Mathématiques et sciences humaines
Mathematics and social sciences

186 | Été 2009
Varia

A different approach to multiple correspondence analysis (MCA) than that of specific MCA

Un traitement de l'analyse des correspondances multiples (ACM) différent de celui de l'ACM spécifique

Odysseas E. Moschidis

Electronic version
URL: http://journals.openedition.org/msh/11091
DOI: 10.4000/msh.11091
ISSN: 1950-6821

Publisher
Centre d'analyse et de mathématique sociales de l'EHESS

Printed version
Date of publication: 12 October 2009
Number of pages: 77-88
ISSN: 0987-6936

Electronic reference
Odysseas E. Moschidis, « A different approach to multiple correspondence analysis (MCA) than that of specific MCA », *Mathématiques et sciences humaines* [Online], 186 | Été 2009, Online since 15 October 2009, connection on 19 April 2019. URL : http://journals.openedition.org/msh/11091 ; DOI : 10.4000/msh.11091

Math. Sci. hum / Mathematics and Social Sciences (47e année, n° 186, *2009*(2), p. 77-88)

A DIFFERENT APPROACH TO MULTIPLE CORRESPONDENCE ANALYSIS (MCA) THAN THAT OF SPECIFIC MCA

Odysseas E. MOSCHIDIS[1]

RÉSUMÉ – Un traitement de l'analyse des correspondances multiples (ACM) différent de celui de l'ACM spécifique
Dans l'analyse des correspondances multiples, la contribution de chaque variable nominale à l'inertie est différente selon le nombre des modalités, ou des catégories de cette variable. Habituellement, pour les variables ayant beaucoup de modalités (ou catégories) il y a des modalités peu fréquentes (les « classes faibles ») qui contribuent à l'inertie de la variable correspondante de façon disproportionnée. Souvent, ces modalités contribuent fortement à la détermination des premiers axes factoriels ce qui a pour conséquence de ne pas correctement représenter le problème étudié. L'analyse spécifique des correspondances multiples traite le problème des modalités peu fréquentes en les supprimant. C'est-à-dire, qu'elle les ignore purement et simplement dans le calcul des distances entre les individus [Le Roux B., 1999 ; Le Roux B., Rouanet H., 2004].

Dans cet article, nous traitons ce problème d'une façon différente. Nous maintenons les modalités faibles dans l'analyse. Ce que réalise notre analyse, c'est de remplacer la métrique du χ^2 (d^2, χ^2)*, par une nouvelle métrique qui prend également en compte le nombre de modalités de chaque variable : c'est aussi d'attribuer un effet raisonnable aux modalités faibles et d'équilibrer toutes les variables nominales.*

En outre, nous tenons compte, uniformément, des modalités faibles, qu'elles dérivent de beaucoup ou de peu de variables, bien que la plupart des cas « dangereux » sont ceux des variables qui ont beaucoup de modalités. Seules les variables à deux modalités ne sont pas concernées.

MOTS-CLÉS – Analyse des correspondances multiples, Analyse spécifique des correspondances multiples, Coefficient d'ajustement, Nouvelle métrique

SUMMARY – *In multiple correspondence analysis, each nominal variable affects the analysis with a different amount of inertia, depending on the number of its modalities or categories. Usually in variables with many modalities – categories created infrequent (weak classes) modalities which contribute disproportionally to the inertia of the corresponding variable. Often these modalities contribute heavily to the determination of the first factorial axes and as a result this cannot clearly represent the investigated problem. Specific multiple correspondence analysis deals with the problem of infrequent (weak) modalities by removing them. That is, it simply ignores them in the calculation of distances between individuals [Le Roux B., 1999; Le Roux B., Rouanet H., 2004].*

In this paper we deal with this problem in a different manner. We keep the weak modalities in the analysis. Replacing the $X^2(d^2x^2)$ *metric by a new metric which also takes into account the number of modalities of each variable, a reasonable effect of the weak modalities and a balancing of all the nominal variables is achieved in the analysis.*

We also encounter uniformly the weak modalities, whether they derive from many or few variables, even though the most "dangerous" case is the one variables where have many modalities. Only variables of two modalities are not affected.

KEYWORDS – Adjustment coefficient, Multiple correspondence analysis, New metric, Specific MCA

[1] General manager and President of General Hospital of Thessaloniki, fmos@uom.gr

1. INTRODUCTION

We deal with the processing of a table of the form (individuals × nominal variables) using multiple correspondence analysis.

It is clear that with the metric X^2 as a mechanism which accents the similarity of the individuals, the latter is affected by the number of modalities of each variable. Variables with few categories create columns with heavy weights compared to the other variables with many categories, which create columns with lower weights.

As a result, the similarity of two individuals in relation to a variable with few categories is *a priori* more powerful than their similarity in relation to a variable with many categories.

Initially we deal with this problem in order to let the nominal variables with a different number of modalities or categories contribute equivalently to the similarity of the individuals. For this reason, we introduce a "balance" or an adjustment coefficient and a new metric $\left(d^2_{2X^2}\right)$.

2. PROPOSAL OF THE NEW METRIC $d^2_{x^2}$

Suppose that $E_1, E_2, ..., E_S$, (S) are the nominal variables or questions with $p_1, p_2, ..., p_S$ respectively the number of modalities or the potential responses.

We define as a *balance* or an *adjustment coefficient* of the question E_k the number:

$$I_k^* = \frac{1}{p_k - 1}, \tag{1.1}$$

Taking into account the adjustment coefficient I_k^*, we define a new metric $d^2_{2X^2}$, by the formula:

$$d^2_{2X^2}\left(a_i, a_j\right) = \sum_{k=1}^{s-m} \frac{1}{p_k - 1} \cdot \left(\frac{1}{f.\lambda_k} + \frac{1}{f.\mu_k}\right), \tag{1.2}$$

where the individuals α_i and α_j give m common responses ($m \leq S$) to S questions, $f \cdot \lambda_k$ is the weight of the modality (column) λ_k of the question E_k which is selected by the individual α_i and respectively $f \cdot \mu_k$ is the weight of the modality (column) μ_k of the question E_k which is selected by the individual α_j.

With the following example we want to show the impact of the new metric $d^2_{2X^2}$ on the balance of the variables.

Suppose the table 0-1 which corresponds to the nominal variables E_1, with 3 modalities we have E_2, with 5 modalities, which are answered by 100 individuals.

Table 1. The table 0-1 of the nominal variables E_1 and E_2

	E_1			E_2					Weight
α_1	1	0	0	0	0	1	0	0	1/100
α_2	0	0	1	0	0	0	1	0	1/100
α_3	0	1	0	0	0	0	0	1	1/100
…	…	…	…	…	…	…	…	…	…
Weight	0.2	0.1	0.2	0.1	0.1	0.1	0.1	0.1	1

With the metric of X^2 it is:

$$d_{X^2}^2(\alpha_1,\alpha_2) = \underbrace{\left(\frac{1}{0.2} + \frac{1}{0.2}\right)}_{\text{Contribution of } E_1} + \underbrace{\left(\frac{1}{0.1} + \frac{1}{0.1}\right)}_{\text{Contribution of } E_2} = 10 + 20$$

this means that in question E_1, it seems that the individuals are much more similar than question E_2, while with the new metric it is:

$$d_{2X^2}^2(\alpha_1,\alpha_2) = \underbrace{\frac{1}{3-1}\left(\frac{1}{0.2} + \frac{1}{0.2}\right)}_{\text{Contribution of } E_1} + \underbrace{\frac{1}{5-1}\left(\frac{1}{0.1} + \frac{1}{0.1}\right)}_{\text{Contribution of } E_2} = 5 + 5$$

this means that in each question we have the individuals with equal degree of similarity.

For each variable, the new metric $d_{2X^2}^2$ functions as an adjustment coefficient of the weights of its columns, so it finally becomes a *balancing* measure of the similarity of the individuals against all the questions (variables).

3. THE NEW METRIC $d_{2X^2}^2$ AND THE INERTIA

In this paragraph we investigate the consistence of the metric $d_{2X^2}^2$ in the total inertia I_o of the cloud of responses, which is equal to the inertia of the cloud of the individuals.

The inertia of the cloud of responses through the metric of X^2 as well as the implication of the latter in multiple correspondence analysis, has been investigated by J.-P. Benzécri [1980(b)], Greenacre [1993], B. Escofier [1988], L. Lebart, A. Morineau, M. Piron [2000]. Below we give the basic conclusions.

The inertia I_j, of the response *j* (or modality *j*), of question *q* (or variable *q*) is given by the formula

$$I_j = \frac{1}{S} \cdot \left(1 - \frac{K_j}{n}\right), \tag{2.1}$$

where (S) is the number of the responses, (K_j) is the weight of column J and (n) is the number of the individuals.

Consequently, the smaller number K_j is, that is the more infrequent it is as a response (or the weaker a modality is), the greater is its inertia. This results in a deformation of the real picture of analysis since this weak modality contributes decisively in the creation of the factorial axis.

The inertia I_q of question q is the sum of inertias of its responses and is given by the formula

$$I_q = \frac{1}{S}(m - 1), \tag{2.2}$$

where m is the number of potential responses to in question q. Consequently, the more responses a question has, the more this question contributes to the total inertia of the cloud of the responses. That is, questions with many responses contribute with great inertia. For this reason the aforementioned authors suggest that if we want the questions to have the same weight, they must have about the same number of responses [Lebart L., 1995].

Finally, the total inertia I_o of the cloud of responses, which is the sum of the inertias of all the questions, is given by the formula

$$I_o = \frac{P}{S} - 1, \tag{2.3}$$

where p is the number of potential responses, to all questions, as a result the more answers there are, the greater appears the total inertia.

Now we investigate the corresponding magnitudes with the new metric $d^2_{2X^2}$. Initially we compute the distance $d^2_{2X^2}(j,g)$ of the response j from the center of weight g of the cloud of responses.

It is

$$d^2_{2X^2}(j,g) = \sum_{i-1}^{n} \frac{1}{m-1} \cdot \frac{1}{\left(\frac{1}{n}\right)} \cdot \left(\frac{k_{ij}}{k_j} - \frac{1}{n}\right)^2 =$$

$$= \frac{n}{m-1} \sum_{i=1}^{n} \left(\frac{k_{ij}}{k_j} - \frac{1}{n}\right)^2 = \frac{n}{m-1} \sum_{i=1}^{n} \left(\frac{k_{ij}^2}{k_j^2} - \frac{2k_{ij}}{nk_j} + \frac{1}{n^2}\right) =$$

$$= \frac{n}{m-1}\left(\frac{1}{k_j^2} \sum_{i=1}^{n} k_{ij}^2 - \frac{2}{nk_j} \sum_{i-1}^{n} \frac{1}{n}\right) = \qquad \left(\text{it is } k_{ij}^2 = k_{ij}, \text{ because } k_{ij} = 0 \text{ or } 1\right)$$

$$= \frac{n}{m-1}\left(\frac{1}{k_j^2}\sum_{i=1}^{n} k_{ij} - \frac{2}{nk_j} + \frac{1}{n}\right) = \frac{n}{m-1}\left(\frac{1}{k_j^2}\ k_j - \frac{2}{n} + \frac{1}{n}\right) =$$

$$= \frac{n}{m-1}\left(\frac{1}{k_j} - \frac{1}{n}\right)$$

In other words $d^2_{2X^2}(j,g) = \frac{n}{m-1} \cdot \left(\frac{1}{k_j} - \frac{1}{n}\right).$ (2.4)

So, the smaller the weight of response *j*, the greater the distance from the center of weight *g*. This is canceled by the divider *m* - 1.

For the inertia Γ_j of the response *J* we have

$$\Gamma_j = \left(\frac{k_j}{nS}\right) \cdot d^2_{2X^2}(j,g) = \frac{kj}{nS} \cdot \frac{n}{m-1}\left(\frac{1}{k_j} - \frac{1}{n}\right) =$$

$$\frac{k_j}{S(m-1)} \cdot \left(\frac{1}{k_j} - \frac{1}{n}\right) = \frac{1}{S(m-1)} \cdot \left(1 - \frac{k_j}{n}\right)$$

That is $\Gamma_j = \frac{1}{S(m-1)} \cdot \left(1 - \frac{k_j}{n}\right).$ (2.1′)

So, the inertia created by an infrequent response (small k_j) is smaller because of the divider *m* – 1 (2.1′). Consequently, the contribution of infrequent responses in the creation of the factorial axes is moderated. This also limits the possibility that the real picture is deformed from the infrequent responses.

The inertia Γ_q of question *q* is the sum of the inertias of its responses

$$\Gamma_q = \sum_{j=1}^{m} \Gamma_j = \sum_{j=1}^{m} \frac{1}{S(m-1)}\left(1 - \frac{k_j}{n}\right) = \frac{1}{S(m-1)}\sum_{j=1}^{m}\left(1 - \frac{k_j}{n}\right) =$$

$$= \frac{1}{S(m-1)}\left(m - \frac{\sum_{j=1}^{m} k_j}{n}\right) = \frac{1}{S(m-1)}\left(m - \frac{n}{n}\right) = \frac{1}{S(m-1)}(m-1) = \frac{1}{S}$$

that is $\Gamma_q = \frac{1}{S}$ (2.2′)

So we believe that we obtain an important result: the inertia of a response is not dependent on the number of its potential responses. Instead questions have equal inertia $\left(\Gamma_q = \frac{1}{S}\right)$. It is significant to rise the specific importance of each response in the question. This results from formula (2.1′), where its inertia depends on its weight (k_j).

It is obvious that the total inertia I_o of the cloud of responses, sum of the inertias of all the questions, is $I_o = \sum_{q=1}^{s} I_q = \sum_{q=1}^{s} \frac{1}{S} = 1$

that is $$I_o = 1 \tag{2.3'}$$

It is also clear that the role and the ability which gives this new metric $d^2_{2X^2}$ is efficient and useful in the analysis of questions with a different number of responses.

4. TRANSFORMATION OF INITIAL DATA

In order to proceed to a concrete application of the proposed approach of MCA and drax attention to the differences from classic MCA, we must develop a program wih the new proposed metric $d^2_{2X^2}$, in which every variable contributes to the analysis with equal inertia, without dependence on the number of its classes. There is another possibility which consists in transforming the initial data in the logical table and then applying classic MCA, in such a way that in the new transformed table, classic MCA with the classic χ^2 metric, lets variables contribute with equal inertia that is independent from the number of their classes. This is what we have chosen to represent.

Initial table of 0-1 data

	E_1	...	K_{ij} E_m	...	E_s	Sums of rows	Center of weight (g)
α_1							
α_2							
α_i			x			S	$\frac{S}{nS} = \frac{1}{n}$
α_n						nS	
			K_j				

Variable E_m consists of P_m classes, where K_{ij} is equal to 0 or 1 and this assigns of individual α_i to class j of E_m variable.

In order to create a new table we divide each E_m variable with number $P_m - 1$ and then we proceed in the same manner for the rest of the variables.

The new Transformed table is:

	E_1	…	E_m	…	E_s	Sums of rows	Center of weight (g)
α_1							
α_2							
			$\frac{K_{ij}}{p_m - 1}$			$\frac{1}{p_1-1} + \frac{1}{p_2-1} + \ldots + \frac{1}{p_m-1} + \ldots + \frac{1}{p_s-1} = A$	
α_i							$\frac{A}{nA} = \frac{1}{n}$
α_n							
						nA	
			$\frac{K_j}{p_m - 1}$				

We calculate consecutively, using the classic χ^2 metric the $d^2_{x^2}(j,g)$ distance of class j from the center of weight g, the inertia I^*_j of class j, the inertia I^*_m of the variable E_m and finally the total inertia $I^*_{o\lambda}$.

We have:

$$d^2_{x^2}(j,g) = \sum_{i=1}^{n} \cdot \frac{1}{\left(\frac{1}{n}\right)} \cdot \left(\frac{\frac{k_{ij}}{p_m - 1}}{\frac{k_j}{p_m - 1}} - \frac{1}{n} \right)^2 =$$

$$= n \sum_{i=1}^{n} \left(\frac{k_{ij}}{k_j} - \frac{1}{n} \right)^2 = n \sum_{i=1}^{n} \left(\frac{1}{k_j} - \frac{1}{n} \right) \quad \text{(as 2.4)} \tag{3.1}$$

and :

$$I_{j*} = \left(\frac{\frac{K_j}{p_m - 1}}{nA} \right) d^2_{x^2}(j,g) = \frac{K_j}{(p_m - 1) \cdot n \cdot A} \; n \left(\frac{1}{K_j} - \frac{1}{n} \right) = \frac{1}{(p_m - 1) \cdot A} \left(1 - \frac{K_j}{n} \right) \tag{3.2}$$

Therefore :

$$I^*_m = \sum_{j=1}^{P_m} I^*_j = \frac{1}{(p_m - 1)A} \sum_{j-1}^{P_m} \left(1 - \frac{K_j}{n} \right)$$

$$= \frac{1}{(p_m - 1)\,A} \left(p_m \frac{n}{n} \right) = \frac{1}{(p_m - 1)A} (p_m - 1) = \frac{1}{A} \tag{3.3}$$

each question contributes with same inertia equal to $1/A$, independent from the number of their classes.

Finally we have:

$$I_{o\lambda}^{*} = \sum_{m=1}^{S} I_{\mu} = \frac{S}{A} \quad (3.4)$$

From the above, we see that both proposals on one hand of the elaboration of the logical table with the new metric $d_{x^2}^2$, and on the other hand of the elaboration of the new transformed table with the classic $d_{x^2}^2$ are equal.

5. THE APPLICATION EXPLOITS REAL DATA

the answers of 68 citizens are summed up in 3 nominal variables:

E_1 ,with 7 classes ($E_{11,}$ E_{12}, …, E_{17}), E_2 with 5 classes(E_{21}, E_{22}, …, E_{25}) and E_3 with 3 classes (E_{31}, E_{32}, E_{33}).

E_{11} is a very weak class (chosen only by 3 citizens). This class contributes with 39,5 % (CTR=395) (cf. Table 2) in the creation of 4th factorial axis (it does not contribute significantly neither in the creation of 1st nor in the 2nd nor in the 3rd axis) which absorbs an important percentage 9,35 % of the total inertia, while the 1st factorial axis absorbs 12,61 % of the total inertia (cf. Table 3).

Table 2. Co-ordinates, Projections, Contributions

	#G1	COR	CTR	#G2	COR	CTR	#G3	COR	CTR	#G4	COR	CTR
ER11	-318	4	4	-1765	143	97	175	1	2	3320	508	395
ER12	150	1	3	-41	0	2	2702	579	396	-620	30	25
ER13	-1607	119	72	-2045	192	130	-1008	46	34	-456	9	10
ER14	880	323	152	-162	10	7	-174	12	8	121	6	6
ER15	-831	118	68	1281	283	177	158	4	4	677	79	59
ER16	37	0	2	305	44	23	-373	66	35	-427	87	52
ER17	-1020	82	49	-938	69	48	-186	2	3	-1063	89	71
ER21	-1967	117	69	2229	150	100	-1255	47	34	1223	45	37
ER22	-941	152	87	772	103	66	1321	301	193	-275	12	12
ER23	-424	34	19	-460	40	26	14	0	2	-1002	193	138
ER24	735	210	99	548	116	63	-149	8	6	562	122	77
ER25	154	14	7	-675	281	128	-310	59	29	24	0	4
ER31	733	399	152	-132	12	7	495	182	78	-30	0	4
ER32	-91	3	2	566	153	77	-824	324	164	-391	72	43
ER33	-1134	428	212	-510	86	49	221	16	11	555	102	67

Table 3. Table of inertia

Total inertia 4,00000

axe	inertia	% of explanation	sum	Histogramme of characteristic roots
1	0,5043067	12,61	12,61	|***
2	0,4562748	11,41	24,01	|*************************************
3	0,4418907	11,05	35,06	|************************************
4	0,3899426	9,75	44,81	|*******************************
5	0,3672703	9,18	53,99	|******************************
6	0,3390280	8,48	62,47	|***************************
7	0,3143848	7,86	70,33	|*************************
8	0,3049687	7,62	77,95	|*************************
9	0,2557831	6,39	84,35	|*********************
10	0,2434496	6,09	90,43	|********************
11	0,2267011	5,67	96,10	|******************
12	0,1559998	3,90	100,00	|*************

Therefore the class E_{11} alters the real image of the analysis. The above is elated to data analysis with classic M.A.C.

By applying multiple correspondence factor analysis in the proposed transformed table, E_{11} class does not contribute significantly to the creation of 4^{th} factorial axis nor, or course to the creation of previous axes) as it results from Table 4.

Table 3. Co-ordinates, Projections, Contributions of transformed model

	#G1	COR	CTR	#G2	COR	CTR	#G3	COR	CTR	#G4	COR	CTR
ER11	-722	24	6	991	45	13	959	42	23	-949	41	24
ER12	454	16	5	980	76	21	-1172	108	56	1001	79	44
ER13	-1311	79	22	117	0	2	1733	138	75	369	6	5
ER14	439	80	16	40	0	2	258	27	12	-418	72	31
ER15	-379	24	6	-257	11	3	-1043	187	89	-143	3	4
ER16	84	3	3	-267	33	7	-17	0	2	12	0	3
ER17	-609	29	9	-124	1	2	682	36	20	1247	123	69
ER21	-1368	56	24	-774	18	8	-1452	63	53	-2010	122	111
ER22	-569	55	21	141	3	2	-1662	475	349	1350	314	251
ER23	-380	27	10	204	8	4	482	45	32	229	10	10
ER24	369	52	17	-193	14	4	-409	64	40	-1177	536	362
ER25	214	28	9	59	2	2	844	441	234	397	98	56
ER31	938	654	336	638	302	165	-80	4	6	31	0	3
ER32	-68	2	4	-1427	973	628	47	1	4	116	6	8
ER33	-1513	762	512	757	191	137	74	1	4	-204	13	19

The contribution of E_{11} is 2,4 % of 8,8 % of inertia of the 4^{th} factorial axis (cf. Table 5)

Table 5. Inertia table of the transformed model

axe	inertia	% of explanation	sum	Histogramme of characteristic roots
1	0,6074772	18,56	18,56	|***
2	0,5699147	17,41	35,98	|**************************************
3	0,3162822	9,66	45,64	|*********************
4	0,2905260	8,88	54,52	|********************
5	0,2810637	8,59	63,11	|*******************
6	0,2687994	8,21	71,32	|******************
7	0,1811420	5,53	76,85	|************
8	0,1785875	5,46	82,31	|************
9	0,1601374	4,89	87,20	|***********
10	0,1583436	4,84	92,04	|***********
11	0,1460429	4,46	96,50	|**********
12	0,1144109	3,50	100,00	|********

The result of this is that the real explanatory importance of the axis does not change.

It is also remarkable that with the proposed analysis, explanation percentages of the first factorial axes are also generally improved.

We must note that each formula is verified for $I_{o\lambda*}$ is: $S = 3$(number of variables)

$$A = \frac{1}{p_1 - 1} + \frac{1}{p_2 - 1} + \frac{1}{p_3 - 1} = \frac{1}{6} + \frac{1}{4} + \frac{1}{2} = \frac{2+3+6}{12} = \frac{11}{12} = 0{,}916$$

Therefore: $I_{o\lambda*} = \frac{S}{A} = \frac{3}{0{,}916} = 3{,}27273$

Exactly as it is shown in Table 5 of inertia $I_{o\lambda*} = 3{,}27273$

The table of the data in the application

	E_1	E_2	E_3		E_1	E_2	E_3		E_1	E_2	E_3		E_1	E_2	E_3
*I*1	6	4	1	*I*21	6	3	3	*I*41	4	5	1	*I*61	4	5	2
*I*2	4	4	1	*I*22	5	2	3	*I*42	4	5	1	*I*62	4	5	1
*I*3	6	3	1	*I*23	4	4	1	*I*43	4	3	3	*I*63	5	5	2
*I*4	1	4	1	*I*24	7	5	1	*I*44	6	2	1	*I*64	2	3	1
*I*5	6	3	1	*I*25	4	5	1	*I*45	6	5	1	*I*65	1	5	3
*I*6	4	4	1	*I*26	6	5	1	*I*46	6	4	3	*I*66	7	5	3
*I*7	6	5	1	*I*27	5	4	1	*I*47	4	2	2	*I*67	1	5	3
*I*8	6	5	1	*I*28	4	4	1	*I*48	5	4	2	*I*68	2	2	1
*I*9	6	4	2	*I*29	6	5	2	*I*49	6	1	2				
*I*10	6	4	2	*I*30	6	5	2	*I*50	5	5	2				
*I*11	5	2	2	*I*31	7	2	2	*I*51	5	3	1				
*I*12	6	4	2	*I*32	6	5	2	*I*52	2	5	1				
*I*13	6	4	2	*I*33	4	3	2	*I*53	4	4	1				
*I*14	6	3	2	*I*34	6	2	3	*I*54	6	5	1				
*I*15	7	5	2	*I*35	3	5	3	*I*55	2	2	1				
*I*16	5	1	3	*I*36	4	5	2	*I*56	4	4	2				
*I*17	2	4	3	*I*37	7	3	3	*I*57	4	3	3				
*I*18	4	4	3	*I*38	4	4	2	*I*58	5	2	3				
*I*19	3	5	3	*I*39	4	5	1	*I*59	5	4	1				
*I*20	6	2	3	*I*40	4	5	1	*I*60	3	3	2				

6. CONCLUSIONS

With the "balance" or adjustment coefficient and the new metric $d^2_{x^2}$, it became feasible to let nominal variables with a different number of modalities, affect the analysis with the same amount of inertia $\left(I_q^{'} = \frac{1}{S}\right)$. As a result, the specific weight and significance of each modality of the variable is more clearly revealed in the analysis.

Also, the inertia which is created by a weak modality (small k_j) is smaller because of the divider $m-1$ (2.1′). Consequently, the contribution of infrequent responses in the determination of the factorial axes is tempered, that is, the possibility that the real picture is deformed by infrequent responses is limited.

Finally, with the new transformed table and the classic MCA with classic the χ^2 metric, the variables contribute with equal inertia that is independent from the number of their classes. With an application, we show the differences between classic MCA and the new proposed approach.

REFERENCES

BENZECRI J.-P., « Sur l'analyse des tableaux binaires associés à une correspondance multiple ['BinMult'] », *Cahiers de l'analyse des données* 2(1), 1977, p. 55-71, [from a mimeographed note of 1972].

ESCOFIER B., PAGES J., « Analyses factorielles multiples », Paris, Dunod, 1988.

GREENACRE M., *Correspondence Analysis in Practice*, London, Academic Press, 1993.

LEBART L., MORINEAU A., PIRON M., *Statistique exploratoire multidimensionnelle*, Paris, Dunod, 2000.

LE ROUX B., « Analyse spécifique d'un nuage euclidien », *Mathématiques et Sciences humaines* 37, 1999, p. 65-83.

LE ROUX B., ROUANET H., *Geometric Data Analysis from Correspondence Analysis to Structured Data Analysis*, Dordrecht-Boston-London, Kluwer Academic Publisher, 2004.

MOSCHIDIS O., PAPADIMITRIOU I., "Evaluation scales (subjectivity–transformation of the initial degrees in probabilities.)", *Data Analysis Bulletin* 1, 2002, p. 42-51.

MOSCHIDIS O., PAPADIMITRIOU I., "Proposition of new indicators and a new metric for scale evaluation analysis with the methods of multidimensional data analysis", *Data Analysis Bulletin* 2, 2002, p. 65-73.

MOSCHIDIS O., "Evaluation scales: metrics–similarity in table 0-1", *Proceedings of 15th Pan-Hellenic Statistic Congress*, 2002, p. 500-508.

MOSCHIDIS O., "A proposition for coding evaluation scales", *Proceedings of the 16th Pan-Hellenic Statistic Congress*, 2003, p. 337-346.

VAN DE GEER J. P., *Multivariate Analysis of Cate-orical Data: Theory*, Sage, California, 1993, 120 p.