

Multiple correspondence analysis, like its two-way version of simple correspondence analysis, is commonly used to visualise the association between two or more variables. Therefore, its focus is rarely concerned with the inferential aspects of data analysis. Despite this, bootstrap solutions have been proposed (which we will not get into here). One variant of multiple correspondence analysis that provides inferences on the nature of the association is *regularised multiple correspondence analysis*, a technique proposed by

- [255] TAKANE, Y. & HWANG, H. (2006). Regularized multiple correspondence analysis. In *Multiple Correspondence Analysis and Related Methods*, eds. M. Greenacre, J. Blasius, pp. 259 – 279, Boca Raton: Chapman & Hall/CRC.

This variant maximises the association between multiple categorical variables by incorporating a diagonal matrix consisting of a ridge parameter that is then added to the Burt matrix prior to the application of singular value decomposition; see, for example, [254; Equation (11.11)]. Such an approach is conducted through bootstrap resampling and further information on it can be found in

- [256] JOSSE, J., CHAVENT, M., LIQUET, B. & HUSSON, F. (2012). Handling missing values with regularized iterative multiple correspondence analysis. *Journal of Classification*, 29, 91 – 116.
- [257] NIENKEMPER, J. (2013). Regularised Iterative Multiple Correspondence Analysis in Multiple Imputation. Master's Thesis, University of the Free State, Bloemfontein, South Africa.
- [258] JOSSE, J. & WAGER, S. (2016). Bootstrap-based regularization for low-rank matrix estimation. *Journal of Machine Learning Research*, 17(124), 1 – 29.

A *regularised non-symmetrical correspondence analysis* variant was proposed by

- [259] TAKANE, Y. & JUNG, S. (2009). Regularized nonsymmetric correspondence analysis. *Computational Statistics and Data Analysis*, 53, 3159 – 3170.

Regularised correspondence can be performed in R using the package `denoiseR`. For details, see

- [260] JOSSE, J., SARDY, S. & WAGER, S. (2016). `denoiseR`: Regularized low rank matrix estimation. R version 1.0, <https://cran.r-project.org/web/packages/denoiseR/index.html>
- [261] JOSSE, J., SARDY, S. & WAGER, S. (2016). `denoiseR`: A package for low rank matrix estimation. Available on arXiv:1602.01206v1.

### 3.22 Log-Ratio Correspondence Analysis

Rather than performing a correspondence analysis on a contingency table by considering the difference between the observed frequencies and their expected values (typically, under independence), an alternative approach is to utilise the Box-Cox transformation. In doing so, one can effectively analyse the logarithm of the cell frequencies and perform a *log-ratio correspondence analysis* first proposed by

- [262] MEULMAN, J. J. & HEISER, W. J. (1998). Visual display of interaction in multiway contingency tables by use of homogeneity analysis: the 2x2x2x2 case. In J. Blasius & M. J. Greenacre (Eds.), *Visualization of categorical data* (pp. 277–296). New York: Academic Press.
- [263] DE ROOIJ, M. & HEISER, W. J. (2005). Graphical representations and odds ratios in a distance association model for the analysis of cross-classified data. *Psychometrika* 70, 99–123.
- [264] DE ROOIJ, M. & ANDERSON, C.J. (2007). Visualizing, summarizing, and comparing odds ratio structures. *Methodology* 3, 139–148.
- [265] GREENACRE, M. & LEWI, P. (2009). Distributional equivalence and subcompositional coherence in the analysis of compositional data, contingency tables and ratio-scale measurements. *Journal of Classification*, 26, 29–54.

and subsequently discussed in

- [266] GREENACRE, M. (2009). Power transformations in correspondence analysis. *Computational Statistics and Data Analysis*, 53, 3107–3116.
- [267] GREENACRE, M. (2010). Log-ratio analysis is a limiting case of correspondence analysis. *Mathematical Geosciences*, 42, 129–134.
- [268] CAMMINATIELLO, I., D’AMBRA, A. & SARNACCHIARO, P. (2014). The association in a two-way contingency table through log odds ratio analysis: the case of Sarno river pollution. *Springer Plus*, 3, 384–492.
- [269] SARNACCHIARO, P., D’AMBRA, L. & CAMMINATIELLO, I. (2015). Measures of Association and Visualization of Log Odds Ratio Structure for a Two Way Contingency Table. *Australian & New Zealand Journal of Statistics*, 57, 363–376.

More on the link between log-ratio analysis and correspondence analysis was described by [232].

### 3.23 Co-Correspondence Analysis

Another variant of correspondence analysis that has been adapted for the ecological disciplines is *co-correspondence analysis* put forward by

- [270] TER BRAAK, C. J. F. & SCHAFFERS, A. P. (2004). Co-correspondence analysis: A new ordination method to relate two community compositions. *Ecology*, 85, 834 – 846.

With its development of this variant being made in the context of research undertaken in the ecological disciplines research, this area of research has some application appearing in the literature, including

- [271] SCHAFFERS, A. P., RAEMAKERS, I. P., SÝKORA, K. V. & TER BRAAK, C. J. F. (2008). Arthropod assemblages are best predicted by plant species composition. *Ecology*, 89, 782 – 794.
- [272] M’ULLER, J., STADLER, J. & BRANDL, R. (2010). Composition versus physiognomy of vegetation as predictors of bird assemblages: The role of lidar. *Remote Sensing Environment*, 114, 490 – 495.
- [273] CHOW, C.-E. T., KIM, D. Y., SACHDEVA, R., CARON, D. A. & FUHRMAN, J. A. (2014). Top-down controls on bacterial community structure: microbial network analysis of bacteria, T4-like viruses and protists. *The ISME Journal*, 8, 816 – 829.

Co-correspondence analysis can be performed in R using the `cocorresp` function in the recently uploaded 2016 `cocorresp` package and depends on using at least version 2.2-0 of `vegan`. For more details on `vegan` see [105, 106, 107, 121] while

- [274] SIMPSON, G. L. (2009). `cocorresp`: Co-correspondence analysis ordination methods. R version 0.1-9. <https://cran.r-project.org/web/packages/cocorresp/index.html>

describes the `cocorresp` package.

### 3.24 Partial Least-Squares Correspondence Analysis

A variant of correspondence analysis called *partial least-squares correspondence analysis* (PLSCA) was proposed by

- [275] BEATON, D., FILBEY, F. & ABDI, H. (2013). Integrating partial least squares correlation and correspondence analysis for nominal data. In *New Perspectives in Partial Least Squares and Related Methods*, eds H. Abdi, W. Chin, V. Esposito Vinzi, G. Russolillo, L. Trinchera, pp. 81 – 94, New York: Springer Verlag.

PLSCA considers the analysis of two categorical variables coded in disjunctive form that describe the same set of I observations with J and K category, respectively. Like simple correspondence analysis, PLSCA analyses the contingency table related to the two variables, although it does so differently. Broadly speaking, this difference may be simply described as follows. Correspondence analysis centres the contingency matrix by subtracting from the observed counts their expected value under independence. PLSCA

does not centre the contingency table but when computing the contingency table, it normalises the two indicator (disjunctive) matrices by centring them at their mean and dividing by their variance (in much the same way as is done with principal component analysis for numerical data). PLSCA then performs an SVD the contingency table computed by the product of these normalised indicator matrices. PLSCA can be seen as an extension of CA and also as a generalisation of the partial least squares method for analysing correlation (Krishnam et al., 2011). An advantage of this variant is that it has inferential characteristics that make use of bootstrap, permutation and chi-squared tests. More information on this variant may be found by referring to

- [276] BEATON, D., DUNLOP, J. & ABDI, H. (2016). Partial least squares correspondence analysis: A framework to simultaneously analyze behavioural and genetic data. *Psychological Methods*, 21, 621 – 651.

Partial least-squares correspondence analysis can be performed using the R function `tepPLSCA` in the `TEPosition` function; see [254] for more details on this package.

An earlier version of a variant referred to as partial least squares correspondence analysis appeared in the chemometrics literature. See, for example

- [277] FRISVAD, J. C. & NORSKER, M. (1996). Use of correspondence analysis partial least squares on linear and unimodal data. *Journal of Chemometrics*, 10, 677 – 685.

who invited the reader to consider

- [278] TER BRAAK, C. J. F. (1995). Non-linear methods for multivariate statistical calibration and their use in palaeoecology: a comparison of inverse (k-nearest neighbours, partial least squares and weighted averaging partial least squares) and classical approaches. *Chemometrics and Intelligent Laboratory Systems*, 28, 165 – 180.
- [279] TER BRAAK, C. J. F. & JUGGINS, S. (1993). Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia*, 269, 485 – 502.

for a mathematical description of the technique.

### 3.25 Canonical Non-Symmetrical Correspondence Analysis

A variant of canonical correspondence analysis (described in Section 3.2) that takes into account the asymmetric association structure between categorical variables was proposed by

- [280] WILLEMS, P. & GALINDO-VILLARDÓN, M. P. (2004). Canonical non symmetrical correspondence analysis: An alternative in direct gradient analysis. In *Proceedings of the Salamanca Statistics Seminar V: Advances in Descriptive Multivariate Analysis*, pp. 137 – 140, Departamento de Estadística, Universidad de Salamanca, Spain,

which, as the title of their paper suggests, is referred to as *canonical non-symmetrical correspondence analysis*. Like non-symmetrical correspondence analysis (described in Section 3.4), this variant assumes that the expected cell frequencies are modelled by a “Gaussian” log-linear model in terms of the row and column scores (to be estimated). The difference between this approach and the one described in Section 4.2 is that canonical correspondence analysis uses asymmetric measures of association (like NSCA does) to quantify the magnitude of the association instead of using Pearson’s chi-squared statistic. One may consider the following references for a technical, computation and applied description of canonical non-symmetrical correspondence analysis:

- [281] WILLEMS, P. M. & GALINDO-VILLARDÓN, M. P. (2008). Canonical non-symmetrical correspondence analysis: An alternative in constrained ordination. *SORT*, 32, 93 – 112.
- [282] LIBRERO, A. B., WILLEMS, P. M. & GALINDO-VILLARDÓN, M. P. (2015). *cncaGUI: Canonical non-symmetrical correspondence analysis in R*. R version 1.0, <https://cran.r-project.org/web/packages/cncaGUI/index.html>.

### 3.26 Generalised Correspondence Analysis

With the increasing number of variations of correspondence analysis being developed over the past few decades, attempts have been made to provide a general framework which includes many of them as special cases. For example

- [283] BEH, E. J. & LOMBARDO, R. (2018). An algebraic generalisation of some variants of simple correspondence analysis. *Metrika*, 81, 423 – 443.

provided an algebraic generalisation that included the traditional approach, non-symmetrical correspondence analysis and “Freeman-Tukey” correspondence analysis [227] as special cases. Their generalisation also incorporates a variety of different plotting options including those of [71, 76, 80] and fits into a framework that allows for ordinal and/or nominal categorical variables to be analysed.

Other methods for generalising correspondence analysis have been proposed. For example,

- [284] YANAI, H. (1986). Some generalizations of correspondence analysis in terms of projection operators. In *Data analysis and Informatics IV*, eds E. Diday, Y. Escofier, L. Lebart, J. P. Pagés, Y. Schektman, R. Thomassone, pp 193 – 207, Amsterdam: North-Holland.
- [285] CHOULAKIAN, V. (1988). Exploratory analysis of contingency tables by loglinear formulation and generalizations of correspondence analysis. *Psychometrika*, 53, 235–250.
- [286] DE LEEUW, J. (1993). Some generalizations of correspondence analysis. In *Multivariate analysis: Future Directions 2*, eds C. M. Cuadras, C. R. Rao, pp. 359 – 375, Amsterdam: North-Holland.

- [287] VAN DE VELDEN, M. & KIERS, H. A. L. (2005). Rotation in correspondence analysis. *Journal of Classification*, 22, 251–271.
- [288] LORENZO-SEVA, U., VAN DE VELDEN, M. & KIERS, H. A. L. (2009). Oblique rotation in correspondence analysis: a step forward in the search for the simplest interpretation. *British Journal of Mathematical and Statistical Psychology*, 62, 583 – 600.

focus on showing how viewing correspondence analysis from a variety of perspectives leads to similar row/column scoring solutions.

### 3.27 Ordered Correspondence Analysis

Much of the attention given to the development of correspondence analysis, including those variants described above, have been tailored to variables consisting of nominal categories. However, in many practical situations it may be clear, or desirable, for the analyst to collect data based on categories that are ordered. To obtain row/column scores for ordered categorical variables, one may consider any of the following in the literature

- [289] NISHISATO, S. & ARRI, P. S. (1975). Nonlinear programming approach to optimal scaling of partially ordered categories. *Psychometrika*, 40, 525 – 548.
- [290] SCHRIEVER, B. F. (1983). Scaling of order dependent categorical variables with correspondence analysis. *International Statistical Review*, 51, 225 – 238.
- [291] PARSA, A. R. & SMITH, B. S. (1993). Scoring under ordered constraints in contingency tables. *Communications in Statistics (Theory and Methods)*, 22, 3537 – 3551.
- [292] RITOV, Y. & GILULA, Z. (1993). Analysis of contingency tables by correspondence models subject to ordered constraints. *Journal of the American Statistical Association*, 88, 1380 – 1387.
- [293] YANG, K.-S. & HUH, M.-H. (1999). Correspondence analysis of two-way contingency tables with ordered column categories. *Journal of the Korean Statistical Society*, 28, 347 – 358.

These approaches generally enforce, along a chosen axis of a correspondence plot that the order structure of the categories be reflected by the ordered position along that axis. Usually the constraint is imposed upon the first axis so that the order of the categories is visually summarised from left to right on the correspondence plot. Such an imposition of the coordinates relating to the ordered points may not necessarily reflect the true nature of the association. An alternative strategy is to score the categories using *orthogonal polynomials* and obtain *generalised correlations* rather than singular values. Such a variant of *ordered correspondence analysis*, based on the partition of Pearson's chi-squared statistic by Lancaster (1953), was originally made for the analysis of two symmetrically associated categorical variables by

- [294] BEH, E. J. (1997). Simple correspondence analysis of ordinal cross-classifications using orthogonal polynomials. *Biometrical Journal*, 39, 589 – 613.

This variant has subsequently been further refined, applied and extended for non-symmetrical correspondence analysis and the analysis of multiple variables by [5; Chapters 6, 7 & Section 10.8] and

- [295] BEH, E. J. (1998). A comparative study of scores for correspondence analysis with ordered categories. *Biometrical Journal*, 40, 413-429.
- [296] D'AMBRA, L., BEH, E. J. & AMENTA, P. (2005). CATANOVA for two-way contingency tables with an ordinal response using orthogonal polynomials. *Communications in Statistics (Theory and Methods)*, 34, 1755 – 1769.
- [297] BEH, E. J., SIMONETTI B. & D'AMBRA L. (2005). Three ways ordinal NSCA analysis of patient satisfaction evaluation. *Statistica & Applicazioni*, 3, 21 – 28.
- [298] D'AMBRA, L., SIMONETTI, B. & BEH, E. J. (2006). A dimensional reduction method for ordinal three-way contingency tables. In *COMPSTAT 2006: Proceedings in Computational Statistics*, eds A. Rizzi, M. Vichi, pp. 271 – 284, Berlin: Springer-Verlag.
- [299] LOMBARDO, R., BEH, E. J. & D'AMBRA, L. (2007). Non-symmetric correspondence analysis with ordinal variables. *Computational Statistics and Data Analysis*, 52, 566 – 577.
- [300] CORBELLINI, A., RIANI, M. & DONATINI, A. (2008). Multivariate data analysis techniques to detect early warnings of elderly frailty. *Statistica Applicata*, 20, 159 – 178.
- [301] SIMONETTI, B., BEH, E. J. & D'AMBRA, L. (2010). The analysis of dependence for three way contingency tables with ordinal variables: A case study of patient satisfaction. *Journal of Applied Statistics*, 37, 91 – 103.
- [302] LOMBARDO, R. & BEH, E. J. (2010). Simple and multiple correspondence analysis for ordinal-scale variables using orthogonal polynomials. *Journal of Applied Statistics*, 37, 2101 – 2116.
- [303] LOMBARDO, R. & MEULMAN, J.J. (2010). Multiple correspondence analysis via polynomial transformations of ordered categorical variables. *Journal of Classification*, 27, 191 – 210.
- [304] LOMBARDO, R., BEH, E. J. & D'AMBRA, A. (2011). Studying the dependence between ordinal-nominal categorical variables via orthogonal polynomials. *Journal of Applied Statistics*, 38, 2119 – 2132.
- [305] MANTÉ, C., BERNARD, G., BONHOMME, P. & NERINI, D. (2013). Application of ordinal correspondence analysis for submerged aquatic vegetation monitoring. *Journal of Applied Statistics*, 40, 1619 – 1638.

- [306] LOMBARDO, R., BEH, E. J. & KROONENBERG, P. M. (2016). Modelling trends in ordered correspondence analysis using orthogonal polynomials. *Psychometrika*, 81, 325 – 349.
- [307] LOMBARDO, R., KROONENBERG, P. M. & BEH, E. J. (2016). Modelling trends in ordered three-way non-symmetrical correspondence analysis. Proceedings of the 48th Scientific Meeting of the Italian Statistical Society, eds. Pratesi, M., C. Perna, 8 pages, University of Salerno.
- [308] D'AMBRA, A., AMENTA P. & CRISCI A. (2017). Decomposition of the main effects and interaction term by using orthogonal polynomials in multiple non symmetrical correspondence analysis. *Communications in Statistics (Theory and Methods)*, 46, 10179 – 10188.
- [309] LOMBARDO, R. & BEH, E.J. (2017). Three-way correspondence analysis for ordinal-nominal variables. In Proceedings of the Conference of the Italian Statistical Society, eds. A. Petrucci, R. Verde, pp. 613 – 620, Firenze University Press.
- [310] MCGETRICK, J. A., BUBELA, T. & HIK, D. S. (2017). Automated content analysis as a tool for research and practice: a case illustration from the Prairie Creek and Nico environmental assessments in the Northwest Territories, Canada. *Impact Assessment and Project Appraisal*, 35, 139 – 147.
- [311] LOMBARDO, R., CAMMINATIELLO, I. & BEH, E. J. (2018). Assessing satisfaction with public transport service by ordered multiple correspondence analysis. *Social Indicators Research* (in press).

In R, this variant of correspondence analysis can be performed using the functions outlined in [5, 142, 143] and are loosely based on the SPLUS functions described in

- [312] BEH, E. J. (2004). S-PLUS code for ordinal correspondence analysis. *Computational Statistics*, 19, 593 – 612.

Another approach to performing correspondence analysis on a contingency table consisting of ordered categories is using the cumulative chi-squared statistic of Taguchi (1966). For a description of this ordered variant refer to

- [313] D'AMBRA, L., KÓKSOY, O. & SIMONETTI, B. (2009). Cumulative correspondence analysis of ordered categorical data from industrial experiments. *Journal of Applied Statistics*, 36, 1315 – 1328.
- [314] BEH, E. J., D'AMBRA, L. & SIMONETTI, B. (2011). Correspondence analysis of cumulative frequencies using a decomposition of Taguchi's statistic. *Communications in Statistics (Theory and Methods)*, 40, 1620 – 1632.

- [315] SARNACCHIARO, P. & D'AMBRA, A. (2011) Cumulative correspondence analysis to improve the public train transport. *Electronic Journal of Applied Statistical Analysis* 2(1):15–24.
- [316] D'AMBRA, L., BEH, E. J. & CAMMINATIELLO, I. (2014). Cumulative correspondence analysis of two-way ordinal contingency tables. *Communications in Statistics (Theory and Methods)*, 43, 1099 – 1113.
- [317] D'AMBRA, A. (2017), Cumulative correspondence analysis using orthogonal polynomials. *Communications in Statistics (Theory and Methods)*, 46, 2942 – 2954.
- [318] D'AMBRA, L.; AMENTA, P. & D'AMBRA, A. (2018). Decomposition of cumulative chi-squared statistics, with some new tools for their interpretation. *Statistical Methods & Applications*, 27, 297 – 318.

### 3.28 Linearly Constrained Correspondence Analysis

Often, when performing a multiple correspondence analysis on categorical data, there may be additional information available for at least one of the variables. Such constraints are imposed for the indicator matrix of each variable by specifying a matrix of constraints for the rows (individuals) and columns (categories). If no constraints are to be imposed, then these matrices are just identity matrices. One such case is to let the components be linearly constrained with respect to the matrix of relative marginal frequencies of the variable(s) being constrained leading to *linearly constrained correspondence analysis*. For a mathematical description of this variant in the context of two-way (simple) correspondence analysis and multiple correspondence analysis the reader is direct to

- [319] BÓCKENHOLT, U. & BÓCKENHOLT, I. (1990). Canonical analysis of contingency tables with linear constraints. *Psychometrika*, 55, 633 – 639.
- [320] TAKANE, Y., YANAI, H. & MAYEKAWA, S. (1991). Relationships among several methods of linearly constrained correspondence analysis. *Psychometrika*, 56, 667 – 684.
- [321] BÓCKENHOLT, U. & TAKANE, Y. (1994). Linear constraints in correspondence analysis. In *Correspondence Analysis in the Social Sciences*, eds. M. Greenacre, J. Blasius, pp. 112 – 127, London: Academic Press.
- [322] HWANG, H. & TAKANE, Y. (2002). Generalized constrained multiple correspondence analysis. *Psychometrika*, 67, 215 – 228.

A linearly constrained non-symmetrical correspondence analysis variant was described by

- [323] TAKANE, Y. & JUNG, S. (2009). Tests of ignoring and eliminating in nonsymmetric correspondence analysis. *Advances in Data Analysis and Classification*, 3, 315 – 340.

Further constrained approaches to correspondence analysis have also been proposed. See, for example,

- [324] GROENEN, P. J. F. & POBLOME, J. (2003). Constrained correspondence analysis for seriation in archaeology applied to Sagalassos ceramic tablewares. In *Exploratory Data Analysis in Empirical Research*, eds M. Schwaiger, O. Opitz, pp. 90 – 97, Berlin: Springer-Verlag.
- [325] HWANG, H. & TAKANE, Y. (2002). Generalized constrained multiple correspondence analysis. *Psychometrika*, 67, 211 – 224.
- [326] VAN DE VELDEN, M., GROENEN, P. J. F. & POBLOME, J. (2009). Seriation by constrained correspondence analysis: A simulation study. *Computational Statistics & Data Analysis*, 53, 3129 – 3138.
- [327] KOSTOV, B., BÉCUE-BERTAUT, M. & HUSSON, F. (2015). Correspondence Analysis on Generalised Aggregated Lexical Tables (CA-GALT) in the FactoMineR Package. *The R Journal*, 7/1, 109 – 117.

### 3.29 Additional Variants

The articles that we refer to in the previous sections give us only a glimpse into the types of variants of correspondence analysis that is traditionally used to analyse data. There are other types of analysis including what is called *semi-supervised detrended correspondence analysis* by

- [328] KONG, Z. & CAI, Z. (2009). Semi-supervised detrended correspondence analysis algorithm, In *Proceedings of the 2009 Third International Symposium on Intelligent Information Technology Application (Volume 1)*, eds Q. Luo, M. Zhu, pp. 429 – 432, Institute of Electrical and Electronics Engineers (IEEE).

although there is very little known about the ongoing development and application of this variant. Another variant is *multi-block discriminant correspondence analysis* proposed by

- [329] WILLIAMS, L. J., ABDI, H., FRENCH, R., & ORANGE, J. B. (2010). A tutorial on Multi-Block Discriminant Correspondence Analysis (MUDICA): a new method for analyzing discourse data from clinical populations. *Journal of Speech Language and Hearing Research*, 53, 1372 – 1393.

Yet another, more recent, variant of correspondence analysis is *cluster correspondence analysis* which was discussed by

- [330] VAN DE VELDEN, M., D'ENZA, A. I & PALUMBO, F. (2017). Cluster correspondence analysis. *Psychometrika*, 82, 158 – 185.

In R, this variant can be performed using the `clusmca` function in the `clustrd` package:

- [331] MARKOS, A. & D'ENZA, A. I. (2018). clustrd: Methods for joint dimension reduction and clustering.  
R version 1.2.2. <https://cran.r-project.org/web/packages/clustrd/index.html>

An alternative, and earlier, variant of cluster correspondence analysis may also be considered by adopting the method proposed in the geographical sciences by

- [332] LU, Y. & THILL, J.-C. (2003). Assessing the cluster correspondence between paired point locations. *Geographical Analysis*, 35, 290 – 309.
- [333] LU, Y. & THILL, J.-C. (2008). Cross-scale analysis of cluster correspondence using different operational neighborhoods. *Journal of Geographical Systems*, 10, 241 – 261.

There are other members of the correspondence analysis family tree that have been described but have not yet appeared in the literature. At the CARME2015 conference, three additional variants of the technique were discussed. They were *disjoint multiple correspondence analysis*, *fragmented correspondence analysis* and *multinomial correspondence analysis*, and were described by

- [334] VICHI, M. (2015), Disjoint correspondence analysis, Presented at CARME2015 in Naples, Italy.
- [335] LEBART, L. (2015), About fragmented correspondence analysis of texts (variations on the “Inaugural Address” corpus), Presented at CARME2015, Naples, Italy.
- [336] GROENEN, P. & JOSSE, J. (2015), Multinomial correspondence analysis, Presented at CARME2015, Naples, Italy and at IFCS Conference, Vienna, 2015.

respectively. Although an alternative approach to the variant multinomial correspondence analysis seemed to have existed prior to 2015. See

- [337] PAPP, Z. (2010). Constituency focus in Hungary 2009: Towards a new concept of representation? 3rd ECPR Graduate Conference, Dublin City University, Ireland.
- [338] TANAKA, N. (2011). The cross-cultural study on work values: Focus on religion, leisure time, and school education. *Ochanomizu University Global COE Program*, 13, 129 – 136.

## 4 Discussion

As we have highlighted here, the development of correspondence analysis has undergone radical changes since it first came to the attention (albeit, slowly) of the statistical and allied communities. Except for a few key places (especially in Europe), the technical

development of correspondence analysis within the statistical community has been relatively glacial (when compared with developments made in other areas of statistical research); this is especially the case in Australia and New Zealand where much of the perception of correspondence analysis seems to be centred on it being purely a “descriptive” approach to data analysis. Also, since correspondence analysis is (incorrectly) largely perceived to be unrelated to the modelling and inferential issues concerned with categorical data analysis in this part of the world may also explain its relative obscurity. Despite this, the types of analyses being proposed is certainly on the rise. Some involve tweaks to existing variants while others address more fundamental statistical and analytical questions. These include, but are not confined, to the following issues:

- Typically, Pearson’s chi-squared statistic is used in correspondence analysis as the basis for quantifying the association between the categorical variables. However, other measures of association for two- and three- way categorical variables have also been incorporated into the analysis. For two variables, these include, the Freeman-Tukey statistic, the Goodman-Kruskal tau index while the Marcotorchino index (Marcotorchino, 1985), Gray-Williams index (Gray and Williams, 1981) and the Delta index (Lombardo, 2011) may be used for three or more variables; see, for example [5; Chapter 11] for details. Traditionally, when examining the association between the variables of a contingency table, it is often assumed that the nature of the association is symmetric. However, there may be practical reasons to consider an asymmetric structure so that one categorical variable is treated as a predictor variable and another is a response variable. This structure underlies non-symmetrical correspondence analysis and its related methods.
- How the association should be visualised has attracted a lot of attention over the decades. As we have discussed in Section 2.3 there are even arguments for no visualisation to be made. Where a graphical summary of data is required, one may consider variations of plotting systems typically used in the classical approach to correspondence plot. These may include the biplot or even “detrending” an axis. Another approach that has proven to be controversial is to adopt a Carroll-Green-Schaffer plotting system; refer to Carrol et al. (1986, 1987, 1989) and Greenacre (1989) for the source of this controversy. We described this approach in [4, 5] and the interested reader is invited to consider the relevant literature described there.
- Depending on the choice of how to quantify the association between categorical variables, and the structure of the categories, one has a choice of determining how the matrix decomposition is performed. For nominal variables, singular value decomposition (SVD) is the most common method of obtaining the scores/components for each variable. When at least one categorical variable consists of ordered categories one may consider a bivariate moment decomposition (BMD) (Lancaster, 1953; Best and Rayner, 1996, 289) or a hybrid version of SVD and BMD (Rayner and Best, 2000; Beh, 2001) of the data. One may also consider any of the strategies for imposing linear constraints to incorporate auxiliary information in the decomposition.

- One of the key questions asked when analysing multiple categorical variables is how to go about performing a correspondence analysis. The most common approach is to transform a multi-way contingency table into a two-way matrix, typically either into its indicator matrix, Burt matrix or concatenation form (Khangar and Kamalja, 2017). However, in such a transformation it is possible to miss key aspects of the association since only two-way classifications are generally considered. That is, performing a correspondence analysis on a two-way transformation of a multi-way contingency table means truly tri-, quad- and multi- way association terms that reflect the simultaneous analysis of three, four and multi (respectively) variables are missed. For this reason, analysing the data in its multi-way form is advantageous. A simple example of this is where a correspondence analysis is performed on a three-way contingency table; a multi-way correspondence analysis treats the data in its “cube” form rather than converting it into its indicator matrix or Burt matrix form, thereby preserving any multi- variate association structure that may exist between the variables (Beh and Lombardo, 2019).
- Distances between points in a low-dimensional space are typically measured using the Euclidean distance, although variants have been adapted to measure them using Hellinger or Taxicab/Manhattan distances.
- The ongoing theoretical development of the foundations of correspondence analysis brings into the limelight some of the key features of a correspondence analysis, while still preserving the interpretation of the output in terms of something meaningful about the structure of the association in the data. This includes taking into consideration over-dispersion in count data, incorporating extraneous information for a variable, or a geographic location and mathematical properties (including, but not limited, to series expansions and the Box-Cox transformation).
- How to incorporate the structure of the categorical variable, whether it consist of nominal or ordinal categories, or a mix, has been a topic of ongoing research for decades now and has yet to gain the same level of attention as those variants and adaptations designed for the analysis of completely nominal categorical variables.

Aside from its ongoing technical development, correspondence analysis can be adapted for the ever-growing issues surrounding the analysis of “big data”. The issue has been tackled by some although it has yet to gain traction amongst many. Those who have contributed in this area typically approach the problem from a neural network, biomedical or computer science perspective. See, for example,

- [339] LEBART, L. (1997). Correspondence analysis, discrimination, and neural networks. In *Data Science and Related Methods*, eds C. Hayashi, N. Ohsumi, K. Yajima, Y. Tanaka, H. H. Bock, Y. Baba, pp. 423 – 430, Berlin: Springer.
- [340] ROBERTS, J. M. (2000). Correspondence analysis of two-mode network data. *Social Networks*, 22, 65 – 72.

- [341] DE STEFANO, D., RAGOZINI, G. & VITALE, M. P. (2008) Non-symmetrical correspondence analysis to explore dependence structure in social networks. In 7th International Conference on Social Science Methodology: RC33 – Logic and Methodology in Sociology, University of Naples, “Federico II”, Italy, September 2 – 4.
- [342] MURTAGH, F., GANZ, A. & MCKIE, S. (2009). The structure of narrative: the case of film scripts, *Pattern Recognition*, 42, 302 – 312.
- [343] D’ENZA, A. I. & GREENACRE, M. (2011). Multiple correspondence analysis for the quantification and visualization of large categorical data sets. In *Advanced Statistical Methods for the Analysis of Large Data-Sets, Studies in Theoretical and Applied Statistics*, eds A. Di Ciaccio, M. Coli, J. M. Angulo Ibanez, pp. 453 – 463, Berlin: Springer-Verlag.
- [344] D’ESPOSITO, M. R., DE STEFANO, D. & RAGOZINI, G. (2014) On the use of multiple correspondence analysis to visually explore affiliation networks. *Social Networks*, 38, 28 – 40.
- [345] MURTAGH, F. & GANZ, A. (2015). Pattern recognition in narrative: Tracking emotional expression in context, *Journal of Data Mining and Digital Humanities*, vol. 2015, 21 pages.
- [346] MURTAGH, F. (2015). Correspondence factor analysis of big data sets: A case study of 30 million words; and contrasting analytics using Apache Solr and correspondence analysis in R, Available at [arXiv:1507.01529v1](https://arxiv.org/abs/1507.01529v1).
- [347] YANG, Y., CHEN, S. & SHYU, M. (2015). Temporal multiple correspondence analysis for big data mining in soccer videos. 2015 IEEE International Conference on Multimedia Big Data, pp. 64 – 71, Beijing, China.
- [348] ZHU, M., KUSKOVA, V., WASSERMAN, S. & CONTRACTOR N. (2016). Correspondence analysis of multirelational multilevel networks. In *Multilevel Network Analysis for the Social Sciences*, eds E. Lazega, T. A. B. Snijders, pp. 145 – 172, Cham: Springer.
- [349] TEKAIA, F. (2016). Genome data exploration using correspondence analysis. *Bioinformatics and Biology Insights*, 10, 59 – 72.
- [350] KOKOL, P., VOSNER, H. B. & ZELEZNIK, D. (2016). Visualising nursing data using correspondence analysis. *Nurse Researcher*, 24, 38 – 40.
- [351] MURTAGH, F. (2017) Big Textual Data: Lessons and Challenges for Statistics. *SIS 2017 Statistics and Data Science: New Challenges, New Generations*, pp. 719 – 730, Florence, Italy: Firenze University Press.
- [352] MURTAGH, F. & FARID, M. (2017) Contextualizing geometric data analysis and related data analytics: A virtual microscope for big data analytics, *Journal of Interdisciplinary Methodologies and Issues in Science*, 3, 19 pages.

- [353] MURTAGH, F. (2017) Data Science Foundations: Geometry and Topology of Complex Hierarchic Systems.
- [354] STRANGE, K. D. & SUN, Z. (2017). Analyzing relationships in terrorism big data using Hadoop and statistics. *Journal of Computer Information Systems*, 57, 67 – 75.
- [355] MURTAGH, F. (2018). Data mining and big data analytics: Exploiting resolution scale, addressing bias, having analytical focus. *International Journal of Computer & Software Engineering*, 3, 7 pages.

To complement the ongoing technical development and any future movements to their application to “big data” one area of research worth considering is the adaptation of virtual reality technology to provide detailed, and an all immersive, visualisation of the structure of categorical association. The technology available to the consumer has been dominated by the gaming industry with the availability of affordable platforms including the Oculus Rift and HTC Vive that place virtual reality tools in the hands of people of all ages. With this growth in popularity, virtual reality has also attracted considerable attention for its potential usage in a variety of disciplines, including engineering (Hilfert, 2016), medicine (Moglia et al., 2016; Egger et al., 2017) and education (Merchant et al., 2014; Velev and Zlateva, 2017). One may also consider Niehorster (2017) who studied the benefits of virtual reality in performing scientific experimentation. As yet, it seems such technology has not yet gained the same level of attention in the data visualisation communities that lie within the statistics, and allied, disciplines. Irrespective of the what future development and growth of correspondence analysis is to come, it is sure to be exciting.

## References

- Agresti, A. (2002). *Categorical Data Analysis*. John Wiley & Sons. *Inc., Publication*.
- Beh, E. J. and Lombardo, R.(2014). *Correspondence analysis: theory, practice and new strategies*. John Wiley & Sons.
- Beh, E. J. and Lombardo, R.(2019). Multiple and multiway correspondence analysis. *WIREs Computational Statistics*. DOI: 10.1002/wics.1464..
- Beh, E. J. (2001). Partitioning Pearson’s chi-squared statistic for singly ordered two-way contingency tables. *The Australian and New Zealand Journal of Statistics*, 43, 327 – 333.
- Best, D. J. and Rayner, J. C. W. (1996). Nonparametric analysis for doubly ordered two-way contingency tables. *Biometrics*, 52, 1153 – 1156.
- Braga, J., Chabadel, O., Sonan, N.K. and Gueramy, A. (2005). Non-adult dental age assessment: correspondence analysis and linear regression versus Bayesian predictions. *International Journal of Legal Medicine*, 119, 260 – 274.
- Carrol, J. D., Green, P. E. and Schaffer, C. M. (1986). Interpoint distance comparisons in correspondence analysis. *Journal of Marketing Research*, 23, 271 – 280.

- Carrol, J. D., Green, P. E. and Schaffer, C. M. (1987). Comparing interpoint distances in correspondence analysis: A clarification. *Journal of Marketing Research*, 24, 445 – 450.
- Carrol, J. D., Green, P. E. and Schaffer, C. M. (1989). Reply to Greenacre's commentary on the Carroll-Green-Schaffer scaling of two-way correspondence analysis solutions. *Journal of Marketing Research*, 26, 366 – 368.
- De Tibeiro, J. J. S. and Murdoch, D. J. (2010). Correspondence analysis with incomplete paired data using Bayesian imputation. *Bayesian Analysis*, 5, 519 – 532.
- Egger, J., Gall, M., Wallner, J., Boechat, P., Hann, A., Li, X., Chen, X. and Schmalstieg, D. (2017). HTC Vive MeVisLab integration via OpenVR for medical applications. *PLoS ONE*, 12(3): e0173972, 14 pages.
- Van der Heijden, P., de Falguerolles, A., De Leeuw, J. (1989) A combined approach to contingency table analysis using correspondence analysis and loglinear analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 38, 2, 249-273.
- Freeman, M. F. and Tukey, J. W. (1950). Transformations related to the angular and square root. *Annals of Mathematical Statistics*, 21, 607–611.
- Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58, 453 – 467.
- Gifi, A. (1990). *Nonlinear Multivariate Analysis* Wiley.
- Goodman, L. A. and Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, 49, 732 – 764.
- Gray, L.N. and Williams, J.S. (1981). Goodman and Kruskal  $\tau_b$  multiple and partial analogs. *Proceedings of the Social Statistics Section*, 10, 50 – 62.
- Greenacre, M. (2010). *Biplots in Practice*. Bilbao: Fundación BBVA.
- Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis*. London: Academic Press.
- Greenacre, M. J. (1989). The Carroll-Green-Schaffer scaling in correspondence analysis: a theoretical and empirical appraisal. *Journal of Marketing Research*, 26, 358 – 365.
- Haberman, S. J. (1973). The analysis of residuals in cross-classified tables. *Biometrics*, 205 – 220.
- Hilfert, T. and König, M. (2016). Low-cost virtual reality environment for engineering and construction. *Visualization in Engineering*, 4(2), 18 pages.
- Khangar, N. V. and Kamalja, K. K. (2017). Multiple Correspondence Analysis and its applications. *Electronic Journal of Applied Statistical Analysis*, 10 (2) 432–462.
- Kendall, D. G. (1971). Seriation. In *Mathematics in the Archaeological and Historical Sciences*, eds. F. R. Hodson, D. G. Kendall, P. Tauti, pp. 215 – 252, University Press.
- Kroonenberg, P. M. (2008). *Applied Multiway Data Analysis*. Chichester: Wiley.
- Krishnam, A., Williams, L. J., McIntosh, A. R. and Abdi, H. (2011). Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review. *NeuroImage*, 56, 455 – 475.

- Lancaster, (1953). A reconciliation of  $\chi^2$ , considered from metrical and enumerative aspects. *Sankhyā*, 13, 1 – 10.
- Lombardo, R. (2011). Three-way association measure decompositions: the Delta index. *Journal of Statistical Planning and Inference*, 141, 1789 – 1799.
- Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W. and Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, 70, 29 – 40.
- Marcotorchino, F. (1985). Utilisation des Comparaisons par Paires en Statistique des Contingencies: Partie III. *Report F081, Etude du Centre Scientifique*, IBM, France.
- Moglia, A., Ferrari, V., Morelli, L., Ferrari, M., Mosca, F. and Cuschieri, A. (2016). A systematic review of virtual reality simulators for robot-assisted surgery. *European Urology*, 69, 1065 – 1080.
- Nenadić, O. and Greenacre, M. (2007). Correspondence analysis in R, with two- and three-dimensional graphics: The ca package. *Journal of Statistical Software*, 20(3).
- Niehorster, D. C., Li, L. and Lappe, M. (2017). The accuracy and precision of position and orientation tracking in the HTC Vive virtual reality system for scientific research. *i-Perception*, May-June 2017, 1 - 13.
- Nishisato, S. (1994). *Elements of Dual Scaling: An Introduction to Practical Data Analysis*. Hillsdale, N. J.: Lawrence Erlbaum Associates.
- Nishisato, S. (2007). *Multidimensional Descriptive Non-Linear Analysis*. London: Chapman & Hall/CRC.
- Rayner, J. C. W. and Best, D. J. (2000). Analysis of singly ordered two-way contingency tables. *Journal of Applied Mathematics & Decision Sciences*, 4, 83 – 98.
- Tenenhaus, M. and Young, F. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika*, 50, 91 – 119.
- Taguchi, G. (1966). *Statistical Analysis*. Tokyo: Maruzen.
- Taguchi, G. (1974). A new statistical analysis for clinical data, the accumulating analysis, in contrast with the chi-square test. *Saishin Igaku*, 29, 806 – 813.
- Tucker, L.R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31, 279–311.
- Van Der Heijden, P. G. M. and De Leeuw, J. (1985). Correspondence analysis used complementary to loglinear analysis. *Psychometrika*, 50, 429 – 447.
- Van Der Heijden, P. G. M., De Falguerolles, A. and De Leeuw, J. (1989). A combined approach to contingency table analysis using correspondence analysis and log-linear analysis. *Applied Statistics*, 38, 249 – 292.
- Velev, D. and Zlateva, P. (2017). Virtual reality challenges in education and training. *International Journal of Learning and Teaching*, 3, 33 – 37.
- Weber, A. (1929). *The Theory of Location of Industry* Chicago University Press.
- Whittaker, R. H. (1967). Gradient analysis of vegetation. *Biological Review*, 42, 207 – 264.