Journal of Applied Statistics, 2015
Vol. 42, No. 10, 2192–2202, http://dx.doi.org/10.1080/02664763.2015.1023269

Taylor & Francis
Taylor & Francis Group

A generalized analysis of the dependence structure by means of ANOVA

Antonello D'Ambra[a*], Anna Crisci[b] and Pasquale Sarnacchiaro[c]

[a]*Department of Economics, Second University of Naples, Via Gran Priorato di Malta, 81043, Capua (Caserta), Italy;* [b]*Department of Economics, Management and Institutions, University of Naples Federico II, Via Cinthia, 80100, Naples, Italy;* [c]*Department of Economics, University Unitelma Sapienza, Viale Regina Elena 295, 00161, Rome, Italy*

(*Received 26 March 2014; accepted 23 February 2015*)

The multiple non-symmetric correspondence analysis (MNSCA) is a useful technique for analysing the prediction of a categorical variable through two or more predictor variables placed in a contingency table. In MNSCA framework, for summarizing the predictability between criterion and predictor variables, the Multiple-TAU index has been proposed. But it cannot be used to test association, and for overcoming this limitation, a relationship with C-Statistic has been recommended. Multiple-TAU index is an overall measure of association that contains both main effects and interaction terms. The main effects represent the change in the response variables due to the change in the level/categories of the predictor variables, considering the effects of their addition. On the other hand, the interaction effect represents the combined effect of predictor variables on the response variable. In this paper, we propose a decomposition of the Multiple-TAU index in main effects and interaction terms. In order to show this decomposition, we consider an empirical case in which the relationship between the demographic characteristics of the American people, such as race, gender and location (*column variables*), and their propensity to move (*row variable*) to a new town to find a job is considered.

Keywords: multiple non-symmetric correspondence analysis; multiple-TAU index; main effects and interaction term; confidence circles; C-Statistic

1. Introduction

The prediction of a categorical variable (or *criterion variable*), through one or more predictor variables, frequently happens in empirical research. When there are a criterion variable and a predictor variable, and data are placed in a two-way contingency table, the non-symmetric correspondence analysis allows analysing the asymmetric relationship between variables [5]. When the criterion variables are two or more, the relationship structure becomes more complicated. In fact, it depends on both the main effects of each variable and the effect of the interactions. When two criterion variables are considered, the analysis of relationship could be performed

*Corresponding author. Email: antonello.dambra@unina2.it

using the external information [13] or through the ANOVA performed on composite coordinates of multiple non-symmetrical correspondence analysis (MNSCA) [3,4].

Considering this [3,4], the aim of this paper, in MNSCA framework, is to propose a decomposition of the multiple-TAU index into main effects and interaction terms. In particular, for simplicity, we consider the case of three predictor variables, but the results can be generalized for more than three predictors variables.

The paper is organized as follows: after the statistics notation (Section 2), in the third section, a short presentation of the MNSCA is done; particularly, the multiple-TAU index and the C-Statistic are illustrated. In Section 4, the decomposition of the multiple-TAU index into main effects and interaction terms is developed. An empirical study on the propensity to move for job of the US citizens is analysed in Section 5; particularly, the multiple-TAU index decomposition has been used in order to point out the advantages of the proposal.

2. Statistics notation

Let N be a two-way contingency table in which we consider the cross-classification of n statistical units according to four categorical variables Y, A, B and C. Define Y as the first (row) criterion variable consisting of I categories. The columns have been obtained by means of the concatenation of three predictor variables A, B and C of J, K and W categories, respectively. The contingency table dimension are $I \times (J \times K \times W)$ with general term n_{ijkw} ($i = 1, 2, \ldots, I$; $j = 1, 2, \ldots, J$; $k = 1, 2, \ldots, K$ and $w = 1, 2, \ldots, W$). Let $P = (n_{ijkw}/n = p_{ijkw})$ be the relative frequency distribution and let $p_{i\bullet\bullet\bullet} = \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{w=1}^{W} p_{ijkw}$ and $p_{\bullet jkw} = \sum_{i=1}^{I} p_{ijkw}$ be the marginal row and column frequencies, respectively.

3. Multiple non-symmetric correspondence analysis

The MNSCA allows studying the association between row variable and column predictor variables, particularly applying the Generalized Singular Value Decomposition to the matrix $\Pi = \{\pi_{ijkw} = p_{ijkw}/p_{\bullet jkw} - p_{i\bullet\bullet\bullet}\}$ with weights $p_{\bullet jkw}$, we obtain

$$\pi_{ijkw} = \frac{p_{ijkw}}{p_{\bullet jkw}} - p_{i\bullet\bullet\bullet} = \sum_{m=1}^{M} a_{im}\lambda_m b_{jkwm} \tag{1}$$

with the constraints

$$\sum_{i=1}^{I} a_{im}a_{im'} = \begin{cases} 1, m & = m' \\ 0, m & \neq m' \end{cases} \quad \sum_{j=1}^{I}\sum_{k=1}^{J}\sum_{w=1}^{W} p_{\bullet jkw} b_{jkwm} b_{jkwm'} = \begin{cases} 1, m & = m' \\ 0, m & \neq m' \end{cases}$$

and where $M = [\min(I, J \times K \times W) - 1]$, $\lambda_m (m = 1, 2, \ldots M)$ are generalized singular values (arranged in descending order), a_{im} is a general element of the singular vector $\mathbf{a}_m$, b_{jkwm} is an element of the joined singular vector $\mathbf{b}_m$ associated with the predictor variables.

For visualizing the dependence between the row and column categories, the row and column profile coordinates can be computed:

$$f_{im} = a_{im}\lambda_m, \tag{2}$$
$$g_{jkwn} = b_{jkwn}\lambda_m,$$

In MNSCA, the effects of the predictor variables (in term of both main effect and interaction effect) on response variable are mixed. For summarizing the association between criterion and predictor variables, the multiple-TAU can be used [6,7].

$$\tau_{\text{mult}} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{w=1}^{W} p_{\bullet jkw}(p_{ijkw}/p_{\bullet jkw}) - p_{i\bullet\bullet\bullet}{}^2}{1-\sum_{i=1}^{I} p_{i\bullet\bullet\bullet}^2} = \frac{\text{num}(\tau_{\text{mult}})}{1-\sum_{i=1}^{I} p_{i\bullet\bullet\bullet}^2}. \tag{3}$$

We focus our discussion on the numerator of τ_{mult}, since $1-\sum_{i=1}^{I} p_{i\bullet\bullet\bullet}^2$ is independent of $n_{ijkw}(\forall i,j,k,w)$.

Although τ_{mult} is an appropriate measure of predictability, it cannot be used to test association. In order to overcome this limitation for testing the association, the C-Statistic [9] can be used:

$$C = (n-1)(I-1)\tau_{\text{mult}} = (n-1)(I-1)\frac{\sum_{m=1}^{M}\lambda_m^2}{1-\sum_{i=1}^{I} p_{i\bullet\bullet\bullet}^2}. \tag{4}$$

C-Statistic and τ_{mult} are linked [1]; moreover, in case of zero predictability hypothesis (no association between variables), it has been shown that C-Statistic is asymptotically a chi-squared distribution with $\{(I-1)\times[(J\times K\times W)-1]\}$ degree of freedom [9]. When the variables are considered to be symmetrically related, as in the case of Correspondence Analysis, confidence circles (CCs) have been proposed in order to identify those categories that are significant [8]. These CCs are similar to those used in canonical analysis [10]. Another approach to calculate CCs is based on the bootstrap procedure; also for MNSCA, the construction of CCs has been proposed [2].

The $1-\alpha$ CC for the jkwth column coordinate in MNSCA plot has the following radius length:

$$\text{rad}_{jkw} = \sqrt{\frac{\chi^2_{2,\alpha}(1-\sum_{i=1}^{r} p_{i\bullet\bullet\bullet}^2)}{p_{\bullet jkw}(n-1)(I-1)}}. \tag{5}$$

4. Decomposition of the multiple TAU into main effects and interaction term

The main effects represent the change in the response variable due to the change in the levels/categories of the predictor variables.

On the other hand, the interaction effect represents the combined effect of predictor variables on the response variable. In particular, there is an interaction between two predictor variables when the effect of one predictor variable varies as the levels/categories of the other predictor vary.

If the interaction is not statistically significant, it is possible to examine the main effects. Instead, if the interaction is statistically significant, then, it is not appropriate to consider the main effects. As a matter of fact, asserting that two predictor variables interact is the same as affirming that the two variables do not have separate effects [11].

The main aim of this paper is to propose a decomposition of τ_{mult} for separating the main effects and the interaction terms. The proposed approach, in MNSCA framework, starts from the exact reconstruction formula of the contingency table by using eigenvalues and profile coordinates, particularly,

$$p_{ijkw} = p_{\bullet jkw}\left[p_{i\bullet\bullet\bullet} + \sum_{m=1}^{M}\left(\frac{1}{\sqrt{\lambda_m}}\right) f_{im} g_{jkwm}\right]. \tag{6}$$

The coordinates g_{jkwm} computed according to Equation (2) include the main effects and the interactions.

Taking into account the procedure suggested by [3,4], we substitute in Equation (6) the g_{jkwm} $(m = 1, 2, \ldots M)$ with the estimation $\hat{g}_{jkwn}$ obtained by means of the three-way ANOVA in terms of projectors [14].

Let $\mathbf{D} = [\mathbf{D}_A|\mathbf{D}_B|\mathbf{D}_C]$ be the matrix of dummy variables, in which the slices $\mathbf{D}_A$ (factor A), $\mathbf{D}_B$ (factor B) and $\mathbf{D}_C$ (factor C) are obtained as:

$$\begin{aligned}\mathbf{D}_A &= \mathbf{I}_W \otimes \mathbf{1}_{(J\times K)},\\ \mathbf{D}_B &= \mathbf{1}_W \otimes (\mathbf{I}_K \otimes \mathbf{1}_J),\\ \mathbf{D}_C &= \mathbf{1}_{(K\times W)} \otimes \mathbf{I}_J,\end{aligned} \tag{7}$$

where $\mathbf{I}_W$, $\mathbf{I}_K$ and $\mathbf{I}_J$ are identity matrices and $\mathbf{1}$ is a column vector whose elements are all 1. Subsequently, we build the matrix $\mathbf{D}_R$ (resp. $\mathbf{D}_{RA}$, resp. $\mathbf{D}_{RB}$, resp. $\mathbf{D}_{RC}$) by means of the repetition of each row of the matrix $\mathbf{D}$ (resp. $\mathbf{D}_A$, resp. $\mathbf{D}_B$, resp. $\mathbf{D}_C$) as many times as the $n_{\bullet jkw}$.

In this way, considering the ANOVA framework, we can compute the main effects by using the orthogonal projectors:

$$\begin{aligned}&\hat{g}_R\{\hat{g}_{hm}^{A\cup B\cup C}\} = D_R(D'_R D_R)^{-1}D'_R g_R && \text{Main effects,}\\ &\hat{g}_{RA}\{\hat{g}_{hm}^{A}\} = D_{RA}(D'_{RA} D_{RA})^{-1}D'_{RA} g_R && \text{Main effect } A,\\ &\hat{g}_{RB}\{\hat{g}_{hm}^{B}\} = D_{RB}(D'_{RB} D_{RB})^{-1}D'_{RB} g_R && \text{Main effect } B,\\ &\hat{g}_{RC}\{\hat{g}_{hm}^{C}\} = D_{RC}(D'_{RC} D_{RC})^{-1}D'_{RC} g_R && \text{Main effect } C,\end{aligned} \tag{8}$$

$\beta_{lm} = (D'D)^{-1}D'g_{hm}$ where $\mathbf{g}_R$ is a matrix of size $(n \times M)$ whose elements g_{hm} $(h = 1, 2, \ldots, n$ and $m = 1, 2, \ldots, M)$ have been obtained repeating the g_{jkwm} $(\forall j, k, w, m)$ as many times as the $n_{\bullet jkw}$. Also, the values contained $\hat{\mathbf{g}}_R$ (resp. $\hat{\mathbf{g}}_{RA}$, resp. $\hat{\mathbf{g}}_{RB}$, resp. $\hat{\mathbf{g}}_{RC}$) are repeated as in $\mathbf{g}_R$ (resp. $\mathbf{g}_{RA}$, resp. $\mathbf{g}_{RB}$, resp. $\mathbf{g}_{RC}$) and if we consider only the values not identical, they represent the estimations of main effects $\hat{g}_{jkwm}^{A\cup B\cup C}$(resp. $\hat{g}_{jkwm}^{A}$, resp. $\hat{g}_{jkwm}^{B}$, resp. $\hat{g}_{jkwm}^{C}$). Then, in the case of main effects (but the same could be computed for each Main Effect), the fitted values $\hat{g}_{jkwm}^{A\cup B\cup C}$ are inserted in Equation (6):

$$\hat{p}_{ijkw}^{A\cup B\cup C} = p_{\bullet jkw}\left[p_{i\bullet\bullet\bullet} + \sum_{m=1}^{M}\frac{1}{\sqrt{\lambda_m}} f_{im}\hat{g}_{jkwm}^{A\cup B\cup C}\right]. \tag{9}$$

By this way, we obtain a new fitted matrix $\hat{\mathbf{P}}^{A\cup B\cup C}$ (with the same marginal row and column frequencies of $\mathbf{P}$) representing the reconstruction matrix $\mathbf{P}$ with only main effects (ignoring interactions).

Using these results, we obtain the following orthogonal decomposition of num (τ_{mult}):

$$\begin{aligned}\underbrace{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{w=1}^{W}\left[\left(\frac{p_{ijkw}}{p_{\bullet jkw}} - p_{i\bullet\bullet\bullet}\right)^2 p_{\bullet jkw}\right]}_{\text{numerator of } \tau_M} &= \underbrace{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{w=1}^{W}\left[\left(\frac{\hat{p}_{ijkw}^{A\cup B\cup C}}{p_{\bullet jkw}} - p_{i\bullet\bullet\bullet}\right)^2 p_{\bullet jkw}\right]}_{\text{Main Effects}}\\ &\quad + \underbrace{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{w=1}^{W}\left[\left(\frac{p_{ijkw}}{p_{\bullet jkw}} - \frac{\hat{p}_{ijkw}^{A\cup B\cup C}}{p_{\bullet jkw}}\right)^2 p_{\bullet jkw}\right]}_{\text{Interaction effect}}.\end{aligned} \tag{10}$$

In particular, we decompose num (τ_{mult}) in the sum of the main effects $(A \cup B \cup C)$and the interaction effect $(A \times B \times C)$. The main effects could be handily decomposed into the following:

The main effect of the factor A + the main effects of B and C ($B \cup C$) conditioned to the main effect of the factor A (we consider the conditional effect since the design is unbalanced and thus the single effects are no longer orthogonal)

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{w=1}^{W}\left[\left(\frac{\hat{p}_{ijkw}^{A\cup B\cup C}}{p_{\bullet jkw}}-p_{i\bullet\bullet\bullet}\right)^2 p_{\bullet jkw}\right]=\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{w=1}^{W}\left[\left(\frac{\hat{p}_{ijkw}^{A}}{p_{\bullet jkw}}-p_{i\bullet\bullet\bullet}\right)^2 p_{\bullet jkw}\right]$$
$$+\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{w=1}^{W}\left[\left(\frac{\hat{p}_{ijkw}^{(B\cup C)/A}}{p_{\bullet jkw}}-p_{i\bullet\bullet\bullet}\right)^2 p_{\bullet jkw}\right]. \tag{11}$$

The main effect of the factor B + the main effects of A and C ($A \cup C$) conditioned to the effect of the factor B:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{w=1}^{W}\left[\left(\frac{\hat{p}_{ijkw}^{A\cup B\cup C}}{p_{\bullet jkw}}-p_{i\bullet\bullet\bullet}\right)^2 p_{\bullet jkw}\right]=\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{w=1}^{W}\left[\left(\frac{\hat{p}_{ijkw}^{B}}{p_{\bullet jkw}}-p_{i\bullet\bullet\bullet}\right)^2 p_{\bullet jkw}\right]$$
$$+\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{w=1}^{W}\left[\left(\frac{\hat{p}_{ijkw}^{(A\cup C)/B}}{p_{\bullet jkw}}-p_{i\bullet\bullet\bullet}\right)^2 p_{\bullet jkw}\right]. \tag{12}$$

The main effect of the factor C + the main effects of A and B ($A \cup B$) conditioned to the effect of the factor C:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{w=1}^{W}\left[\left(\frac{\hat{p}_{ijkw}^{A\cup B\cup C}}{p_{\bullet jkw}}-p_{i\bullet\bullet\bullet}\right)^2 p_{\bullet jkw}\right]=\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{w=1}^{W}\left[\left(\frac{\hat{p}_{ijkw}^{C}}{p_{\bullet jkw}}-p_{i\bullet\bullet\bullet}\right)^2 p_{\bullet jkw}\right]$$
$$+\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{w=1}^{W}\left[\left(\frac{\hat{p}_{ijkw}^{(A\cup B)/C}}{p_{\bullet jkw}}-p_{i\bullet\bullet\bullet}\right)^2 p_{\bullet jkw}\right]. \tag{13}$$

In case of three predictor variables, we can obtain 13 decompositions of the τ_{mult} (Table 1).

For each decomposition, the C-Statistic that under null hypothesis is distributed as a chi-square can be calculated (for degrees of freedom see Table 2).

Table 1. Decomposition of τ_{mult}.

		$A \cup B \cup C$			+	($A \times B \times$ C)
A	+		$(B \cup C)/A$		+	($A \times B \times$ C)
B	+		$(A \cup C)/B$		+	($A \times B \times$ C)
C	+		$(A \cup B)/C$		+	($A \times B \times$ C)
$A \cup B$			+	$C/(A \cup B)$	+	($A \times B \times$ C)
A	+	B/A	+	$C/(A \cup B)$	+	($A \times B \times$ C)
A/B	+	B	+	$C/(A \cup B)$	+	($A \times B \times$ C)
$A \cup C$			+	$B/(A \cup C)$	+	($A \times B \times$ C)
A	+	C/A	+	$B/(A \cup C)$	+	($A \times B \times$ C)
A/C	+	C	+	$B/(A \cup C)$	+	($A \times B \times$ C)
$B \cup C$			+	$A/(B \cup C)$	+	($A \times B \times$ C)
B	+	C/B	+	$A/(B \cup C)$	+	($A \times B \times$ C)
B/C	+	C	+	$A/(B \cup C)$	+	($A \times B \times$ C)

Table 2. Degrees of freedom for C-Statistic.

$(I-1)\times[(J-1)+(K-1)+(W-1)]$					+	$(I-1)\times[(J\times K\times W)-J-K-W+2]$
$(I-1)\times(W-1)$	+	$(I-1)\times[(K-1)+(J-1)]$			+	$(I-1)\times[(J\times K\times W)-J-K-W+2]$
$(I-1)\times(K-1)$	+	$(I-1)\times[(W-1)+(J-1)]$			+	$(I-1)\times[(J\times K\times W)-J-K-W+2]$
$(I-1)\times(J-1)$	+	$(I-1)\times[(W-1)+(K-1)]$			+	$(I-1)\times[(J\times K\times W)-J-K-W+2]$
$(I-1)^*[(W-1)+(K-1)]$			+	$(I-1)\times(J-1)$	+	$(I-1)\times[(J\times K\times W)-J-K-W+2]$
$(I-1)\times(W-1)$	+	$(I-1)\times(K-1)$	+	$(I-1)\times(J-1)$	+	$(I-1)\times[(J\times K\times W)-J-K-W+2]$
$(I-1)\times(W-1)$	+	$(I-1)\times(K-1)$	+	$(I-1)\times(J-1)$	+	$(I-1)\times[(J\times K\times W)-J-K-W+2]$
$(I-1)\times[(W-1)+(J-1)]$			+	$(I-1)\times(W-1)$	+	$(I-1)\times[(J\times K\times W)-J-K-W+2]$
$(I-1)\times(W-1)$	+	$(I-1)\times(J-1)$	+	$(I-1)\times(W-1)$	+	$(I-1)\times[(J\times K\times W)-J-K-W+2]$
$(I-1)\times(W-1)$	+	$(I-1)\times(J-1)$	+	$(I-1)\times(W-1)$	+	$(I-1)\times[(J\times K\times W)-J-K-W+2]$
$(I-1)\times[(K-1)+(J-1)]$			+	$(J-1)\times(W-1)$	+	$(I-1)\times[(J\times K\times W)-J-K-W+2]$
$(I-1)\times(K-1)$	+	$(I-1)\times(J-1)$	+	$(J-1)\times(W-1)$	+	$(I-1)\times[(J\times K\times W)-J-K-W+2]$
$(I-1)\times(K-1)$	+	$(I-1)\times(J-1)$	+	$(J-1)\times(W-1)$	+	$(I-1)\times[(J\times K\times W)-J-K-W+2]$

5. The empirical study

The aim of the study is to evaluate how demographic characteristics of US citizen, gender (X_1), race (X_2) and location (X_3) influence the inclination of individuals moving to another state (Y_1) in order to find a job. Many recent studies analysed the relationship between these characteristics and the decision to move for job [11]. Y_1 has been classified into four categories: people who do not prefer changing residential country 'No-Move'; people who prefer moving to West American direction 'West America'; people who prefer moving to East American direction 'East America'; and 'Undecided' people. In this way, we can study both propensity to mobility and favourite destination. The data have been placed in a three-way contingency table (Table 3).

In order to analyse the dependence structure, MNSCA has been carried out. The τ_{mult} numerator and the C-Statistic are 0.0805 and 1520.962 (DoF = 21, P-value = 0.000), respectively.

The factorial plan representations (99.64% of explained variability on the first two axes) have been considered (Figure 1). In Figure 1(a), the Y_1's categories have been drawn. In Figure 1(b), in which the categories of predictor variable have been plotted, we notice that *Male* categories are on the left part of the figure and *Female* categories on the right. Considering Figure 1(a) and 1(b) jointly, we observe that the citizen 'White, Female, West – W.F.W' and 'Black, Female, West – B.F.W' prefer to move towards East, while the citizen 'White, Male, East – W.M.E' and 'Black, Male, East – B.M.E' prefer to move towards West.

In order to identify the categories which are statistically significant, the CCs [2] have been computed (Figure 2). In this case, all categories are statistically significant for explaining Y_1.

Table 3. Cross-classification of 4518 US citizen according to *Decision to move*, *Race*, *Gender* and *Location*.

	Black				White				
	Male		Female		Male		Female		
	West	East	West	East	West	East	West	East	Total
Prefer to stay (no-move)	98	42	130	462	184	173	27	240	1356
Prefer to move to West	96	438	61	190	294	437	25	46	1587
Prefer to move to East	18	84	135	394	81	82	88	194	1076
Undecided	26	56	52	136	81	82	20	46	499
Total	238	620	378	1182	640	774	160	526	4518

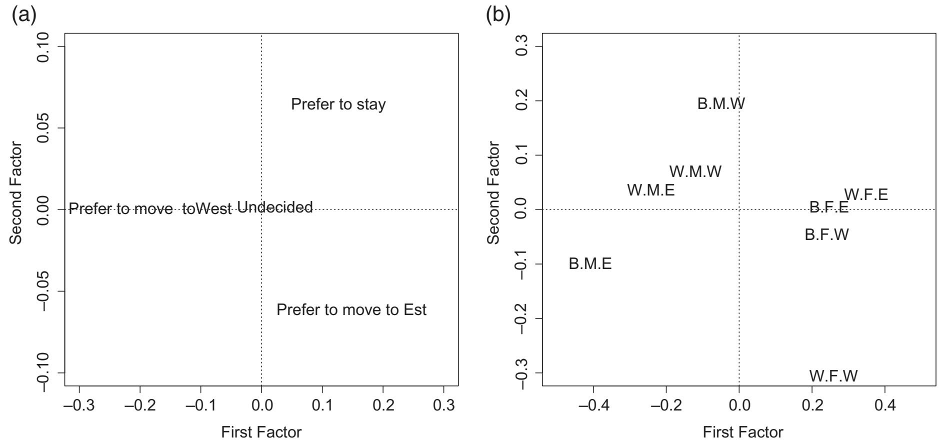

Figure 1. MNSCA for **P** – factorial planes.

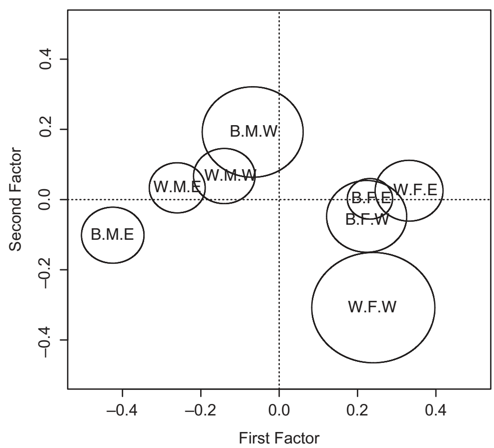

Figure 2. MNSCA for **P** – CCs for the categories of predictor variables.

Table 4. Decomposition of the τ_{mult} in main effects and interaction term.

	Stat. C	DoF	P-Value
Race ∪ Gender ∪ Location	1301.128	9	0.000
Race × Gender × Location	219.834	12	0.000
Total	1520.692	21	0.000

In order to improve the analysis of the dependence structure between criterion and predictor variables, according to Table 1, the τ_{mult} numerator has been decomposed. The decomposition into main effects and interaction term shows that they are statistically significant (Table 4).

For detecting which main effects are statistically significant, further decompositions of τ_{mult} have been performed (Tables 5–7). In particular, we observe that only the main effect *Race* is not significant, while *Gender* and *Location* are statistically significant.

Table 5. τ_{mult} – main effect *Race* vs. others.

	Stat. *C*	DoF	*P*-Value
Race	6.009	3	0.111
(Gender $\cup$ Location)/Race	1295.119	6	0.000
Race $\times$ Gender $\times$ Location	219.834	12	0.000
Total	1520.692	21	0.000

Table 6. τ_{mult} – main effect *Gender* vs. others.

	Stat. *C*	DoF	*P*-Value
Gender	1215.541	3	0.000
(Race $\cup$ Location)/Gender	85.587	6	0.000
Race $\times$ Gender $\times$ Location	219.834	12	0.000
Total	1520.692	21	0.000

Table 7. τ_{mult} – main effect *Location* vs. others.

	Stat. *C*	DoF	*P*-Value
Location	26.886	3	0.000
(Race $\cup$ Gender)/Location	1274.242	6	0.000
Race $\times$ Gender $\times$ Location	219.834	12	0.000
Total	1520.692	21	0.000

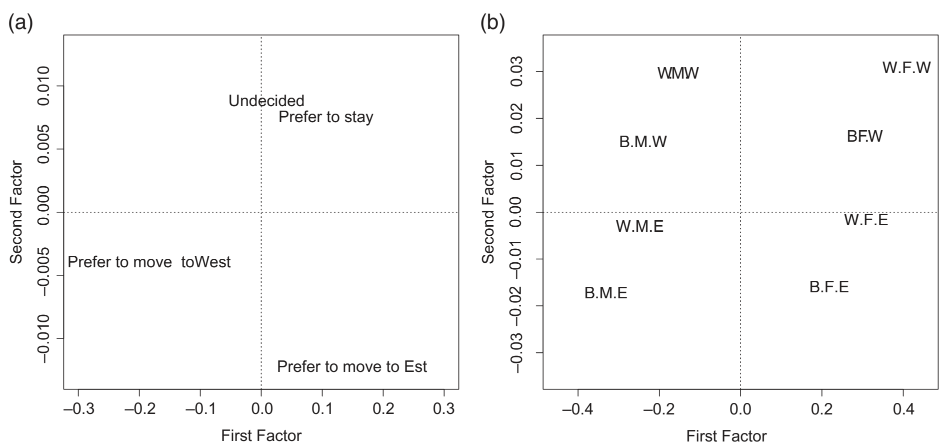

Figure 3. MNSCA for $\hat{\mathbf{P}}^{A\cup B\cup C}$ – factorial planes.

By using Equation (9), we reconstruct the matrix $\hat{\mathbf{P}}^{A\cup B\cup C}$ (main effects without interaction) and we perform an MNSCA (Figure 3). Also in this case, we retain two factors (99.89% of explained variability). Looking at Figure 3(b), we observe that *Male* categories are on the left part; the contrary happens for *Female* categories (as in Figure 1(b)). Moreover, *West* categories are on the high part while *East* categories are on the law part of the plan.

Table 8. Matrix of the squared distances between the columns of $\hat{\mathbf{P}}^{A\cup B\cup C}$.

	B.M.W	B.M.E	B.F.W	B.F.E	W.M.W	W.M.E	W.F.W	W.F.E
B.M.W	–							
B.M.E	0.103	–						
B.F.W	0.552	0.651	–					
B.F.E	0.455	0.552	0.103	–				
W.M.W	0.093	0.195	0.461	0.366	–			
W.M.E	0.017	0.093	0.559	0.461	0.103	–		
W.F.W	0.644	0.743	0.093	0.195	0.552	0.651	–	
W.F.E	0.546	0.644	0.017	0.093	0.455	0.552	0.103	–

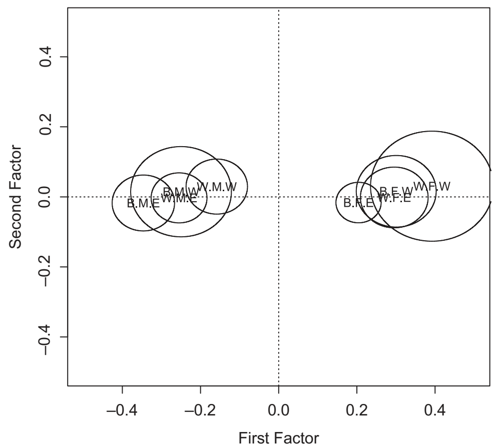

Figure 4. MNSCA for $\hat{\mathbf{P}}^{A\cup B\cup C}$ – CCs for categories column.

Later, we calculated the squared distances between columns of $\hat{\mathbf{P}}^{A\cup B\cup C}$ (Table 8) using the following formula:

$$d^2(jkw; jkw') = \sum_{i=1}^{r} \left(\frac{\hat{p}_{ijkw}^{A\cup B\cup C}}{p_{\bullet jkw}} - \frac{\hat{p}_{ijkw'}^{A\cup B\cup C}}{p_{\bullet jkw'}} \right)^2 (\forall\, w \neq w'). \tag{14}$$

We point out that the squared distances between two categories of a main effect are equal independently from the categories of other two variables.

For example, the squared distance between the categories of *Location* is 0.103; in fact, $d^2(\mathrm{BME};\mathrm{BMW}) = d^2(\mathrm{BFE};\mathrm{BFW}) = d^2(\mathrm{WME};\mathrm{WMW}) = d^2(\mathrm{WFE};\mathrm{WFW}) = 0.103$.

In order to detect the categories of the predictor variables that influence the criterion variable, the CCs have been represented (Figure 4). As can be seen, all categories are statistically significant.

Performing an MNSCA on the matrix $\hat{\mathbf{P}}^{A\times B\times C}$ we analyse the interaction effects. Particularly, we observe that 'B.M.W' prefer to stay (No-move), while the 'W.F.W' prefer to move towards East and 'B.F.W' prefer to move towards West.

Also in this case, we construct the CCs for the categories column and we represent the categories not statistically significant.

By comparing the plot of MNSCA on the original table (Figure 1) with the plot of interaction (Figure 5), we can observe how the joint effect of the race, gender and location (interaction effect) leads to a different behaviour of the women in terms of e propensity to move. In particular, the

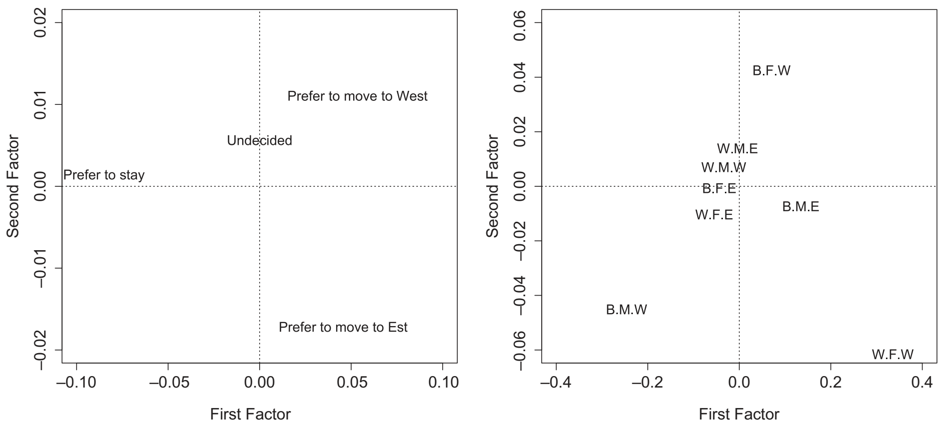

Figure 5. MNSCA for $\hat{\mathbf{P}}^{A\times B\times C}$ – factorial planes.

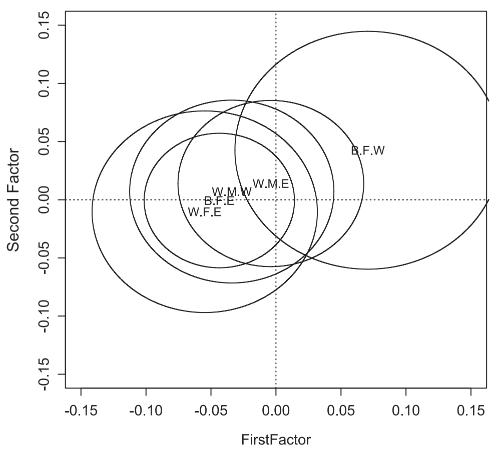

Figure 6. MNSCA for $\hat{\mathbf{P}}^{A\times B\times C}$ – CCs for categories column not statistically significant.

black women of the *West* do prefer to move, but remain within the *West* America. While, in the original plot, where we consider both the main effects and interaction term, this relation is not clear or not identified (Figure 6).

6. Discussion

In this paper, an extension of MNSCA, obtained through the decomposition of the Multiple-TAU index, has been proposed. This decomposition allows separating the main effects and interaction term in the Multiple-TAU index. In this way, it is possible to have the following advantages: to take into account only the statistically significant components, to examine in depth the effect of each predictor variable on criterion variable, to investigate the impact of each modality of statistically significant predictor variable and to detect if there are combined effects of criterion variables that could not be identified with the classical approach.

This proposal has been used to analyse the relationship between the demographic characteristics (Race, Gender and Location) of US citizen and their propensity to move for job. Through this approach, integrating the results obtained with the traditional statistical methods, we have shown that a deepened point of view for the phenomenon can be obtained.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] R.J. Anderson and J.R. Landis, *Catanova for multidimensional contingency tables: Nominal scale response*, Comm. Statist. Theory Methods 9 (1980), pp. 1191–1206.

[2] E.J. Beh and L. D'Ambra, *Non-symmetrical correspondence analysis with concatenation and linear constraints*, Aust. N. Z. J. Stat. 52(1) (2010), pp. 27–44.

[3] A. D'Ambra and A. Crisci, *Ellipses confidence in decomposition multiple non symmetrical correspondence Analysis*, Comm. Statist. Theory Methods 43(6) (2014), pp. 1209–1221.

[4] L. D'Ambra, A. D'Ambra, and P. Sarnacchiaro, *Visualising main effects and interaction in multiple non symmetric correspondence analysis*, J. Appl. Stat. 39(10) (2012), pp. 2165–2175.

[5] L. D'Ambra and N. Lauro, *Non symmetrical analysis of three way contingency tables*, in *Multiway Data Analysis*, R. Coppi and S. Bolasco, eds., North Holland, Amsterdam, 1989, pp. 301–315.

[6] L.A. Goodman and W.H. Kruskal, *Measures of association for cross-classifications*, J. Amer. Statist. Assoc. 49 (1954), pp. 732–764.

[7] L.N. Gray and J.S. Williams, *Goodman and Kruskals Tau B: Multiple and partial analogy*, Proceedings of the Social Statistics Section, American Statistical Association, 1975, pp. 444–448.

[8] L. Lebart, A. Morineau, and K.M. Warwick, *Multivariate Descriptive Statistical Analysis*, Wiley, New York, 1984.

[9] R.J. Light and B.H. Margolin, *An analysis of variance for categorical data*, J. Amer. Statist. Assoc. 66 (1971), pp. 534–544.

[10] K.V. Mardia, J.T. Kent, and J.M. Bibby, *Multivariate Analysis*, Academic Press, London, 1982.

[11] J.C. Moore and E. Krupat, *Relationship between source status, authoritarianism and conformity in social setting*, Sociometry 34 (1971), pp. 122–134.

[12] R.H. Murrillo, L.S. Ott, M.T. Owyang, and D. Whaken, *Patterns of interstate migration in United States from the survey of income and program participation*, Fed. Reserve Bank St. Louis Rev. 93(3) (2011), pp. 169–185.

[13] Y. Takana and T. Shibayana, *Principal component analysis with external information on both subjects and variables*, Psychometrika 55 (1991), pp. 151–158.

[14] K. Takeuchi, H. Yanay, and B.N. Mukherjee, *The Foundation of Multivariate Analysis. A Unified Approach by means of Projection onto Linear Subspaces*, Wiley Eastern Limited, New Delhi, 1981, pp. 138–156.