

ALTERNATIVE METHODS TO MULTIPLE CORRESPONDENCE ANALYSIS IN RECONSTRUCTING THE RELEVANT INFORMATION IN A BURT'S TABLE

Sergio Camiz¹ and Gastão Coelho Gomes^{2*}

Received September 11, 2015 / Accepted February 20, 2016

ABSTRACT. In this work, the reconstruction of the Burt's table, Greenacre (1988)'s Joint Correspondence Analysis (*JCA*), and Gower & Hand (1996)'s Extended Matching Coefficient (*EMC*) are compared to Multiple Correspondence Analysis (*MCA*) in order to check the quality of the methods. In particular, for the whole table, the ability is considered separately the diagonal, and the off-diagonal tables, that is the ability to describe either each character's distribution or the interaction between pairs of characters, or both. The theoretical aspects are discussed first, and finally the results obtained in an application are shown and discussed.

Keywords: Correspondence Analysis, Multiple Correspondence Analysis, Joint Correspondence Analysis, Extended Matching Coefficient, Singular Value Decomposition.

1 INTRODUCTION

Both Multicriteria Analysis and Multicriteria Decision Models [4, 37] are tools largely adopted in operational research, in particular when dealing with Knowledge Discovering in Databases [14]. In present days it may be of high interest to deal with qualitative unstructured data, whose treatment may be more complex. Studies in this framework are found in recent operational research literature [16, 17, 30, 40]. In this context, data reduction through exploratory multidimensional scaling may contribute to clarify the data at hand, by revealing structures and factors. In particular, factors, together with the observed characters most associated to them, may lead to a consistent dimension reduction and at the same time to the ability to select the most appropriate characters to take into account for further, more focused investigation. Thus, the identification of the proper dimension of a data table may be a topic of investigation *per se*.

In exploratory multidimensional scaling the identification of the proper dimension of the solution is the basis to define a threshold between relevant information and residuals. The relevant information is also tied to the possibility of interpreting the factors according to the paradigms of the

*Corresponding author.

¹Dipartimento di Matematica, Sapienza Università di Roma, Italy. E-mail: sergio.camiz@uniroma1.it

²Departamento de Métodos Estatísticos, Instituto de Matemática, Universidade Federal do Rio de Janeiro, Brazil. E-mail: gastao@im.ufrj.br

methods at hand: in the linear case, the percentage of explained inertia is the most widely used. Thus, to take into account a large share of inertia is the most evident rough method that may be used and a higher-dimensional solution is normally preferred rather than a smaller one only if its corresponding inertia is significantly larger than a smaller-dimensional solution. Tied to this aspect, the reconstruction of the original data table according to a lower rank matrix is of relevance, since a good reconstruction of the data obtained this way is helpful to better understand to what extent the reduction in dimension, through the use of factors, is a reasonable approximation of the original data.

In this paper, we consider the special case of qualitative data, that are usually summarized by the so-called Burt's matrix, the super-contingency table that cross-tabulates all characters taken into account. Multiple Correspondence Analysis (*MCA*) [7, 19] is the best known exploratory factor analysis method to deal with it, but alternative methods are proposed in literature, based on different rationale. Critics to *MCA* emphasize the misuse of the chi-square metrics [18], a metrics that for contingency tables finds its rationale in the partition of the chi-square in independent components [27, 35, 44], thus on the deviation from the expectation. Now, to compute distances between lines in both the indicator matrix and its square, the Burt's table, such metrics is tremendously biased by the obvious (squared) differences between levels belonging to the same character [20, 21]. In addition, rare levels raise their importance in the computation but enhance aspects that reveal being useless, since the chi-square statistics may not be applied to these tables; hence, its use is hardly justifiable [18]. Eventually, it is known since long the problem of the underestimation of the inertia explained by the factors, that deserve being re-evaluated in some way [2, 8, 19]. Indeed, applying the chi-square metrics to such a table, the highest contributions to the total inertia result from the block-diagonal tables crossing each character with itself, which is information without noticeable value, since this way the expected values provide the maximum deviation from expectation. Along with this, a problem results in the unpredictability of the partial data reconstruction of the Burt's table, as already put in evidence by [10]. The problem may be relevant in qualitative discriminant analysis [39], which is based on *MCA* coordinates. Here, a bad reconstruction would prevent the reduction of the number of factors to take into account.

In this paper, we consider two different alternatives: the Joint Correspondence Analysis (*JCA*, [20]) and the Principal Component Analysis (*PCA*, [26]) of the Extended Matching Coefficient matrix (*EMC*, [18]), which are suggested by the corresponding authors as a solution. It must be noted that the aims of both *MCA* and *EMC* are the same: to perform a kind of *PCA* on qualitative data, i.e. to find uncorrelated quantitative characters giving independent ordinations to the units, that optimize the explained inertia. It is the definition of inertia that is different, bearing in mind that *EMC* aims at fixing the emphasis given to the rare levels of *MCA* by the chi-square metrics. Since the solutions of both methods are based on the eigendecomposition, their different dimensional solutions are encapsulated, so that the reconstruction of the original data may be seen as a sum of 1-rank independent tables. On the opposite, *JCA* aims at finding the best reconstruction of the off-diagonal tables of the Burt's table only, generalizing the Simple

Correspondence Analysis to the simultaneous analysis of several 2-way contingency tables. The idea is analogous to the one of [42] to fit the off-diagonal elements of a correlation matrix and is obtained by numerical optimization. Therefore, this time different dimensional solutions are not encapsulated.

To compare the results, we consider the reconstruction quality only. At the end, these methods will be applied to a very small table, taken from studies in linguistics [32].

2 THEORETICAL FRAMEWORK

2.1 Singular Value Decomposition

We may ground our further discussion on the well-known Singular Value Decomposition (SVD, [1, 19]) theorem, which states

Theorem 1. Any real matrix X may be decomposed as $X = U\Lambda^{1/2}V'$, with Λ the diagonal matrix of the real non-negative eigenvalues of XX' , U the orthogonal matrix of the corresponding eigenvectors, and V the matrix of eigenvectors of $X'X$ (with the same eigenvalues), with both constraints $U'U = I$ and $V'V = I$.

This theorem corresponds to the reconstruction formula of an S -rank matrix $X \in \mathbb{R}^{r \times c}$

$$x_{ij} = \sum_{\alpha=1}^S \sqrt{\lambda_{\alpha}} u_{i\alpha} v_{j\alpha}, \quad \forall i \in (1, r), j \in (1, c) \tag{1}$$

or, in vector notation

$$X = \sum_{\alpha=1}^S \sqrt{\lambda_{\alpha}} \mathbf{u}_{\alpha} \mathbf{v}'_{\alpha},$$

on which the Eckart & Young's theorem [13] is based:

Theorem 2. (Eckart and Young) The s -rank reconstruction of any real matrix X , with $s < S$, the rank of X , once its singular values are sorted in decreasing order,

$$X \approx \sum_{\alpha=1}^s \sqrt{\lambda_{\alpha}} \mathbf{u}_{\alpha} \mathbf{v}'_{\alpha} = \widehat{H}$$

is the best one in the least-squares sense.

This means that, for every $s < S$, the matrix \widehat{H}_s solves the problem to approximate a matrix X by another matrix H of lower rank at the best in the least-squares sense, thus by minimizing

$$\sum_{i=1}^r \sum_{j=1}^c (x_{ij} - h_{ij})^2 = \text{trace}((X - H)(X - H)') \tag{2}$$

As it is trivial that $\text{trace}(\widehat{H}_s \widehat{H}'_s) = \sum_{\alpha=1}^s \lambda_{\alpha}$, the minimum of (2) reached by \widehat{H}_s equals

$$\text{trace}((X - \widehat{H}_s)(X - \widehat{H}_s)') = \sum_{\alpha=s+1}^S \lambda_{\alpha}.$$

It is well-known that Principal Component Analysis (PCA, [26]) finds its rationale in this theorem, once the data table is standardized according to $z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{n\sigma_j}}$, with \bar{x}_j and σ_j the average and the standard deviation of the j -th character; indeed, this way $Z'Z = cor(X) = C$, the matrix of correlations between the columns of X . Thus, given the PCA of the correlation matrix C , with Λ and V as their diagonal matrix of eigenvalues and unit matrix of eigenvectors respectively, and given U as the unit eigenvectors of ZZ' , the reconstruction formula (1) becomes

$$x_{ij} = \left(\sum_{\alpha=1}^r \sqrt{\lambda_\alpha} u_{i\alpha} v_{j\alpha} \right) \sqrt{n}\sigma_j + \bar{x}_j, \quad \forall i \in (1, r), j \in (1, c) \tag{3}$$

For correspondence analysis, we shall adopt the Generalized Singular Values Decomposition (GSVD, [1, 19]), in which two other matrices are involved:

Theorem 3. *Given two real positive definite matrices M and N , any real matrix X may be decomposed as $X = \tilde{U} \Lambda^{1/2} \tilde{V}'$, under constraints $\tilde{U}' M \tilde{U} = I$ and $\tilde{V}' N \tilde{V} = I$.*

The solution is given by the SVD of the matrix $\tilde{X} = M^{1/2} X N^{1/2} = \tilde{F} \Lambda^{1/2} \tilde{G}'$, with $\tilde{F}' \tilde{F} = I$, $\tilde{G}' \tilde{G} = I$, $\tilde{U} = M^{-1/2} \tilde{F}$, and $\tilde{V} = N^{-1/2} \tilde{G}$. It results that $\tilde{U} \tilde{U}' = M^{-1}$ and $\tilde{V} \tilde{V}' = N^{-1}$ respectively.

In this case the minimization problem (2) becomes

$$\sum_{i=1}^r \sum_{j=1}^c (\tilde{x}_{ij} - \tilde{h}_{ij})^2 = \text{trace} ((\tilde{X} - \tilde{H})(\tilde{X} - \tilde{H})')$$

with $\tilde{H} = \sum_{\alpha=1}^s \sqrt{\lambda_\alpha} \tilde{f}_\alpha \tilde{g}'_\alpha$, with $s \leq S = \text{rank}(\tilde{X})$.

Using the matrix

$$\hat{H}_s = M^{-1/2} \tilde{H} \tilde{N}^{-1/2} = \sum_{\alpha=1}^s \sqrt{\lambda_\alpha} (M^{-1/2} \tilde{f}_\alpha) (N^{-1/2} \tilde{g}'_\alpha)', \tag{4}$$

as $\text{trace} (\hat{H} \hat{H}') = \sum_{\alpha=1}^s \lambda_\alpha$, the minimization problem is solved as:

$$\sum_{l=1}^r \sum_{k=1}^c \left(\sum_{i=1}^r \sum_{j=1}^c m_{li}^{1/2} (x_{ij} - \hat{h}_{ij})^2 n_{jk}^{1/2} \right)^2 = \text{trace} (M(X - \hat{H}_s)N(X - \hat{H}_s)') = \sum_{\alpha=s+1}^S \lambda_\alpha. \tag{5}$$

In the particular case in which both M and N are diagonal, (5) simplifies to

$$\sum_{i=1}^r \sum_{j=1}^c m_{ii} n_{jj} (x_{ij} - \hat{h}_{ij})^2 = \text{trace} (M(X - \hat{H}_s)N(X - \hat{H}_s)') = \sum_{\alpha=s+1}^S \lambda_\alpha. \tag{6}$$

Therefore, the exploratory analysis paradigm states that the most relevant information is tied to the largest eigenvalues and the non-relevant to the least ones. The problem of distinguishing

among them, that is to identify at least a tentative cutpoint of either the singular- or the eigenvalues sequence, remains a crucial issue, that did not find a univocal solution so far (for PCA see, e.g., [11, 25, 36]). In Simple Correspondence Analysis (SCA, [7, 19]), it seems more easily solved, since the special chi-square metrics adopted allows a useful solution and an easy interpretation of the results, and regarding MCA we shall see that [5, 6] propose an interesting method.

2.2 Simple Correspondence Analysis

Let N an $r \times c$ contingency table, with $n = n_{..}$ the table grand total, $X = N/n$ the table of relative frequencies $p_{ij} = n_{ij}/n$, $\mathbf{r} = (p_{1.}, \dots, p_{r.})'$ the vector of row marginal profile, $\mathbf{c} = (p_{.1}, \dots, p_{.c})'$ the vector of column marginal profile, and $D_r = \text{diag}(\mathbf{r})$, $D_c = \text{diag}(\mathbf{c})$ the corresponding diagonal matrices. Let $E = \mathbf{r}\mathbf{c}'$ represents the table with equal marginal profiles of X under the independence hypothesis. The SCA of N results from the application of GSVD to the matrix X with the real positive definite matrices represented by the diagonal matrices D_r^{-1} and D_c^{-1} . In this particular case the decomposition results $\tilde{X} = D_r^{-1/2} X D_c^{-1/2} = \tilde{F} \Lambda^{1/2} \tilde{G}'$ and both matrices $\tilde{X}\tilde{X}'$ and $\tilde{X}'\tilde{X}$ have a trivial eigenvalue $\lambda_1 = 1$ to which the eigenvectors $\tilde{f}_1 = (\sqrt{r_i})$ and $\tilde{g}_1 = (\sqrt{c_j})$ correspond respectively. It can also be shown that the other non-trivial eigenvalues are always contained between 0 and 1. If we take off the summation the trivial eigenvalue, the s -rank reconstruction formula ($s \leq S = \min(r, c)$) may be little transformed:

$$\begin{aligned}
 p_{ij} &\approx \hat{h}_{ij,s} = r_i c_j \sum_{\alpha=1}^s \sqrt{\lambda_\alpha} (r_i^{1/2} \tilde{f}_{i\alpha}) (c_j^{1/2} \tilde{g}_{j\alpha}) \\
 &= r_i c_j \left(1 + \sum_{\alpha=2}^s \sqrt{\lambda_\alpha} (r_i^{-1/2} \tilde{f}_{i\alpha}) (c_j^{-1/2} \tilde{g}_{j\alpha}) \right)
 \end{aligned}
 \tag{7}$$

Incidentally, we observe that, in order to produce graphics with simultaneous symmetrical representation of both rows and columns, the SCA eigenvectors are usually rescaled by defining as coordinates the vectors φ_α and ψ_α given by $\varphi_{i\alpha} = \lambda_\alpha^{1/2} r_i^{-1/2} \tilde{f}_{i\alpha}$ and $\psi_{j\alpha} = \lambda_\alpha^{1/2} c_j^{-1/2} \tilde{g}_{j\alpha}$, respectively. In this case the reconstruction formula (7) becomes

$$p_{ij} \approx \hat{h}_{ij,s} = r_i c_j \left(1 + \sum_{\alpha=2}^s \frac{1}{\sqrt{\lambda_\alpha}} \varphi_{i\alpha} \psi_{j\alpha} \right).$$

Thus, depending on which coordinates one chooses, the reconstruction formula for N becomes:

$$N = n \mathbf{r} \mathbf{c}' + n D_r F \Lambda^{1/2} G' D_c = n \mathbf{r} \mathbf{c}' + n D_r \Phi \Lambda^{-1/2} \Psi' D_c.
 \tag{8}$$

Returning to the optimization problem (6), we may explicitly write the residual as

$$\text{trace} \left(D_r^{-1} (X - \hat{H}_s) D_c^{-1} (X - \hat{H}_s)' \right) = \sum_{i=1}^r \sum_{j=1}^c \frac{(x_{ij} - \hat{h}_{ij,s})^2}{r_i c_j} = \sum_{\alpha=s+1}^S \lambda_\alpha.
 \tag{9}$$

Now, if we consider the 1-rank reconstruction, that is $\hat{H}_1 = E = \mathbf{rc}'$, the residual takes the particular form

$$\sum_{i=1}^r \sum_{j=1}^c \frac{(x_{ij} - \hat{h}_{ij,1})^2}{r_i c_j} = \sum_{i=1}^r \sum_{j=1}^c \frac{(x_{ij} - r_i c_j)^2}{r_i c_j} = \sum_{\alpha=2}^S \lambda_{\alpha}. \tag{10}$$

that is the sum of squared deviations of the observed values from the expected ones under independence divided by the expected ones. This is commonly called the *inertia* of the table X that is, up to the factor n , the chi-square of N ; indeed, multiplying (10) by n , we get

$$n \sum_{i=1}^r \sum_{j=1}^c \frac{(p_{ij} - r_i c_j)^2}{r_i c_j} = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - n r_i c_j)^2}{n r_i c_j} = \chi^2 = n \sum_{\alpha=2}^S \lambda_{\alpha}. \tag{11}$$

Hence, the decomposition of the inertia along the non-trivial eigenvectors of SCA leads to the decomposition of the chi-square along each corresponding dimension:

$$\chi^2 = n(\lambda_2 + \dots + \lambda_S) = \chi_2^2 + \dots + \chi_S^2$$

α equals $\chi_{\alpha}^2 = n\lambda_{\alpha}$.

Based on [44], this result led [35] to check for significance the partial chi-squares associated to each eigenvalue, χ_{α}^2 with $df = (r + c - 2\alpha - 1)$ degrees of freedom, to detect if there are linear ordinations of both rows and column levels that explain the deviation from expectation. Indeed, [28] proved that the first eigenvalue is larger than the corresponding chi-square and [29] proved that the distribution of the eigenvalues is that of those of a Wishart matrix $(W(m_1 - 1)(m_2 - 1), Min(m_1 - 1, m_2 - 1), I)$. Nevertheless, [15] finds that the sum of the smallest eigenvalues may be used as a test for their nullity, that leads to the Malinvaud's [31] stopping rule, to test the residuals of each reduced rank solution. Indeed, one may test for significance

$$\tilde{Q}_s = \sum_{ij} \frac{(n_{ij} - n\hat{h}_{ij,s})^2}{nr_i c_j} = \chi^2 - \sum_{\alpha=2}^s \chi_{\alpha}^2 = n \sum_{\beta=s+1}^S \lambda_{\beta},$$

asymptotically chi-square-distributed with $(r - \alpha - 1) \times (c - \alpha - 1)$ degrees of freedom.

2.3 Multiple Correspondence Analysis

Let us consider now a qualitative data table X with n observations, with Q as nominal characters and J as the total number of levels, that is $J = \sum_{i=1}^Q l_i$, where l_i is the number of levels of the i -th character. It is well-known that MCA of such a matrix is nothing but a generalization of SCA and it is based on SCA of either the indicator matrix Z , whose rows are the units and the columns are all the J levels of the Q considered variables, or the so-called Burt's table $B = Z'Z$ that gathers all contingency tables obtained by cross-tabulating all the variables in Z , including the diagonal tables obtained by crossing each variable with itself. The idea is to adopt for both matrices the same optimized decomposition of SCA , namely the $GSVD$ of either $Q^{-1}ZD_r^{-1}$ or

$$Q^{-2}D_r^{-1}BD_r^{-1} = Q^{-2}D_r^{-1}Z'ZD_r^{-1}. \tag{12}$$

Indeed, it is evident that the latter is the square of the previous, so that they share the eigenvectors, and the singular values in Burt's case are the squares of those of the indicator matrix case: $\nu_\alpha = \mu_\alpha^2$. The identity of the eigenvalues allows identical interpretation of the resulting factors. Thus, it makes no difference to perform *MCA* on either matrix. On the other side, whereas the total inertia of Z is $I_z = \frac{J-Q}{Q}$, the one of B equals $\sum \nu_\alpha = \sum \mu_\alpha^2$. In both cases, the chi-square metrics is adopted so that the interpretation of results ought to be done once again in terms of deviations from expectation. This point deserves some special attention, since the deviation refers to all contingency tables gathered in the Burt's table, including the diagonal ones. The problem is that such diagonal matrices, that "theoretically" would indicate maximum deviation, in this case are just the expected ones, as they cross each character with itself.

As *SCA*, given a Burt matrix B , *MCA* may be defined as the weighted least-squares approximation of B by another matrix H of lower rank, minimizing

$$n^{-1} Q^{-2} \text{trace} \left(D_r^{-1} (B - H) D_r^{-1} (B - H)' \right). \tag{13}$$

Notice how (13) derives from (10). In terms of the subtables, this may be rewritten as

$$\begin{aligned} & n^{-1} Q^{-2} \text{trace} \left(D^{-1} (B - H) D^{-1} (B - H)' \right) = \\ & = n^{-1} \sum_{i=1}^Q \sum_{j=1}^Q \text{trace} \left(D_i^{-1} (N_{ij} - H_{ij}) D_j^{-1} (N_{ij} - H_{ij})' \right), \end{aligned}$$

where H is the supermatrix of the H_{ij} . Introducing the norm notation

$$\|A - B\|_{ij}^2 = \text{trace} \left(D_i^{-1} (A - B) D_j^{-1} (A - B)' \right)$$

the minimization can be written as

$$n^{-1} \sum_{i=1}^Q \sum_{j=1}^Q \|N_{ij} - H_{ij}\|_{ij}^2. \tag{14}$$

In *MCA* the identification of the true dimension is particularly difficult, despite the *MCA* is a *SCA* of a particular table, because the chi-square test has no sense. Indeed, for B a chi-squared statistic may again be calculated as if it were a contingency table, and this simplifies as

$$\chi_B^2 = 2 \sum_{i=1}^Q \sum_{j=1}^{i-1} \chi_{ij}^2 + n(J - Q),$$

where χ_{ij}^2 is the chi-squared statistic for the off-diagonal subtable $N_{ij} = Z'_i Z_j$ crossing the i -th and the j -th characters, but without the possibility to make a test. Unfortunately neither Q_α nor \tilde{Q}_α computed on the indicator matrix Z are chi-square distributed [5], since Z is composed by 0's and 1's, and it is dramatically inflated by the diagonal tables without any real meaning.

The high number of eigenvalues of the *MCA*, and their corresponding low values, were criticized by the same Benzécri [8] that suggests to reevaluate those larger than their average. Indeed, by applying both *SCA* and *MCA* to the same two characters, by partitioning the Burt's table $Z'Z$ into submatrices it can be shown (ibid.) the relation $\mu_\alpha = \frac{1 \pm \sqrt{\lambda_\alpha}}{2}$ that holds among μ_α , the eigenvalues of Z , and λ_α , those of the *SCA* of the contingency table crossing the two characters. In this case, it is evident that to the eigenvalues $\lambda_\alpha = 0$ of *SCA* correspond eigenvalues $\mu_\alpha = \frac{1}{2}$ of Z and $\nu_\alpha = \frac{1}{4}$ of B , whereas to the other two correspond, one of which larger and the other smaller than $\frac{1}{2}$ and $\frac{1}{4}$ respectively. To generalize this argument to several characters results in admitting to limit attention in *MCA* only to the eigenvalues larger than their mean, that is $\mu \geq \bar{\mu}_\alpha = \frac{1}{Q}$.

The argument is discussed in detail by both Benzécri [8] and Greenacre [20, 21]. Both authors suggest, in order to get a measure of relative importance of each factor, to re-evaluate the eigenvalues larger than the mean (equal to $\frac{1}{Q}$) according to the formula

$$\rho(\mu_\alpha) = \left(\frac{Q}{Q-1}\right)^2 (\mu_\alpha - \bar{\mu})^2, \quad \mu_\alpha \geq \bar{\mu} = \frac{1}{Q}.$$

Benzécri [8] suggests to consider as total inertia the sum of the re-evaluated eigenvalues and to take as percentage of explained inertia the ratio $\frac{\rho(\mu_\alpha)}{\sum_\alpha \rho(\mu_\alpha)}$. This results in a dramatic re-evaluation of the relative importance of the first eigenvalues. On the opposite, Greenacre [20] bases his arguments on the unusefulness to take into account the diagonal block matrices and the utility to limit attention only to the total off-diagonal inertia of the table, that is the sum of squared (non-re-evaluated) eigenvalues minus the diagonal inertia; represented as:

$$\frac{Q}{Q-1} \left(\sum_{\mu_\alpha > 1/Q} \mu_\alpha^2 - \frac{J-Q}{Q^2} \right).$$

Experiments show that the Greenacre's reevaluation is always limited to a share of the total inertia of Burt's table even by taking into consideration all the eigenvalues larger than the mean. This does not affect the interpretation of the factors, that essentially depends upon the eigenvectors – and thus to the contributions of both levels and characters to them –, but affects more the quality of representation of these elements on the factor subspaces, that varies according to the percentage of inertia attributed to each one. Indeed, this is a point that would deserve some specific consideration, in particular in deciding which reevaluation may be better taken into account. In the following, we shall call *adjusted MCA* the one with re-evaluated inertia, therefore with the coordinates recalculated accordingly.

The reduction in number of the dimension, thanks to both Benzécri's and Greenacre's reevaluations, does not solve the problem of the true dimension of the table. To this question, an answer comes from Ben Ammou & Saporta [5, 6]: they suggest to estimate the significance of the eigenvalues of *MCA* according to their distribution. If the characters are independent,

$$\sum_{\beta=1}^{J-Q} \mu_\beta = \frac{J-Q}{Q} \quad \text{and} \quad S_{\mu^2} = \sum_{\beta=1}^{J-Q} \mu_\beta^2 = \frac{J-Q}{Q^2} + \frac{\sum_{i \neq j} \phi_{ij}^2}{Q^2}$$

with $n_{..}\phi_{ij}^2 \approx \chi_{(l_i-1)(l_j-1)}^2$, thus,

$$E[n_{..}\phi_{ij}^2] = E[\chi_{ij}^2] = (l_i - 1)(l_j - 1)$$

so the expectation of the variance S_{μ}^2 of the eigenvalues is

$$\sigma^2 = E[S_{\mu}^2] = \frac{1}{n_{..}Q^2(J - Q)} \sum_{i \neq j} (l_i - 1)(l_j - 1).$$

Roughly, one may assume that the interval $\frac{1}{Q} \pm 2\sigma$ should contain about 95% of the eigenvalues. Indeed, since the kurtosis of the set of eigenvalues is lower than for a normal distribution, the actual proportion is larger than 95%.

2.4 Joint Correspondence Analysis

Greenacre [20] criticizes the *MCA* approach, stating that it is not a natural generalization of *SCA*, and proposes as more natural his *Joint Correspondence Analysis (JCA)*. With it, he overcomes the *MCA*'s useless fitting of the diagonal subtables of *B* which contribute with the term $n(J - Q)$ to the total inertia. Hence, he takes as more natural measure of total inertia the sum $\sum \sum_{q \neq s} \chi_{qs}^2$ of the inertias of the off-diagonal tables. This suggests an alternative generalization of correspondence analysis which fits only the off-diagonal tables, analogous to factor analysis where values on the diagonal of the covariance or correlation matrix are of no direct interest.

Indeed, the proposed redefinition of the total inertia, by removing the diagonal block-matrices, would fix an important bias due to the application to the Burt's table of the chi-square metrics, since the diagonal structure of the diagonal block-matrices represents a very high fictitious deviation from the expected values, that *MCA* analyzes as if it were a true deviation. On this basis, opposite to the current use, this kind of analysis is not really suitable.

Greenacre [20] proposes his *Joint Correspondence Analysis (JCA)* as a weighed least-squares approximation aiming at minimizing

$$n^{-1} \sum_{i=1}^Q \sum_{j=1}^{i-1} \|N_{ij} - H_{ij}\|_{ij}^2, \tag{15}$$

instead of (14) with the corresponding $\chi_j^2 = \sum_{i=1}^Q \sum_{j=1}^{i-1} \chi_{ij}^2$, sum of the chi-squares of all off-diagonal tables, that unfortunately may not be checked for significance.

In order to get the solution, he proposes an alternating least-squares algorithm, based on the reformulation of (15) as follows:

$$n^{-1} \sum_{i=1}^Q \sum_{j=1}^{i-1} \|N_{ij} - H_{ij}\|_{ij}^2 = n^{-1} \sum_{i=1}^Q \sum_{j=1}^{i-1} \|N_{ij} - n \mathbf{r}_i \mathbf{r}_j' - L_{ij}\|_{ij}^2 \tag{16}$$

with \mathbf{r}_i the diagonal of the *i*-th block-diagonal matrix. Calling *H* and *L* the supermatrices gathering the H_{ij} and L_{ij} respectively, [20] states the equivalence of the rank-*K* solution of *L* which

satisfies the normal equations in the minimization of the second term of (16) with the rank- $(K + 1)$ matrix $H = \mathbf{r} \mathbf{r}' + L$ which satisfies minimizing (15), with \mathbf{r} the supervector gathering the Q vectors \mathbf{r}_i .

The matrix approximation L of rank K is of the form $L = nDXD_\beta X'D$, where the $J \times K$ matrix X is normalized as $X'DX = QI$, with $D = \text{diag}(\mathbf{r})$. The matrix X of parameters has rows corresponding to the categories of the variables and columns corresponding to the dimensions of the solution, that must be chosen in advance. The diagonal matrix D_β contains a scale parameter for each dimension. This form of L and the normalization conditions are chosen to generalize the bivariate case (8). The parameter matrix X is partitioned row-wise according to the variables as X_1, \dots, X_Q , where X_q is $J_q \times K$, so that the submatrices of L are $L_{qs} = nD_q X_q D_\beta X'_s D_s$. There are also inherent centering constraints on X of the form $X'r = 0$ due to the orthogonality with the dimension defined by the trivial solution. It is evident that the dimension of the solution must be chosen in advance.

Thus [20] proposes the approximate reconstruction of the whole matrix $B - n \mathbf{r} \mathbf{r}'$, namely

$$B - n \mathbf{r} \mathbf{r}' \approx nDXD_\beta X'D + C,$$

where C is a block diagonal matrix with submatrices C_{qq} , $q = 1, \dots, Q$ down the diagonal and zeros elsewhere. Here, each C_{qq} is composed by dummy parameters which effectively allow perfect fitting of the submatrices on the diagonal of $B - n \mathbf{r} \mathbf{r}'$, thereby eliminating their influence on the model of interest. The minimization of

$$\begin{aligned} B - n \mathbf{r} \mathbf{r}' &= 2n^{-1} \sum_{i=1}^Q \sum_{j=1}^{i-1} \|N_{ij} - n \mathbf{r}_i \mathbf{r}_j' - L_{ij}\|_{ij}^2 \\ &+ n^{-1} \sum_{k=1}^Q \|N_{kk} - n \mathbf{r}_k \mathbf{r}_k' - L_{kk} - C_{kk}\|_k^2. \end{aligned} \tag{17}$$

is equivalent to minimizing (16) because the latter set of terms in (17) can always be made zero by setting $C_{ii} = N_{ii} - n \mathbf{r}_i \mathbf{r}_i' - L_{ii}$.

The algorithm proposed by [20] to minimize (17) can be performed iteratively by alternating between the variables in C and those in X and D_β as follows:

1. fix the dimension K of the solution.
2. initiate the algorithm with an analysis of the full Burt matrix B , that is

$$B - n \mathbf{r} \mathbf{r}' \approx nDXD_\beta X'D. \tag{18}$$

3. limiting attention to the first K dimensions, say the first K columns of X $\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(K)}$, (18) can be rewritten as

$$B - n \mathbf{r} \mathbf{r}' \approx \sum_{k=1}^K n\beta_k D \mathbf{x}_{(k)} \mathbf{x}'_{(k)} D. \tag{19}$$

so that, if all quantities except the $\beta_k (k = 1, \dots, K)$ are regarded as fixed, the problem reduces to a simple weighted least-squares regression (see [20], for further details).

4. Keeping X and D_β fixed, set

$$C_{ii} = N_{ii} - n \mathbf{r}_i \mathbf{r}_i' - n D_i X_i D_\beta X_i' D_i \quad (i = 1, \dots, Q).$$

5. Keeping C fixed, minimize with respect to X and D_β : this is achieved by performing a correspondence analysis on the table $B_* = B - C$, that is the Burt's matrix with modified submatrices on its diagonal, setting X equal to the first K vectors of optimal row or column parameters and the diagonal of D_β equal to the square roots of the first K principal inertias respectively.

6. Iterate the last two steps until convergence.

In the special case $Q = 2$, the problem reduces to fitting the single off-diagonal submatrix N_{12} . It is relevant to mention that the initial solution described above is optimal and provides the simple correspondence analysis of $N = N_{12}$ exactly. It is also noteworthy to quote [22, pag. 148]'s alternative proposal to reevaluate the inertias along the axes by estimating them through a weighted least-squares regression. Indeed, it means to estimate only once the β_k s in (19), where the \mathbf{x}_k s are the eigenvectors issued by *MCA*, once the dimension of the solution K is decided. We are not convinced by this solution, since the objections that concern the use of the chi-square metrics remain unaltered.

2.5 The Extended Matching Coefficient

JCA was introduced by [20] as a way to drop the excessive attention given to the diagonal of the Burt's matrix by *MCA*, that indeed does not deserve any interest, but with a different optimization aim, that is to reconstruct at the best the off-diagonal subtables. In this case, the chi-square metrics is justified, whereas we have already stated that in classical *MCA* it has no theoretical justification. For this reason, we are not convinced that some inertia reevaluation of *MCA*, as the quoted ones, may be a solution. On the opposite, we find interesting to explore Gower & Hand's [18] proposal to drop the chi-square metrics in favor of a more simple one: the *Extended Matching Coefficient*. Indeed, for two units, they define it as the number of common levels across all characters. Therefore, given the indicator matrix Z , ZZ' would give us a similarity matrix to deal with; indeed, given its size, the Burt's table B as such is its corresponding in the dual space, so that it is sufficient to perform the *SVD* of the centered Burt's matrix, that is

$$Q^{-2}B = Q^{-2}Z'Z \tag{20}$$

to be compared with (12). Now, the reconstruction formula (3) holds, but this time the layers may not be interpreted in terms of deviations from expectation, that is not taken into account by the method, but merely as contribution to the reconstruction, that is in this case the frequencies in the Burt's cells. Unfortunately, no stopping rule is available for this method so far, thus we considered both Cattell's [12] and brokenstick [3] tests.

3 AN APPLICATION TO THE KIND OF WORDS

To show in detail the different behavior of the different analyses in practice, we refer first to a data set taken from [32], consisting in 2000 words taken from four different kind of periodic reviews (*Childish (TC)*, *Review (TR)*, *Dissemination (TD)*, and *Scientific Summary (TS)*), classified according to their grammatical kind (*Verb (WV)*, *Noun (WN)*, and *Adjective (WA)*) and the number of internal layers (*Two- (L2)*, *Three- (L3)*, and *Four and more layers (L4)*), as a measure of the word complexity. In Table 1 the Burt's table that results by crossing the three characters is reported.

Table 1 – Burt's table of the words' type example.

	L2	L3	L4	WN	WV	WA	TC	TR	TD	TS
L2	1512	0	0	788	483	241	433	385	399	295
L3	0	375	0	203	23	149	64	82	86	143
L4	0	0	113	62	9	42	3	29	21	60
WN	788	203	62	1053	0	0	229	284	273	267
WV	483	23	9	0	515	0	174	133	125	83
WA	241	149	42	0	0	432	97	79	108	148
TC	433	64	3	229	174	97	500	0	0	0
TR	385	82	29	284	133	79	0	496	0	0
TD	399	86	21	273	125	108	0	0	506	0
TS	295	143	60	267	83	148	0	0	0	498
	L2	L3	L4	WN	WV	WA	TC	TR	TD	TS

In Table 2 are reported the eigenvalues of the three SCAs of the contingency data tables that cross the three characters two by two: the eigenvalues, the percentage of corresponding inertia, and the p -value associated to the chi-square calculated for the corresponding one-dimensional reconstruction, that, in this case, is identical to the Malinvaud's test, since each solution is 2-dimensional.

Table 2 – SCA of the three contingency data tables of the three characters two by two. In the columns, the eigenvalues, the percentage of inertia, and the p -value of the chi-square associated to the factors.

N.	Words - Levels			Publications - Words			Publications - Levels		
	Eigen	%	p -value	Eigen	%	p -value	Eigen	%	p -value
1	.0925	99.98	.0000	.0253	80.53	.0000	.0619	98.82	.0000
2	.0000	0.02	.8625	.0061	19.47	.0022	.0007	1.18	.4771

In two cases, the tests do not attribute to the second factors any real meaning, since the p -value is larger than 5%, whereas in the case of the table type of publication – kind of words the second factor is also significant.

In Figure 1 the results of the three SCAs are represented too: it must be pointed out that the vertical position of the items is significant only for the second graphic. Indeed, the inspection of this factor plane shows an arch pattern due to a Guttman effect [9, 24]; the same, the interpretation is straightforward: for the first table, both verbs and nouns seem to have in general less syllables than the adjectives; for the second, the variation in use of the words according to the higher complexity of the publication: verbs for the childish, nouns for reviews and disseminations, adjectives for scientific summaries; for the third, the more complicated words (3 and more syllables) in scientific summaries than in all others. It is noteworthy in the second table the opposite pattern of verbs and adjectives, the first reducing while the publication is of higher level and the second raising; this explains clearly the observed Guttman effect. The position of long words very elongated on the second axis of both the first and the third analyses, in the latter case also with review is explained by the shortness of the verbs and its scarce presence in childish publications, but it is not significant. We may ground our comparisons on this interpretation of the data. Running *MCA*, the pattern of eigenvalues is represented in Table 3, in which are reported the singular values of Z , their percentage to their total (that equals $\frac{I-Q}{Q} = 2.33$), the cumulate percentage, the eigenvalues of the Burt's matrix, corresponding to the explained inertia, and the cumulate inertia.

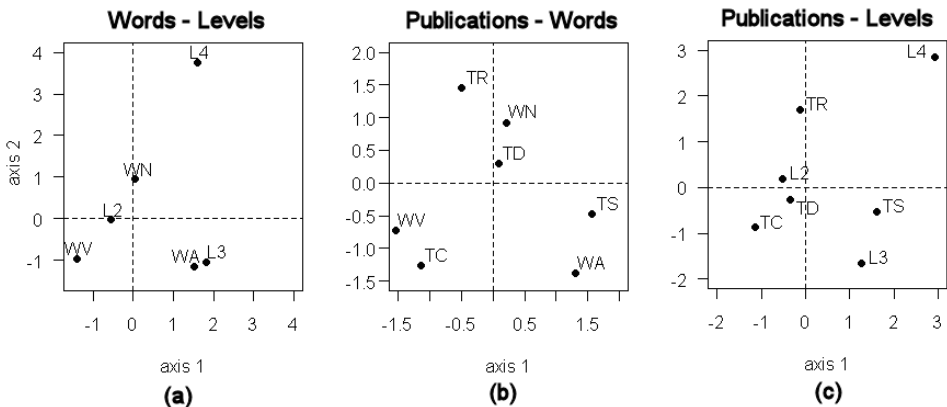


Figure 1 – Words’ type example: The pair of characters levels on the three two-way SCAs: (a) Words vs. Levels; (b) Publications vs. Words; (c) Publications vs. Levels.

In addition, on the table are reported the re-evaluated inertia and its percentages and cumulated ones according to both [8] and [20], limited to the only three singular values larger than $1/Q = 1/3$, with the totals in the following row. In both cases, the first dimension’s re-evaluated inertia is by far larger than the others. If we apply the Ben Ammou & Saporta’s [5, 6] estimation of the average singular value distribution under independence, we find that the standard deviation is $\sigma = 0.0159364$, so that the confidence interval at 95% level is $(0.30146 < \lambda < 0.36521)$. As a consequence, only the first singular value is outside the confidence interval and should be considered significant. As a matter of facts, the second one is very close to the threshold (0.3640): this is consistent with the fact that only one of the 2-dimensional tables has a significant second eigenvalue.

Table 3 – MCA singular values, percentage to the total and cumulate percentage, eigenvalues, and cumulate inertia of the Burt’s table of words’ type example. Then re-evaluated inertia and percentages according to both [8] and [20].

N	Sing. value	% Inertia	Cumul. %	Eigen.	Cum. inertia	Re-ev. Inertia	Benzécri’s		Greenacre’s	
							%	Cum.%	%	Cum.%
1	0.4896	20.98	20.98	0.2397	0.2397	0.0549	95.91	95.91	88.36	88.36
2	0.3640	15.60	36.58	0.1325	0.3722	0.0021	3.69	99.60	3.40	91.76
3	0.3434	14.72	51.30	0.1179	0.4901	0.0002	0.40	100.00	0.37	92.13
4	0.3300	14.14	65.44	0.1089	0.5990	0.0572	100.00		92.13	
5	0.3084	13.22	78.66	0.0951	0.6941					
6	0.2728	11.69	90.35	0.0744	0.7685					
7	0.2252	9.65	100.00	0.0507	0.8192					

In Figure 2a the distribution of all character levels on the plane spanned by the first two factors of MCA is represented. Indeed, the patterns of all the characters’ levels repeat fairly well the same in the three two-way tables: thus it may be taken as a sign of coherence between the individual SCAs and MCA.

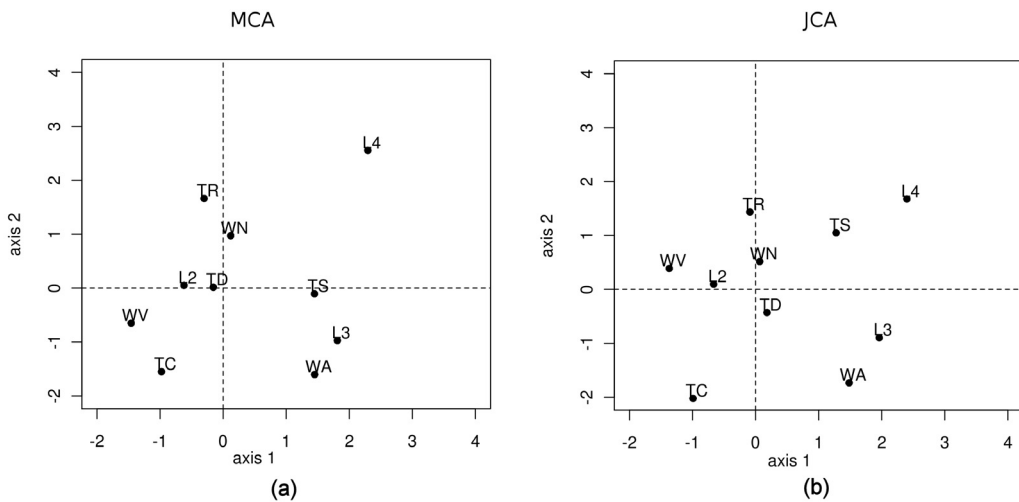


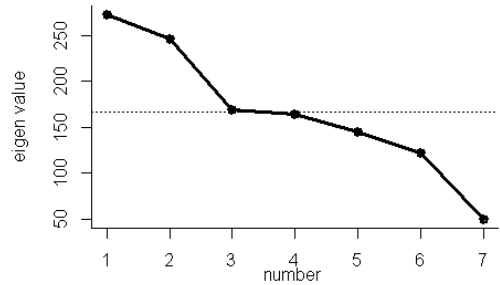
Figure 2 – Words’ type example: representation of the three-character levels on the plane spanned by the first two factors: (a) MCA; (b) JCA.

It may be observed that the similarity is good even on the second dimension, albeit not significant, whereas on the plane the Guttman effect appears again in good evidence. This may also depend upon the magnitude of the first two eigenvalues, that is sufficiently high to state that the three characters share around either 48% or 36% of the first and second factor respectively. Concerning the inertia reevaluation, this does not affect the interpretation of the single factors but in case of the spaces, since it acts as different multiplicative constants on the factors.

Let us now discuss the results of the *JCA* carried out on the same example. In the 2-dimensional solution¹ the axes inertias are 0.2488 and 0.0272, with a proportion of 90.15% and 9.85%, respectively: considering only the first axis as significant, we may observe in Figure 2*b* a pattern of levels nearly identical to the one of *MCA*. Some differences appear on the second axis, in which are noticeable the very different positions of verbs and childish publications on the negative side and of long words and summaries on the positive one, but, once again, this may not be considered significant.

Table 4 – Eigenvalues, percentages of explained and cumulate inertia of the analysis of *EMC* on Words’ type example. On the right, the pattern of the eigenvalues. The dotted line represents their average (166.66).

N.	Eigen	%	Cum %
1	272.7187	23.38	23.38
2	245.3787	21.03	44.41
3	168.3971	14.43	58.84
4	163.5518	14.02	72.86
5	145.3435	12.46	85.32
6	121.7007	10.43	95.75
7	49.5341	4.25	100.00



Eventually, we got the results of *EMC*. The seven ($J - Q$) non-zero eigenvalues and the corresponding percentages of explained and cumulate inertia are reported on Table 4. They are reported also in the figure nearby, where the average (166.66) corresponds to the dotted line. Thus, one may identify two major eigenvalues that summarize 44% of the total inertia, three others around the mean and two minor ones. Here, the Cattell test would suggest two factors, whereas the brokenstick considers random even the first one, since the threshold to consider it non-random would be 37.04%. These contradictory results lead us to compare them with the previous ones, therefore considering the first dimension as the “true” one, but also taking into account the second, at least for the graphical representation. In Figure 3 all levels are plotted on the plane spanned by the first two factors: indeed, the pattern of levels along the first axis is somehow similar to the ones resulting from both *MCA* and *JCA* but not so much: both *L4* and *L3* and even more *WA* and *WN* are exchanged, slightly modifying the interpretation of the results. On the opposite, the pattern along the second factor is so different that no agreement seems to be possible. In both cases, differences may result from the fact that here rare levels are found close to the centroid and the frequent ones are far away, whereas in the chi-square-based methods the opposite occurs. Indeed, this is the case of both *L2* and *WN* that have the highest marginal values, whereas *L4*, with the lowest ones, is set toward the center.

Let us look now at the one-dimensional reconstruction, as resulting by the *SCAs* of the three individual tables, by the *MCA*, and by Greenacre’s *JCA* as reported in Table 5. The comparison of the *SCA* one-dimensional solutions with the original tables shows that the amount of the

¹The *R* package *ca*, that we used, gave a diagnostic when asked of running the 1-dimensional solution.

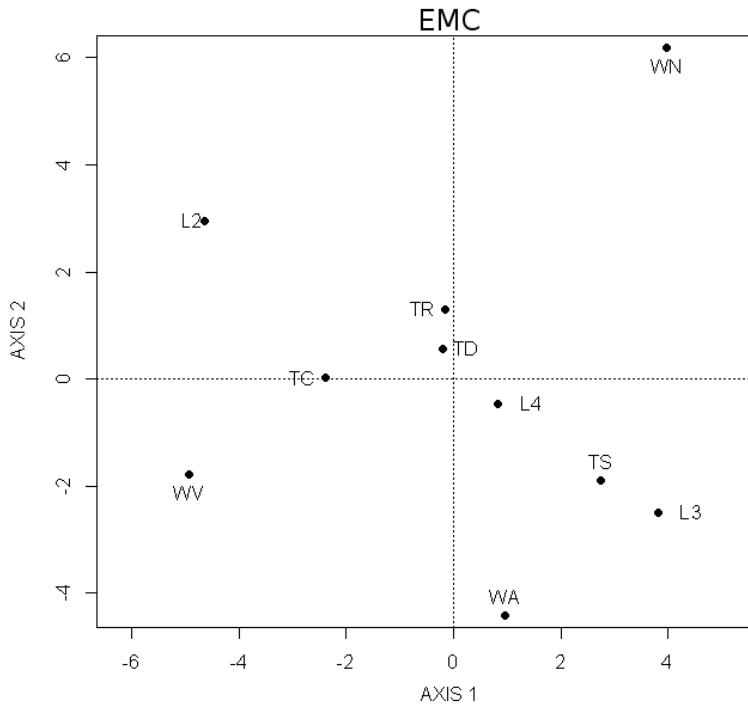


Figure 3 – Words’ type example: representation of the three-character levels on the plane spanned by the first two factors of the centered *PCA* on the Burt’s table, corresponding to the Extended Matching Coefficient.

accumulate absolute residuals is in good agreement with the quality of the solution, as represented by the corresponding chi-square. For this reason, the low quality of the reconstruction of the table crossing kind of words with the type of publications depends on the significance of the second dimension of the *SCA* of this table, that here is not taken into account. At first glance, it is evident the high difference in the cumulate absolute residuals of *MCA* in respect to the other solutions, that is an important sign of the limits of *MCA* in respect to *JCA*.

Indeed, the quality of *JCA* one-dimensional reconstruction is in all cases acceptable, so that it is possible to observe a synthetic graphical representation of the three tables that is realistic. Finally, looking at the first layer obtained by *EMC* we find a behavior somehow comparable with the first layer of *MCA*: much worst for the first table, much better for the second and relatively equal for the third. This may also depend on the different way that this method uses to reconstruct the data table, as each layer does not represent a deviation from expectation but rebuilds the table anew. Thus, a better reconstruction must be expected through a larger number of factors. We did it by comparing the sum of the absolute differences in the partial reconstructions obtained by increasing the solution dimension: this could be done for the whole 7 factors of both *MCA* and *EMC* and only for the first 3 above the mean for both adjusted *MCA* and *JCA*. The results are given in Table 6.

Table 5 – Original two-way contingency tables of words’ type example and their reconstruction according to the first dimension of SCAs, *MCA*, adjusted *MCA*, *JCA*, and *EMC* with the corresponding cumulate absolute residuals.

Original Contingency Tables													
	WN	WV	WA		TC	TR	TD	TS		TC	TR	TD	TS
L2	788	483	241	L2	433	385	399	295	WN	229	284	273	267
L3	203	23	149	L3	64	82	86	143	WV	174	133	125	83
L4	62	9	42	L4	3	29	21	60	WA	97	79	108	148
SCA First Layer													
	WN	WV	WA		TC	TR	TD	TS		TC	TR	TD	TS
L2	788	483	241	L2	435	382	400	296	WN	253	257	267	276
L3	204	23	149	L3	60	89	85	141	WV	165	144	127	79
L4	61	9	42	L4	5	25	22	61	WA	82	96	112	142
SCA cumulate absolute residuals													
	2				29					134			
MCA First Layer													
	WN	WV	WA		TC	TR	TD	TS		TC	TR	TD	TS
L2	770	559	183	L2	492	409	401	211	WN	249	257	264	283
L3	216	-24	183	L3	13	69	82	211	WV	219	155	145	-3
L4	67	-20	66	L4	-5	18	23	76	WA	32	84	97	219
MCA cumulate absolute residuals													
	304				342					397			
Adjusted MCA First Layer													
	WN	WV	WA		TC	TR	TD	TS		TC	TR	TD	TS
L2	783	471	258	L2	433	385	399	295	WN	229	284	273	267
L3	206	39	130	L3	64	82	86	143	WV	174	133	125	83
L4	63	6	44	L4	3	29	21	60	WA	97	79	108	148
Adjusted MCA cumulate absolute residuals													
	78				67					166			
JCA First Layer													
	WN	WV	WA		TC	TR	TD	TS		TC	TR	TD	TS
L2	783	484	245	L2	435	391	393	293	WN	259	260	266	269
L3	207	29	139	L3	53	82	87	153	WV	160	136	136	82
L4	63	2	48	L4	12	24	25	52	WA	81	100	104	147
JCA cumulate absolute residuals													
	44				64					134			
EMC First Layer													
	WN	WV	WA		TC	TR	TD	TS		TC	TR	TD	TS
L2	630	595	287	L2	477	381	391	262	WN	178	256	259	360
L3	334	-72	114	L3	12	88	88	187	WV	234	134	139	7
L4	89	-8	32	L4	10	27	27	49	WA	88	106	108	131
EMC cumulate absolute residuals													
	631				219					390			

Table 6 – Absolute residuals of the reduced dimensional reconstructions of both the Burt's table and the two-way off-diagonal ones according to *MCA*, adjusted *MCA* and *JCA* respectively: to 0 correspond the deviations from independence.

N	MCA			Adjusted MCA			JCA			EMC		
	Total	Diag.	Off	Total	Diag.	Off	Total	Diag.	Off	Total	Diag.	Off
	8906	7000	953	8906	7000	953	8906	7000	953			
1	7557	5470	1044	6879	6263	308	6629	6149	240	7849	5363	1243
2	7378	4303	1537	6588	6116	236	6206	5916	145	5950	3907	1022
3	7089	3463	1813	6510	6080	215	5836	5800	18	5185	3129	1028
4	5949	2805	1572							3961	2172	895
5	3675	1720	977							2143	1080	531
6	2335	877	729							513	394	60
7	0	0	0							0	0	0

In Table 6 are reported the cumulate absolute residuals of reconstructions of *MCA*s, both normal and adjusted, *JCA*, and *EMC*: they are both total and partitioned according to the diagonal matrices and the off-diagonal ones. In this latter case, the residuals are divided by two, that is the sum of the residuals of the individual 2×2 contingency tables, that form either triangular off-diagonal sub-matrix. The residuals for 0-dimension are the deviations from independence and the following are reported for all the allowed dimensions: $7 = J - Q$ for both *MCA* and *EMC* and 3 for both adjusted *MCA* and *JCA*, that corresponds to the number of singular values of the Burt's table larger than the average.

The first row reports the deviations in respect to the independence, that for *EMC* does not make any sense. For each method, the first column represents the inertia of the whole Burt's table reconstruction: it is always descendant, as it should be expected, although with different slope: in this respect, *EMC* performs best by far. Indeed, the same occurs for what concerns the reconstruction of the diagonal tables: once again the *EMC*'s performance is the best, albeit not as for the total table. Both *MCA* and *EMC* eventually rebuild totally the Burt's table, as expected. The surprises arise looking at the off-diagonal tables reconstruction: here, the *MCA* reconstruction is dramatically bad and problematic: indeed, all partial reconstructions are worst than the independence, that is the estimated frequencies are further from the observed than those due the independence, but the last one. That is, the first 5 dimensions, instead of improving the estimation, get it even worse! In this respect, *EMC* performs much better, as it is constantly decreasing.

If we look now at both adjusted *MCA* and *JCA*, we notice that, for what concerns the diagonal submatrices, they perform very badly, even worst than *MCA*, but this ought to be expected, specifically for *JCA*, in which the diagonal submatrices are intentionally neglected. On the other side, the improvement in the reconstruction of the off-diagonal ones is incredibly better, even in respect to *EMC*, with an excellent performance of *JCA*.

4 CONCLUSION

This study started with the aim to understand to what extent the *JCA* [20] could be of help in identifying the true dimension of an analysis concerning a set of qualitative data. In this sense, the confidence interval proposed by Ben Ammou & Saporta [5, 6] seems a better answer to this problem, that in the proposed example resulted in agreement with the most one-dimensional solution of the *SCAs* applied to the two-way tables.

During the study, the problem of the data reconstruction not only showed that *MCA* is bad in reconstructing the whole data table, in respect to *EMC*, even in what concerns the diagonal submatrices, but mostly concerning the off-diagonal ones, that are even more biased: the reconstruction of the two-way off-diagonal tables is for the most reduced-dimensional solutions worst than the independence table. Indeed, only redefining the inertia according to the adjusted *MCA*, a suitable reconstruction may be performed, albeit far from optimality, that is much better approached by *JCA*. It is interesting to note that, concerning the off-diagonal tables the adjusted *MCA* seems to perform better than *EMC*, a result that should be further studied. Eventually, the performance of *JCA*, as expected, is by no means the most suitable to deal with the off-diagonal tables, that is on the study of the relations between pairs of characters. As for the interpretation of the factors, *JCA* is not very different from *MCA*, whereas the method's differences of *EMC* impose a different interpretation that may be further studied. Thus, *JCA* seems the most promising development of *MCA* and its properties deserve some further deepening, including the three available programs [21, 41, 43]: indeed, a direct comparison of these results with those obtained through Greenacre's [22] inertia evaluation through regression, may provide further insights on both methods, albeit our critics on the use of chi-square metrics for the whole Burt's table remain.

Indeed, no direct comparison of the results is strictly correct, since the methods considered in this work use different metrics, either chi-square or *EMC*, and/or optimize different criteria, as described. In addition, *JCA* solutions are not nested. These aspects deserve being taken into account while interpreting the results obtained by the different methods. Eventually, address the study in a different framework such as maximum-likelihood estimation, may be a fruitful alternative.

ACKNOWLEDGEMENTS

We want to thank the anonymous referees that significantly helped in improving our paper. This work was mostly carried out during the reciprocal visits of both authors in the framework of the bilateral agreement between Sapienza Università di Roma and Universidade Federal do Rio de Janeiro, of which both authors are the scientific responsible. The first author was also granted by his Faculty of belonging, the Facoltà d'Architettura ValleGiulia of Sapienza and *FAPERJ* of Rio de Janeiro. All institutions grants are gratefully acknowledged.

REFERENCES

- [1] ABDI H. 2007. Singular Value Decomposition (*SVD*) and Generalized Singular Value Decomposition (*GSVD*). In: SALKIND N (Ed.), *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage: pp. 940–945.
- [2] ABDI H & VALENTIN D. 2007. Multiple Correspondence Analysis. In: SALKIND N (Ed.), *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage: pp. 651–657.
- [3] BARTON DE & DAVID FN. 1956. Some Notes on Ordered Random Intervals. *Journal of the Royal Statistical Society, Series B*, **18**(1): 79–94.
- [4] BELTON V & STEWART T. 2002. *Multiple Criteria Decision Analysis*. Boston, Kluwer Academic Publishers.
- [5] BEN AMMOU S & SAPORTA G. 1998. Sur la normalité asymptotique des valeurs propres en ACM sous l’hypothèse d’indépendance des variables. *Revue de Statistique Appliquée*, **46**(3): 21–35.
- [6] BEN AMMOU S & SAPORTA G. 2003. On the connection between the distribution of eigenvalues in multiple correspondence analysis and log-linear models. *REVSTAT-Statistical Journal*, **1**(0): 42–79.
- [7] BENZÉCRI JP ET AL. 1973-82. *L’Analyse des données*, Tome 2. Paris: Dunod.
- [8] BENZÉCRI JP. 1979. Sur les calcul des taux d’inertie dans l’analyse d’un questionnaire. *Les Cahiers de l’Analyse des Données*, **4**(3): 377–379.
- [9] CAMIZ S. 2005. The Guttman Effect: its Interpretation and a New Redressing Method. *Τετραδιά Αναλύσης Δεδομένων (Data Analysis Bulletin)*, **5**: 7–34.
- [10] CAMIZ S & GOMES GC. 2013. Joint Correspondence Analysis Versus Multiple Correspondence Analysis: A Solution to an Undetected Problem. In: GIUSTI A et al. (eds.), *Classification and Data Mining, Studies in Classification, Data Analysis, and Knowledge Organization*, Berlin, Springer: pp. 11–18.
- [11] CAMIZ S & PILLAR VD. (In press). Identifying the True Dimension in Principal Component Analysis. Submitted to *Community Ecology*.
- [12] CATTELL RB. 1966. The scree test for the number of factors. *Multivariate Behavioural Research*, **1**: 245–276.
- [13] ECKART C & YOUNG G. 1936. The approximation of one matrix by another of lower rank. *Psychometrika*, **1**: 211–218.
- [14] FRAWLEY WJ, PIATETSKY-SHAPIRO G & MATHEUS CJ. 1992. Knowledge Discovery in Databases: An Overview, *Artificial Intelligence Magazine*, **13**(3): 57–70.
- [15] GHOMARI A. 1983. *Analyse Canonique d’une table de contingence. Étude asymptotique par échantillonnage*. Thèse de l’Université de Pau et des Pays de l’Adour (France).

- [16] GÓES ART, STEINER MTA & PENICHE RA. 2015. Classification of Power Quality Considering Voltage Sags in Distribution Systems Using KDD Process. *Pesquisa Operacional*, **35**(2): 329–352.
- [17] GOMES LFAM & DE ANDRADE RM. 2012. Performance evaluation in assets management with the AHP. *Pesquisa Operacional*, **32**(1): 31–53.
- [18] GOWER JC & HAND DJ. 1006. *Biplots*. London, Chapman and Hall.
- [19] GREENACRE MJ. 1983. *Theory and Application of Correspondence Analysis*. London: Academic Press.
- [20] GREENACRE MJ. 1988. Correspondence analysis of multivariate categorical data by weighted least squares. *Biometrika*, **75**: 457–467.
- [21] GREENACRE MJ. 2006. From Simple to Multiple Correspondence Analysis. In: GREENACRE & BLASIUS (Eds.), pp. 41–76.
- [22] GREENACRE MJ. 2007. *Correspondence Analysis in Practice*, 2nd Ed., London, Chapman and Hall.
- [23] GREENACRE MJ & BLASIUS J. (Eds.) 2006. *Multiple Correspondence Analysis and Related Methods*. Dordrecht (The Netherlands): Chapman and Hall (Kluwer).
- [24] GUTTMAN L. 1941. The Quantification of a Class of Attributes: a Theory and Method of Scale Construction. In: HORST P (Ed.) *The Prediction of Personal Adjustment*. New York, Social Science Research Council.
- [25] JACKSON DA. 1993. Stopping rules in principal component analysis: a comparison of heuristical and statistical approaches. *Ecology*, **74**: 2204–2214.
- [26] JOLLIFFE IT. 2002. *Principal Components Analysis*. Berlin, Springer.
- [27] LANCASTER HO. 1949. The derivation and partition of χ^2 in certain discrete distributions. *Biometrika*, **36**: 117–129.
- [28] LANCASTER HO. 1963. Canonical Correlations and Partitions of χ^2 . *Quarterly Journal of Mathematics*, **14**(1): 220–224.
- [29] LEBART L. 1976. The significance of Eigenvalues issued from Correspondence Analysis. *COMPSTAT*, Vienna, Physica Verlag: pp. 38–45.
- [30] LOPES YG & DE ALMEIDA AT. 2013. A multicriteria decision model for selecting a portfolio of oil and gas exploration projects. *Pesquisa Operacional*, **33**(3): 417–441.
- [31] MALINVAUD E. 1987. Data analysis in applied socio-economic statistics with special consideration of correspondence analysis. Marketing Science Conference, Joy en Josas: HEC-ISA.
- [32] NARDY MNS. 2007. A sintaxe no interior das palavras – efeitos de gênero na língua escrita contemporânea. PhD Thesis in Linguistics. Rio de Janeiro, Faculdade de Letras da Universidade Federal de Rio de Janeiro.

- [33] NENADIC O & GREENACRE M. 2006. *Computation of multiple correspondence analysis, with code in R*. In: GREENACRE AND BLASIUS (Eds), (2006), pp. 523–551.
- [34] NENADIC O & GREENACRE M. 2007. Correspondence analysis in R, with two- and three-dimensional graphics: the *ca* package. *Journal of Statistical Software*, **20**(3): 1–13.
- [35] ORLÓCI L. 1978. *Multivariate Analysis in Vegetation Research*, 2nd ed.. Den Haag: Junk.
- [36] PERES-NETO PR, JACKSON DA & SOMERS KM. 2005. How Many Principal Components? Stopping Rules for Determining the Number of Non-trivial Axes Revisited. *Computational Statistics and Data Analysis*, **49**: 974–997.
- [37] ROY B & BOUYSSOU D. 1985. *Aide à la décision: méthodes et cas*. Paris, Economica.
- [38] R-PROJECT. 2009. <http://www.r-project.org/>
- [39] SAPORTA G. 1975. *Liaison entre plusieurs ensembles de variables et codage de données qualitatives*. Thèse de troisième cycle, Paris VI, Université Pierre et Marie Curie.
- [40] SASSI RJ. 2012. An Hybrid Architecture for Clusters Analysis: Rough Sets Theory and Self-Organizing Map Artificial Neural Network. *Pesquisa Operacional*, **32**(1): 139–163.
- [41] TATENENI K & BROWNE MW. 2000. A Noniterative Method of Joint Correspondence Analysis. *Psychometrika*, **65**(2): 157–165.
- [42] THOMSON GH. 1934. Hotelling's method modified to give Spearman's ρ . *Journal of Educational Psychology*, **25**: 366–374.
- [43] VERMUNT JK & ANDERSON C. 2005. Joint Correspondence Analysis (JCA) by Maximum Likelihood. *European Journal of Research Methods for the Behavioral and Social Sciences*, **1**(1): 18–26.
- [44] WILLIAMS EJ. 1952. Use of scores for the analysis of association in contingency tables. *Biometrika*, **39**: 274–289.