

metodološki zvezki - Advances in Methodology and Statistics

ISSN 1854-0031

[Home](#) | [Articles](#) | [Editorial Board](#) | [Instructions for Authors](#) | [About](#) | [Archive](#) | [News](#)



Scope

Metodološki zvezki - Advances in Methodology and Statistics is an interdisciplinary peer reviewed journal that deals with methodology and statistics. The journal promotes the scientific development of methodology and research, using quantitative, qualitative and mixed methods. It also seeks to publish new statistical methods and models, as well as innovative applications of statistical methods to address practical problems.

Applications and development of statistics and methodology in areas that include life sciences (such as biology, medicine and bioinformatics) and social sciences (such as sociology, psychology and economics) are equally appreciated. Papers from other areas are also considered. In addition to original research reports, the journal

publishes review articles.

Authors are also encouraged to enhance the reproducibility of their research by submitting software code and data, which will be published as supplementary materials on the journal's website.

Indexed by

SCOPUS
EBSCO
ECONIS
STMA-Z Statistical
Theory
PROQUEST

© MZ 2013; Partially based on Free Website Templates Thanks

metodološki zvezki - Advances in Methodology and Statistics

Articles 2007 - Volume 4

[Home](#) | [Articles](#) | [Editorial Board](#) | [Instructions for Authors](#) | [About](#) | [Archive](#)

Issue I: Statistics

Bronislaw Ceranka and Malgorzata Graczyk: A-Optimal Chemical Balance Weighing Design under Certain Condition. [Abstract](#), [Paper](#)

Maurizio Brizzi: Indices of Skewness Derived from a Set of Symmetric Quantiles: A Statistical Outline with an Application to National Data of E.U. Countries. [Abstract](#), [Paper](#)

Rosa Arboretti Giancristofaro and Stefano Bonnini: Permutation Tests for Heterogeneity Comparisons in Presence of Categorical Variables with Application to University Evaluation. [Abstract](#), [Paper](#)

Claudio Quintano et al.: Estimating Poverty in the Italian Provinces using Small Area Estimation Models. [Abstract](#), [Paper](#)

Katarina Čobanovič et al.: Use of Trellis Graphics in the Analysis of Results from Field Experiments in Agriculture. [Abstract](#), [Paper](#)

Gaj Vidmar: Statistically Sound Distribution Plots in Excel. [Abstract](#), [Paper](#)

Issue II: Methodology and Applications

Samo Pavlin and Tina Kogovšek: Qualitative and Quantitative Approaches to the Measurement of Complex Social Phenomena: A Case of Occupational Professionalisation. [Abstract](#), [Paper](#)

Juergen H.P. Hoffmeyer-Zlotnik and Uwe Warner: How to Survey Education for Cross-National Comparisons: The Hoffmeyer-Zlotnik/Warner-Matrix of Education. [Abstract](#), [Paper](#)

Rohangiz Modaber Dabagh: Authorship Attribution and Statistical Text Analysis. [Abstract](#), [Paper](#)

Michele Gallo: The Scaling Problems in Service Quality Evaluation. [Abstract](#), [Paper](#)

Marisa Civardi and Franca Crippa: Becoming Economically Independent: Weakening or Strengthening Parental Ties? [Abstract](#), [Paper](#)

Vladimir Batagelj et al.: Analysis of the Customers' Choice Networks: An Application on Amazon Books and CDs Data. [Abstract](#), [Paper](#)

Aina Maria Capo et al.: Networks of PhD Students and Academic Performance: A Comparison across Countries. [Abstract](#), [Paper](#)

Irena Ograjenšek: Improving Student In-Depth Understanding of Statistics: The Case for Offering Historical Insights in Statistics Courses. [Abstract](#), [Paper](#)

[2015](#) | [2014](#) | [2013](#) | [2012](#) | [2011](#) | [2010](#) | [2009](#) | [2008](#) | [2007](#) | [2006](#) | [2005](#) | [2004](#)

© MZ 2013; Partially based on Free Website Templates

The Scaling Problems in Service Quality Evaluation

Michele Gallo¹

Abstract

In service quality evaluation we have to treat data having different kinds of scales. In order to obtain a measure of the service quality level a conventional ordinal rating scale for each attribute of a service is used. Moreover additional information on the customers or on the objective characteristics of the service is available (interval, ordinal and or categorical scale). In the latter the importance or weight assigned to the different items must be also considered (compositional scale). To analyze these different kinds of data particular precaution should be used, a transformation of quality level perceived (expected) data in quantitative scale is carried out before a multidimensional data analysis. In literature more techniques are proposed for the quantification of ordinal data preserving the original characteristics. The aims of this paper are to analyze different ways to quantify ordinal data, and illustrate how the additional information on the customers or on the service could be used in the multidimensional analysis as external information.

1 Introduction

Customer satisfaction has become a vital concern for companies and organizations in their efforts to improve product and service quality, and maintain customer loyalty within a highly competitive market place. It is conceptualized as an affectively laden “fulfillment response” to service received (Oliver, 1997). To obtain a measure of satisfaction is not a simple meter because satisfaction is mostly due to physiologically conditioned assessments. It reflects both emotional and cognitive elements (Oliver 1993). In the last decade, more Customer Satisfaction Indexes (CSI) have been proposed (e.g., USA, Fornell et al., 1996; European Union, ECSI Technical Committee, 1998) where the structure of all CSI are continually undergoing review and subject to modifications.

If the structure of CSI(s) are continually subject to modifications, the core of the model is in most respects standard. It is encased within a system of cause and

¹ Department of Social Science, University of Naples - L'Orientale, L.go S. Giovanni Maggiore 20, 80156 Naples, Italy; mgallo@unior.it

effect running from the antecedents of overall customer satisfaction – quality expectation, quality perceived, perceived value and image – to the consequences of overall customer satisfaction – customer loyalty and customer complaints. Eklöf (2000) proposed to distinguish between perceived product quality and perceived service quality where perceived service quality is the evaluation of recent consumption experience of associated services like customer service, conditions of product display, range of services and products etc. On this basis, we posit that service quality is primarily an antecedent of customer satisfaction (Fornell et al., 1996).

In this work, we have not taken care of the problems associated with the choice of the “better” CSI measurement framework. In the following just the SERVQUAL model (Parasuraman et al., 1988) is considered for measuring the service quality (SQ). Anyway most of the results can be generalized to the other models (D’Ambra and Gallo, 2006). The SERVQUAL have a structure based on a set of attributes and dimensions. Where each attribute is evaluated by an item and sets of items giving the evaluation of the dimensions. The item has the same ordinal rating scale with a ranging on seven scores. Moreover an importance or weight (C) is attached to each dimension (or attribute) that is principally used to weigh the gap between performance perception (R) and service quality expectation (E). Whereas the weight could be analyzed independently to obtain information on the nature and causes of the interrelationships between the quality dimensions (or attributes). In this case, we should consider the constraint ratio scale of the importance data. Likewise, in these studies some additional information is available which permits us to investigate the degree of satisfaction of homogeneous customer clusters (gender, age, profession, level of education, etc.) with respect to some objective characteristics of the service (for example in hospital: procedure, illness, etc.). The inclusion of additional information in the multidimensional analysis of SQ data could be used to obtain more accurate results.

Let $\mathbf{R}_{N,J}$ ($\mathbf{E}_{N,J}$) be the matrix of perceived (expected) results where N are the customers and J are the ordinal variables (for each item there is one variable) with the same number of categories ℓ . And let $\mathbf{C}_{N,D}$ be the matrix of importance with D variables. The nature of data should be considered before we carry out a multidimensional analysis. In particular, perception (expectation) evaluation has an ordinal scale. This scale establishes an explicit rank, but not all arithmetic transformations are meaningful because the distances between points on an ordinal scale are not meaningful. The importance data has a constrained ratio scale. For these data intervals between values and ratios of values are meaningful but the constrained of the unit-sum of the composition scale causes more problems (Aitchison, 1986). Therefore additional information could have different kinds of scales: as categorical (or nominal) scale, where there is no explicit ranking on the

category labels; interval scale, where the distances between data are meaningful but where zero is not meaningful; and where there is an ordinal and ratio scale.

The principal purpose of this paper is to point out the necessity to transform the row data before we carry out some multidimensional statistics analysis, while the transformation should respect the original nature of the data. Moreover, the inclusion of additional information could be particularly appropriate in the framework of SQ , because it allows us to know the degree of satisfaction of homogeneous subject clusters evaluating the SQ in a more precise and objective way. In Section 2, we consider the quantification of perceived (expected) data. In particular, after a brief review of some technique for the optimal scaling of ordinal data, where the optimal scaling is defined in terms of the correlation matrix of quantified variables. We propose a new way to quantify the perceived (expected) data based on the conservation of the different subjective scales of each customer. In Section 3, we define the properties of compositional data and propose a logcontrast transform of the matrix $\mathbf{C}_{N,D}$ in order to perform multivariate analysis. In Section 4, a presentation of how the additional information could be included into a multidimensional analysis of SQ data is given.

2 Some approaches to the optimal quantification of ordinal data

In literature different approaches are proposed to quantify the ordinal data. Most of them are based on a loss function to attain a minimum between the transformed variables, so one possibility is to use the mean squared euclidean distance between the transformed variables and one hypothetical common variable. A generic loss function is

$$J^{-1} \sum_{j=1}^J SSQ(\mathbf{X}\boldsymbol{\beta}_j - \varphi_j(h_j)) \quad (2.1)$$

where \mathbf{X} is a matrix of basis vectors of order $(N \times s)$, $\boldsymbol{\beta}_j$ is a vector of s loadings and $\varphi_j(h_j)$ may be any non linear function of the variable h_j ($j=1, \dots, J$). Equation (2.1) is used by Kruskal and Shepard (1974), Young, de Leeuwe and Takane (1976), and many others. De Leeuwe, van Rijkevorsel (1980) and Gifi (1982) use the following alternative loss function

$$J^{-1} \sum_{j=1}^J SSQ(\mathbf{X} - \varphi_j(h_j)) \quad (2.2)$$

where the weights $\boldsymbol{\beta}_j$ are incorporated into function $\varphi_j(\cdot)$. Equation (2.2) is used because the treatment of missing data becomes more simple, and both variables

with a single quantification and variables with multiple ones can be analyzed simultaneously. We interpret $\varphi_j(\cdot)$ as an approximation of a function based on a representation of some finite amount of data on that function. In the present case, we would like to incorporate into the analysis the underlying monotone structure of data, so that each variable is treated as ordinal. Moreover the rank-one restrictions are included in (2) which implies that the quantifications in s dimensional space of each variable h_j become proportional to each other (Gifi, 1982). For the generic variable h_j , we can write $\varphi_j(h_j) = \mathbf{G}_j \mathbf{q}_j \boldsymbol{\beta}_j'$, where \mathbf{q}_j is a vector of single category quantification, $\boldsymbol{\beta}_j$ a vector of loadings and \mathbf{G}_j is a matrix which indicates the category of the j^{th} variable. Here we propose a B -spline transformation because it has more attractive proprieties (van Rijckevorsel and De Leeuwe, 1988). Nevertheless, we could use a Fuzzy coding or a Monotone-spline (van Rijckevorsel, 1987; Winsberg and Ramsay, 1983).

2.1 B -Spline transformation

The B -spline takes a variable as input and produces more than one variable as output. By B -spline a variable h_j ($h_j \in [a, b]$ with $a \neq b$) is partitioned into a number of intervals, where two boundary points, called knots, restrict every interval: $a \leq \kappa_1 < \dots < \kappa_K \leq b$, with exactly r "interior knots" (knots that meet another knot on both sides) and $K - r$ "exterior knots". Let piecewise coding function be a positive function between 0 and 1 on some contiguous part on range $[a, b]$. For each variable h_j there exists s different piecewise coding functions represented by correspondingly s column-vectors $\mathbf{G}_k(h)$ $k = 1, \dots, s$, where the s column-vectors $\mathbf{G}_k(h)$ are collected in a pseudo-indicator matrix \mathbf{G} , where

$$\sum_{k=1}^s \mathbf{G}_{ik} = 1.$$

A B -spline is a particular piecewise coding function with functional pieces of degree v , that is positive on exactly $v + 1$ consecutive intervals, with an overlap of exactly v intervals with the next B -spline, all intervals defined by the knot sequence κ , with $v \leq r$. The order of B -spline is equal to $v + 1$, while the number of B -splines needed to code a variable with knot sequence κ is equal to $s = v + r$. B -splines of zero degree are the most used for the good properties of the crisp coding (van Rijckevorsel, 1988). More coding functions and relative characteristics are illustrated in Schumaker (1981). To have a global transformation we use a linear combination of B -spline $\varphi(h^j) = \sum_{k=1}^s \alpha_k \mathbf{B}_k(h^j)$, where the unknown spline coefficient α_i is the only determining parameter of $\varphi(\cdot)$ (Van

Rijkevorsel, 1987). Nevertheless, other methods to determine the optimal coefficient are proposed in literature (De Boor, 1978; Schumaker, 1981).

2.2 A quantification of perceived/expected data for the multidimensional analysis

In Section 2.1 we have considered an approach based on the difference between the variables where each categorical value of the j^{th} column of $\mathbf{R}_{N,J}$ ($\mathbf{E}_{N,J}$) is substituted from the correspondent score of the vector \mathbf{q}_j . In this way categories of the same variable are quantified similarly and categories of dissimilar variables are quantified differently ($\mathbf{q}_j \neq \mathbf{q}_{j'}$ with $j \neq j'$). This characteristic is not required in the quantification of perceived (expected) data. An additional effect of this kind of transformation is that costumers with the same rate for an item obtain the same quantification whereas a customer with the same rate for different items obtains dissimilar quantifications.

An alternative approach is based on the difference between the “psychological” scale used with respect to different customers. Following this approach, the quantification procedure has to preserve the different origin of the measuring system of each customer, the different distance between two points on the scale as the non-linear distance between two subsequent points. Finally, by the time that a SERVQUAL questionnaire is drawn up each customer has his own specific reference system. This is the same for each item of the questionnaire and it does not change at the moment but only after a long time.

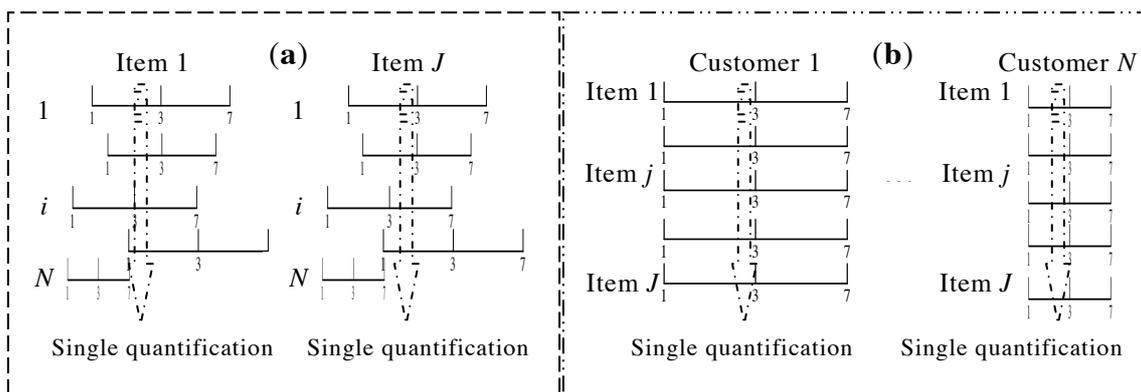


Figure 1: The approach based on the differences between the variables (a) and the single quantification based on the difference between the subjective scale of the customers (b).

To give a graphic example, Figure 1 compares the approach presented in the first section (a) and the alternative one based on the difference between the

customers (b). It is clear that by the first approach non homogenous customers have the same quantification, while in the second one the difference between the customers is respected. Moreover, each customer has a single quantification for all the items of the questionnaire and, in respect to the specific reference system, each customer has a quantification that preserves the subjective origin of the measuring system and the different distance between two points on the scale as the non-linear distance between two subsequent points.

In order to obtain a quantification that present the characteristics of the alternative approach we propose the following strategy. Before each generic row of matrix $\mathbf{R}_{N,J}$ ($\mathbf{E}_{N,J}$) is crisp coding in a matrix \mathbf{G}_i ($J \times \ell$) in accordance with Section 2.1. But the j^{th} row (with $j=1,\dots,J$) of \mathbf{G}_i is the coding of the rate that the i^{th} subject has given to the different items. Afterwards an Alternative Least Squares (ALS) algorithm is engaged to minimize the following loss function

$$N^{-1} \sum_{i=1}^N SSQ(\mathbf{X} - \mathbf{G}_i \mathbf{q}_i \boldsymbol{\beta}_i') \quad (2.3)$$

with $\hat{\mathbf{1}}_s \mathbf{X} = \mathbf{0}_J$, $\mathbf{X}' \mathbf{X} = \mathbf{J}$ and $\boldsymbol{\beta}_i' \mathbf{D}_i \boldsymbol{\beta}_i = 1$, where $\hat{\mathbf{1}}_s$ is a s dimensional vector of unit, $\mathbf{0}_J$ is a J dimensional vector of zero, \mathbf{X} a matrix of order ($J \times s$), $\boldsymbol{\beta}_i$ is a vector of score, \mathbf{q}_i is a vector of single category quantification for each single customer, and $\mathbf{D}_i = \mathbf{G}_i' \mathbf{G}_i$. Equation (2.3) presents the rank-one restrictions ($\mathbf{Y}_i = \mathbf{q}_i \boldsymbol{\beta}_i'$) that implies the quantifications in s dimensional space of each customer. To minimize the loss function (4) the ALS algorithm is proposed to search the optimal solution through the satisfaction of the two centroid principles with respect to \mathbf{X} and each \mathbf{Y}_i (van Rijckevorsel, 1987). The algorithm is given by the following step:

- *Step 0* Initialize the matrix \mathbf{X} by a singular decomposition analysis of \mathbf{R} (\mathbf{E}) so that $\hat{\mathbf{1}}_s \mathbf{X} = \mathbf{0}$ and $\mathbf{X}' \mathbf{X} = \mathbf{J}$
- *Step 1* Estimate the matrix $\hat{\mathbf{Y}}_i = \mathbf{D}_i^{-1} \mathbf{G}_i \mathbf{X}$, $i \in N$
- *Step 2* Estimate the vector $\hat{\boldsymbol{\beta}}_i = \hat{\mathbf{Y}}_i' \mathbf{D}_i \mathbf{q}_i / \mathbf{q}_i' \mathbf{D}_i \mathbf{q}_i$, $i \in N$
- *Step 3* Estimate the vector of single quantification $\hat{\mathbf{q}}_i = \hat{\mathbf{Y}}_i' \hat{\boldsymbol{\beta}}_i / \hat{\boldsymbol{\beta}}_i' \hat{\boldsymbol{\beta}}_i$, $i \in N$
- *Step 4* Update the matrix $\hat{\mathbf{Y}}_i = \hat{\mathbf{q}}_i \hat{\boldsymbol{\beta}}_i'$, $i \in N$
- *Step 5* Estimate the matrix $\mathbf{X} = N^{-1} \sum_{i=1}^N \mathbf{G}_i \hat{\mathbf{Y}}_i$
- *Step 6* Center and orthonormalize the matrix \mathbf{X}
- *Step 7* Go to *Step 1* until the convergence criterion is reached.

By the single quantification \hat{q}_i of each customer we are coding the perceived (expected) matrix into the quantified matrix \mathbf{R}^* (\mathbf{E}^*). In this way, the subjective scale of each customer has been respected. Further a multidimensional analysis of the gap between the performance and expectation, or the performance and expectation data, could be carried out separately.

The B -spline takes a variable as input and produces more than one variable as output. By B -spline a variable h_j ($h_j \in [a, b]$ with $a \neq b$) is partitioned into a number of intervals, where two boundary points, called knots, restrict every interval: $a \leq \kappa_1 < \dots < \kappa_K \leq b$, with exactly r “interior knots” (knots that meet another knot on both sides) and $K - r$ “exterior knots”.

Let piecewise coding function be a positive function between 0 and 1 on some contiguous part on range $[a, b]$. For each variable h_j exists s different piecewise coding functions represented by correspondingly s column-vectors $\mathbf{G}_k(h)$ $k = 1, \dots, s$, where the s column-vectors $\mathbf{G}_k(h)$ are collected in a pseudo-indicator matrix \mathbf{G} , where $\sum_{k=1}^s \mathbf{G}_{ik} = 1$. A B -spline is a particular piecewise coding function

with functional pieces of degree v , that is positive on exactly $v + 1$ consecutive intervals, with an overlap of exactly v intervals with the next B -spline, all intervals defined by the knot sequence κ , with $v \leq r$. The order of B -spline is equal to $v + 1$, while the number of B -splines needed to code a variable with knot sequence κ is equal to $s = v + r$. B -splines of zero degree are the most used for the good properties of the crisp coding (van Rijckevorsel, 1988).

More coding functions and relative characteristics are illustrated in Schumaker (1981). To have a global transformation we use a linear combination of B -spline

$$\phi(h^j) = \sum_{k=1}^s \alpha_k \mathbf{B}_k(h^j),$$

where the unknown spline coefficient α_i is the only

determining parameter of $\phi(\cdot)$ (Van Rijckevorsel, 1987). Nevertheless, other methods to determine the optimal coefficient are proposed in literature (De Boor, 1978; Schumaker, 1981).

3 Compositional data

The compositional data have particular properties that pose special problems for imputation and they can rarely be analyzed with the usual multivariate statistical methods. For each row of matrix $\mathbf{C}_{N,D}$ we define $\tilde{c}_1, \dots, \tilde{c}_D$ as positive quantities with the same measurement scale $\tilde{\mathbf{c}} = (\tilde{c}_1, \dots, \tilde{c}_D)$ $\tilde{c}_1 \geq 0, \dots, \tilde{c}_D \geq 0$ and $\|\tilde{\mathbf{c}}\|$ the trace of $\tilde{\mathbf{c}}$. The vector $\tilde{\mathbf{c}}$ is the basis of compositional data and $\mathbf{c} = \tilde{\mathbf{c}}/\|\tilde{\mathbf{c}}\|$ is a composition vector.

More generally, we define $\mathbf{C}_{N,D}$ a compositional data matrix if all elements are positive and each row is constrained to the unit-sum $\mathbf{C}\hat{\mathbf{1}}_D = \hat{\mathbf{1}}_N$ where $\hat{\mathbf{1}}_D$ and $\hat{\mathbf{1}}_N$ are vectors of units of D and N dimension, respectively. Let $\mathbf{Q} = \left[\mathbf{M}_N - \hat{\mathbf{1}}_N \hat{\mathbf{1}}_N' \right]$ be the product between $1/N$ and the usual centering projector then $\mathbf{C}'\mathbf{Q}\mathbf{C}$ is the covariance matrix of \mathbf{C} called *crude covariance matrix* (Aitchison, 1986). The unit-sum constraint for each row of \mathbf{C} implies four difficulties: 1) Negative bias, 2) Subcomposition, 3) Basis, 4) Null correlation. Each row and column of $\mathbf{C}'\mathbf{Q}\mathbf{C}$ has zero-sum: $\hat{\mathbf{1}}_D \mathbf{C}'\mathbf{Q}\mathbf{C} = \hat{\mathbf{0}}_D$ where $\hat{\mathbf{0}}_D$ is a D dimensional vector of zero. Therefore each variable has a covariance sum equivalent to negative variance (the first difficulty). No-relationship exists between the crude covariance matrix and the crude subcomposition covariance one. Therefore the variation of subcomposition can substantially influence the covariance (the second difficulty). Likewise in the subcomposition, it is not easy to select a basis $\tilde{\mathbf{c}}$ for the composition (which is the third difficulty). Like the *crude covariance matrix*, each row and column of the *crude correlation matrix* of \mathbf{C} has a zero-sum. Therefore the correlation between two variables is not free to range over the usual interval $[-1,1]$. The negative bias causes a radical difference from the standard interpretation of correlation between variables. Zero correlation between two ratios does not mean that there is no association (the latter difficulty). Moreover the uninterpretable crude covariance structure is not the only problem of compositional data. Unfortunately, compositional data often exhibit curvature when standard multivariate methods are employed.

Aitchison (1986) richly described the properties of compositional data and proposed an alternative form of logratio, where the more useful is based on a geometric mean $g(\mathbf{c})$. Replacing the natural non-negative condition by the following stronger assumption of the strict positive quantities: $\tilde{w}_{i1} > 0, \dots, \tilde{w}_{iD} > 0$ (see Gallo, 2003); Aitchison (1982) proposes to transform each element of \mathbf{C} (c_{ij}) in the logratio $\log[c_{ij}/g(\mathbf{c})]$, because the relative matrix of centred logratio \mathbf{Z} , with generic element $z_{ij} = \log[c_{ij}/g(\mathbf{c})]$ is adequate for a low-dimensional description of compositional variability. Moreover, a generalization of the logratios – called *logcontrasts* – have particular and researched properties in compositional data analysis. Logcontrast of \mathbf{c} is any loglinear combination $\mathbf{u}'\log \mathbf{c} = u_1 \log c_1 + \dots + u_D \log c_D$ with $u_1 + \dots + u_D = 0$, where of logcontrast with the geometric mean $g(\mathbf{c})$ presents the property: $\mathbf{u}'\log \mathbf{c} = \mathbf{u}'\log(\mathbf{c}/g(\mathbf{c}))$.

The study of composition is essentially concerned with the relative magnitudes of c_{i1}, \dots, c_{iD} rather than their absolute values. In this case, ratios between components are meaningful, and those ratios are independent from the arbitrary total. Moreover any logcontrast is scale free: $\mathbf{u}'\log \mathbf{c} = \mathbf{u}'\log k\mathbf{c}$ (with $k > 0$).

Aitchison (1986) richly describes how the logcontrast transformation is adequate to resolve the difficulties of compositional data. Barceló-Vidal, Martín-Fernández and Pawlowsky-Gòahn, (2001) show, from a mathematical point of view, that this transformation is not arbitrary.

4 External information in CS analysis

Before a multidimensional analysis of the matrix \mathbf{R}^* we could include the additional information available on the customers or process (Takane and Shibayama, 1991). In this way the additional information defines a priori levels of sampling hierarchical structures, since it permits us to investigate how well structures supplied by the a priori information can account for the data.

Let \mathbf{H} (N, Q with Q the number of predictor categories, $Q \leq N$) be the external informational matrix, which can take a variety of forms. We can consider the following decomposition model of the matrix \mathbf{R}^* :

$$\mathbf{R}^* = \mathbf{H}\mathbf{T} + \mathbf{E} \quad (4.1)$$

where $\hat{\mathbf{T}} = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{R}^*$ is the estimated coefficient matrix and \mathbf{E} the error matrix. Each term of the model is column-wise orthogonal which implies that the sum of squares of \mathbf{T} is decomposed into the sum of squares of the components of (5). The problem of estimating $\hat{\mathbf{T}}$ is equivalent to minimizing $SS(\mathbf{E}) = tr(\mathbf{E}'\mathbf{E})$ where $\hat{\mathbf{E}} = \mathbf{R}^* - \mathbf{H}\hat{\mathbf{T}} = \mathbf{R}^* - \mathbf{P}_H\mathbf{T}$ and $\mathbf{P}_H = \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'$ is the orthogonal projection operator onto the subspace spanned by the column vectors of \mathbf{H} , so that (4) is decomposed into two additive components: $\mathbf{R}^* = \mathbf{P}_H\mathbf{R}^* + \mathbf{P}_H^\perp\mathbf{R}^*$ where \mathbf{P}_H^\perp is the orthogonal projection operator that is orthogonal to \mathbf{P}_H . Once the data matrix is decomposed according to the additional information (External Analysis), multidimensional analysis is carried out on $\mathbf{P}_H\mathbf{R}^*$ and $\mathbf{P}_H^\perp\mathbf{R}^*$ separately (Internal Analysis). An analysis of $\mathbf{P}_H\mathbf{R}^*$ allows us to incorporate the external information into the analysis, whereas the analysis of $\mathbf{P}_H^\perp\mathbf{R}^*$ allows us to exclude them. Incorporating the external information we have the evaluation of the performance perception in a more precise and objective way because we have the degree of satisfaction of homogeneous customer clusters. Differently, incorporating the second additive component (analysis on $\mathbf{P}_H^\perp\mathbf{R}^*$) we have the evaluation of the performance perception while excluding the influence of the external information. Similarly we carry out an analysis with the external information on the matrices \mathbf{E}^* , $\mathbf{R}^* - \mathbf{E}^*$, \mathbf{C}^* and so on.

It is also possible to include the external information before the quantification of the raw data (D'Ambra *et al.*, 2002) to obtain a more parsimonious

representation of the data. Nevertheless we have to respect the original scale of the external information and it is not always possible if the external information has different scale systems.

5 Conclusion and perspectives

The central theme of this paper is the quantification of expected (perceived) data. Different scaling methods are proposed in literature to quantify these data (see Zanella, 1999) and most of them have a large number of researched properties. Nevertheless, here is a new approach to scaling expected (perceived) data because the preservation of the subjective scale of each customer is necessary for an accurate multivariate analysis. Moreover a strategy that preserves the rule and the properties of the row data that we collect with a *SQ* analysis is a secondary aim of this paper. The strategy that we propose is based on the following steps:

- quantification of expected (perceived) data by an approach that preserves the subjective scale of each customer,
- inclusion of the additional information available on the customers or process,
- research of the latent factors by more independent analysis.

As further developments, we are comparing the different scaling methods with the appreciable monotone property. For example, the *M*-splines are proportional to *B*-splines and a basis of integrated *M*-splines (*I*-spline) have the characteristics of a probability distribution function. Moreover we are checking on real data for the benefit to use the approach that we have proposed.

Relative to the importance data, the compositional scale of the kind of data should be considered because the analysis of these data without transformation give misleading information (Aitchison, 1986). Logcontrast transformation is the most used for compositional data afterward this data could be used in *SQ* analysis in accordance to the strategy given before.

Acknowledgement

August Viglione is thanked for revising the original text. The referees are also thanked for many useful comment. The present paper is financially supported by the University of Naples – L’Orientale (Department of Social Science 60% fund 2006).

References

- [1] Aitchison, J. (1982): The statistical analysis of compositional data (with discussion). *Journal of the Royal Statistical Society*, **44**, 139-177.
- [2] Aitchison, J. (1986): *The Statistical Analysis of Compositional Data*. London: Chapman and Hall.
- [3] Barceló-Vidal, C., Martín-Fernández, J.A., and Pawłowsky-Glahn, V. (2001): *Mathematical Foundations of Compositional Data Analysis*. Proceedings of IAMG'01 – The sixth annual conference of the International Association for Mathematical Geology. Electronic publication.
- [4] D'Ambra, L., Amenta, P., and Gallo, M. (2002): *Riflessioni sulla Valutazione dei Servizi di Day Surgery nel contesto dell'Analisi Multidimensionale dei dati*. In Frosini et al. (Eds): Vita & Pensieri, 153-165.
- [5] D'Ambra, L., Amenta, P., and Gallo, M. (2006): La valutazione della Customer Satisfaction. In Carpita et al. (Eds): *Guerini Studio*, 267-290.
- [6] De Boor, C. (1978): *A practical Guide to Spline*. Berlin: Springer-Verlag.
- [7] De Leeuw, J. and van Rijckeversel, J. (1980): Homals en Princals. In Diday et al. (Eds): *Data Analysis and Informatics*.
- [8] ECSI Technical Committee (1998): *European Customer Satisfaction Index, Foundation and Structure for Harmonized National Pilot Projects*. Report Prepared for the ECSI Steering Committee.
- [9] Eklöf, J.A. (2000): *European Customer Satisfaction Index Pan-European Telecommunication Sector Report Based on the Pilot Studies 1999*. Stockholm, Sweden. European Organization for Quality and European Foundation for Quality Management.
- [10] Fornell, C., Johnson, M.D., Anderson, E.W., Cha, J., and Bryant, B.E. (1996): The American customer satisfaction index. Nature, purpose and findings. *Journal of Marketing*, **60**, 7-18.
- [11] Gallo, M. (2003): Partial Least Squares for compositional data: An approach based on the splines. *Italian Journal of Applied Statistics*, **15**, 349-358.
- [12] Gifi, A. (1982): *Princals User's Guide*. Leiden: Department of Data Theory.
- [13] Kruskal, J.B. and Shepard, R.N. (1974): A nonmetric variety of linear factor analysis. *Psychometrika*, **39**.
- [14] Oliver, R.L. (1993): A conceptual model of service quality and service satisfaction: compatible goals, different concepts. In *Advances in Service Marketing and Management: Research and Practice*, **2**, JAI Press.
- [15] Parasuraman, A., Zeithman, V.A., and Berry, L.L. (1994): Reassessment of expectations as a comparison standard in measuring service quality: Implications for further research. *Journal of Marketing*, **58**, 111-124.
- [16] Ramsay, J.O. (1982): When data are functions. *Psychometrika*, **47**, 379-396.

-
- [17] Schumaker, L.L. (1981): *Spline Functions*. New York: John Wiley & Sons.
- [18] van Rijckevorsel, J. and De Leeuw, J. (1988): *Component and Correspondence Analysis Dimension Reduction by Functional Approximation*. New York: John Wiley & Sons.
- [19] van Rijckevorsel, J. (1987): *The Application of Fuzzy Coding and Horseshoes in Multiple Correspondence Analysis*. DSMO Press.
- [20] Winsberg, S. and Ramsay, J.O. (1980): Monotonic transformations to additivity using splines. *Biometrika*, **67**, 669 - 674.
- [21] Winsberg, S. and Ramsay, J.O. (1983): Monotone spline transformations for dimension reduction. *Psychometrika*, **48**.
- [22] Young, F.M., De Leeuwe, J., and Takane, Y. (1976): Regression with qualitative variables: an alternating least squares method with optimal scaling features. *Psychometrika*, **41**.
- [23] Young, F.M., Takane, Y., and De Leeuwe, J. (1978): The principal components of mixed measurement level multivariate data: an alternating least squares methods with optimal scaling features. *Psychometrika*, **46**, 357-388.
- [24] Zanella, A. (1999): A Stochastic model for the analysis of customer satisfaction: some theoretical aspects. *Statistica*, LIX.