Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal. (2011), Vol. 4, Issue 1, 1 – 22
e-ISSN 2070-5948, DOI 10.1285/i20705948v4n1p1

A GRAPHICAL TOOL TO COMPARE GROUPS OF SUBJECTS ON CATEGORICAL VARIABLES

Marica Manisera[*]

Department of Quantitative Methods, University of Brescia, Italy.

Received 20 February 2010; Accepted 17 June 2010
Available online 26 April 2011

Abstract: *This paper proposes a graphical statistical tool easy to interpret that can be used to compare the attitude of different groups of subjects (individuals or organizations) with respect to categorical variables. The construction of the proposed graph is based on the combination of (i) an unusual application of the Nonlinear Principal Components Analysis, oriented to quantify categorical variables and focused on the so-called Projected Centroid Plot, (ii) the Inferential Confidence Intervals, and (iii) a nonparametric bootstrap study. An application investigates the quality of work in social cooperatives by exploring the relations between quality of work and characteristics of workers (gender, age, education, membership) and cooperatives (geographical area, type – A or B, dimension – in terms of number of workers). Results easily show how the groups of workers perceive the different aspects of the quality of work.*

Keywords: *Nonlinear Principal Components Analysis, Projected Centroids, Bootstrap, Inferential Confidence Intervals*

1. Introduction

Categorical data are common in many research fields, in marketing, education, genetics, social, economical, behavioural, and biomedical sciences, and many others. For example, in the social and economical sciences subjective data like individuals' attitudes and perceptions (e.g. customer satisfaction) are often collected through the administration of questionnaires, with several items referring to different aspects of the concept being measured. Responses usually indicate the degree of agreement with each statement, with higher scores reflecting a higher degree of agreement. Consequently, the variables resulting from the questionnaire are ordered

[*] E-mail: manisera@eco.unibs.it

categorical (i.e., ordinal) variables. Attention has to be paid to the treatment of ordinal variables, because we cannot assume a priori that the distances between the categories are equal. In analyzing ordinal variables it should be taken into account that the categories of the variable have a fixed a priori order, but this should not be taken to imply that the differences between numeric labels of the categories should be maintained; therefore, models for continuous data should not be used.

In recent years, most scientists and statisticians have realized that it is unnecessary and often inappropriate to use techniques for continuous data when dealing with categorical data; this led to an increase in the development and use of specialized statistical methods and models for categorical data (see, for example, [1]). Also from a data analysis perspective, statistical methods as well as graphical representations must take into account the quantitative or categorical nature of the analysed variables. Moreover, some of the plots suitable for ordinal variables are not appropriate for nominal variables.

This paper proposes a graphical statistical tool easy to interpret that can be used in a broad variety of situations whenever the focus is on the comparison of different groups of subjects (individuals, organizations, etc.) with respect to both nominal and ordinal categorical variables. Therefore, we consider to have one (or more) grouping variable(s), classifying subjects into groups, and we want to compare the attitudes of such groups with respect to one (or more) analysed categorical variable(s). For example, we want to compare the level of job satisfaction of males and females. In multivariate data analysis, several graphs representing categorical data by groups have been proposed (see, for example, [30]); however, when the number of categories of the analysed variable(s) and/or the number of groups of the grouping variable(s) are large, such graphs become difficult to read; moreover, some of them often require analysed variables measured on at least ordinal scales, like the box plots by groups. The proposed graphical tool can also be used when the number of categories of analysed and grouping variables is quite large and the analysed categorical variables are nominal or ordinal. Its use can be extended to quantitative data, when the aim is to compare the position of different groups on those variables.

Although the proposed graphical tool can be constructed in several situations, we consider to start from a data matrix (subjects x variables) obtained by the administration of questionnaires. The construction of this graph is based on the combination of (*i*) an unusual application of the N*on*L*inear* P*rincipal* C*omponents* A*nalysis* (NL-PCA: [14], [26]); (*ii*) the I*nferential* C*onfidence* I*ntervals* (ICIs: [32], [15]); and (*iii*) a nonparametric bootstrap study ([12], [13]).

For each analysed categorical variable, the idea is to represent the position of groups on that variable by points, with associated intervals helping the interpretation of the different positions and, in particular, allowing a graphic test of the statistical differences. In the literature, an exploratory plot derived by the standard use of NL-PCA with points representing groups already exists (the P*rojected* C*entroids* P*lot*, PCP: [25]) and ICIs have been proposed as an inferential graphical tool to test statistical differences ([32], [15]). The original contribution of this paper does not only consists in the use of an unusual application of NL-PCA to derive the PCP as well as the completion of the PCP, with the representation of some elements, helping the interpretation of the position of points. The most important contribution is the introduction of inferential issues by combining the PCP with ICIs and, in particular, by using the nonparametric bootstrap procedure in order to obtain the desired intervals.

In the present paper, the proposed graph was used to analyse real data coming from the survey on the Italian Social Cooperatives called ICSI^{2007} ([8]), with the aim of evaluating the quality of work in social cooperatives. The perceptions of different groups of workers, with respect to categorical variables of quality of work, are investigated. In particular, the relations between quality of work and characteristics of workers (gender, age, education, membership) and cooperatives (geographical area, type – A or B, dimension – in terms of number of workers) are explored.

2. Methods

Starting from a data matrix (subjects x variables), we propose to construct a graphical tool to compare groups on categorical variables by combining (*i*) the application of the NL-PCA, devoted to transform categorical variables into quantitative ones with metric properties (allowing the computation of group mean values) and focused on the graphical representation of "projected centroids"; (*ii*) the ICIs, an inferential graphical tool allowing to test the null hypothesis of equal means (for pairwise comparisons) by simply checking the overlapping of two intervals, and (*iii*) a nonparametric bootstrap study, usually used to assess some stability issues in the NL-PCA context, but here especially used to obtain bootstrap standard deviations to be used in the construction of intervals.

2.1 The NonLinear Principal Components Analysis (NL-PCA) in brief

NL-PCA is the technique chosen to take the categorical nature of variables into account. It is the nonlinear equivalent of classical PCA (see, for example, [16], [34]) and it simultaneously reduces the dimensionality of the data and transforms categorical variables into quantitative ones, by means of *optimal scaling* that assigns optimal quantifications to the original categories ([14], [26]). We consider a $n \times m$ data matrix $\mathbf{H}=[\ \mathbf{h}_1\ |\dots|\ \mathbf{h}_j\ |\dots|\ \mathbf{h}_m]$, where m is the number of variables (or items) observed on n subjects. The j-th item has k_j categories contained in vector $\mathbf{c}_j'=(1,\dots,k_j)$, $j=1,\dots,m$. Each categorical variable $\mathbf{h}_j$ defines a $n \times k_j$ binary indicator matrix $\mathbf{G}_j$ such that
$\mathbf{h}_j = \mathbf{G}_j\ \mathbf{c}_j$. The dimensionality reduction consists in an orthogonal projection from the $\mathcal{R}_m$ space to the $\mathcal{R}_p$ space, with $p<<m$.

The optimisation problem is solved by minimizing the loss function $\sigma = \sum_j Sq(\mathbf{X}-\mathbf{q}_j\mathbf{a}'_j) = \sum_j Sq(\mathbf{X}-\mathbf{G}_j\mathbf{y}_j\mathbf{a}'_j)$, where $Sq(\cdot)$ stands for the sum of squared elements of a matrix or a vector, $\mathbf{X}$ is the $n \times p$ matrix containing the scores of the n subjects on the p dimensions (components) of $\mathcal{R}_p$, $\mathbf{q}_j$ is the $n \times 1$ vector of the j-th quantified variable (transformation of the original variable $\mathbf{h}_j$) and $\mathbf{y}_j$ is the $k_j \times 1$ vector of the category quantifications (quantifications of categories $\mathbf{c}_j$), $\mathbf{a}_j$ is the $p \times 1$ vector of the component loadings corresponding to the j-th variable. The solution is found by identifying the optimal values for $\mathbf{X}$, $\mathbf{a}_j$ e $\mathbf{y}_j$ by means of an Alternating Least Squares (ALS) algorithm that minimizes σ with respect to $\mathbf{X}$ (for fixed $\mathbf{a}_j$) and with respect to $\mathbf{a}_j$ (for fixed $\mathbf{X}$), with a further internal ALS loop, alternate over $\mathbf{y}_j$ and $\mathbf{a}_j$. The minimization process is constrained, because orthonormalization constraints are imposed to avoid trivial solutions ([14]).

NL-PCA finds category quantifications that are optimal in the sense that the overall variance accounted for in the transformed variables, given the number p of components, is maximized. In the optimal scaling process, information in the original categorical data is retained in the optimal

quantifications, depending upon the optimal transformation function (or scaling level) that can be chosen for each variable separately ([26], [20]).
In NL-PCA, all variables are transformed according to a *single* scaling level; when a single scaling level is chosen for a variable, each category of the variable receives only one quantification, valid for all dimensions. The j-th transformed variable $\mathbf{q}_j$ can be written as $\mathbf{q}_j = \mathbf{G}_j\ \mathbf{y}_j$ and the m transformed variables $\mathbf{q}_j$, j=1,…,m, are enclosed in the $n \times m$ matrix $\mathbf{Q}$, where the original scores for the individuals are replaced by the quantification of the category a subject scored in.
By contrast, when a *multiple* scaling level is chosen for a variable, that variable receives multiple quantifications (a separate quantification for each dimension). The multiple quantifications obtained by the k_j categories of variable j are contained in the $k_j \times p$ matrices $\mathbf{Y}_j$ and not just one but multiple (one per dimension) transformed data matrix $\mathbf{Q}_s$, s=1,…,p, are obtained.
In NL-PCA, categorical variables can be quantified by means of different *single* scaling levels, differing on the level of information contained in the original categorical variables and maintained in the transformed variables. The least restrictive level (requiring less restrictions and retaining the least amount of information) is the *nominal* scaling level, able to preserve in the category quantifications only grouping information in the original categories, allowing for a non-monotonic transformation. The *ordinal* and *spline ordinal* scaling levels preserve grouping and ordering information, resulting in a monotonic transformation (usually, monotone non-decreasing transformation, with reference to the original categories, are derived from a weighted monotonic regression process ([18], [19], [2]). Both nominal and ordinal transformations can also be obtained by spline transformations ([29]), which require the estimation of a lower number of parameters and result in smoother transformations (but at the cost of lower fit) than their non-spline counterparts.
Finally, the *numerical* scaling level is the most restrictive level, preserving not only grouping and ordering, but also interval information, resulting in a linear transformation (choosing numerical scaling level for all variables, NL-PCA results are equal to classical PCA results).
NL-PCA is useful when dealing with categorical variables, but also with numerical variables when they are supposed to be related by nonlinear relations.

2.2 The Projected Centroids Plot (PCP)

NL-PCA can be applied in order to obtain composite indicators of latent variables ([9]), but in this paper, we focus on a further aim of NL-PCA, that is the graphical representation of the analysed variables and the relations between variables and subjects. We refer to the *vector model* ([26]) that represents, in the same low-dimensional $\mathcal{R}_p$ space, (optimally scaled) variables by vectors and subjects by points (Figure 1a). Subjects can also be grouped according to a grouping variable (for example, gender) and represented by centroids; each centroid corresponds to one group and its coordinates are given by averaging the coordinates of the subjects belonging to that group. For example, in Figure 1a the n subjects are represented by n single points but also by five centroids, corresponding to five categories A, B, C, D, E of a grouping variable. At the same time, centroids represent both groups of subjects and categories of a variable (the grouping variable). When variables are represented by points (i.e., the centroids) associated to their categories, we are in the framework of the *centroid model* ([26]), in which variables are quantified according to a multiple scaling level and their categories receive a distinct quantification for each dimension in the solution. When the multiple scaling level is adopted for

all the variables in analysis, a multiple correspondence analysis, or homogeneity analysis or dual scaling, is being performed ([4], [14], [27]).
To interpret the relations between subjects and variables, each point or centroid can be projected onto the vectors[1] representing variables. Considering each variable separately, this projection allows to identify similarities and differences in the attitudes of (groups of) subjects, with reference to the considered variable. When the dimensionality reduction leads to two or three-dimensional spaces, graphical representations are guaranteed. However, it is possible to represent the projection of centroids also with higher dimensional spaces by the P*rojected* C*entroids* P*lot* (PCP; [25]).
In the PCP (Figure 1b), each variable is represented by a straight (usually vertical) line and the groups of subjects by points onto that line. The position of points results from the projection of centroids onto the variable vector in the NL-PCA solution space. The centroids to be projected belong to variables treated with a multiple nominal scaling level while the variables on which centroids are projected are necessarily treated with single scaling levels. The projected centroids of variable l on variable j, $j \notin J$, (where J is the index set recording which variables have multiple scaling level) are given by $\mathbf{Y}_l\mathbf{a}_j(\mathbf{a}_j'\mathbf{a}_j)^{-1/2}$.
The PCP allows one to identify particular groups in the data that stand out on selected variables ([26]). In Figure 1b, the same five centroids A, B, C, D, E represented in Figure 1a are projected onto the variable “var2”. This projection is the same as the one made in interpreting the biplot in Figure 1a, but now the projections are shown on a straight line representing the variable “var2”. It is also possible to represent more than one active variable by more straight parallel lines on which the centroids of one grouping variable are projected. Alternatively, more straight parallel lines can represent the same active variable on which centroids of more grouping variables are projected (see Section 3). The PCP in Figure 1b shows how each group, created according to the grouping variable, scores on the variable “var2”. The position of each group can be evaluated with respect to other groups (points close each other represent groups with similar attitude - on average - on that variable), to the general mean (corresponding to the zero quantification, because quantified variables are standardized) and to the variable categories, that we propose to represent on the same graph by their quantifications. The evaluation of differences and similarities can be done in terms of metric distance. For example, Figure 1b shows that, on average, subjects belonging to D and E groups have similar position (with reference to the variable “var2”), and the same holds for subjects of B and C groups, while subjects of A group differ from every other group (on average).
The present paper proposes to use a PCP resulting from a NL-PCA solution in which $p=m-1$, where p is the number of dimensions maintained in the solution and m is the number of the analysed variables. This application of NL-PCA allows to keep a very large amount of variability of the original variables in the final solution: in other words, the loss of information is negligible. The attention is therefore not on the dimensionality reduction but only on the variable quantification, obtained by optimal scaling. The advantage of this approach is to maintain nearly the total amount of information in the new representation of the data, given by the PCP.

[1] Usually, in the biplot of variables and subjects (Figure 1a), the vector corresponding to the j-th variable originates at the origin of the axes coordinates and ends in the point with coordinates given by the px1 vector $\mathbf{a}_j$, where p is the number of components in the NL-PCA solution; actually, each vector originates in the point with coordinates given by $-\mathbf{a}_j$, and the projection of points and centroids must take into account this extension.

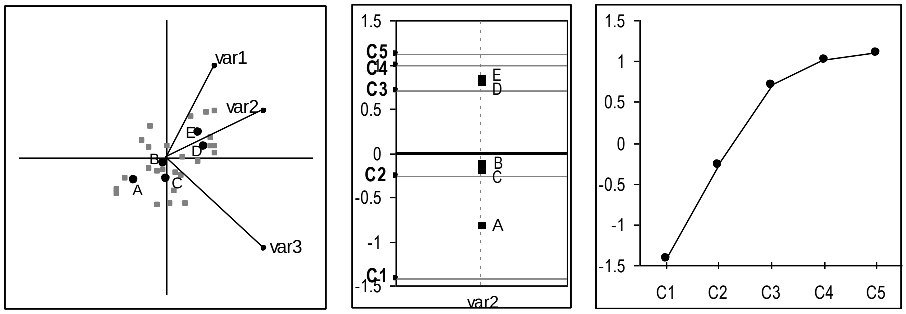


Figure 1. (a) Biplot of subjects (▪ points), three variables “var1”, “var2”, and “var3” (vectors) and centroids (A, B, C, D, E points); (b) projection of centroids on the variable “var2” and (c) transformation plot of the variable “var2”.

The interpretation of the PCP can be improved by assigning a meaning (related to the original categories of the variable) to the line representing the variable on which centroids are projected. This is easily achieved by analyzing the transformation plots, one per variable, in which the (horizontal) *x*-axis displays the original categories $\mathbf{c}_j$ of the variable and the (vertical) *y*-axis the corresponding category quantifications $\mathbf{y}_j$. The transformation plot shows the nonlinear transformation allowing the quantification of each original variable (for example, Figure 1c reports the transformation of the variable “var2”). Quantifications $\mathbf{y}_j$ can be used to assign a meaning, related to the original categories of the variables, to the vertical line in the PCP (Figure 1b). This allows not only to study the uniformity or diversity in the attitudes of groups of subjects, but also to interpret the position of group points with reference to the original categories of the variable and evaluate the metric distance among groups and between groups and categories.

It is interesting to note, for example, that D and E groups are located (on average) between the categories C3 and C4 of the variable “var2”, and both B-C and A groups are close to the category C2 of the variable “var2”, even if the distance from C2 is larger for A than B-C.

Moreover, since quantified variables are standardized (zero mean and unit variance), the zero on the vertical line of PCP represents the mean quantification. This indicates where the subject mean is located, with respect to the original categories of the variable: for example, Figure 1b suggests that the subject mean is between the categories C2 and C3 (though closer to C2). In addition, the position of groups can be evaluated not only with respect to each other, but also with respect to the mean of the whole set of subjects: Figure 1b shows that A, B, and C groups are below, while D and E are above the general mean.

The group size obviously impacts on the position of centroids: a larger group size implies a higher contribute of that group to the general subject mean and the corresponding centroid will be closer to zero. It should be noted that the frequency distributions of variables do have an effect on the optimal quantification of categories and, therefore, on the NL-PCA results. In particular, categories with very low marginal frequencies tend to receive quantifications similar or equal to those of adjacent categories, suggesting a recoding of that variable by merging categories with equal quantifications. In the presence of low marginal frequencies, NL-PCA results can show some instability: this question has been faced in different ways in the literature ([21]).

Since the NL-PCA is a multivariate data analysis technique, the final solution takes into account all the variables in the analysis and the relations among them. In this paper, the objective is to study the relations between some variables, on one hand, and the groups of subjects defined by some grouping variables, on the other hand. We want that such relations do not have effect on the NL-PCA solution. For this reason, the grouping variable are treated as passive or supplementary variables; therefore, their quantifications are computed in a second moment, when the ALS algorithm has already converged. Handling a variable as supplementary ensures that it does not influence the solution, but it can be displayed in the solution for illustrative purposes. Also, more important for this paper, selecting the multiple scaling level for a grouping variable, it is possible to analyse multivariate data on a group level, rather than on an individual level.

As mentioned before, the PCP is already present in the CATPCA program of SPSS ([25]). In this paper the PCP is derived by un unusual application of NL-PCA and extended via the representation of the quantified categories of the active variable on the vertical axis and the mean quantification by the horizontal line corresponding to zero. But the most important extension is the introduction of inferential issues by the confidence intervals associated to centroids, as explained in the next subsection 2.3.

2.3 Completing the PCP by Bootstrap ICIs

The NL-PCA is a descriptive data analysis technique and it was developed from an exploratory point of view. In this sense, the PCP resulting from the application of NL-PCA cannot give an answer to the following question: "Are the group means (centroids) statistically different or not?". Therefore, the PCP allows to interpret the position of the centroids with respect to each other, to the quantifications of the original categories, to the general subject mean, but it does not allow to draw inferential conclusions on those positions. In order to introduce inferential issues on PCP, in this paper we propose to complete the PCP with the construction of the I*nferential* C*onfidence* I*ntervals* (ICIs: [32], [15]), a graphic test of statistical difference designed to avoid common interpretative problems associated with the null hypothesis statistical testing. Graphed confidence intervals can be used for overlap pairwise comparisons as an inferential graphical tool at the stated significance level only after reducing their widths: the reduced statistical intervals have been named ICIs by Tryon ([32]) and thanks to the reduction, nonoverlapping ICIs are algebraically equivalent to a null hypothesis statistical test at the stated significance level. When dealing with large samples, the two ICIs at the approximate level α corresponding to groups A and B can be defined as ([15]):

$$\mathrm{CI_A} = [m_\mathrm{A} \pm z_{\alpha/2}\, e_\mathrm{AB}\, s_\mathrm{A}] \text{ and } \mathrm{CI_B} = [m_\mathrm{B} \pm z_{\alpha/2}\, e_\mathrm{AB}\, s_\mathrm{B}] \qquad (1)$$

where m_A and m_B are the sample means, $s_\mathrm{A} = \sqrt{d_\mathrm{A}^2 / n_\mathrm{A}}$ and $s_\mathrm{B} = \sqrt{d_\mathrm{B}^2 / n_\mathrm{B}}$ are the estimated standard errors with d_A^2 and d_B^2 the unbiased sample variances, $z_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard Normal distribution and $e_\mathrm{AB} = s_\mathrm{AB} / (s_\mathrm{A} + s_\mathrm{B})$ with $s_\mathrm{AB} = \sqrt{s_\mathrm{A}^2 + s_\mathrm{B}^2}$ is the estimated ratio ε_AB, $1/\sqrt{2} \le \varepsilon_\mathrm{AB} \le 1$, necessary to reduce the width of confidence intervals, in order to reach the equivalence between nonoverlapping ICIs and null hypothesis statistical test at the stated significance level. If the ICIs to be represented on the same graph are more than two, $g_\mathrm{AB} = z_{\alpha/2}\, e_\mathrm{AB}$ can be replaced by the mean g of the g_AB's computed over all the possible pairs (A,B) ([15]).

Because the reduction of the standard confidence intervals depends on the unknown population variances that must be estimated, for small or moderated sample sizes the ICIs should be extended using the *t*-test ([32]) or the Welch-test ([7]).
In the present paper, we propose to use the projected centroids corresponding to two groups as m_A and m_B in formulae (1) and to conduct a bootstrap study (with a number R of replications) in order to obtain the bootstrap estimates for standard errors of the NL-PCA projected centroids, to be used as s_A and s_B in the construction of ICIs. The obtained inferential confidence intervals can be referred to as Bootstrap ICIs (BICIs). Results can be nicely represented on the PCP by intervals associated with each single centroid; the comparison between groups of subjects can be easily achieved by checking the overlapping between the BICIs associated with the corresponding projected centroids.
We chose a nonparametric bootstrap study ([12], [13]) in line with recent works introducing inferential issues to NL-PCA studies. For example, the nonparametric bootstrap procedure has been used ([24], [21]) in order to establish the stability of the results of the NL-PCA solution.

3. Case study and Results

The PCP with BICIs described in Section 2 was applied to real data coming from the survey on the Italian Social Cooperatives called ICSI^{2007} ([8]) in order to study the quality of work in social cooperatives. The attention was devoted to groups of workers defined by individual characteristics and characteristics of the social cooperative in which they are employed. These groups were compared with respect to some subjective variables of quality of work.
Although objective aspects of the quality of work in social cooperatives (for example referring to the characteristics of the contract) could also have been considered, without loss of generality this paper considers only subjective aspects of the quality of work, referring to workers' perceptions and attitudes (see, for example, [3], [5], [17]). It should be noted that also the objective variables are usually measured by categorical variables originated by the administration to the workers of questionnaires; therefore, they can also be analysed by the proposed graphical tool.
The position of groups of workers - with different characteristics (gender, education, age, membership) employed in cooperatives of different geographical area, type (A or B) and dimension, in terms of number of workers - was analysed, with reference to the following variables of quality of work: DISTRIBUTIONAL and PROCEDURAL FAIRNESS, importance of interpersonal relations in the workplace (IMPORTANCE OF RELATIONS), effort put into the work and required by job tasks (EFFORT), organizational commitment (COMMITMENT), PAY and JOB SATISFACTION, loyalty to the cooperative (LOYALTY). The study involved 3,914 workers, obtained by excluding the subjects who answered "I do not know" to the distributional fairness item from the 4,134 workers included in the survey. All the considered variables result from the submission of single-item scales in the questionnaire and, except for LOYALTY, which is a nominal variable, all the variables are ordinal. The detailed description of the analysed variables and corresponding response scales are reported in Table 1, while their frequency distributions are in Figure 2.

Table 1. The eight variables of quality of work considered in the analysis.

Variable	Question	Response scale
DISTRIBUTIONAL FAIRNESS	Do you think that your pay is fair in general?	from 1 = "much less than fair" to 7 = "much more than fair" (with 4 = "fair" and ? = "I do not know")
PROCEDURAL FAIRNESS	The cooperative correctly behaves with respect to you (fair procedures)	from 1 = "not at all agree" to 7 = "strongly agree"
IMPORTANCE OF RELATIONS	How much important are the interpersonal relations in the workplace?	from 1 = "not at all important" to 7 = "strongly important"
EFFORT	How much effort do you usually put into your work?	from 1 = "at all" to 7 = "very much"
COMMITMENT	How much does the cooperative involve you in its mission, to recognize your work and motivate you to make it as better as possible?	1= "never"; 2 = "rarely"; 3 = "sometimes"; 4 = "often"; 5 = "always"
PAY SATISFACTION	How satisfied are you with your pay?	from 1 = "very dissatisfied" to 7 = " very satisfied" (with 4 = "neither dissatisfied nor satisfied")
JOB SATISFACTION	How satisfied are you with your job?	from 1 = "very dissatisfied" to 7 = " very satisfied" (with 4 = "neither dissatisfied nor satisfied")
LOYALTY	What are your future intentions with reference to this cooperative?	1= "stay as longer as possible, because I like my work and the workplace"; 2 = "stay as longer as possible, because I don't have any alternative"; 3 = "stay but not for long"; 4 = "leave as soon as possible"

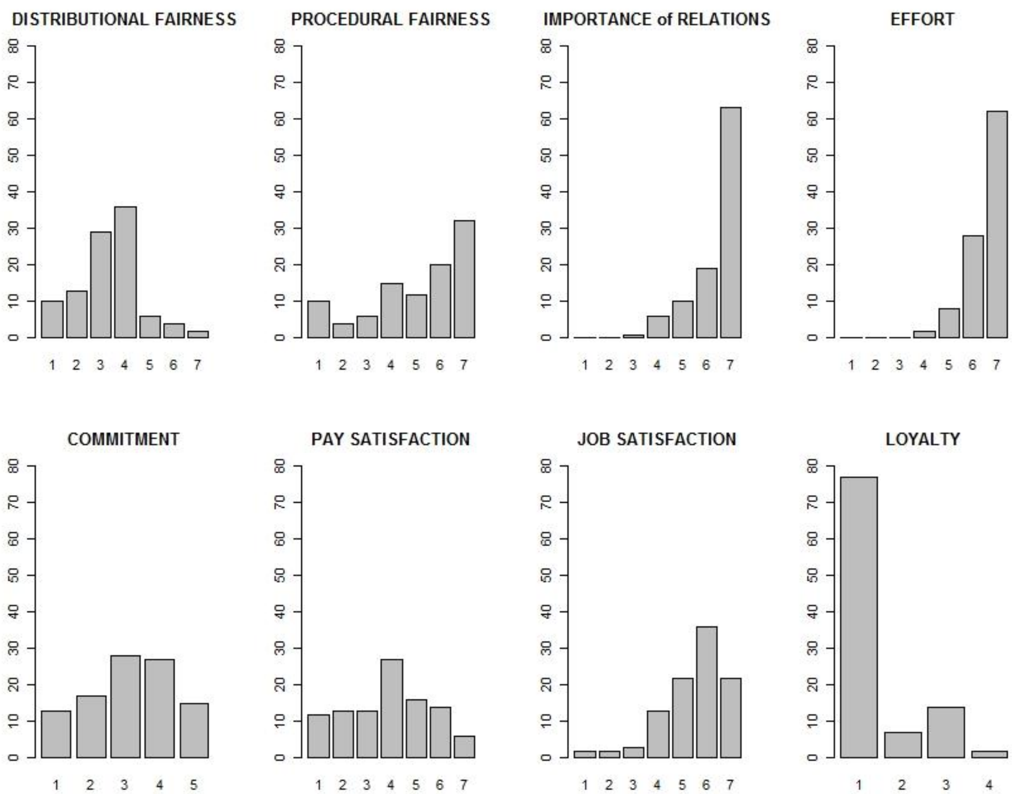


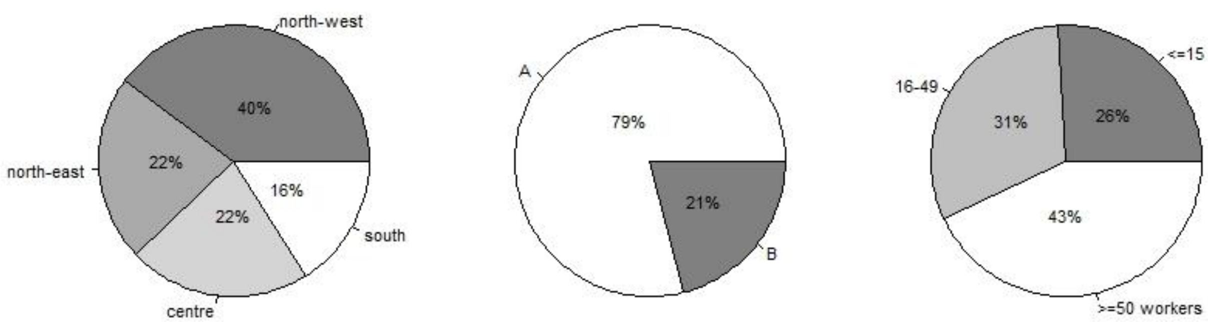
Figure 2. Frequency distributions of the 3,914 workers according to the quality of work variables (% values).

The seven grouping variables used to classify workers in groups refer to both cooperative and worker characteristics (Table 2).

Table 2. The seven grouping variables.

Variable	**Response categories**
Cooperative characteristics	
GEOGRAPHICAL AREA	north-west; north-east; centre; south and islands (of Italy)
TYPE	A type (providing health, social or educational services) and B type (integrating disadvantaged people into the labour market)
DIMENSION	≤15; 16-49; ≥50 paid workers
Worker characteristics	
GENDER	male; female
EDUCATION	middle school; diploma; university (M.S. degree and higher)
AGE	≤30; 31-40; ≥50 years old
MEMBERSHIP	member; nonmember

Figures 3 and 4 show the frequency distributions of the 3,914 workers according to the seven grouping variables, separating cooperative (Figure 3) from worker (Figure 4) characteristics. The frequency distributions over the whole sample of 4,134 workers are substantially the same as the ones shown in Figures 2-4.



GEOGRAPHICAL AREA | TYPE | DIMENSION

Figure 3. Frequency distributions of the 3,914 workers according to the cooperative characteristics.

Firslty[2], NL-PCA with $p=m-1$ was performed on the original sample of 3,914 workers. The variance accounted for in the final solution resulted 95.3%. Except for LOYALTY, which was optimally scaled according to the nominal transformation, the ordinal transformation was chosen to scale the other seven variables of quality of work. The resulting transformation plots are represented in Figure 5. These plots are useful to assign a meaning to the vertical axis of the PCP with BICIs. In particular, the vertical axis limits of the grey areas correspond to the vertical axis limits set in the next Figures 6-13, where PCPs with BICIs are displayed.

[2] The entire analysis was performed using the R 2.9.2 software.

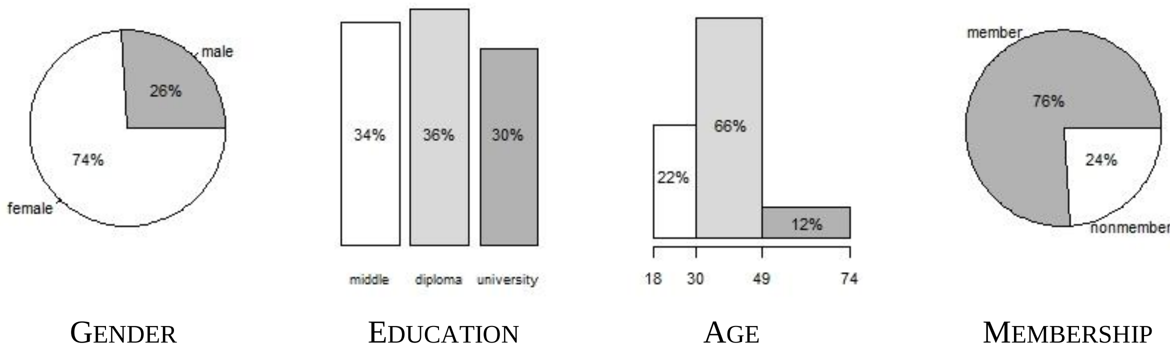


Figure 4. Frequency distributions of the of the 3,914 workers according to the worker characteristics.

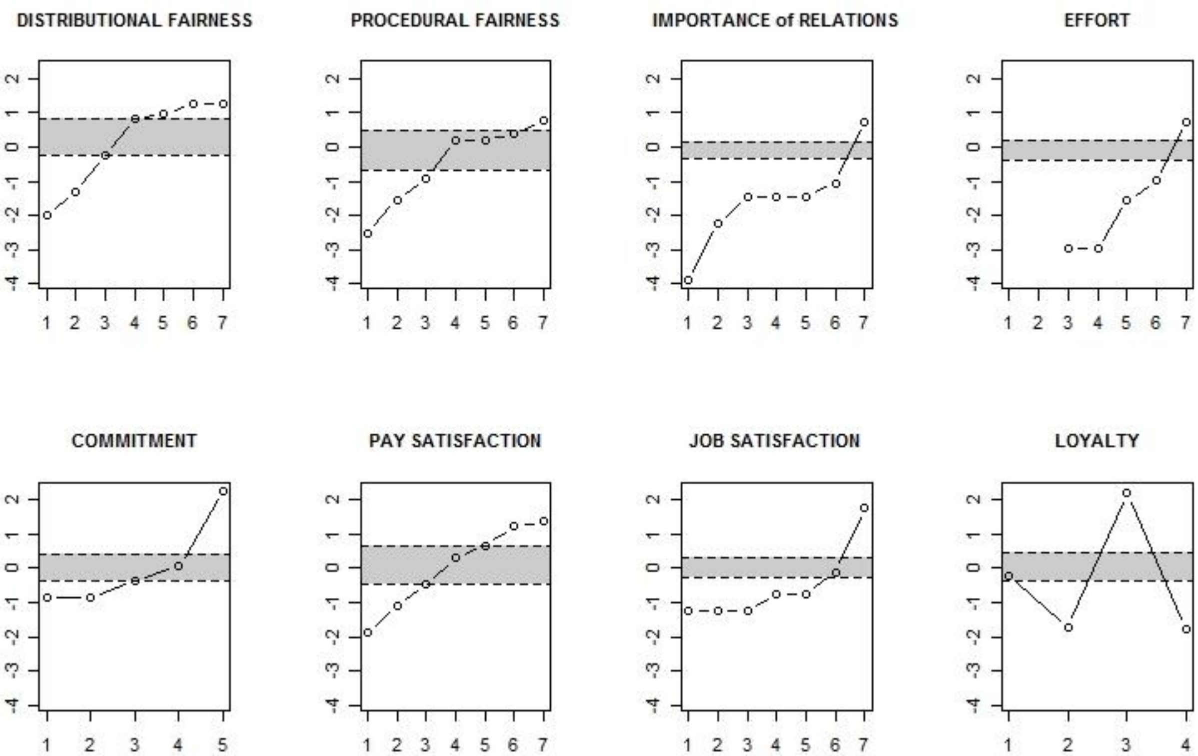


Figure 5. Transformation plots of the eight variables of quality of work (the original categories and the optimal quantifications are displayed in *x*-axis and *y*-axis, respectively).

The seven grouping variables were analysed as supplementary or passive variables, scaled by a multiple nominal transformation. The projected centroids of the grouping variables onto the quality of work variables were computed.

In order to obtain their associated BICIs and consequently the final graphical representations, a bootstrap study was conducted, with R=1,000 replications. For the construction of BICIs, we considered α=0.05 and referred to the Normal distribution (the plots of the R projected centroids of each group resemble a Normal curve and the group sample sizes were always larger than 422 over the R=1,000 replications).

Figures 6-13 show the projected centroids and their associated BICIs of the seven considered grouping variables onto the eight categorical variables under study. In order to make the comparisons among groups clearer, representations have been enlarged as much as possible by using different scales on the vertical axes in Figures 6-13 and an incomplete representation of the quantified categories on the vertical axes. However, it is still possible to recover distances between centroids and categories by looking at the transformation plots in Figure 5, where grey areas are depicted to indicate the scales used on the vertical axes in Figures 6-13. Moreover, grey areas suggest that projected centroids and their associated BICIs are compressed around the mean quantification (zero), especially for some variables like IMPORTANCE OF RELATIONS and JOB SATISFACTION.

Figure 6 refers to DISTRIBUTIONAL FAIRNESS, whose response scale ranged from 1 = much less than fair to 7 = much more than fair (with 4 = fair), as displayed in Table 1. On the vertical axis the quantifications assigned to (some of) the original categories are displayed. The horizontal solid line, corresponding to the zero quantification (i.e., the mean quantification), informs that the 3,914 workers scored, on average, between categories 3 and 4. In other words, they consider in general their pay less than fair. This result is also visible in the transformation plot of DISTRIBUTIONAL FAIRNESS (Figure 5). BICIs associated to the projected centroids corresponding to the four geographical areas show that the workers employed in social cooperatives in the south as well as in the north-west of Italy perceive more fairness than the workers in the north-east and the centre. Distributional fairness perceived by workers in B type cooperatives is significantly higher than in A cooperatives. The perceived distributional fairness significantly increases as the dimension of the cooperative decreases. Looking at the groups of workers defined by individual characteristics, Figure 6 shows that females and members perceive their pay less fair than males and nonmembers, respectively; moreover, the perceived distributional fairness significantly increases as the worker educational level decreases (and the age increases).

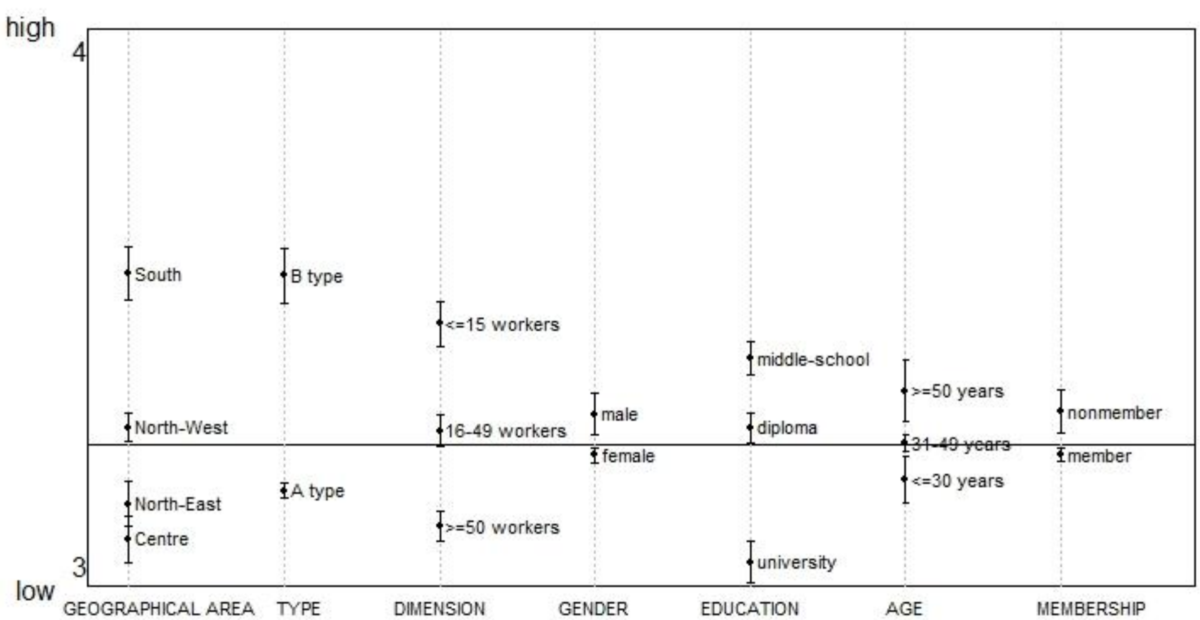


Figure 6. Projected centroids with associated BICIs of seven grouping variables (GEOGRAPHICAL AREA, TYPE, DIMENSION, GENDER, EDUCATION, AGE, MEMBERSHIP) onto DISTRIBUTIONAL FAIRNESS.

The study of PROCEDURAL FAIRNESS shows that all workers scored on average between categories 3 and 4-5 (which received equal quantifications) on a response scale from 1 = not at all agree to 7 = strongly agree. The projected centroids and their associated BICIs onto PROCEDURAL FAIRNESS are represented in Figure 7. Groups defined by cooperative characteristics show substantially the same perceptions examined with reference to DISTRIBUTIONAL FAIRNESS (Figure 6). The most evident distinction with respect to the perceived distributional fairness is that the differences between pairs of groups defined by each worker characteristic are now not significant (except for members, who surprisingly perceive less procedural fairness than nonmembers). This was expected, because procedural fairness regards the justice in the behaviour of the cooperative, while distributional fairness refers to the justice in the individual pay: it seems reasonable that males and females, for example, evaluate the cooperative at the same manner but their pay in different ways.

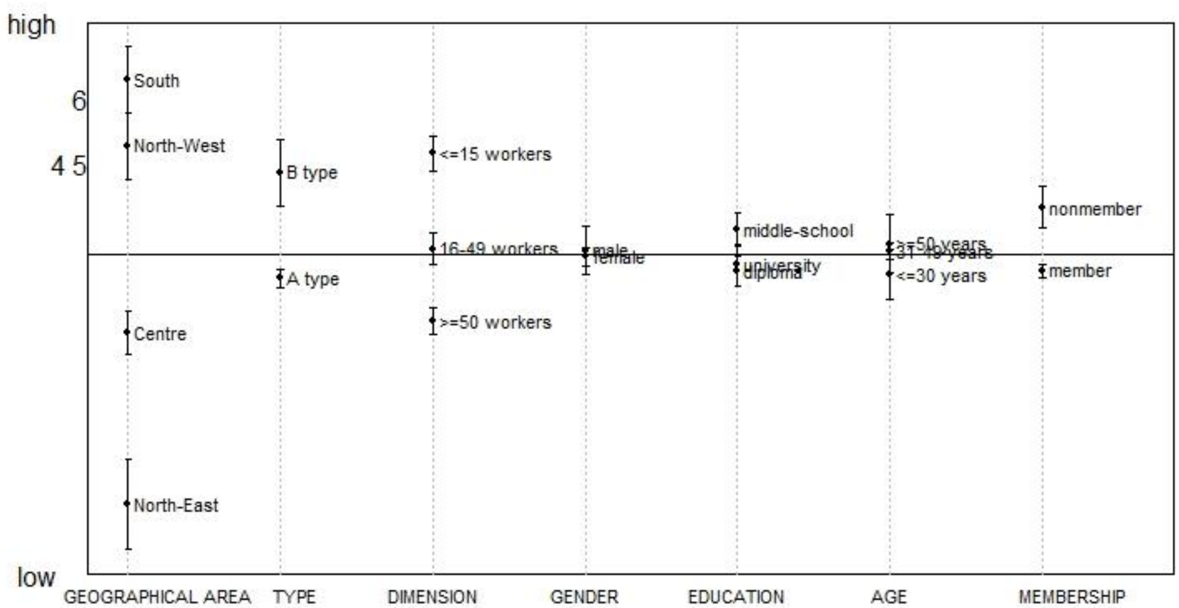


Figure 7. Projected centroids with associated BICIs of seven grouping variables (GEOGRAPHICAL AREA, TYPE, DIMENSION, GENDER, EDUCATION, AGE, MEMBERSHIP) onto PROCEDURAL FAIRNESS.

The interpersonal relations at workplace are considered important by all workers: on average, they scored between categories 6 and 7 of IMPORTANCE OF RELATIONS (on a scale from 1 = not at all important to 7 = strongly important). From the graphical analysis of the PCP with BICIs (Figure 8) some significant differences appear: workers in A type cooperatives and females consider relations more important than workers in B cooperatives and males, respectively. This could be related to the different job of such workers: in the kind of services offered by the A cooperatives, the interpersonal relations play a fundamental role, much more important than in the integration of disadvantaged people into the labour market offered by B cooperatives; moreover, females are more often employed in direct contact with end users than males and this could be related to a perception of higher importance of relations.

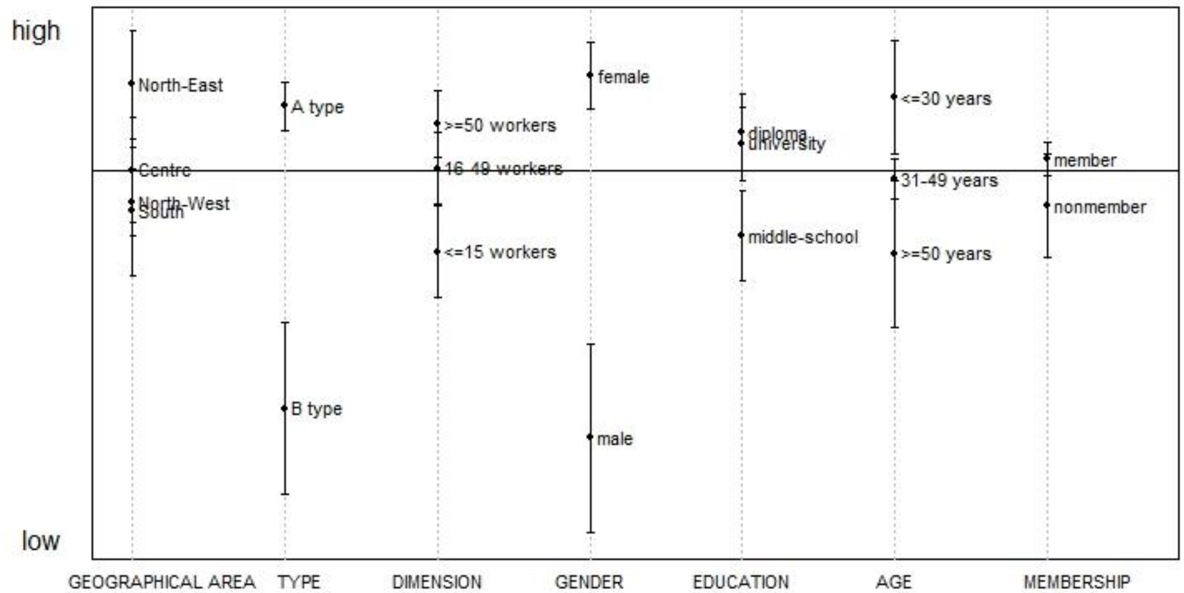


Figure 8. Projected centroids with associated BICIs of seven grouping variables (GEOGRAPHICAL AREA, TYPE, DIMENSION, GENDER, EDUCATION, AGE, MEMBERSHIP) onto IMPORTANCE OF RELATIONS.

All workers declare to put much effort on their job (they scored on average between 6 and 7 of EFFORT on the ordinal scale from 1 to 7). It is interesting to note (see Figure 9) that workers in A type cooperatives, females, and members declare a significantly higher effort than workers in B cooperatives, males, and nonmembers, respectively. Moreover, the declared effort significantly increases as the dimension of the cooperative increases.

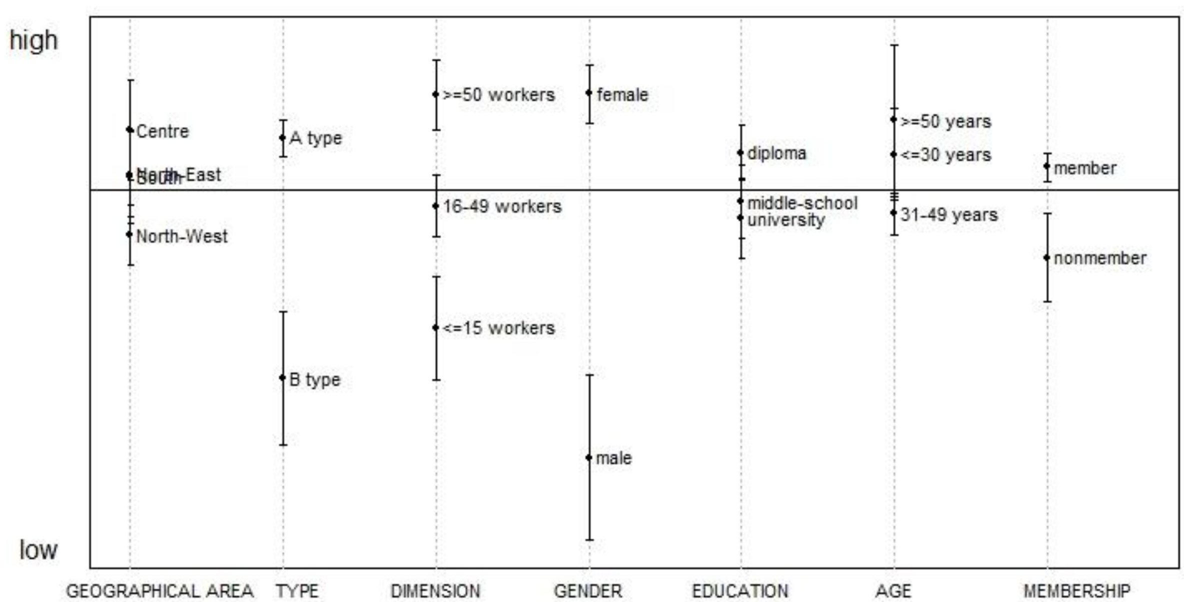


Figure 9. Projected centroids with associated BICIs of seven grouping variables (GEOGRAPHICAL AREA, TYPE, DIMENSION, GENDER, EDUCATION, AGE, MEMBERSHIP) onto EFFORT.

The ICSI^{2007} workers are involved in the mission of the cooperative and identify with the cooperative rather often: the mean quantification of COMMITMENT is between categories 3 (sometimes) and 4 (often), though much closer to 4. Figure 10 shows that the workers employed in the south of Italy are more involved than the workers in the north and the centre, and members are more involved than nonmembers, as expected. The organizational commitment significantly increases as the dimension of the cooperative decreases.

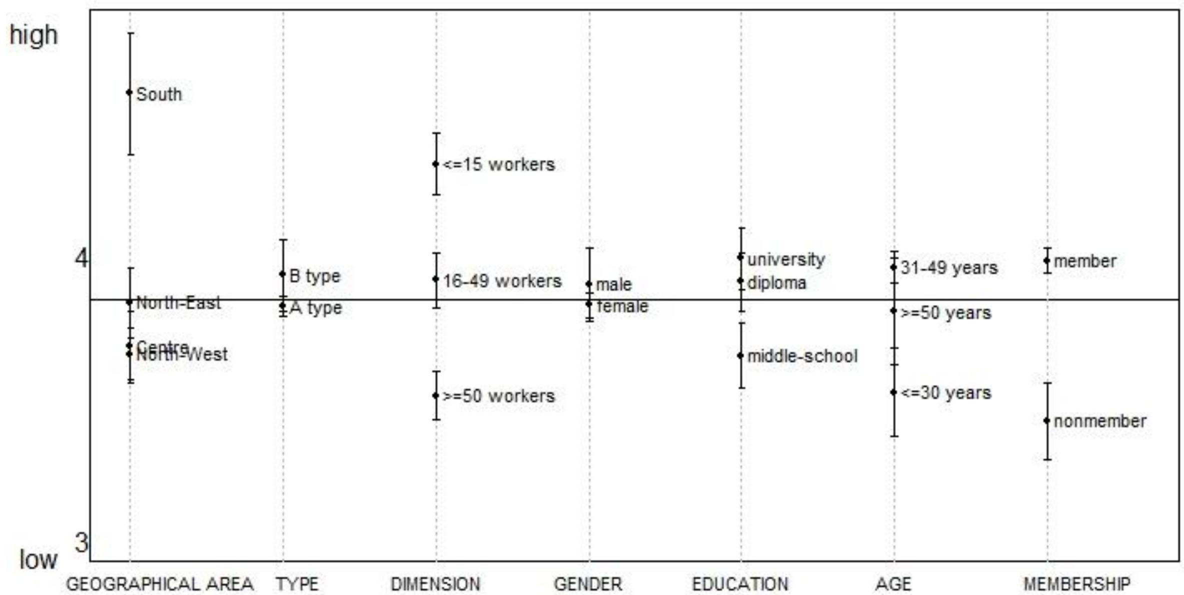


Figure 10. Projected centroids with associated BICIs of seven grouping variables (GEOGRAPHICAL AREA, TYPE, DIMENSION, GENDER, EDUCATION, AGE, MEMBERSHIP) onto COMMITMENT.

With reference to PAY SATISFACTION, the mean quantification is between categories 3 and 4 (on a scale from 1 = very dissatisfied to 7 = very satisfied). Figure 11 shows that the workers employed in the south of Italy are significantly more satisfied with their pay than workers in the north, while workers in the centre of Italy are the least satisfied with pay. Workers in B type cooperatives are significantly more satisfied with pay than workers in A type cooperatives. Both these results could be related to the lack of labour alternatives in the south of Italy and for the disadvantaged workers employed in B type cooperatives. Satisfaction with pay significantly increases as the dimension of the cooperative decreases. Looking at the groups of workers defined by each individual characteristic, there are no significant differences between males and females, while the workers with higher educational level are less satisfied with pay than workers with a lower educational level. Younger workers and members are less satisfied with pay than elder people and non-members, respectively.

Although the workers of ICSI^{2007} are not very satisfied with pay, they are more satisfied with job in general. This means that the pay satisfaction is not a very important driver of the overall job satisfaction for these workers ([10]).

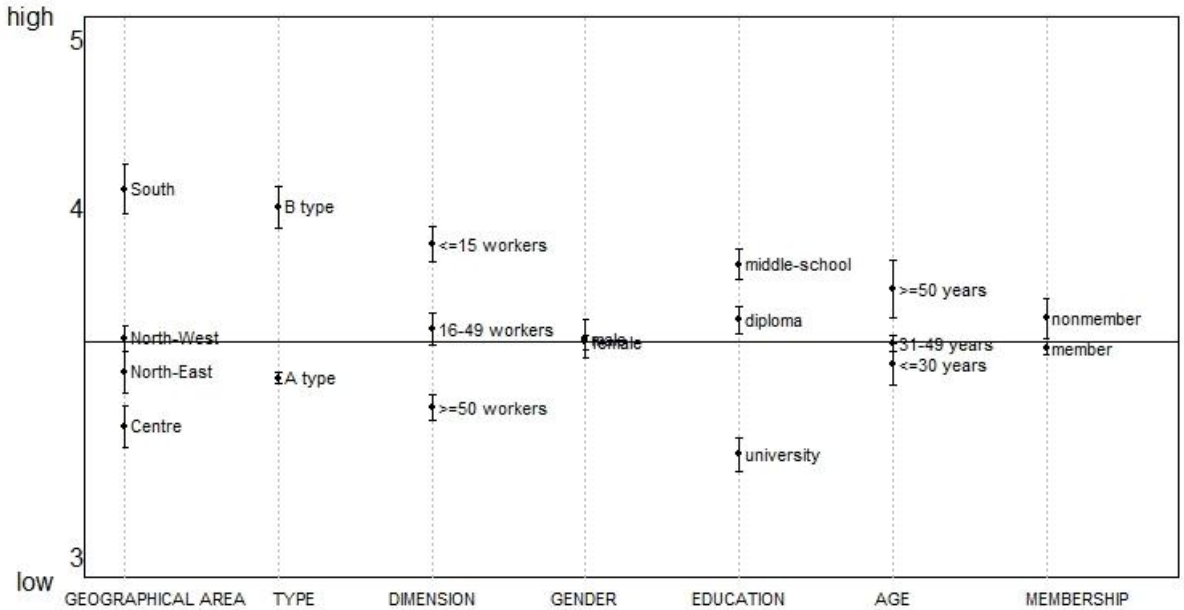


Figure 11. Projected centroids with associated BICIs of seven grouping variables (GEOGRAPHICAL AREA, TYPE, DIMENSION, GENDER, EDUCATION, AGE, MEMBERSHIP) onto PAY SATISFACTION.

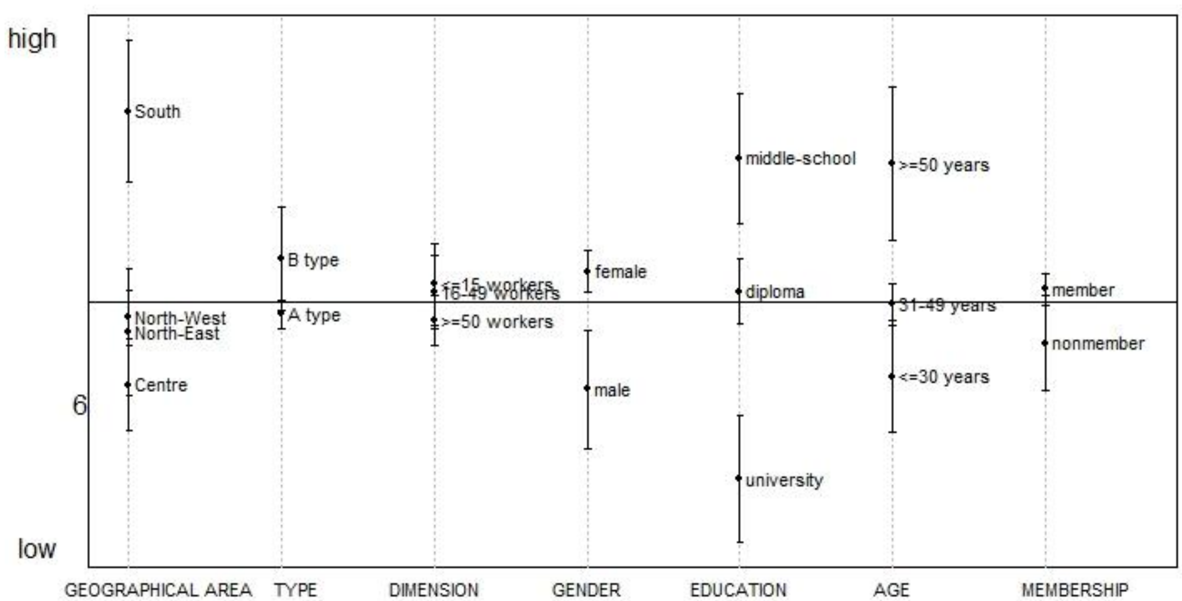


Figure 12. Projected centroids with associated BICIs of seven grouping variables (GEOGRAPHICAL AREA, TYPE, DIMENSION, GENDER, EDUCATION, AGE, MEMBERSHIP) onto JOB SATISFACTION.

In fact, the mean quantification of JOB SATISFACTION is between the categories 6 and 7 (on a scale from 1=very dissatisfied to 7=very satisfied). As visible in Figure 12, the groups defined by the worker characteristics show notably differences with respect to job satisfaction: females are more satisfied than males; job satisfaction decreases as the educational level of workers increases; younger workers are less satisfied than elder workers. With reference to the cooperative characteristics, workers employed in the south cooperatives are more satisfied than workers employed in the rest of the Country. Still, the lack of labour alternatives seems to play a role in determining a higher job satisfaction for those workers. The analysis of LOYALTY shows that the ICSI^{2007} workers intend to stay at the cooperative as longer as possible, mostly because they like work and workplace. The PCP with BICIs referred to LOYALTY (Figure 13) shows that there are no significant differences among the considered groups, except for workers with the highest educational level and the youngest, less loyal than workers with lower educational level and more aged, respectively. There is a relation between the loyalty and job satisfaction perceived by each group of workers: the most faithful groups are exactly the most satisfied with job in general. It should be noted that the interpretation of the response categories of LOYALTY is difficult because while categories 1, 3 and 4 could be considered ordered (from the highest to the lowest level of loyalty), category 2 explains a concept of loyalty-nonloyalty ([6])[3].

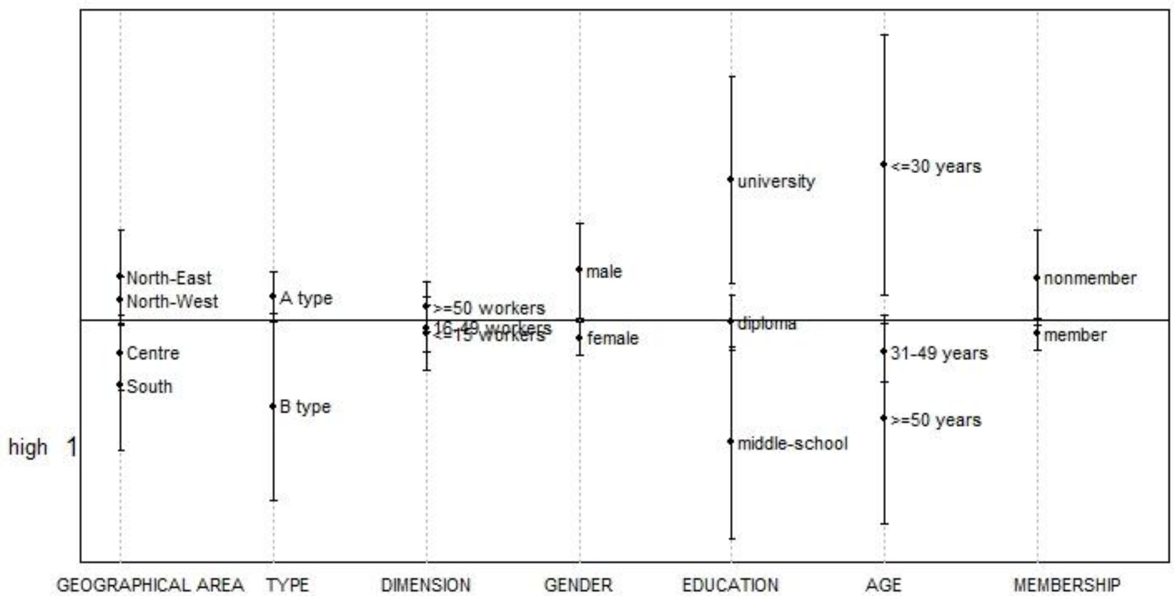


Figure 13. Projected centroids with associated BICIs of seven grouping variables (GEOGRAPHICAL AREA, TYPE, DIMENSION, GENDER, EDUCATION, AGE, MEMBERSHIP) onto LOYALTY.

Concluding, results show that, with reference to the groups defined by the worker characteristics:

[3] With the aim of improving the interpretation of results referred to LOYALTY, NL-PCA was applied to the original sample by scaling all variables (including LOYALTY) by the ordinal scaling level. Results show that quantifications assigned to categories of LOYALTY obviusly change ($\mathbf{q}_j$=[-0.447, -0.328, 2.286, 2.286] for *j*=LOYALTY), due to the ordering restrictions: category 2 of loyalty-nonloyalty lies between categories 1 (high level of loyalty) and 3-4 (moderate and low levels of loyalty, which received tied quantifications), though much closer to category 1. The centroids of the seven grouping variables projected onto LOYALTY respect the ordering obtained in the analysis with LOYALTY nominally scaled, but now they are all located between categories 2 and 3, though much closer to 2 (and 1). Since results on the PCPs of all the other seven variables do not substantially change and the interpretation of the PCP of LOYALTY does not improve, we decided to apply the bootstrap study and the construction of ICIs to the NL-PCA solution with LOYALTY nominally scaled.

- the differences between males and females are often negligible; significant differences refer to the higher importance of the interpersonal relations, the higher effort and the higher job satisfaction that females showed with respect to males;
- as the educational level of workers increases (and their age decreases), distributional fairness, pay and job satisfaction significantly decrease;
- members perceive a lower level of distributional and procedural fairness than nonmembers, but they are more involved in the mission of the cooperative and declare to put more effort in their work.

With reference to the groups defined by the cooperative characteristics:

- the workers employed in the south (and islands) of Italy show a peculiar position on the considered subjective variables of quality of work: with respect to the workers employed in the north and in the centre, they perceive a higher level of fairness, are more involved in the mission of the cooperative, and are more satisfied with pay and job in general;
- workers employed in A type cooperatives perceive less fairness, give more importance to the interpersonal relations, put more effort, and are less satisfied with pay than workers in B cooperatives;
- as the cooperative dimension decreases, the distributional and procedural fairness perceived by workers increases, as well as the involvement on the mission of the cooperative and the satisfaction with pay. This confirms the idea that in the small cooperatives the level of involvement in the mission is higher and the sharing of ideals, decisions and information is more common than in larger organizations.

4. Discussion

The results of the present study showed that the proposed graphical tool is easy to read and helps the interpretation of the existing relations in the data. It shows differences and similarities among the different groups of subjects (workers) with respect to the categorical variables of quality of work. It considers the position of groups on average and with respect to a specified categorical variable.

In Section 3 we considered PCPs with BICIs where one single active variable and more grouping variables were represented in the same plot. A PCP with BICIs can also be constructed with several parallel lines representing several active variables where the centroids of one single grouping variable are projected. However, in this case the reading of the (quantified) categories on the line representing each active variable becomes more difficult. Moreover, there is the need to keep the same scale for the several active variables. This can make the differences among some groups not visible anymore, when the active variables have very different ranges.

Having applied NL-PCA with $p=m-1$, where p is the number of dimensions in the solution and m is the number of active variables in the analysis, the loss of information is negligible. The optimal quantification of the categorical variables allows to compute metric distances among groups. The bootstrap study allows to overcome the descriptive nature of the NL-PCA procedure and to construct BICIs. Thanks to the use of BICIs, inferential arguments have been introduced in the analysis and it is immediate to note whether the differences between two groups are statistically significant or not.

Usually, the aim of PCA is to reduce data dimensionality. In this paper, the focus is only on the transformation of categorical variables obtained by NL-PCA. Because the attention is then on a graph considering each variable separately, the solution can be defined on a high dimensional space. We chose to keep $p=m-1$ principal components in the final solution in order to obtain a solution with the maximum variance accounted for. This solution will have some redundancy of information, unless the rare case in which correlations between the (transformed) variables are all very low. A smaller number p could be chosen from time to time in order to better satisfy the trade-off between interpretability and variance accounted for, but we proposed $p=m-1$ because (*i*) we are not interested in the principal components and in their interpretation, but in each variable separately and (*ii*) the bootstrap study requires a fixed p, and the choice made on the original sample could not be ideal in each of the R bootstrap samples.

Obviously, the NL-PCA solution with $p=m$ would guarantee no loss of information, but in that case all variables in analysis would be transformed according to a linear quantification, because there are no degrees of freedom in the NL-PCA optimisation problem. The optimal transformations maximizing the sum of the $p=m$ eigenvalues of the correlation matrix among the m transformed variables are linear. The linear quantification, based on the hypothesis of equal distance between the categories, could be also considered as an alternative quantification method. But although it is widespread and in some contexts leads to results close to the ones obtained with more refined techniques, the hypothesis of equal distance between the categories is often unrealistic (see, among others, [23]). Moreover, when $p=m$ the projected centroids of one group on a certain variable can be easily computed averaging that transformed variable over the subjects belonging to that group and the use of centroids as m_A and m_B in formulae (1) is straightforward. In this case, the standard errors necessary to construct ICIs can be derived from the sample variances, and our proposal would simply be an application of ICIs. Moreover, in that case, the transformation does not take into account the relations among all the variables in the analysis, but it is a sort of univariate transformation. Thus, other (univariate) quantification procedures could be taken into account and compared (for example, the indirect quantification based on the Normal cumulative function [31] or the one based on the Negative Exponential cumulative function [28]). In this way, however, the original idea of projecting centroids onto vectors softens up.

The graphical representation of projected centroids refers to one active variable at a time, but takes into account the relations between that variable and the other variables under study, because optimal quantifications are assigned to the original categories with the aim of maximizing the variance accounted for in the final p-dimensional solution (given by the sum of the first p eigenvalues of the correlation matrix among the transformed variables). Therefore, the PCP remains a multivariate tool, even if points are represented on a straight line.

This is a very attractive property, because one PCP is able to reveal information on the position of groups of subjects on one variable, taking into account the relations of that variable with all the others.

For example, with reference to the case study described in Section 3, Figures 5 and 7 show that categories 4 and 5 of PROCEDURAL FAIRNESS received the same quantification, because workers scoring 4 and 5 on that variable gave similar answers to all the other quality of work variables in the analysis. This can be used when interpreting the position of centroids projected on PROCEDURAL FAIRNESS.

The interpretation of the PCP must take into account this ability of the plot to consider the multivariate structure of the data, being aware that it makes sometimes the interpretation tricky.

This especially happens when the multivariate structure of the data leads to optimal transformations masking in some ways the relation between the analysed variable and one or more grouping variables. For example, when (*i*) two or more categories of a categorical variable receives equal quantifications and/or (*ii*) the optimal transformations are non-monotonic because of the presence of nonlinear (especially non-monotonic) relations among variables.

As an example, consider one categorical variable with 5 categories, scaled according to an optimal transformation assigning a unique quantification q_1 to categories 1 and 2 and a unique quantification $q_2 \neq q_1$ to categories 3, 4, and 5. Consider now gender as grouping variable and suppose that males only scored on categories 1 and 3; the position of the corresponding centroid on the PCP will be between q_1 and q_2 and, therefore, cannot directly reveal the dependency existing between gender and the categorical variable.

However, the assignment of the same quantification to different categories means that subjects scoring on those categories gave similar answers to all the other variables under study. Therefore, confounding categories 1 and 2, on one hand, and 3, 4, and 5, on the other hand, in the example is not a serious drawback, especially when the meaning of the original categories allows a practical interpretation, like when dealing with ordinal variables.

The interpretation of the PCP is not straightforward in the presence of nominal variables, when quantifications are tied and the merging of categories could have no meaning, and when the optimal transformations are non-monotonic and the position of centroids is not able to reveal the true relation among the analysed variable and the grouping variable.

The risk of finding a situation difficult to interpret decreases as the number of components p retained in the NL-PCA solution increases. In fact, the choice on the number p is not only related to the well-known trade-off between the amount of information maintained in the solution and the interpretability of the components, but also to the ability (or the need) of the NL-PCA to take into account the multivariate relations among variables. In fact, with a notably reduction of dimensionality, the optimal transformations must be able to catch almost all the linear and nonlinear relations among variables in order to maximize the variance accounted for. Instead, as mentioned before, in the extreme situation with p=m the solution lies on a m-dimensional space and there is no need to transform variables to maximize the variance accounted for; the resulting PCPs can be considered univariate.

Another way to use the proposed idea is to project centroids directly on the principal components. In this case, the components need to be interpreted and a small number of components is usually used. When this number equals 2 or 3, it is not necessary to construct the PCP; the projection can directly be made in the biplot representing the reduced space and the idea of BICIs can be replaced by other tools, like convex hulls ([21]), even if they have not exactly the meaning of statistical test on difference.

Future research will (*i*) extend the proposed graphical tool in order to include the new developments on ICIs (see, for example, [33], [22]) and (*ii*) study the various definition of bootstrap intervals ([11], [13]), in order to look for a solution directly interpretable as a graphic test of statistical difference like ICIs. This solution could be used for the construction of the PCP with intervals but also of the graphical tools introducing inferential issues on the NL-PCA solution when it can be represented in 2 or 3 dimensions.

Acknowledgements

The author wishes to thank Paola Zuccolotto from University of Brescia for her helpful suggestions.

References

[1]. Agresti, A. (2002). *Categorical data analysis (2nd ed.)*. Hoboken: Wiley.
[2]. Barlow, R.E., Bartholomew, D.J., Bremner, J.M., and Brunk, H.D. (1972). *Statistical inference under order restrictions*. New York: Wiley.
[3]. Benz, M. (2005). Not for the profit, but for the satisfaction? Evidence on worker well-being in non-profit firms. *Kyklos*, 58(2), 155-176.
[4]. Benzécri, J.P. (1992). *Correspondence analysis handbook*. New York: Marcel Dekker.
[5]. Borzaga, C., Depedri, S. (2005). Interpersonal Relations and Job Satisfaction: Some Empirical Results. In *Economics and Social Interaction. Accounting for Interpersonal Relations*, eds. B. Gui and R. Sugden. Cambridge: Cambridge University Press, 132-153.
[6]. Borzaga, C., Tortia, E.C. (2006). Worker Motivations, Job Satisfaction, and Loyalty in Public and Nonprofit Social Services. *Nonprofit and Voluntary Sector Quarterly*, 35(2), 225-248.
[7]. Carpita, M. (2006). On the Inferential Confidence Intervals for pairwise comparisons. In *Proceedings XLIII Scientific Meeting SIS*, Torino, 14 - 16 June 2006, 563-566.
[8]. Carpita, M. (Ed.) (2009). *La qualità del lavoro nelle cooperative sociali. Misure e modelli statistici*. Milano: Franco Angeli.
[9]. Carpita, M., Manisera, M. (2006). Un'analisi delle relazioni tra equità, motivazione e soddisfazione per il lavoro. In *Valutare la qualità. I servizi di pubblica utilità alla persona*, eds. M. Carpita, L. D'Ambra, M. Vichi, G. Vittadini. Milano: Guerini.
[10]. Carpita, M., Zuccolotto, P. (2007). Mining the Drivers of Job Satisfaction using Algorithmic Variable Importance Measures. In *Metodi, Modelli e Tecnologie dell'Informazione a Supporto delle Decisioni, Vol. I: Metodologie*, eds. L. D'Ambra, P. Rostirolla, M. Squillante. Milano: FrancoAngeli, 63-70.
[11]. Davidson, A.C, Hinkley, D.V. (1997). *Bootstrap methods and their applications*. New York: Cambridge University Press.
[12]. Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans*. Philadelphia, Pennsylvania: Society for Industrial and Applied Mathematics.
[13]. Efron, B., Tibshirani, R.J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
[14]. Gifi, A. (1990). *Nonlinear multivariate analysis*. Chichester: Wiley.
[15]. Goldstein, H., Healy, M.J.R. (1995). The graphical presentation of a collections of means. *Journal of the Royal Statistical Society*, A, 158(1), 175-177.
[16]. Jolliffe, I.T. (2002). *Principal Component Analysis (2nd ed.)*. New York: Springer.
[17]. Kahneman, D., Krueger, A.B. (2006). Developments in the measurement of subjective well-being. *The Journal of Economic Perspectives*, 20, 3-24.
[18]. Kruskal, J.B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, 1-28.

[19]. Kruskal, J.B. (1964). Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29, 115-129.
[20]. Linting, M., Meulman, J.J., Groenen, P.J.F., and Van der Kooij, J.J. (2007). Nonlinear Principal Components Analysis: Introduction and Application. *Psychological Methods*, 12, 336-358.
[21]. Linting, M., Meulman, J.J., Groenen, P.J.F., and Van der Kooij, A.J. (2007). Stability of Nonlinear Principal Components Analysis: An empirical study using the balanced bootstrap. *Psychological Methods*, 12, 359-379.
[22]. Marasini, D., Migliorati, S. (2009). On the use of inferential confidence intervals. *Statistica & Applicazioni*, VII(1), 1-12.
[23]. Marbach, G. (1974). Sulla presunta equidistanza degli intervalli nelle scale di valutazione. *Metron*, XXXII, 1-4.
[24]. Markus, M.T. (1994). *Bootstrap confidence regions in nonlinear multivariate analysis*. Leiden: DSWO Press.
[25]. Meulman, J.J., Heiser, W.J., and SPSS (2004). *SPSS Categories 13.0. Chicago: SPSS Inc.*
[26]. Meulman, J.J., Van der Kooij, A.J., and Heiser, W.J. (2004). Principal components analysis with nonlinear optimal scaling transformations for ordinal and nominal data. In *Handbook of quantitative methodology for the social sciences*, ed. D. Kaplan, London: Sage, 49-70.
[27]. Nishisato, S. (1980). *Analysis of categorical data: dual scaling and its applications.* Toronto: University of Toronto Press.
[28]. Portoso, G. (2004). On the Choice between the Exponential and the Normal in the Indirect Scaling. In *Proceedings XLII Scientific Meeting SIS*, Bari, 9 - 11 June 2004, 551-554.
[29]. Ramsay, J.O. (1988). Monotone regression splines in action. *Statistical Science*, 3, 425-441.
[30]. Riani, M., Zani, S. (2007). Rappresentazioni grafiche e trasformazioni di dati multidimensionali. In *Analisi dei dati e data mining per le decisioni aziendali*, eds. S. Zani and A. Cerioli, Milano: Giuffrè, 173-214.
[31]. Torgerson, W.S. (1958). *Theory and methods of scaling*. New York: Wiley.
[32]. Tryon, W.W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: an integrated alternative method of conducting null hypothesis statistical test. *Psychological Methods*, 6(4), 371-386.
[33]. Tryon, W.W., Lewis, C. (2009). Evaluating Independent Proportions for Statistical Difference, Equivalence, Indeterminacy, and Trivial Difference Using Inferential Confidence Intervals. *Journal of Educational and Behavioral Statistics*, 34(2), 171-189.
[34]. Zani, S. (2000). *Analisi dei dati statistici, Vol. I*. Milano: Giuffrè.