How to improve the Quality Assurance System of the Universities: a study based on compositional analysis

Bertaccini B., Gallo M., Simonacci V., and Menini T.

Abstract The National Agency for the Evaluation of Universities and Research (ANVUR) has for some decades defined the criteria for systematically evaluating student satisfaction. The analysis of these data presents various difficulties both in terms of data collection and analysis. The aim of this work is to propose Candecomp/Parafac for a compositional analysis, which is able to capture the multidimensional aspects of the phenomenon taking into account its ordinal nature and the temporal characteristics of data collection.
Abstract L'Agenzia nazionale per la valutazione delle università della ricerca (ANVUR) definisce da qualche decade i criteri per valutare in modo sistematico la soddisfazione degli studenti. L'analisi di tali dati presenta diverse difficoltia in termini di raccolta dati che di analisi. Scopo del presente lavoro è proporre il Candecomp/Parafac per un'analisi composizionale, il quale, rispettando le caratteristiche temporali della raccolta dei dati e la natura ordinale degli stessi, è in grado di cogliere gli aspetti multidimensionali del fenomeno.

Key words: compositional data, log-ratios, quality, student satisfaction.

Bruno Bertaccini
Department of Statistics, Computer Science, Applications ”G. Parenti”, University of Florence
e-mail: bruno.bertaccini@unifi.it

Michele Gallo
Department of Human and Social Sciences, University of Naples ”L'Orientale”, Italy.
e-mail: mgallo@unior.it

Violetta Simonacci
Department of Human and Social Sciences, University of Naples ”L'Orientale”, Italy.
e-mail: vsimonacci@unior.it

Tullio Menini
Department of Human and Social Sciences, University of Naples ”L'Orientale”, Italy.
e-mail: menini@unior.it

1 Introduction

Since the adoption of Law No. 370/99 in 1999, universities are directly responsible for the system of students' opinion polling on the educational modules undertaken. The main goal is to handle the information collected in a synthetic, efficient and effective way in order to obtain an actual review of the didactical organization and a general improvement of the courses offered by the university. The National Agency for the Evaluation of Universities and Research (ANVUR), established in 2006, defines the criteria to systematically assess student satisfaction on the educational activities as part of the Quality Assurance System of the Universities. Student surveys on the educational training provided, become a necessary step for the accreditation of Universities and single courses of study.

In student satisfaction analysis it is important to keep in mind that the rating expressed by each student does not only reflect the quality of service but also her/his perception, personal experiences, inclinations and socio-cultural background. Each student uses a subjective scale which influences his evaluation of the attributes of education. To address this issue a ratio-based approach such as compositional data analysis can be useful. In order to be coherent, the analysis of this kind of data should also take into account the variability of its many observable attributes, which are generally collected with multiple item questionnaires. Consequently, when student satisfaction is observed across the academic years the use of multilinear techniques as Candecomp/Parafac (CP) is advisable.

Proceeding from these considerations, a compositional analysis by CP is proposed to evaluate student satisfaction. Thus, a short review of the method proposed is given in Sect 2, while in Sect 3 the case study is presented.

2 Compositional analysis: short review

Compositional data were extensively studied in the 1980s by Aitchison, for more details see (1). The turning point of this methodology was the possibility to adopt a transformation in a log-scale to move from a constrained space, defined simplex, to the real unconstrained euclidean space. This important finding contributed towards the use of compositional data in several disciplines. In recent years, a further development of the theoretical framework (namely the principle of working in coordinates (2)) laid the basis for a practical use of compositional data in additional fields of studies as evaluation of academic educational quality (5).
Following (4) for three-way data, where on the first way there are I students, on the second way there are J aspects (variables) and on the third way there are K academic years, the starting three-way array $\underline{\mathbf{V}}$ $(I \times J \times K)$ is transformed in a new array $\underline{\mathbf{Z}}$. Here the frontal slices are obtained as $\mathbf{Z}_k = \mathbf{L}_k \mathbf{P}_J^{\perp}$, where $\mathbf{L}_k$ is the k-th frontal slice of the array $\underline{\mathbf{L}}$ with generic element $log(\dot{v}_{ijk})$. In addition, $\mathbf{P}_J^{\perp} = \left(\mathbf{I} - \mathbf{1}^T \mathbf{1}/J\right)$ represents the idempotent row centering matrix where $\mathbf{I}$ is an identity matrix of order J and $\mathbf{1}$ is a J dimensional vector of 1s. A three-way analysis can now be performed

on $\underline{\mathbf{Z}}$ where the generic row vector $\mathbf{z}_{ik}$ is a centered log-ratio (*clr*) defined as:

$$\mathbf{z}_{ik} = clr(\dot{\mathbf{v}}_{ik}) = \left[\log\frac{\dot{v}_{i1k}}{g(\dot{\mathbf{v}}_{ik})}, \ldots, \log\frac{\dot{v}_{iJk}}{g(\dot{\mathbf{v}}_{ik})}\right] \quad \text{with } g(\dot{\mathbf{v}}_{ik}) = \sqrt[J]{\prod_{j=1}^{J} \dot{v}_{ijk}} \tag{1}$$

As for principal component analysis on two-way arrays, the use of *clr*-coefficients allows for a direct application of the CP model on the array $\underline{\mathbf{Z}}$ without further concerns. We thus obtain the three loading matrices $\mathbf{A}$ $(I \times F)$, $\mathbf{B}$ $(J \times F)$, and $\mathbf{C}$ $(K \times F)$, where F is the number of components extracted. This decomposition can be written as follows:

$$\mathbf{Z}_I = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T + \mathbf{R}_I \tag{2}$$

$\mathbf{Z}_I$ $(I \times JK)$ and $\mathbf{R}_I$ $(I \times JK)$ identify the array of *clr*-coordinates and the array of residuals respectively, unfolded with respect to the first mode. In other words they are the horizontal concatenation of the $k = 1, \ldots, K$ matrices $\mathbf{Z}_k$ and $\mathbf{R}_k$ of dimension $(I \times J)$, i.e. frontal slices. The symbol $\odot$ identifies the Khatri-Rao product. The parameters contained in the three loading matrices are often estimated in a least squares sense by use of the Alternating Least Squares algorithm, an iterative procedure for which the objective function is:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} = \|\mathbf{Z}_I - \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T\|^2 \tag{3}$$

where $\|\cdot\|$ is the Frobenius norm.

Of course, the CP results are expressed in $clr-$coordinates, thus it is important to keep in mind that they need to be translated back into compositional terms for proper interpretation, see (3) and (6) for more details.

3 Case study

Following the ANVUR criteria, the University of Florence monitors the students' opinions on teaching quality through questionnaires. Questionnaires have the same item structure, however, some items are excluded for students who did not attend courses. In any case, they were electronically administered for each teaching and all responses were kept anonymous. Only data collected from 2012 to 2017 were considered.

In general several univariate statistics are used to analyze the data and they are available at https://valmon.disia.unifi.it/sisvaldidat/unifi. In this work a CP analysis of the compositional structure is proposed in order to investigate the logconstrasts between the different characteristics of service. Proceeding in the fashion provides several advantages compared to others statistical methods: 1) it exposes the preference structure between items and allows to properly visualize it by means of *ad hoc* graphical tools and interpretational rules; 2) it always guarantees a coherent outcome whether the questionnaire is really measuring a specific construct or not; 3) the results are the same even if the number of items considered should change (subcompositional coherence). The full case study will be discussed at conference.

References

[1] Aitchison, J.: The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability, Chapman and Hall Ltd., London (1986) (Reprinted in 2003 with additional material by The Blackburn Press

[2] Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal C.: Isometric logratio transformations for compositional data analysis. Mathematical Geology, 35, 279-300 (2003).

[3] Gallo, M: Log-Ratio and Parallel Factor Analysis: An Approach to Analyze Three-Way Compositional Data. Advanced Dynamic Modeling of Economic and Social Systems, 209.

[4] Gallo, M: Tucker3 model for compositional data. Communications in Statistics - Theory and Methods, 44(21), 4441-4453.

[5] Simonacci V., Gallo, M.: Statistical tools for student evaluation of academic educational quality. Quality and Quantity, 51(2), 565-579 (2017).

[6] Simonacci V., Gallo, M.: Detecting public social spending patterns in Italy using a three-way relative variation approach. Social Indicators Research, 1-14 (2018).