Multiple Correspondence Analysis

Hervé Abdi[1] & Dominique Valentin

1 Overview

Multiple correspondence analysis (MCA) is an extension of correspondence analysis (CA) which allows one to analyze the pattern of relationships of several categorical dependent variables. As such, it can also be seen as a generalization of principal component analysis when the variables to be analyzed are categorical instead of quantitative. Because MCA has been (re)discovered many times, equivalent methods are known under several different names such as optimal scaling, optimal or appropriate scoring, dual scaling, homogeneity analysis, scalogram analysis, and quantification method.

Technically MCA is obtained by using a standard correspondence analysis on an indicator matrix (*i.e.*, a matrix whose entries are 0 or 1). The percentages of explained variance need to be corrected, and the correspondence analysis interpretation of interpoint distances needs to be adapted.

[1]In: Neil Salkind (Ed.) (2007). *Encyclopedia of Measurement and Statistics*. Thousand Oaks (CA): Sage.
Address correspondence to: Hervé Abdi
Program in Cognition and Neurosciences, MS: Gr.4.1,
The University of Texas at Dallas,
Richardson, TX 75083–0688, USA
E-mail: herve@utdallas.edu http://www.utd.edu/~herve

2 When to use it

MCA is used to analyze a set of observations described by a set of nominal variables. Each nominal variable comprises several levels, and each of these levels is coded as a binary variable. For example gender, (F *vs.* M) is one nominal variable with two levels. The pattern for a male respondent will be 0 1 and 1 0 for a female. The complete data table is composed of binary columns with one and only one column taking the value "1" per nominal variable.

MCA can also accommodate quantitative variables by recoding them as "bins." For example, a score with a range of −5 to +5 could be recoded as a nominal variable with three levels: less than 0, equal to 0, or more than 0. With this schema, a value of 3 will be expressed by the pattern 0 0 1. The coding schema of MCA implies that each row has the same total, which for CA implies that each row has the same *mass*.

3 An example

We illustrate the method with an example from wine testing. Suppose that we want to evaluate the effect of the oak species on barrel-aged red Burgundy wines. First, we aged wine coming from the same harvest of Pinot Noir in six different barrels made with two types of oak. Wines 1, 5, and 6 were aged with the first type of oak, whereas wines 2, 3, and 4 were aged with the second. Next, we asked each of three wine experts to choose from two to five variables to describe the wines. For each wine and for each variable, the expert was asked to rate the intensity. The answer given by the expert was coded either as a binary answer (*i.e.,* fruity *vs.* non-fruity) or as a ternary answer (*i.e.,* no vanilla, a bit of vanilla, clear smell of vanilla). Each binary answer is represented by 2 binary columns (*e.g.,* the answer "fruity" is represented by the pattern 1 0 and "non-fruity" is 0 1). A ternary answer is represented by 3 binary columns (*i.e.,* the answer "some vanilla" is represented by the pattern 0 1 0). The results are presented in Table 1 (the same data are used to illustrate STATIS and Multiple factor analysis, see the respective entries). The goal of the analysis is twofold. First

we want to obtain a typology of the wines and second we want to know if there is an agreement between the scales used by the experts. We will use the type of type of oak as a supplementary (or illustrative) variable to be projected on the analysis after the fact. Also after the testing of the six wines was performed, an unknown bottle of Pinot Noir was found and tested by the wine testers. This wine will be used as a supplementary observation. For this wine, when an expert was not sure of how to use a descriptor, a pattern of response such .5 .5 was used to represent the answer.

4 Notations

There are K nominal variables, each nominal variable has J_k levels and the sum of the J_k is equal to J. There are I observations. The $I \times J$ indicator matrix is denoted $\mathbf{X}$. Performing CA on the indicator matrix will provide two sets of factor scores: one for the rows and one for the columns. These factor scores are, in general scaled such that their variance is equal to their corresponding eigenvalue (some versions of CA compute row factor scores normalized to unity).

The grand total of the table is noted N, and the first step of the analysis is to compute the probability matrix $\mathbf{Z} = N^{-1}\mathbf{X}$. We denote $\mathbf{r}$ the vector of the row totals of $\mathbf{Z}$, (*i.e.,* $\mathbf{r} = \mathbf{Z1}$, with $\mathbf{1}$ being a conformable vector of 1's) $\mathbf{c}$ the vector of the columns totals, and $\mathbf{D_c} = \text{diag}\{\mathbf{c}\}$, $\mathbf{D_r} = \text{diag}\{\mathbf{r}\}$. The factor scores are obtained from the following singular value decomposition:

$$\mathbf{D_r}^{-\frac{1}{2}}\left(\mathbf{Z} - \mathbf{rc}^{\mathsf{T}}\right)\mathbf{D_c}^{-\frac{1}{2}} = \mathbf{P\Delta Q}^{T} \tag{1}$$

($\mathbf{\Delta}$ is the diagonal matrix of the *singular* values, and $\mathbf{\Lambda} = \mathbf{\Delta}^2$ is the matrix of the *eigenvalues*). The row and (respectively) columns factor scores are obtained as

$$\mathbf{F} = \mathbf{D_r}^{-\frac{1}{2}}\mathbf{P\Delta} \qquad \text{and} \qquad \mathbf{G} = \mathbf{D_c}^{-\frac{1}{2}}\mathbf{Q\Delta}\,. \tag{2}$$

The squared (χ^2) distance from the rows and columns to their respective barycenter are obtained as

$$\mathbf{d_r} = \text{diag}\left\{\mathbf{FF}^{\mathsf{T}}\right\} \qquad \text{and} \qquad \mathbf{d_c} = \text{diag}\left\{\mathbf{GG}^{\mathsf{T}}\right\}\,. \tag{3}$$

Table 1: Data for the barrel-aged red burgundy wines example. "Oak Type" is an illustrative (supplementary) variable, The wine *W?* is an unknown wine treated as a supplementary observation.

		Expert 1								Expert 2										Expert 3					
Wine	Oak Type	fruity		woody			coffee			red fruit		roasted		vanillin			woody			fruity		butter		woody	
W1	1	1	0	0	0	1	0	1		1	0	0	1	0	0	1	0	1		0	1	0	1	0	1
W2	2	0	1	0	1	0	1	0		0	1	1	0	0	1	0	1	0		0	1	1	0	1	0
W3	2	0	1	1	0	0	1	0		0	1	1	0	1	0	0	1	0		0	1	1	0	1	0
W4	2	0	1	1	0	0	1	0		0	1	1	0	1	0	0	1	0		1	0	1	0	1	0
W5	1	1	0	0	0	1	0	1		1	0	0	1	0	0	1	0	1		1	0	0	1	0	1
W6	1	1	0	0	1	0	0	1		1	0	0	1	0	1	0	0	1		1	0	0	1	0	1
W?	?	0	1	0	1	0	.5	.5		1	0	1	0	0	1	0	.5	.5		1	0	.5	.5	0	1

The squared *cosine* between row i and factor ℓ and column j and factor ℓ are obtained respectively as:

$$o_{i,\ell} = \frac{f_{i,\ell}^2}{d_{r,i}^2} \qquad \text{and} \qquad o_{j,\ell} = \frac{g_{j,\ell}^2}{d_{c,j}^2} \; . \tag{4}$$

(with $d_{r,i}^2$, and $d_{c,j}^2$, being respectively the i-th element of $\mathbf{d_r}$ and the j-th element of $\mathbf{d_c}$). Squared cosines help locating the factors important for a given observation or variable.

The *contribution* of row i to factor ℓ and of column j to factor ℓ are obtained respectively as:

$$t_{i,\ell} = \frac{f_{i,\ell}^2}{\lambda_\ell} \qquad \text{and} \qquad t_{j,\ell} = \frac{g_{j,\ell}^2}{\lambda_\ell} \; . \tag{5}$$

Contributions help locating the observations or variables important for a given factor.

Supplementary or illustrative elements can be projected onto the factors using the so called *transition* formula. Specifically, let $\mathbf{i}_{\text{sup}}^{\mathsf{T}}$ being an illustrative row and $\mathbf{j}_{\text{sup}}$ being an illustrative column to be projected. Their coordinates $\mathbf{f}_{\text{sup}}$ and $\mathbf{g}_{\text{sup}}$ are obtained as:

$$\mathbf{f}_{\text{sup}} = \left(\mathbf{i}_{\text{sup}}^{\mathsf{T}}\mathbf{1}\right)\mathbf{i}_{\text{sup}}^{\mathsf{T}}\mathbf{G}\boldsymbol{\Delta}^{-1} \text{ and } \mathbf{g}_{\text{sup}} = \left(\mathbf{j}_{\text{sup}}^{\mathsf{T}}\mathbf{1}\right)\mathbf{j}_{\text{sup}}^{\mathsf{T}}\mathbf{F}\boldsymbol{\Delta}^{-1} \; . \tag{6}$$

Performing CA on the indicator matrix will provide factor scores for the rows and the columns. The factor scores given by a CA program will need, however to be re-scaled for MCA, as explained in the next section.

The $J \times J$ table obtained as $\mathbf{B} = \mathbf{X}^{\mathsf{T}}\mathbf{X}$ is called the *Burt matrix* associated to $\mathbf{X}$. This table is important in MCA because using CA on the Burt matrix gives the same factors as the analysis of $\mathbf{X}$ but is often computationally easier. But the Burt matrix also plays an important theoretical rôle because the eigenvalues obtained from its analysis give a better approximation of the inertia explained by the factors than the eigenvalues of $\mathbf{X}$.

5 Eigenvalue correction for multiple correspondence analysis

MCA codes data by creating several binary columns for each variable with the constraint that one and only one of the columns gets the value 1. This coding schema creates *artificial* additional dimensions because *one* categorical variable is coded with *several* columns. As a consequence, the inertia (*i.e.*, variance) of the solution space is artificially *inflated* and therefore the percentage of inertia explained by the first dimension is severely *underestimated*. In fact, it can be shown that all the factors with an eigenvalue less or equal to $\frac{1}{K}$ simply code these additional dimensions ($K = 10$ in our example).

Two corrections formulas are often used, the first one is due to Benzécri (1979), the second one to Greenacre (1993). These formulas take into account that the eigenvalues smaller than $\frac{1}{K}$ are coding for the extra dimensions and that MCA is equivalent to the analysis of the Burt matrix whose eigenvalues are equal to the squared eigenvalues of the analysis of $\mathbf{X}$. Specifically, if we denote by λ_ℓ the eigenvalues obtained from the analysis of the indicator matrix, then the corrected eigenvalues, denoted ${}_c\lambda$ are obtained as

$${}_c\lambda_\ell = \begin{cases} \left[\left(\frac{K}{K-1}\right)\left(\lambda_\ell - \frac{1}{K}\right)\right]^2 & \text{if } \lambda_\ell > \frac{1}{K} \\ 0 & \text{if } \lambda_\ell \leq \frac{1}{K} \end{cases} . \tag{7}$$

Using this formula gives a better estimate of the inertia, extracted by each eigenvalue.

Traditionally, the percentages of inertia are computed by dividing each eigenvalue by the sum of the eigenvalues, and this approach could be used here also. However, it will give an *optimistic* estimation of the percentage of inertia. A better estimation of the inertia has been proposed by Greenacre (1993) who suggested instead to evaluate the percentage of inertia relative to the average inertia of the off-diagonal blocks of the Burt matrix. This average

Table 2: Eigenvalues, corrected eigenvalues, proportion of explained inertia and corrected proportion of explained inertia. The eigenvalues of the Burt matrix are equal to the squared eigenvalues of the indicator matrix; The corrected eigenvalues for Benzécri and Greenacre are the same, but the proportion of explained variance differ. Eigenvalues are denoted by λ, proportions of explained inertia by τ (note that the average inertia used to compute Greenacre's correction is equal to $\overline{\mathscr{I}} = .7358$).

	Indicator Matrix		Burt Matrix		Benzécri Correction		Greenacre Correction	
Factor	${}_{\mathrm{I}}\lambda$	τ_{I}	${}_{\mathrm{B}}\lambda$	τ_{B}	${}_{\mathrm{Z}}\lambda$	τ_{Z}	${}_{\mathrm{c}}\lambda$	τ_{c}
1	.8532	.7110	.7280	.9306	.7004	.9823	.7004	.9519
2	.2000	.1667	.0400	.0511	.0123	.0173	.0123	.0168
3	.1151	.0959	.0133	.0169	.0003	.0004	.0003	.0004
4	.0317	.0264	.0010	.0013	0	0	0	0
$\sum$	1.2000	1	.7822	1	.7130	1	.7130	.9691

inertia, denoted $\overline{\mathscr{I}}$ can be computed as

$$\overline{\mathscr{I}} = \frac{K}{K-1} \times \left(\sum_{\ell} \lambda_{\ell}^2 - \frac{J-K}{K^2} \right) \tag{8}$$

According to this approach, the percentage of inertia would be obtained by the ratio

$$\tau_{\mathrm{c}} = \frac{{}_{\mathrm{c}}\lambda}{\overline{\mathscr{I}}} \text{ instead of } \frac{{}_{\mathrm{c}}\lambda}{\sum {}_{\mathrm{c}}\lambda_{\ell}} \,. \tag{9}$$

6 Interpreting MCA

As with CA, the interpretation in MCA is often based upon proximities between points in a low-dimensional map (*i.e.,* two or three dimensions). As well as for CA, proximities are meaningful only between points from the same set (*i.e.,* rows with rows, columns with columns). Specifically, when two row points are close to each other they tend to select the same levels of the nominal variables. For the proximity between variables we need to distinguish two cases. First, the proximity between levels of *different* nominal variables means that these levels tend to appear together in the observations. Second, because the levels of the *same* nominal variable cannot occur together, we need a different type of interpretation for this case. Here the proximity between levels means that the groups of observations associated with these two levels are themselves similar.

6.1 The example

Table 2 lists the corrected eigenvalues and proportion of explained inertia obtained with the Benzécri/Greenacre correction formula. Tables 3 and 4 give the corrected factor scores, cosines, and contributions for the rows and columns of Table 1. Figure 1 displays the projections of the rows and the columns. We have separated these two sets, but, because the projections have the same variance, these two graphs could be displayed together (as long as one

Table 3: Factor scores, squared cosines, and contributions for the observations (I-set). The eigenvalues and proportions of explained inertia are corrected using Benzécri/Greenacre formula. Contributions corresponding to negative scores are in italic. The mystery wine (Wine ?) is a supplementary observation. Only the first two factors are reported.

			Wine 1	Wine 2	Wine 3	Wine 4	Wine 5	Wine 6	Wine ?
F	${}_{c}\lambda$	$\%_{c}$	*Factor Scores*						
1	.7004	95	0.86	−0.71	−0.92	−0.86	0.92	0.71	0.03
2	.0123	2	0.08	−0.16	0.08	0.08	0.08	−0.16	−0.16
F			*Squared Cosines*						
1			.62	.42	.71	.62	.71	.42	.04
2			.01	.02	.01	.01	.01	.02	.96
F			*Contributions* ×1000						
1			177	*121*	*202*	*177*	202	121	–
2			83	*333*	83	83	83	*333*	–

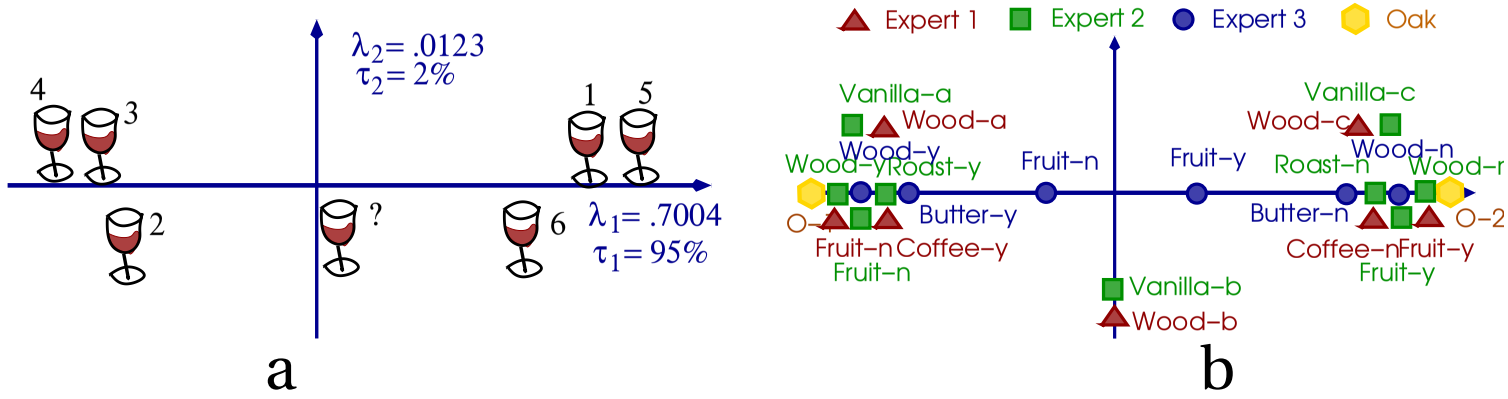

Figure 1: Multiple Correspondence Analysis. Projections on the first 2 dimensions. The eigenvalues (λ) and proportion of explained inertia (τ) have been corrected with Benzécri/Greenacre formula. (a) The I set: rows (*i.e.*, wines), wine ? is a supplementary element. (b) The J set: columns (*i.e.*, adjectives). Oak 1 and Oak 2 are supplementary elements. (the projection points have been slightly moved to increase readability). (Projections from Tables 3 and 4).

Table 4: Factor scores, squared cosines, and contributions for the for the variables (J-set). The eigenvalues and percentages of inertia have been corrected using Benzécri/Greenacre formula. Contributions corresponding to negative scores are in italic. Oak 1 and 2 are supplementary variables.

			Expert 1							Expert 2								Expert 3							
			fruity		woody			coffee		red fruit		roasted		vanillin			woody		fruity		butter		woody		Oak	
			y	n	1	2	3	y	n	y	n	y	n	1	2	3	y	n	y	n	y	n	y	n	1	2
F	${}_{c}\lambda$	$\%_{c}$	*Factor Scores*																						
1	.7004	95	.90	−.90	−.97	.00	.97	−.90	.90	.90	−.90	−.90	.90	−.97	.00	.97	−.90	.90	.28	−.28	−.90	.90	−.90	.90	.90	−.90
2	.0103	2	.00	.00	.18	−.35	.18	.00	.00	.00	.00	.00	.00	.18	−.35	.18	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
F			*Squared Cosines*																							
1			.81	.81	.47	.00	.47	.81	.81	.81	.81	.81	.81	.47	.00	.47	.81	.81	.08	.08	.81	.81	.81	.81	1.00	1.00
2			.00	.00	.02	.06	.02	.00	.00	.00	.00	.00	.00	.02	.06	.02	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
F			*Contributions* ×1000																							
1			58	*58*	*44*	0	44	*58*	58	58	*58*	*58*	58	*44*	0	44	*58*	58	6	*6*	*58*	58	*58*	58	–	–
2			0	0	83	*333*	83	0	0	0	0	0	0	83	*333*	83	0	0	0	0	0	0	0	0	–	–

keeps in mind that distances between point are meaningful only within the same set). The analysis is essentially uni-dimensional, with Wines 2, 3, and 4 being clustered on the negative side of the factors and Wines 1,5, and 6 on the positive side. The supplementary (mystery) wine does not seem to belong to either clusters. The analysis of the columns shows that the negative side of the factor is characterized as being non fruity, non-woody and coffee by Expert 1, roasted, non fruity, low in vanilla and woody for Expert 2, and buttery and woody for Expert 3. The positive side, here gives the reverse pattern. The supplementary elements indicate that the negative side is correlated with the second type of oak whereas the positive side is correlated with the first type of oak.

7 Alternatives to MCA

Because the interpretation of MCA is more delicate than simple CA, several approaches have been suggested to offer the simplicity of interpretation of CA for indicator matrices. One approach is to use a different metric than χ^2, the most attractive alternative being the Hellinger distance (see entry on distances and Escofier, 1978; Rao, 1994). Another approach, called *joint correspondence analysis*, fits only the off-diagonal tables of the Burt matrix (see Greenacre, 1993), and can be interpreted as a factor analytic model.

References

[1] Benzécri, J.P. (1979). Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. *Cahiers de l'Analyse des Données*, **4**, 377–378.

[2] Clausen, S.E. (1998). *Applied correspondence analysis*. Thousand Oaks (CA): Sage.

[3] Escofier, B. (1978). Analyse factorielle et distances répondant au principe d'équivalence distributionnelle. *Revue de Statistiques Appliquées*, **26**, 29–37.

[4] Greenacre, M.J. (1984). *Theory and applications of correspondence analysis*. London: Academic Press.
[5] Greenacre, M.J. (1993). *Correspondence analysis in practice*. London: Academic Press.
[6] Rao, C. (1995). Use of Hellinger distance in graphical displays. In E.-M. Tiit, T. Kollo, & H. Niemi (Ed.): Multivariate statistics and matrices in statistics. Leiden (Netherland): Brill Academic Publisher. pp. 143–161.
[7] Weller S.C., & Romney, A.K. (1990). *Metric scaling: Correspondence analysis*. Thousand Oaks (CA): Sage.

Acknowledgments

Many thanks to Szymon Czarnik and Michael Greenacre for pointing out a mistake in a previous version of this paper.