Electronic Journal of Applied Statistical Analysis: Decision Support Systems and Services Evaluation
EJASA:DSS (2010). Vol 1, Issue 1, 2 – 15
e-ISSN 2037-3627, DOI 10.1285/i2037-3627v1n1p2

THE ROLE OF SIMPLE COMPONENT ANALYSIS IN THE CONTEXT OF THE EXPLORATORY METHODS. AN HEALTHCARE SERVICES EVALUATION

Gabriella Milone[*]

Dipartimento di Chimica "Paolo Corradini",
Università degli studi di Napoli Federico II, Italy.

Received 07 January 2010. Accepted 27 January 2010
Available online 17 December 2010

Abstract*: Among the exploratory techniques, Principal Component Analysis has the best properties in the study of relations between original variables, but in customer satisfaction applications it provides all positive correlations (the first component is an average or a sum of the scores).This feature entails trivial results of little interest that cannot help in decision-making, or even less, rotations (varimax, etc.) that can improve the interpretation of data structure. The aim of this paper is to highlight, via a comparison of methods, the role of Simple Component Analysis to improve the interpretability, over and above the lack of some desirable property (variance explained, etc.).This comparison will be supported by an application to real data on Patient satisfaction in a hospital in Naples.*

Keywords*: Principal Component Analysis, Simple Component Analysis, Rotation criteria, Patient Satisfaction*

1. Introduction

Today when we want to measure the performance of a Public service, the focus is moving in the direction of the Customer Satisfaction. This is a useful implementation of the logic of quality in the design and delivery of public services.
In health care companies, for example, increasing the patient satisfaction may increase the use of the system, attract new customers, improve the image of the company in Public Sector and, consequently, ensure an improvement in long term financial results.

[*] E-mail: gamilone@unina.it

This paper focuses on the explorative statistical analysis. It starts from a critical review, based on a comparison of the methodological instruments of this type of analysis (in particular Principal Component Analysis and techniques of rotation), and some scientific developments. After a brief methodological review, the paper highlights the contribution of a recent methodology for exploratory data analysis, Simple Component Analysis, to be used in the evaluation process of Customer Satisfaction. This approach, as explained below, is extremely useful in the application contexts to which we have referred.

2. Methodological developments based on different strategies of analysis

From a viewpoint of the exploratory techniques, the Principal Component Analysis (PCA) has optimal properties for determining the latent variables in the study of relations between original variables. However, in many scientific applications, such as those based on data regarding satisfaction, the correlations between the variables are all positive (or negative) and, consequently, the first principal component is equivalent to the average or to the sum of the values of the observed items. For these reasons the results of PCA are not easily interpretable.

Consequently, the PCA technique may be unsatisfactory when used in contexts such as those related to customer satisfaction. To overcome such problems, several authors have proposed alternative methods (sub-optimal) that are more easily interpretable, in comparison to the traditional approaches.

PCA leads to results that are trivial, not usable by public decision-makers. In order to improve the interpretability of results, the most recent literature has proposed methodological developments based on different strategies of analysis. In particular, it points to Least Absolute Shrinkage and Selection Operator (LASSO) [10], [12], Hausman Principal Component Analysis [4], Simple Component Analysis [20], [21], Simple Principal Components [21], Simple Component Analysis based on RV Coefficient, SCA-RV (by analogy with the usual statistical vocabulary, RV = Vectorial Correlation, invoking the usual notation Bravais-Pearson’s correlation coefficient linear "r") [5], over the alternative that sees the use of means of rotation. These approaches all produce pseudo-optimal, but better interpretable, results.

2.1 Principal Component Analysis

Principal Component Analysis is a well known technique in multivariate analysis [16] that seeks to uncover the underlying structure of a relatively large set of variables.

PCA involves a mathematical procedure that transforms a number of correlated numeric variables into a corresponding number of uncorrelated variables (artificial variables) called principal components, of which the first most representative can be considered (relevant PCA).

In this way a further objective of PCA is to reduce the dimensionality of the data set and to identify new meaningful underlying variables. It is useful when, analyzing data on a large number of variables, one may believe that there is some redundancy in those variables. In this case, redundancy means that some of the variables are correlated with one another, possibly because they are measuring the same construct. Because of this redundancy, it is believed that it should be possible to reduce the observed variables into a smaller number of components that will account for most of the variance showed by the observed variables. A principal component is a linear combination of optimally-weighted observed variables.

We describe the steps for getting PCA:

- it is assumed that the multidimensional data has been collected in a data matrix, in which the rows are associated with *n* statistical units (sample) and the *p* columns with the original variables.
- PCA is performed on the symmetric covariance matrix or on the correlation matrix. The latter case is equivalent to standardize the data first, so eliminating the influence of the different units measurement of the original variables.

The mathematical technique used in PCA is called eigen analysis, looking for the eigenvalues and eigenvectors of a square symmetric matrix (covariance or correlation matrix) [17]. The eigenvector associated with the largest eigenvalue has the same direction as the first principal component. The eigenvector associated with the second largest eigenvalue determines the direction of the second principal component. The sum of the eigenvalues equals the trace of the square matrix and the maximum number of eigenvectors equals the number of rows (or columns) of this matrix. We can make the so called scree plot of the eigenvalues, to get an indication of the importance of each eigenvalue. The exact role of each eigenvalue (or a range of eigenvalues) to the "explained variance" can also be queried.

There are two methods to extract the number of components, based on the graphical analysis of the eigenvalues plot.

1. If the points on the graph tend to level out, these eigenvalues are usually considered close enough to zero, so that they can be ignored.
2. Limit variance accounted. The fraction of the total variance, accounted for by the extracted components.

To understand the meaning of "total variance" as it is used in a PCA, remember that the observed variables are standardized in the course of the analysis. This means that each variable is transformed so that it has zero mean and unit variance. The "total variance" in the data set is simply the sum of the variances of these observed variables. Because they have been standardized, each observed variable contributes one unit of variance to the "total variance" in the data set. Because of this, the total variance in a principal component analysis will always be equal to the number of observed variables.

PCA procedure can be geometrical interpreted as the projection of the multivariate data on the space spanned by the eigenvectors. The first component (C_1) extracted in a PCA accounting for a maximal amount of total variance in the observed variables (X_i):

$$C_1 = \alpha_{11}X_1 + \alpha_{21}X_2 + \alpha_{31}X_3 + ... + \alpha_{p1}X_p = \boldsymbol{\alpha}_1^T \mathbf{X} \qquad (1)$$

is the linear combination of the variables such that $\boldsymbol{\alpha}_1^T \boldsymbol{\alpha}_1 = 1$. We can show that Var($C_1$) is equal to first eigenvalue of the correlation matrix.

The second component (C_2) extracted in a PCA is the linear combination of the variables:

$$C_2 = \alpha_{12}X_1 + \alpha_{22}X_2 + \alpha_{32}X_3 + ... + \alpha_{p2}X_p = \boldsymbol{\alpha}_2^T \mathbf{X} \qquad (2)$$

where:

$\boldsymbol{\alpha}_2{}^T\boldsymbol{\alpha}_2 = 1;\ \boldsymbol{\alpha}_1{}^T\boldsymbol{\alpha}_2 = 0$.

C_2 has two important characteristics:

- this component will account for a maximal amount of variance in the data set that was not accounted for by the first component; again under typical conditions, this means that the second component will be correlated with some of the observed variables that did not display strong correlations with C_1;
- it will be uncorrelated with the first component; if you were to compute the correlation between C_1 and C_2, that correlation would be zero.

The remaining components that are extracted in the analysis display the same two characteristics: each component accounts for a maximal amount of variance in the observed variables that was not accounted for by the preceding components, and is uncorrelated with all of the preceding components. A principal component analysis proceeds in this fashion, with each new component accounting for progressively smaller and smaller amount of variance (this is why only the first few components are usually retained and interpreted). When the analysis is complete, the resulting components will display varying degrees of correlation with the observed variables, but are completely uncorrelated with one another.

2.2 Rotation Criteria

Classical exploratory methods make an extraction of common factors in order to create a unique solution assuming that:

- the common factors are mutually uncorrelated;
- the common factors may be arranged in descending order of importance (decreasing eigenvalues).

However, these assumptions have several drawbacks. On the one hand, they are arbitrarily imposed on the data to uniquely determine the solution and do not appear realistic, we have no reason to believe that the latent factors are always independent of one another. The extracted factors posses usually a complex structure, which is hard to interpret because each variable is often associated with multiple factors.

In order to obtain a factorial model that is more easily interpretable, one may transform the solution found, taking into account the number of fixed factors and the communality (variance explained) of each observed variable.

In this section, attention will be devoted to single stage rotation methods [8] that, in order to attain simplicity, minimize a smooth function of factor pattern coefficients. Both orthogonal and oblique rotations will be considered, although oblique rotations are probably more appropriate in most practical situations.

Particular importance will be given to situations where a perfect cluster solution is inappropriate and more complex patterns are required.

Most of these are virtually unknown but are very interesting. Methods for row standardization of a factor matrix prior to rotation in order to improve the solution will also be examined.

Analytic rotation methods involve the post multiplication of an input $p \times m$ factor loading matrix, $\mathbf{A}$, by a $m \times m$ matrix, $\mathbf{T}$, to yield a rotated primary factor pattern matrix,

$$\boldsymbol{\Lambda} = \mathbf{AT}$$

that minimizes a continuous function, $f(\mathbf{\Lambda})$, of its factor loadings. This function is intended to measure the complexity of the pattern of loadings in $\mathbf{\Lambda}$. By minimizing the complexity function, $f(\mathbf{\Lambda})$, the rotation procedure yields a rotated matrix $\mathbf{\Lambda}$ with a simpler pattern of loadings.
Let the factor correlation matrix after rotation be represented by $\mathbf{\Phi}$. In orthogonal rotation the transformation matrix is required to satisfy the $m(m\text{-}1)/2$ constraints

$$\mathbf{\Phi} = \mathbf{T}'\mathbf{T} = \mathbf{I} \qquad (3)$$

defining factors that are uncorrelated and have unit variances. In direct oblique rotation [9], a complexity function, $f(\mathbf{\Lambda})$, is also minimized but now a smaller number, m, of constraints

$$\mathrm{Diag}(\mathbf{\Phi})=\mathrm{Diag}(\mathbf{I}) \qquad (4)$$

is imposed, defining factors possibly correlated but still having unit variances. This process defines a factor pattern, $\mathbf{\Lambda}$, that directly minimizes the complexity criterion.
Thus orthogonal and oblique rotations involve the same problem of minimizing a complexity criterion, and only the imposed constrains differ. It is appealing to make use of complexity functions that are suitable for both orthogonal and oblique rotation. Since fewer constraints are imposed in oblique rotation, it is generally possible to obtain a lower value of the complexity function and consequently greater simplicity of the factor pattern than in orthogonal rotation.
All rotation criteria to be considered are expressed as complexity functions to be minimized to yield a simple pattern of loadings. All of these complexity functions have a greatest lower bound of zero.

3 The contribution of Simple Component Analysis to the evaluation of patient satisfaction

Simple Component Analysis is an alternative to the traditional exploratory analysis procedures, in terms of contribution to interpretation: its methodological approach is, in fact, based on the criterion of obtaining a more simple and more interpretable result, even at the cost of give up optimal solutions.
The purpose of SCA is to replace the first q principal components with an improved system of q so called “simple components”, which are better interpreted. However, though SCA sacrifices some of the features of classical optimality of PCA, if the loss of variability extracted is minimal, and correlations between components are low, it may be attractive to use this approach whenever the interpretation of results becomes easier.
There are three different approaches to remove the trade-off between the PCA optimality and SCA simplicity:

- to seek a compromise between simplicity and optimality, using criteria that get the optimality, incorporating at the same time aspects of simplicity [13];
- to seek the simple solution of optimal systems of components, thus obtaining a “simple optimal system” [13];
- seek the optimum solution of simple systems, thus achieving an “optimal simple system” [8], [21], [20].

Of course, whichever approach will be preferred, it is necessary to define what is meant by optimality and simplicity.
Let be (U_1, …, U_q) an orthonormal system of components, (X_1, …, X_p) with $q \leq p$, a set of original variables and (Z_1, …, Z_p) the corresponding standardized ones. A system of components (U_1, …, U_q) is called optimal if it maximizes the explained variability of the original variables, getting the maximum of one of the following criteria:

$$Opt_1(P) = \frac{\sum_{k=1}^{q} Var(U_k)}{\sum_{k=1}^{q} Var(Z_k)} \tag{5}$$

$$Opt_2(P) = \frac{\sum_{j=1}^{p} (Var(X_j) - Var(X_j / U_1, ...U_q)}{\sum_{K=1}^{q} Var(Z_k)} \tag{6}$$

$$Opt_3(P) = \frac{Var(U_1) + \sum_{k=2}^{p} (Var(U_k / U_1, ...U_{k-1})}{\sum_{K=1}^{q} Var(Z_k)} \tag{7}$$

In our case, the system (U_1, …, U_q) is the one defined by the loadings concerning the first q relevant principal components.
Let now introduce the concept of simplicity of a system of standardized random variables (Y_1, …. , Y_p). The characteristic elements of the simplicity are:

- to have more than one block-component in case of an approximate block-structure in the correlation matrix C with rank q ($q \leq p$);
- the weighting scheme of variables should be “really” simple: all variables involved in a block component should have the same weight. For the difference-components, we would like to a have “proper” contrasts of variables: all positive weights should be equal, all negative weights should be equal, and the sum of all weights should be zero.

In formal terms, a system of q components will be called "simple", with b blocks, if its matrix of generating loadings **P** satisfies the following requirements:

- each of the first b columns of **P** is proportional to a vector with values 0 and 1. Moreover, these b columns should form a partition of the variables (these are the b block-components, where each variable belongs to one and only one block);
- the last (q-b) columns of **P** are proportional to vectors which contain m times the value h, h times the value -m and p-m-h times the value 0 (m and h are strictly positive integers such that $m+h \leq p$, and m and h might be different for each column); this leads to (q-b) so called difference-components (as the sum of loadings is zero for each component, these differences correspond to the contrasts of the variables);

- (optional) we obtain then "within-block" difference-components, as in case of an exact block-structure in the correlation matrix.

The first stage of SCA classifies *p* variables into *b* disjoint blocks. The approximate block-structure in the correlation matrix leads to a maximum within block correlations and, in the meantime, to a minimum between block correlations. Several authors [20] have dealt this problem with an agglomerative hierarchical procedure based on a dissimilarity measure between clusters called "median" linkage, alternative to the possible "single" or "complete" linkages.
With reference to the loading matrix corresponding to the *b* simple block-components of the first stage, the second step of the algorithm is based on a suitable difference-component shrinkage procedure of the sequential first components of the residual variables obtained by regressing, step by step, the original variables on the first ($j-1$) simple components.
Recent developments have modified the first stage criterion [5], using the so called RV vectorial correlation coefficient [18], given by:

$$\mathrm{RV}(V,Z)=\frac{tr(\mathbf{W}_V\mathbf{D}\mathbf{W}_Z\mathbf{D})}{\sqrt{tr(\mathbf{W}_V\mathbf{D})^2\,tr(\mathbf{W}_Z\mathbf{D})^2}} \tag{8}$$

where $\mathbf{W}_V = \mathbf{V}\mathbf{Q}_1\mathbf{V}'$ and $\mathbf{W}_Z = \mathbf{Z}\mathbf{Q}_2\mathbf{Z}'$ are the scalar product matrices associated to the vector variables *V* and *Z*, respectively, with $\mathbf{Q}_1$ and $\mathbf{Q}_2$ metric matrices and $\mathbf{D}$ a weighting diagonal matrix. It provides a measure of the similarity of the two configurations *V* and *Z*, taking also into account the possibly distinct metrics to be used on them.

4. Patient Satisfaction in a local public health service

Often in patient satisfaction analysis the first component of PCA quite corresponds to overall size, so it can be considered as a sum or an average of the original variables. To show how the SCA analysis produces easier interpretable results compared to PCA we use a dataset relative to the patient satisfaction in a Hospital of the City of Naples.
Data collected is based on 35 items measuring five latent dimensions [3]. Item V1 concerns the "Managing Hospitalization" dimension, items V2 – V9 measure the "Medical Assistance" dimension, items V10 – V15 are related to the "Nursing" dimension, items V16 – V26 to the "Organization" dimension and items V27 – V35 to "Tangible aspects and location" dimension.
The nominal scales used in the questionnaire range from 1 to 6 and are easy to understand, for each instance and given meaning.
Before analyzing the data, it was decided to pre-treat them, in order to ensure the quality of information that will be extracted. In particular, the following operations were performed:

- assessing the quality of the data, responding to the requirements defined by Eurostat documentation in evaluating the quality of statistics produced by the member countries of the European Community; this concerns the dimensions of relevance, accuracy, timeliness, transparency, comparability, consistency, completeness;
- the treatment of missing data, using deterministic techniques (deductive imputation, mean imputation);
- Thurstone scaling, which respects the ordinal nature of the data through a transformation from ordinal to metrical [22], [6].

Table 1. Description of Variables

DIMENSIONS	VARIABLES	
Managing Hospitalization	**V1**	Reception
Medical Assistance	**V2**	Competence of doctors
	V3	Information about the disease and treatments
	V4	Availability of doctors
	V5	Regularity and frequency of medical
	V6	Medical care
	V7	Information on the risks of therapy
	V8	Information received about surgery and / or diagnostic
	V9	Privacy by doctors
Nursing	**V10**	Competence of nursing
	V11	Availability of nurses
	V12	Regularity and frequency of nursing care
	V13	Readiness of the nursing staff
	V14	Ability to calm
	V15	Privacy by nurses
Organization	**V16**	Availability staff
	V17	Staff ability to offer customized services
	V18	Food Quality
	V19	Food Quantity
	V20	Mealtimes
	V21	Choice menu
	V22	Personalized menus
	V23	Entertainment and socialization
	V24	Visiting
	V25	Common Areas
	V26	Quality spaces
Tangible aspects and location	**V27**	Identifiability and decorum staff
	V28	Hospital environments
	V29	Clean Rooms
	V30	Comfort of beds
	V31	Clean linen
	V32	Restrooms
	V33	Therapy in post-resignation
	V34	Dietary recommendations for post-resignation
	V35	Physical activity for post-resignation

4.2 Results of SCA and comparison

The comparison of the different exploratory techniques prove that the SCA technique made it possible to obtain a preliminary analysis to highlight the links between the different relevant variables. These results may support the assessment phase of real Patient Satisfaction in which one can propose and validate a specific statistical model.

The following tables show the results of the SCA.

Table 2. Percentage of extracted variability

Extracted variability PCA	54.51 %
Extracted variability SCA	50.42 %
Optimality SCA	92.50 %

Table 3. Optimality Criteria used in SCA

SIMPLE COMPONENT ANALYSIS	
Optimality criterion	Corrected sum of variances
Clustering procedure	Median linkage
Number of block-components	2
Number of diff.-components	3

Table 4. SCA components

Dimensions	Var.	B1	B2	D3	D4	D5
Managing Hospitalization	***V1***	1	0	13	0	0
Medical Assistance	***V2***	1	0	13	5	0
	V3	1	0	13	5	0
	V4	1	0	13	5	0
	V5	1	0	13	0	13
	V6	1	0	13	0	0
	V7	1	0	13	5	0
	V8	1	0	13	5	0
	V9	1	0	0	5	-7
Nursing	***V10***	1	0	-12	-6	0
	V11	1	0	-12	-6	-7
	V12	1	0	0	-6	-7
	V13	1	0	-12	-6	-7
	V14	1	0	0	-6	-7
	V15	1	0	0	-6	-7

Dimensions	Var.	B1	B2	D3	D4	D5
Organization	***V16***	1	0	13	-6	-7
	V17	1	0	13	-6	-7
	V18	1	0	13	-6	0
	V19	1	0	13	-6	0
	V20	1	0	-12	5	-7
	V21	1	0	-12	5	-7
	V22	1	0	-12	5	-7
	V23	0	1	0	0	0
	V24	1	0	-12	5	-7
	V25	0	1	0	0	0
	V26	1	0	-12	5	-7
Tangible aspects and location	***V27***	0	1	0	0	0
	V28	0	1	0	0	0
	V29	1	0	0	0	13
	V30	1	0	-12	0	13
	V31	1	0	-12	0	13
	V32	1	0	-12	0	13
	V33	1	0	-12	0	13
	V34	1	0	-12	0	13
	V35	1	0	0	5	0
		26%	6.%	8%	7%	5%

Table 5. Comparison of results

ACP					VARIMAX					SCA				
C1	**C2**	**C3**	**C4**	**C5**	**C1**	**C2**	**C3**	**C4**	**C5**	**B1**	**B2**	**D1**	**D2**	**D3**
0,57	-0,25	0,07	-0,01	-0,08	0,52	0,19	0,17	0,04	0,29	1	0	13	0	0
0,55	-0,36	0,35	-0,08	0,13	0,73	0,08	-0,01	0,19	0,13	1	0	13	5	0
0,62	-0,41	0,40	-0,12	-0,02	0,85	0,11	0,10	0,07	0,03	1	0	13	5	0
0,62	-0,36	0,40	-0,02	0,02	0,83	0,15	0,06	0,09	0,02	1	0	13	5	0
0,56	-0,25	0,13	0,13	-0,21	0,54	0,18	0,19	-0,12	0,23	1	0	13	0	13
0,62	-0,33	0,19	0,09	-0,10	0,67	0,24	0,11	-0,06	0,16	1	0	13	0	0
0,46	-0,31	0,38	-0,11	-0,01	0,66	0,03	0,12	-0,03	-0,06	1	0	13	5	0
0,58	-0,37	0,27	-0,08	0,06	0,70	0,19	0,07	-0,02	0,03	1	0	13	5	0
0,49	-0,16	0,17	-0,15	0,21	0,41	0,15	0,11	0,06	0,03	1	0	0	5	-7
0,68	0,14	-0,34	0,01	0,07	0,13	0,69	0,30	0,04	0,15	1	0	-12	-6	0
0,68	0,11	-0,38	0,02	0,14	0,16	0,78	0,20	0,09	0,14	1	0	-12	-6	-7
0,68	0,12	-0,36	0,14	0,13	0,16	0,80	0,17	0,04	0,12	1	0	0	-6	-7
0,67	0,12	-0,36	-0,05	0,35	0,16	0,82	0,10	0,21	0,04	1	0	-12	-6	-7
0,70	0,08	-0,33	0,09	0,22	0,21	0,80	0,11	0,12	0,12	1	0	0	-6	-7
0,57	0,00	-0,22	0,02	0,22	0,16	0,50	0,13	-0,02	0,17	1	0	0	-6	-7
0,72	-0,15	-0,11	0,18	0,10	0,46	0,56	0,06	0,06	0,28	1	0	13	-6	-7
0,66	-0,17	-0,19	0,18	0,12	0,35	0,46	0,04	0,08	0,46	1	0	13	-6	-7
0,57	-0,21	-0,34	0,36	-0,16	0,20	0,25	0,10	-0,03	0,90	1	0	13	-6	0
0,57	-0,21	-0,34	0,36	-0,16	0,20	0,25	0,10	-0,03	0,90	1	0	13	-6	0
0,39	0,45	0,14	-0,26	0,21	0,09	0,25	0,23	0,68	-0,08	1	0	-12	5	-7
0,39	0,49	0,24	-0,12	0,34	0,10	0,19	0,13	0,76	-0,02	1	0	-12	5	-7
0,25	0,51	0,31	-0,07	0,32	0,00	0,00	0,09	0,74	0,01	1	0	-12	5	-7
0,19	0,33	0,31	0,32	-0,04	0,10	0,08	0,04	0,21	-0,09	0	1	0	0	0
0,16	0,31	0,29	-0,04	0,41	0,00	0,03	-0,10	0,46	-0,08	1	0	-12	5	-7
0,15	0,26	0,20	0,13	-0,21	0,02	-0,01	0,14	0,16	0,07	0	1	0	0	0
0,32	0,26	0,09	-0,08	0,29	0,06	0,05	0,13	0,51	0,19	1	0	-12	5	-7
0,22	0,42	0,32	0,71	-0,14	0,03	0,05	0,05	0,09	0,08	0	1	0	0	0
0,19	0,41	0,32	0,71	-0,14	0,00	0,03	0,05	0,05	0,06	0	1	0	0	0
0,42	0,06	0,11	-0,03	-0,22	0,23	0,06	0,32	0,04	0,19	1	0	0	0	13
0,53	0,37	-0,03	-0,25	-0,45	0,08	0,19	0,79	0,04	0,03	1	0	-12	0	13
0,59	0,29	-0,04	-0,32	-0,50	0,19	0,18	0,83	0,07	0,09	1	0	-12	0	13
0,55	0,16	-0,06	-0,31	-0,13	0,25	0,25	0,52	0,26	0,09	1	0	-12	0	13
0,59	0,30	-0,13	-0,20	-0,25	0,14	0,29	0,64	0,22	0,17	1	0	-12	0	13
0,45	0,37	0,03	-0,32	-0,48	0,10	0,10	0,79	0,11	-0,02	1	0	-12	0	13
0,56	-0,09	0,28	-0,09	0,04	0,52	0,10	0,17	0,24	0,14	1	0	0	5	0
27.8%	**8.6%**	**6.9%**	**5.9%**	**5.3%**	**14.6%**	**12.8 %**	**9.2%**	**7.1%**	**6.8%**	**26.1%**	**6.0%**	**7.6%**	**6.8%**	**4.8%**

It may be observed that there is no significant difference between the first components extracted with either PCA or SCA. However, an interesting contribution comes from the information matrix of the *simple difference-components* (D3, D4, D5), showing strengths and weaknesses on which the hospital management should take action to improve the service.

These aspects are summarized in the following tables.

Table 6. Comparison of results for Difference component 2 (D2)

AREAS	Evaluation by D2	
	Positive	**Negative**
Managing Hospitalization	Reception	
Medical Assistance	Competence of doctors	
	Information about the disease and treatments	
	Regularity and frequency of medical	
	Regularity and frequency of medical	
	Medical care	
	Information on the risks of therapy	
	Information received about surgery and / or diagnostic	
Organization	Availability staff	Mealtimes
	Staff ability to offer customized services	Choice menu
	Food Quality	Personalized menus
	Food Quantity	Visiting
		Common Areas
		Quality spaces
Nursing		Competence of nursing
		Availability of nurses
		Readiness of the nursing staff
Tangible aspects and location		Comfort of beds
		Clean linen
		Clean Rooms
		Therapy in post-resignation
		Dietary recommendations for post-resignation

Table 7. Comparison of results for Difference component 3 (D3)

AREAS	Evaluation by D3	
	Positive	**Negative**
Managing Hospitalization		
Medical Assistance	Competence of doctors	
	Information about the disease and treatments	
	Regularity and frequency of medical	
	Information on the risks of therapy	
	Information received about surgery and/or diagnostic	
	Privacy by doctors	
Organization	Mealtimes	Availability staff
	Choice menu	Staff ability to offer customized services
	Personalized menus	Food Quality
	Visiting	Food Quantity
	Quality spaces	
Nursing		Competence of nursing
		Availability of nurses
		Regularity and frequency of nursing care
		Readiness of the nursing staff
		Ability to calm
		Privacy by nurses
Tangible aspects and location	Physical activity for post-resignation	

Table 8. Comparison of results for Difference component 4 (D4)

AREAS	Evaluation by D4	
	Positive	**Negative**
Managing Hospitalization		
Medical Assistance	Competence of doctors	
	Information about the disease and treatments	
	Regularity and frequency of medical	
	Information on the risks of therapy	
	Information received about surgery and / or diagnostic	
	Privacy by doctors	
Organization	Mealtimes	Availability staff
	Choice menu	Staff ability to offer customized services
	Personalized menus	Food Quality
	Visiting	Food Quantity
	Quality spaces	
Nursing		Competence of nursing
		Availability of nurses
		Regularity and frequency of nursing care
		Readiness of the nursing staff
		Ability to calm
		Privacy by nurses
Tangible aspects and location	Physical activity for post-resignation	

5. Conclusion

The critical analysis of the techniques useful for the exploratory assessment of Customer Satisfaction leads to consider the SCA to be a valid alternative to the traditional tools of analysis (PCA and rotation criteria): despite to their pseudo optimal performance, they make the interpretation of results easier and not trivial. This work is accompanied by an application on real data, collected from a hospital in Naples, of the techniques here discussed (PCA, rotation criteria and SCA), providing a useful comparison and allowing an enrichment of the exploratory phase.
PCA is a useful tool when the number of observed variables is very large. Furthermore, in case of variables with high levels of correlation, PCA doesn't produce fully interpretable results. In order to overcome this problem of interpretability, SCA focuses on simplicity and seeks optimal simple components. It introduces the aspect of simplicity leading to a sufficiently large number of block components in the system and to a simple weighting scheme.
A lot of vectorial correlation coefficients proposed in the literature do not respect all the cited properties [1]. Several proposals in the literature considered the use of RV coefficients for variable selection, like, for example, the selection of subsets of variables in the context of PCA [4],[15]. A comparative study of the performance effects of several correlation matrices within SCA is under investigation.

Another interesting application of this technique is related to the methods of the adaptive modeling procedures (e.g., partial least squares, discriminant analysis, canonical correlation analysis, etc.) where the generated linear combinations of variables are often not easily interpretable.

References

[1]. Amenta, P. (1993). Il coefficiente di correlazione lineare tra matrici di dati nel contesto multivariato. In *Proceeding of XVII Convegno A.M.A.S.E.S., Ischia.*

[2]. Basilevsky, A. (1994). *Statistical Factor Analysis and Related Methods*, New York: John Wiley.

[3]. Babakus, E., Mangold, G. (1992).Adapting the Servqual scale to hospital services: an empirical investigation. *Health Services Research*, 26(6), 767–786.

[4]. Bonifas, I., Escoufier, Y., Gonzalez, P., Sabatier, R. (1984). Choix de variables en analyse en composantes principales. *Revue de Statistique Appliquée*, 23, 5–15.

[5]. Choulakian, H.A., D'Ambra, L., Simonetti, B. (2005). Hausman principal component analysis. In *Proceeding of 29th Annual International Conference of the German Classification Society* (GfKl 2005), Magdeburg : Germany.

[6]. Ciavolino, E., Dahlgaard J.J. (2009). Simultaneous equation model based on generalized maximum entropy for studying the effect of the management's factors on the enterprise performances. *Journal of Applied Statistics*, 36 (7), 801–815.

[7]. Gallo, M., Amenta, P., D'Ambra, L. (2006). Simple component analysis based on RV Coefficient. In *Data Analysis, Classification and the Forward Search*, 1, 93–102, eds. Zani, S., Cerioli, A., Riani, M., Vichi, M.. Berlin: Springer.

[8]. Hausman, R.E. (1982). Constrained multivariate analysis. In *Optimization in Statistics*, eds. Zanckis, S.H. and Rustagi, J.S.. Amsterdam: North Holland.

[9]. Jolliffe, I.T. (1989). Rotation of ill-defined principal components. *Applied Statistics*, 38, 139 – 147.

[10]. Jolliffe, I.T., Uddin, M. (2003). A modified principal component technique based on the Lasso. *Journal of Computation and Graphical Statistics*, 12, 531 – 547.

[11]. Jolliffe, I.T. (1995). Rotation of principal components: choice of normalization constraints. *J. Appl. Statist*., 22, 29–35.

[12]. Jolliffe, I.T. (2002). *Principal Component Analysis* (2nd ed.), New York: Springer.

[13]. Jolliffe, I.T., Uddin, M. (2000). The simplified component technique: An alternative to rotated principal components. *J. Comput. Graph. Statist*., 9, 689–71.

[14]. Kaiser, H.F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrik*a, 23, 187–20.

[15]. Mori, Y., Iizuka, M., Tarumi, T., Tanaka, Y. (1999). Variable selection in principal component analysis based on a subset of variables. *Bulletin of the International Statistical Institute 52nd Session*, volume Contributed Papers Book 2, 333–334.

[16]. Morrison, D.F. (1976) .*Multivariate Statistical Methods*, New York: McGraw-Hill.

[17]. Ramsay, J., Ten Berge, J., Styan, G. (1984). Matrix correlation. *Psychometrika*, 49, 403–423.

[18]. Robert, P., Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods:the rv-coefficient. *Applied Statistics*, 25, 257–265.

[19]. Rousson, V., Gassen, T. (2003). Some case studies of simple component analysis. Manuscript on http://www.unizh.ch/biostat/Manuscripts.
[20]. Rousson, V., Gasser, Th. (2003). Simple Component Analysis. *Appl. Stat*., 53, part 4, 539-555.
[21]. Vines, S.K. (2000). Simple principal components. *Appl. Stat*., 49, 441–451.
[22]. Zanella, A. (2001). Measures and models of customer satisfaction: the underlying conceptual construct and a comparison of different approaches, Sixth TQM World Congress, Saint Petersburg.