WIKIPEDIA

Multiple correspondence analysis

In statistics, **multiple correspondence analysis** (**MCA**) is a data analysis technique for nominal categorical data, used to detect and represent underlying structures in a data set. It does this by representing data as points in a low-dimensional Euclidean space. The procedure thus appears to be the counterpart of principal component analysis for categorical data.[1][2] MCA can be viewed as an extension of simple correspondence analysis (CA) in that it is applicable to a large set of categorical variables.

Contents

As an extension of correspondences analysis

MCA is performed by applying the CA algorithm to either an indicator matrix (also called *complete disjunctive table* – CDT) or a *Burt table* formed from these variables.[3] An indicator matrix is an individuals × variables matrix, where the rows represent individuals and the columns are dummy variables representing categories of the variables.[4] Analyzing the indicator matrix allows the direct representation of individuals as points in geometric space. The Burt table is the symmetric matrix of all two-way cross-tabulations between the categorical variables, and has an analogy to the covariance matrix of continuous variables. Analyzing the Burt table is a more natural generalization of simple correspondence analysis, and individuals or the means of groups of individuals can be added as supplementary points to the graphical display.

In the indicator matrix approach, associations between variables are uncovered by calculating the chi-square distance between different categories of the variables and between the individuals (or respondents). These associations are then represented graphically as "maps", which eases the interpretation of the structures in the data. Oppositions between rows and columns are then maximized, in order to uncover the underlying dimensions best able to describe the central oppositions in the data. As in factor analysis or principal component analysis, the first axis is the most important dimension, the second axis the second most important, and so on, in terms of the amount of variance accounted for. The number of axes to be retained for analysis is determined by calculating modified eigenvalues.

Recent works and extensions

In recent years, several students of Jean-Paul Benzécri have refined MCA and incorporated it into a more general framework of data analysis known as Geometric data analysis. This involves the development of direct connections between simple correspondence analysis, principal component analysis and MCA with a form of cluster analysis known as

Euclidean classification.[5]

Two extensions have a great practical use.

- It is possible to include, as active elements in the MCA, several quantitative variables. This extension is called factor analysis of mixed data (see below).
- Very often, in questionnaires, the questions are structured in several issues. In the statistical analysis it is necessary to take into account this structure. This is the aim of multiple factor analysis which balances the different issues (i.e. the different groups of variables) within a global analysis and provides, beyond the classical results of factorial analysis (mainly graphics of individuals and of categories), several results (indicators and graphics) specific of the group structure.

Application fields

In the social sciences, MCA is arguably best known for its application by Pierre Bourdieu,[6] notably in his books *La Distinction*, *Homo Academicus* and *The State Nobility*. Bourdieu argued that there was an internal link between his vision of the social as spatial and relational --- captured by the notion of field, and the geometric properties of MCA.[7] Sociologists following Bourdieu's work most often opt for the analysis of the indicator matrix, rather than the Burt table, largely because of the central importance accorded to the analysis of the 'cloud of individuals'.[8]

Multiple correspondence analysis and principal component analysis

MCA can also be viewed as a PCA applied to the complete disjunctive table. To do this, the CDT must be transformed as follows. Let y_{ik} denote the general term of the CDT. y_{ik} is equal to 1 if individual i possesses the category k and 0 if not. Lets denote p_k, the proportion of individuals possessing the category k. The transformed CDT (TCDT) has as general term:

$$x_{ik} = y_{ik}/p_k - 1$$

The unstandardized PCA applied to TCDT, the column k having the weight p_k, leads to the results of MCA.

This equivalence is fully explained in a recent book of Jérôme Pagès.[9] It plays an important theoretical role because it opens the way to the simultaneous treatment of quantitative and qualitative variables. Two methods simultaneously analyze these two types of variables: factor analysis of mixed data and, when the active variables are partitioned in several groups: multiple factor analysis.

This equivalence does not mean that MCA is a particular case of PCA as it is not a particular case of CA. Its only means that these methods are closely linked one another as they belong to the same family: the factorial methods.

Software

There are numerous software of data analysis that include MCA, such as STATA and SPSS. The R package FactoMineR (http://factominer.free.fr/) also features MCA. This software is related to a book describing the basic methods for performing MCA .[10]

References

1. Le Roux; B. and H. Rouanet (2004). *Geometric Data Analysis, From Correspondence Analysis to Structured Data Analysis*. Dordrecht. Kluwer: p.180.

2. Greenacre, Michael and Blasius, Jörg (editors) (2006). *Multiple Correspondence Analysis and Related Methods*. London: Chapman & Hall/CRC.

3. Greenacre, Michael (2007). *Correspondence Analysis in Practice, Second Edition*. London: Chapman & Hall/CRC.

4. Le Roux, B. and H. Rouanet (2004), Geometric Data Analysis, From Correspondence Analysis to Structured Data Analysis, Dordrecht. Kluwer: p.179

5. Le Roux; B. and H. Rouanet (2004). *Geometric Data Analysis, From Correspondence Analysis to Structured Data Analysis*. Dordrecht. Kluwer.

6. Scott, John & Gordon Marshall (2009): Oxford Dictionary of Sociology, p. 135. Oxford: Oxford University Press

7. Rouanet, Henry (2000) "The Geometric Analysis of Questionnaires. The Lesson of Bourdieu's La Distinction", in Bulletin de Méthodologie Sociologique 65, pp. 4–18

8. Lebaron, Frédéric (2009) "How Bourdieu "Quantified" Bourdieu: The Geometric Modelling of Data", in Robson and Sanders (eds.) Quantifying Theory: Pierre Bourdieu. Springer, pp. 11-30.

9. Pagès Jérôme (2014). *Multiple Factor Analysis by Example Using R*. Chapman & Hall/CRC The R Series London 272 p

10. Husson F., Lê S. & Pagès J. (2009). *Exploratory Multivariate Analysis by Example Using R*. Chapman & Hall/CRC The R Series, London. ISBN 978-2-7535-0938-2

External links

- Le Roux, B. and H. Rouanet (2004), Geometric Data Analysis, From Correspondence Analysis to Structured Data Analysis at Google Books: [1] (https://books.google.com/books?id=a6bDBUF58XwC&lpg=PP1&dq=geometric%20data%20analysis&hl=no&pg=PP1#v=onepage&q=&f=false)
- Greenacre, Michael (2008), *La Práctica del Análisis de Correspondencias*, BBVA Foundation, Madrid, available for free download at the foundation's web site [2] (https://web.archive.org/web/20100325141907/http://www.fbbva.es/TLFU/tlfu/esp/publicaciones/libros/fichalibro/index.jsp?codigo=300)
- FactoMineR (http://factominer.free.fr/) A R software devoted to exploratory data analysis.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Multiple_correspondence_analysis&oldid=867395638"