Statistica & Applicazioni - Vol. XV, n. 1, 2017, pp. 3-18

NON-SYMMETRICAL CORRESPONDENCE ANALYSIS TO EVALUATE HOW AGE INFLUENCES THE ADDICTION DISCOURSES

Enrico Ciavolino*
Antonello D'Ambra**
Claudia Venuleo*
Marta Vernai*

SUMMARY

The current paper aims to present the method of Non-Symmetrical Correspondence Analysis (NSCA) based on Emerson's orthogonal polynomials, which takes into account, in efficient way, the ordinal structure of the data. The extension of NSCA is the so called Singly Ordered Non-Symmetrical Correspondence Analysis version (SONSCA), that, by taking into account the ordinal structure in the table, improving the interpretation ability of the analysis. The methods was applied to 40 in-depth interviews, gathered with people in treatment for their problems with addiction. NSCA and SONSCA are used to evaluate if the classes of age of the subjects interviewed influence the addiction thematic categories that characterize their discourses.The work provides insight on NSCA and SONSCA methods and how they could be applied in the psychological context and in particular to study the dependence between ordinal and categorical variables reported in a contingency table.

Keywords: *Non-Symmetrical Correspondence Analysis (NSCA), Emerson's Orthogonal Polynomials, Singly Ordered Non-Symmetrical Correspondence Analysis (SONSCA), Addiction Thematic Categories, Classes of Age.*

1. INTRODUCTION

Scientific research in many disciplines including psychology, biology, ecology, epidemiology, and so on, often concerns the prediction of one categorical variable from another.

Correspondence analysis-CA (Fisher, 1940; Beh and Lombardo, 2014; Lombardo and Beh, 2016) or Non-symmetrical correspondence analysis (D'Ambra and Lauro, 1989) for a two-way contingency tables are a popular statistical tools for graphically identifying the nature of the association between categorical variables (Ciavolino, Redd, Evrinomy, Falcone, Fini, Kadianaki, Kullasepp, Mannarini, Matsopoulos, Mossi *et al*., 2017; Kamalja and Khangar, 2017). The key difference between the symmetrical and non-symmetrical versions of correspondence analysis rests on the

* Dipartimento di Storia, Società e Studi sull'Uomo - Università del Salento - Via di Valesio - 73100, LECCE (e-mail: enrico.ciavolino@unisalento.it; claudia.venuleo@unisalento.it; marta.vernai@hotmail.it).
** Dipartimento di Economia - Università della Campania ''Luigi Vanvitelli'' - Gran Priorato di Malta, CAPUA (e-mail: antonello.dambra@unicampania.it).

measure of the association used to quantify the relationship between the variables. In particular, the Pearson's chi-square statistic (symmetric case) or Goodman-Kruskal's tau statistic (not-symmetric case) are used to study the association between categorical variables, but these statistics can not be perform when we consider an ordinal structure of the categories (D'Ambra, 2017; Agresti, 2007; Barlow, Bartholomew, Bremner and Brunk, 1972).

We will illustrate the usefulness of a specific variant of correspondence analysis applied to 40 in depth interviews gathered with people in treatment for their problems with addiction. The interviews aimed to gain a deeper insight in how members describe, understand and face their problems with substance or behavioural addiction. We wonder whether the age of the interviewed relate to the thematics that characterize their discourses.

We used a contingency table with addiction thematic categories as rows and range of age as columns. A contingency table of thematic categories by age is a special case of the situation with two categorical variables, in which the distribution of one variable, the criterion variable (in this case is represented by categories) is assessed conditional on the values of the other variable, the predictor (here: age of subjects). Thus the evaluation of the relative subjects? discourses may also be described as a prediction problem: ''Given that we know to which range of age a subject belongs, how well can we predict his or her discourses?''.

This paper is organized as follows: in the second section, NSCA with its aims, SONSCA and the differences between the two techniques, were treated; in the third section, orthogonal polynomials and the method of decomposition were described; in the fourth section, the empirical study was explained, adding a description of participants of the research and the instrument used; in the fifth section, results were described and the discussion on the study was presented in the sixth section. The conclusion can be found in the last section.

2. NSCA AND SONSCA

NSCA is a useful tool for graphically depicting the relationship between variables when one is assumed to be logically dependent from the other one (Lombardo, Beh, D'Ambra, 2011). The aim of NSCA is predicting the value of a row variable from the categories of the other (column) variable, in other words the aim is the reduction of uncertainty about the value of the response variable from the knowledge of the predictor variable.This reduction of uncertainty can be measured with the Goodman and Kruskal's tau which is explicitly sensitive to the asymmetry of the situation (Goodman and Kruskal, 1954; Light and Margolin, 1971).

NSCA is primarily an exploratory tool, but the basic measures of increase in predictability can be tested as well. Its results can be fairly easily understood. NSCA involves decomposing the non-symmetric measure of association given by the tau index (Goodman, 1981; Light and Margolin, 1971), using generalized singular value decomposition (GSVD, Takane and Shibayama, 1991). When NSCA is used to study the association between categorical variables, it can perform poorly when dealing with ordinal structures (D'Ambra, 2017; Agresti, 2007; Barlow *et al.*, 1972).

In the case where both variables are ordered a special partition of the tau index (D'Ambra, Beh and Amenta, 2005), using what is referred to as bivariate moment decomposition, or BMD (Best and Rayner, 1996; Davy, Rayner, Beh *et al.*, 2003), can be considered (doubly ordered NSCA), while only one of the categorical variables exhibits an ordinal structure we refer to the methodology defined as singly ordered NSCA (SONSCA) using what is referred to as hybrid decomposition, or HD (Best and Rayner, 1996; Davy *et al.*, 2003). It is possible to find two kind of SONSCA: in the case where the predictor variable is ordered we refer to the analysis as SONSCA1, while the analysis for an ordinal-scale response variable is referred to as SONSCA2.

In particular, the methodology used in this paper is the SONSCA1, which will be called generic SONSCA, which allows combining the eigenvectors (calculated on the nominal variable) and the orthogonal polynomials (computed on the ordinal variable).It allows the user to visualize and identify the primary causes (if they exist) of the dependent relationship between the categories of two variables where one of them is assumed to be an ordinal categorical variable (Lombardo *et al.*, 2011).

In the next two subsections the theoretical formalization of NSCA and SONSCA are proposed.

2.1 *NSCA*

Let $\mathbf{N}$ be a two-way contingency table in which you consider the joint distribution of one ordinal categorical variable I (rows) and one nominal categorical variable J (columns), where the $(i,j)^{th}$ cell entry is given by the absolute frequencies n_{ij} for $i = 1, \ldots, I$ and $j = 1, \ldots, J$, where n is the total number of observations. Moreover, we define $n_{i+} = \sum_{j=1}^{J} n_{ij}$ and $n_{+j} = \sum_{j=1}^{J} n_{ij}$ the rows and columns marginal respectively. The $(i,j)^{th}$ element of the probability matrix $\mathbf{P}$ is defined as $p_{ij} = n_{ij}/n$ so that $\sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij} = 1$. Moreover, let $\mathbf{D}_I$ and $\mathbf{D}_J$ represent the diagonal matrices of row and column marginal probabilities p_{i+} and p_{+j}, respectively. We compute the dependent of column from row through the Goodman-Kruskal's tau index.

Let be $\mathbf{H}$ a matrix where the generic element, defined as $h_{ij} = [(p_{ij}/p_{i+}) - p_{+j}]$, for $i = 1, \ldots, I$ and $j = 1, \ldots, J$, is the difference the conditional predictor of j-th response given the joint proportion of predictor variable p_{ij}/p_{i+} and the inconditionl marginal proportion of the j-th response category p_{+j}. By using the ANOVA notation, the Total Sum of Squares (TSS) is equal to:

$$TSS = \frac{n}{2} - \frac{1}{2n}\sum_{j=1}^{J} n_{+j}^2 = \frac{n}{2} TSS^*$$

(where $TSS^* = (1 - \sum_{j=1}^{J} p_{+j}))$ and it can be decomposed according to the sum of the Within (group) Sum of Squares (WSS):

$$WSS = \sum_{i=1}^{I}\left(\frac{n_{i+}}{2} - \frac{1}{2n_{i+}}\sum_{j=1}^{J} n_{ij}^2\right) = \frac{n}{2} - \frac{1}{2}\sum_{i=1}^{I}\frac{1}{n_{i+}}\sum_{j=1}^{J} n_{+j}^2$$

and Between (group) Sum of Squares (BSS):

$$BSS = \frac{1}{2}\left(\sum_{i=1}^{I}\frac{1}{n_{i+}}\sum_{j=1}^{J} n_{ij}^2\right) = \frac{n}{2}BSS^*$$

(where $BSS^* = \sum_{i=1}^{I}\sum_{j=1}^{J} = \frac{p_{ij}^2}{p_{i+}}$). The TSS decomposition into several components allows us to measure the ''explained variation'' of the response variable which is attribute to a single factor of variation in a categorical framework. This measure is given by

$$R^2 = \frac{BSS}{TSS} = \frac{\left(\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{n_{i+}}\sum_{i=1}^{I} n_{ij}^2\right) - \frac{1}{n}\sum_{j=1}^{J} n_{+j}^2}{n - \frac{1}{n}\sum_{i=1}^{I} n_{+j}^2}$$

and represent a ''proportion of variation explained'' interpretation. This R^2 is identical to the measure of association. The Goodman-Kruskal (data) proposed an index for two categorical variables by considering:

$$\tau_{GK} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} p_{i+}h_{ij}^2}{1 - \sum_{j=1}^{J} p_{+j}^2} = \frac{num(\tau_{GK})}{1 - \sum_{j=1}^{J} p_{+j}^2} = \frac{BSS^*}{TSS^*}$$

The numerator of the Goodman-Kruskal index, $num(\tau_{GK})$, will be the focus of our discussion here, since the denominator is independent of the any of the joint cell proportions of the table $\mathbf{N}$. To determine the structure of the dependence between two categories variables, one may consider the NSCA. The NSCA attempts to predict column categories given row categories of a contingency table. Measure of predictive power rows on columns is given collectively by:

$$\mathbf{H} = \mathbf{D}_I^{-1}\mathbf{Q}_{D_I}^T\mathbf{P} = \mathbf{Q}_{D_I}\mathbf{D}_I^{-1}\mathbf{P} \tag{1}$$

where:

$$\mathbf{Q}_{D_I} = \mathbf{I} - \mathbf{1}(\mathbf{1}^T\mathbf{D}_I\mathbf{1})^{-1}\mathbf{1}^T\mathbf{D}_I$$

is the orthogonal projector onto the null space of $\mathbf{1}$ (the vector of ones) in metric $\mathbf{D}_I$. The matrix $\mathbf{Q}_{D_I}$ eliminates the row marginal effect from the entire relationship between rows and columns. The $ij - th$ element of $\mathbf{H}$ is equal to:

$$h_{ij} = \frac{p_{ij}}{p_{i+}} - p_{+j}$$

For dimension reduction, we may use GSVD of (1) with metric matrices $\mathbf{D}_I$ and $\mathbf{I}$.

This written as:

$$GSVD(\mathbf{H})_{D_I;I} \Rightarrow \mathbf{H} = \mathbf{UDV}^T \tag{2}$$

with $\mathbf{U}^T\mathbf{D}_I\mathbf{U} = \mathbf{I}$ and $\mathbf{V}^T\mathbf{V} = \mathbf{I}$?, where $\mathbf{U}$ and $\mathbf{V}$ are singular vectors of dimension $I \times M$ and $J \times M$, respectively, $\mathbf{D}$ is diagonal of the singular values d_m (arranged in descending order such that $d_1 \geq d_2 \geq \ldots \geq d_M \geq 0$) with $m = 1, \ldots, M$ and $M = rank(\mathbf{H})$. In particular:

$$||\mathbf{H}||^2_{D_I} = trace(\mathbf{H}^T\mathbf{D}_I\mathbf{H}) = \sum_{i=1}^{I}\sum_{j=1}^{J} p_{i+}h_{ij}^2 = \sum_{m=1}^{M} d_m^2 = BSS^*$$

The significance overall predictability $\mathbf{H}$ can be tested using the CATANOVA method (Light and Margolin, 1971).

$$C = (n-1)(J-1)\tau_{GK} = (n-1)(J-1)\frac{BSS^*}{TSS^*}$$

Under the zero predictability hypothesis (note that zero predictability also implied no association, i.e independence) $H_0 : p_{ij} = p_{+j}$. Light and Margolin (1971) showed that the C-statistic is asymptotically chi-squared distribution with $(I-1)(J-1)$ degree of freedom. In particular, Light and Margolin (1971), Onukogu (1985) and Singh (1996), introduced this statistic to obtain a variance analysis procedure for contingency tables, commonly referred to as CATANOVA.

2.2 *SONSCA*

Using Emerson's orthogonal polynomials it is possible to decompose the BSS quantity in different components, each of which represents a different power of the supposed relationship between row and column (linear, quadratic, etc.). The advantage of using orthogonal polynomials relies in the fact that the power information is considered in the analysis and the resulting scoring scheme allows a clear interpretation of the linear, quadratic or higher order trend components. In order to compute the Emerson's orthogonal polynomials for the rows of the matrix $\mathbf{P}$ we use the general recurrence formula (Emerson, 1968; Beh, 2001; Beh and Lombardo, 2014):

$$a_u = S_u[(s_I(i) - T_u)a_{u-1}(i) - V_u a_{u-2}(i)] \tag{3}$$

where

$$T_u = \sum_{i=1}^{I} p_{i+}s_I(i)a_{u-1}^2(i)$$

$$V_u = \sum_{i=1}^{I} p_{i+}s_I(i)a_{u-1}^2(I)a_{u-2}^2(i)$$

and

$$S_u = \left\{ \sum_{i=1}^{I} p_{i+} s_I(i)^2 a_{u-1}^2(I) - T_u^2 V_u^2 \right\}^{-\frac{1}{2}}$$

These polynomials are placed into a matrix $\mathbf{A}[I \times (I-1)]$ in which the general term is $a_u(i)$, for $u = 1, \ldots, I$. Let us assume that first column of $\mathbf{A} = \{a_0(i); i = 1, \ldots, I-1\}$ is trivial orthogonal polynomials whose elements are all equal to one and, also, that a $a_{-1}(i) = 0$ (for $i = 1, \ldots, I-1$). These polynomials are orthogonal with respect to the matrix $\mathbf{D}_I$:

$$\mathbf{A}^T \mathbf{D}_I \mathbf{A} = \mathbf{I}_{I-1}$$

The general recurrence relation of (3) is used when the column categories are ordered, to reflect this ordinal structure, a set of ordered column score $s_I(i)$ (for $i = 1, \ldots, I$) is used. Considering the matrices $\mathbf{A}$ and $\mathbf{V}$ $[J \times M]$ are the row orthogonal polynomials of matrix $\mathbf{P}$ according to (2) and right singular vectors of matrix $\mathbf{H}$ respectively, where $\mathbf{A}^T \mathbf{D}_I \mathbf{A} = \mathbf{I}_{I-1}$ and $\mathbf{V}^T \mathbf{V} = \mathbf{I}_M$. The total inertia BSS can be partitioned into location, dispersion and higher order components by means of the Hybrid Decomposition (HD) (Beh, 2001):

$$HD(\mathbf{H}) \Rightarrow \mathbf{H} = \mathbf{AZV}^T$$

where the matrix $\mathbf{Z}$ is obtained by:

$$\mathbf{Z} = \{z_{um}\} = \mathbf{A}^T \mathbf{D}_I^{\frac{1}{2}} \mathbf{HV}$$

and $BSS^* = \sum_{u=1}^{I-1} \sum_{m=1}^{M} z_{um}^2$ (D'Ambra *et al.*, 2005), this procedure is called Singly Ordered Non-Symmetrical Correspondence Analysis version 1 (SONSCA1). Therefore the marginal rows of the matrix $\bar{\mathbf{Z}} = \mathbf{Z}^2 \frac{(n-1)(J-1)}{TSS^*}$ we obtain the power component of the relationships to association columns from rows. Moreover, each component has a chi-squared distribution with appropriate order of degree of freedom. Using orthogonal polynomials, the i-th and j-th column category coordinates on the u-th and m-th axes can be defined by $\mathbf{F} = \mathbf{AZ}$ and $\mathbf{G} = \mathbf{VZ}^T$, so that:

$$BSS^* = trace(\mathbf{F}^T \mathbf{D}_I \mathbf{F}) = trace(\mathbf{G}^T \mathbf{G})$$

3. EMPIRICAL STUDY

3.1 *Participants*

We collected 40 semi-structured interviews: 5 to the users of a rehabilitative community, 20 to the users of a public health service for the treatment of addiction, 5 to the members of an Alcoholics Anonymous (AA) and 10 to the members of Gambling Anonymous (GA) group. Most of the respondents were male and older than 25 years; there was only one female member of the gambling anonymous group and no females were present among the users of the rehabilitative community. Gender, age and kind of addiction motivating the request of help are illustrated in the Table 1.

TABLE 1. - *Subjects' socio-anagraphic information*

Variables	Modalities	Gamblers n = 10	Alcoholics n = 5	Health Service n = 20	Community n = 5	Total
Gender	*Male*	9 (90%)	3 (60%)	12 (60%)	5 (100%)	29 (72,5%)
	Female	1 (10%)	2 (40%)	8 (40%)	0 (0%)	11 (27,5%)
Age	*18-25*	0 (0%)	0 (0%)	1 (5%)	1 (20%)	2 (5%)
	26-35	0 (0%)	0 (0%)	6 (30%)	2 (40%)	8 (20%)
	36-45	3 (30%)	2 (40%)	7 (35%)	1 (20%)	13 (32,5%)
	46-55	2 (20%)	3 (60%)	3 (15%)	1 (20%)	9 (22,5%)
	Over 55	5 (50%)	0 (0%)	3 (15%)	0 (0%)	8 (20%)
Addiction	*Alcohol*	0 (0%)	5 (100%)	2 (10%)	1 (20%)	8 (20%)
	Drug	0 (0%)	0 (0%)	13 (65%)	4 (80%)	17 (42,5%)
	Gambling	10 (100%)	0 (0%)	5 (25%)	0 (0%)	15 (37,5%)

This distribution reflects either the characteristics of the users of the Italian public health services for the treatment of addiction (ISTAT, 2015; ODP, 2011) and of the anonymous self-groups and socio-rehabilitative community that we met.

3.2 *The instrument*

An in-depth semi-structured open-ended interview was proposed to detect subjects' implicit theories about what addiction is, how can be explained and what motivates the request for help. The individual interviews were begun by asking for standard biographical information, followed by a more open-ended portion in which the subjects were asked to describe the way they choose their personal view of problem gambling or alcohol overuse and the reasons which motivate the search for help. The interview was composed of three questions:

- How do you describe your problem with..?
- How do you explain your problem with..?
- What does it motivate your request for help?

The interviews were exploratory, allowing respondents to touch on any topic meaningful to them, but the topic guide ensured that all relevant topics were covered.

In each context, the members were informed about the general aim of the research, the voluntary nature of participation and the anonymity of responses. In the case of social-rehabilitative communities and of the AA and GA self help-group this presentation was made in the opening of a weekly meeting by the research group, after that consent was obtained for an information session through the contact person in each group.

In the case of Italian public services for the treatment of substance and gambling addiction (SerT), the users give before the consent to the therapist to be contacted for the interview. No incentive was given.

The interviews were conducted by the authors with a team of research assistants. Each interview was conducted individually and took place in a room of the service or of the association (in the case of self-help groups). They lasted between 20 and 60 minutes, and all were tape-recorded. Before each individual interview, according to the ethical code of the Italian Psychology Association (AIP) (http://www.aipass.org/node/26) and the Italian Code on the protection of personal data (Legislative decree No 196/2003), each participant signed an informed consent both for the interview itself and for audio-taping.

3.3 *Data analysis*

In this section, we show the benefits of using orthogonal polynomials to graphically depict and summarize the association between ordinal/nominal variables. The analysis was conducted on a ''subjects' addiction thematic categories x range of age'' matrix, composed of 12 rows and 5 columns.

Specifically, the columns represent classes of age, which are 6, divided in this way: the first class is between 18 and 25; the second one is from 26 to 35; the third one goes from 36 to 45; the fourth is between 46 and 55; the last one concerns subjects over 55.

The 12 rows are represented by categories of words (those with more high frequency in the corpus obtained by the transcripts of the interviewees), gathered on the basis of the similarity of semantic or functional contents.

The addiction thematic categories which characterize the users' discourses about their being in treatment are detected as we can observe in the Table 2. Thus, the aim of our study here is to determine, considered some classes of age, how effective they can influence the type of words used by subjects.

4. RESULTS

Three tables can be described in order to understand the type of relation between variables. Table 3 summarizes the total value of C-statistic and P.value. Therefore since the C-statistic is 219.345, with a p-value of 0.000 and 44 degrees of freedom, there is evidence of an association between the addiction thematic categories used by subjects and ranges of age, a very strong, asymmetric, dependence structure between the variables.

TABLE 2. - *Addiction thematic categories*

Time dimensions	**Context of aid**	**Addiction**	**Daily life**	**Riflessive dimensions**	**Interation**
‘‘now’’-‘‘year’’	‘‘to help’’-‘‘help’’	‘‘to drink’’-‘‘addiction’’	‘‘friend’’-‘‘home’’	‘‘to understand’’-‘‘to think’’	‘‘to ask’’-‘‘to give’’
‘‘to begin’’-‘‘after’’	‘‘community’’	‘‘gambling’’-‘‘gamble’’	‘‘family’’-‘‘son’’	‘‘to know’’-‘‘sense’’	‘‘to say’’-‘‘to leave’’
‘‘day’’-‘‘to become’’		‘‘substance’’-‘‘problem’’	‘‘to work’’-‘‘work’’		‘‘to speak’’
‘‘never’’-‘‘month’’			‘‘wife’’-‘‘person’’		‘‘moment’’-‘‘before’’
		‘‘guy’’-‘‘situation’’			
‘‘always’’-‘‘time’’			‘‘life’’-‘‘to life’’		
‘‘end’’			‘‘money’’		
Passive position	**Loneliness**	**Losing**	**Willingness**	**Normative dimensions**	**Motivational dimensions**
‘‘to go’’-‘‘to arrive’’	‘‘lonely’’	‘‘to lose’’	‘‘to succeed’’	‘‘good’’-‘‘to must’’	‘‘to like’’-‘‘to find’’
‘‘to enter’’-‘‘to look at’’	‘‘nobody’’		‘‘to want’’	‘‘bad’’	‘‘to can’’
‘‘to pass through’’-‘‘to stay’’					
‘‘to come’’-‘‘to see’’					

TABLE 3. - *C-statistic and P.value*

C-Statistic	DoF	p-value
219.345	44	0.000

To determinate how the predictor categories vary in terms of their location, spread, and higher order moments, accross each of the axis, we can partition C-statistic into the z_{um} values, given in tables 4. By considering the z_{um} values in Table 4, the dominant source of variation for the first squared value in due to the linear component since:

$$d_1 = (0.029)^2 + (-0.013)^2 + (0.021)^2 + (-0.014)^2 = 0.001647$$

Similarly, the dominant source of principal inertia associated with the second axis is due to the quadratic (dispersion) component since:

$$d_2 = (-0.019)^2 + (-0.022)^2 + (0.002)^2 + (-0.013)^2 = 0.001018$$

To determine where, and if, significant sources of variation between the age categories exist, we consider the component values in the Table 5, that summarizes the decomposition of the C-statistic into location, dispersion, and higher-order components.

TABLE 4. - *Matrix of z_{um} values*

z_{um}	m=1	m=2	m=3	m=4
u=1	0.029	-0.019	-0.004	0.005
u=2	-0.013	-0.022	-0.004	-0.009
u=3	0.021	0.002	0.014	-0.008
u=4	-0.014	-0.013	0.016	0.006

Table 5 shows the first two dimensions account for 79.88% of the asymmetric association between the variables. The first axis accounts for the majority (49.347%) of this association, while the second axis explains 30.533% of the association. It may also be noted that the third and the fourth axes contribute to explain the 20.12% of the asymmetric association between the variables.

TABLE 5. - *Values of components*

C-Statistics	m=1	m=2	m=3	m=4	Total	% inertia of comp.	DoF	P. value
Location (u=1)	53.435	22.912	1.242	1.706	79.296	36.151	11	0.000
Dispersion (u=2)	11.716	32.185	1.174	5.158	50.232	22.901	11	0.000
Skewness (u=3)	29.738	0.336	12.084	4.448	46.606	21.248	11	0.000
Kurtosis (u=4)	13.352	11.538	15.784	2.537	43.211	19.700	11	0.000
Total	108.241	66.972	30.283	13.849	219.345	100	44	0.000
% inertia of axes	49.347	30.533	13.806	6.314	100			

It can be seen that the most dominant source of variation between the age categories, for the linear component, is due to the location component and skewness. That is, the variation between the column categories is dominated more by their variation in location than their spread (dispersion) of the profiles. The second axis (when $m = 2$) is largely dominated by the dispersion component of the variable and, to a lesser extent, by the dispersion component. For the third axis, only the third (skewness) and fourth (kurtosis) components are significant; the fourth axes (when $m = 4$) is more largely dominated by the dispersion and skewness components.

By observing that all the components have a zero p-value, we may conclude that there is a significant difference in the location and dispersion of the predictor categories. The zero p-value for the error component indicates that there are higher order moments that, while not dominating the variation of the predictor profiles, is still statistically significant (Lombardo, Beh and D'Ambra, 2007).

Moreover, in Table 5, we can observe the percentage of inertia explained by the different components, each of which represents a different power of the supposed relationship between row and column; the first four, from linear to quadratic, are considered. We can see that all components appear significant and the most dominant contribution to the total inertia of the data is due to the component associated with the linear polynomial of the columns.

To understand which column categories (ranges of age) are dominant contributors, we can calculate the absolute and percentage contribution of each predictor category to the statistic. It can be shown, in Table 5, that the dominant contributor is the linear component, which accounts for 36.151%, followed by the dispersion component and the cubic component, which respectively accounts for 22.901% and 21.248%. It may also be noted that the fourth component contributes for 19.700%.

The contribution of the row and column categories are reflected in the classical NSCA plot of Figure 2 and the SONSCA plot of Figure 2.

The NSCA plot (Figure 1) shows how the different classes of age predict the use of different thematic categories. The second, the third and the fourth classes of age (26-35; 36-45; 46-55) place near the following addiction thematic categories: ''Context of aid'', ''Motivational Dimension'', ''Willing'', ''Loneliness'', ''Losing'', ''Reflexive Dimension'' and ''Temporal Dimensions''; thus the above mentioned classes of age seem to predict discourses focused on own being in treatment, and related to the will to reflect on feeling (loneliness) and critical event (losing) in motivating the request for help. The discourses appear to be situated in the time; who speak choose to talk about what is happened before and what happens now to represent their problem and the will to be helped.

''Normative Dimensions'' and ''Daily life'' categories are shared from both the users from 36 to 55 years (second and third class of age) and the younger users (from 18 to 25; first class of age). The closeness of the two categories suggest discourses about what is (or what was) good and what bad in own life. However, while the discourses of the users from 36 to 55 integrate the focus on these aspects with a reflexive attitude toward emotions, feeling and reasons of change, the discourses of the youngest users suggest a more passive position; the subject appears a little child, without apparent feelings, desires, or purposes mediating his arriving, staying or going out the relationships he talks about, but a normative dimensions splitting the world in good and bad things that have to be done ot not.

The fifth class of age (over 55) predicts the use of the thematic categories ''Addiction'' and ''Interaction with the others'', as if the relational dimension of own life revolved around the addictive object, which seems to define the reason and the ultimate purpose of our action and interaction with other people.

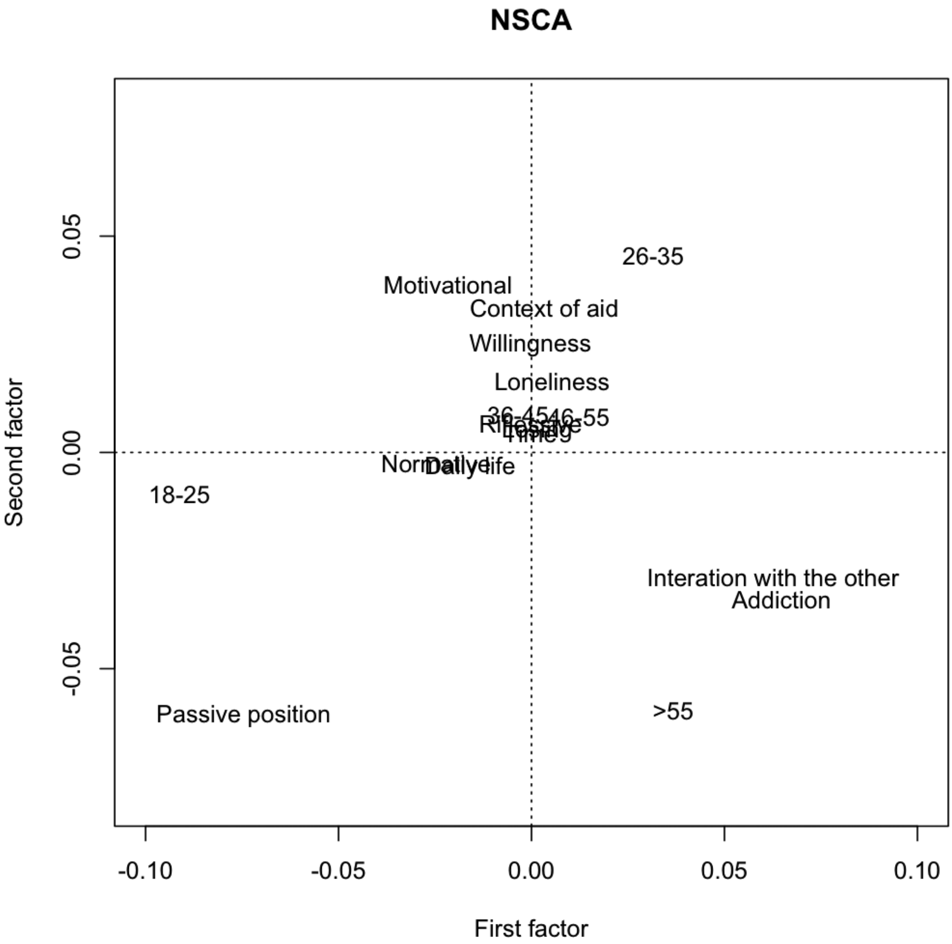

FIGURE 1. - *NSCA*

Figure 2 shows the singly ordered correspondence plot obtained by performing SONSCA1. In this case too, we can see how different classes og age predict different kind of discourses, but it shows also some differences compared to the NSCA plot.

The subjects belong to the first class of age (18-25) appear now to predict the use of ''Motivation category''; the belonging to the second class (26-35) predicts the use of words suggesting a ''Passive position'' toward events and circumstances.

Discourses of the users from 36 to 55 (third and fourth class of age) appear to be more closed to the ''Normative Dimensions'' and ''Daily life'' categories, and more distant from the ''Context of aid'' category (compared to NSCA output).

The reference to what was is good and what is bad in their own life goes together with discourses on how addiction organized their interaction with the other and feeling of loneliness and losing.

Finally, subjects over 55 appear to be characterized by the use of words referring to the context of aid and the request for help.

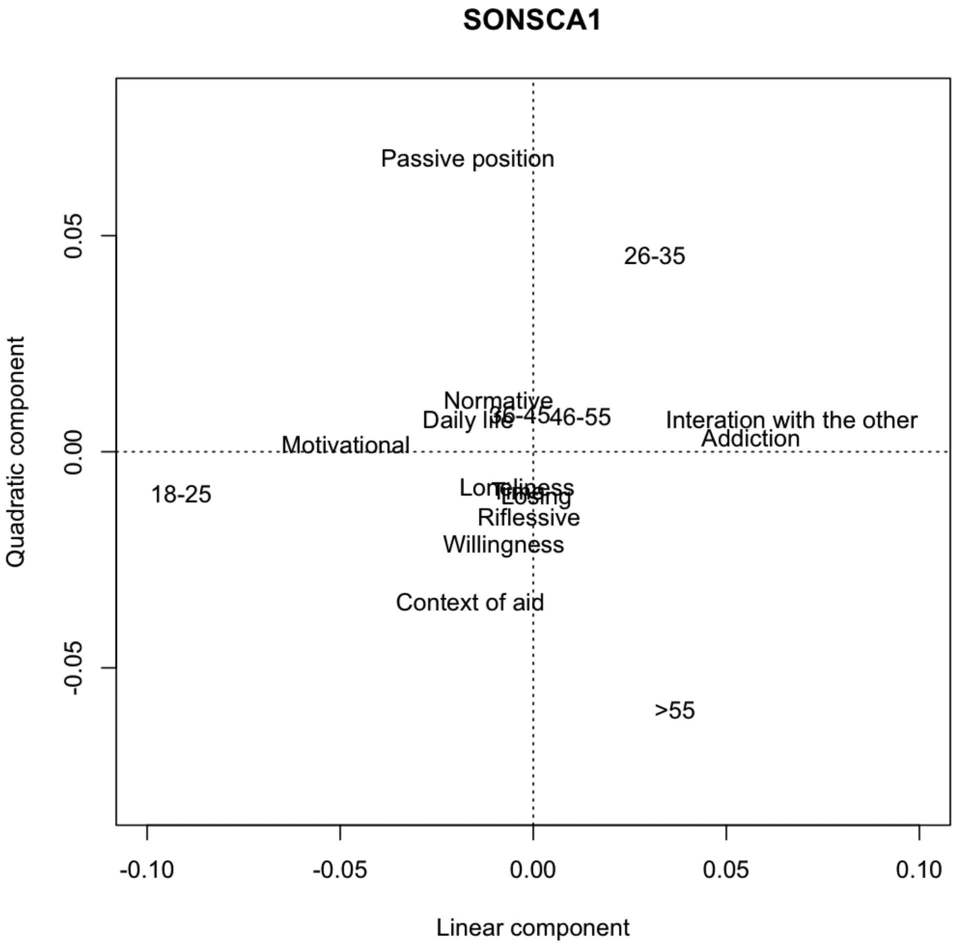

FIGURA 2. - *SONSCA*

5. DISCUSSION AND CONCLUSIONS

The focus of this paper has been on how NSCA and SONSCA can investigate the dependence structure between addiction thematic categories and classes of age. Recent papers that describe the use of orthogonal polynomials for correspondence analysis, have shown how they can be an important tool for identifying sources of association in two-way contingency tables with one (Beh, 2001) or two ordinal variables (Beh, 1997, 1998). We have showed how NSCA and SONSCA can be fruitfully used to analyze data sets arising from psychological studies.

Both the methods allow us to observe how different classes of age of users predict the use of different addiction thematic categories to talk about their own problem and request for help. Furthermore, in both the cases, users from 36 to 45 and from 46 to 55, compared to the others, appear to be likely to make use of words which refer to a reflexive process on what happened in owns life and related feeling.

At the same time, NSCA and SONSCA plots show some differences in the relationship among addiction thematic categories and between thematic categories used by subjects and class of age. Discourses of the younger users appear to be characterized by the reference to normative dimensions in the NSCA plot, by the reference to motivational dimensions in the plot generated by the SONSCA. Discourses of the older users appear to be focused on the relationship between addiction and interaction with others in the NSCA plot, on the context of aid in the SONSCA plot. Passive position appears to be dominant among the users from 26 to 35 in the SONSCA plot.

Differences related to the age maybe reflect cultural differences in the way different generations make sense of their addiction (Chassin, Presson, Rose and Sherman, 2007), different kind of discourses exchange within family or health services (Reinarman, 2005), or also different years of attendance of the service.

Unfortunately, we do not know how long users have been in treatment, but it is reasonable that the discourses of the ‘‘middle age classes’’ users are characterized by more reflexive dimensions, compared to the younger users, not only because of the age, but because these users attend the service and benefits the help the longest. In the case of the users over 55, we can guess that they have behind them a long history of addiction; discourses (focused on the addiction or the context of aid) may reflect the chronicity of the relationship with the substance and with services (Dennis, Scott, Funk and Foss, 2005).

Prior research provides little information about what kind of understanding people in treatment develop about their own problems and what reasons motivate the attendance of a care program. To our knowledge, no prior study investigated (dis)similarities in the discourses produced by users of different ages.

Our study suggests that age of users should be taken in account for a better understanding of the meaning which feed the request for help. Future works will investigate into the results of the technique to analyze a nominal-ordinal contingency table, using both of decompositions (Lombardo *et al.*, 2007): bivariate moment decomposition and singular value decomposition.

REFERENCES

Agresti A. (2007). *An introduction to categorical data analysis*, 2nd ed. Hoboken. NJ, Wiley-Interscience.

Barlow R., Bartholomew D.J., Bremner J.M., Brunk H.D. (1972). *Statistical inference under order restrictions; the theory and application of isotonic regression*. NJ, Wiley-Interscience.

Beh E.J., Lombardo R. (2014). *Correspondence analysis: theory, practice and new strategies*. John Wiley and Sons.

Beh E.J. (1997). Simple correspondence analysis of ordinal cross-classifications using orthogonal polynomials. *Biometrical Journal*, **39**(5), 589-613.

Beh E.J. (1998). A comparative study of scores for correspondence analysis with ordered categories. *Biometrical Journal*, **40**(4), 413-429.

Beh E.J. (2001). Confidence circles for correspondence analysis using orthogonal polynomials. *Advances in Decision Sciences*, **5**(1), 35-45.

Best D., Rayner J. (1996). Nonparametric analysis for doubly ordered two-way contingency tables. *Biometrics*, 1153-1156.

Chassin L., Presson C.C., Rose J., Sherman S.J. (2007). What is addiction? Age-related differences in the meaning of addiction. *Drug and alcohol dependence*, **87**(1), 30-38.

Ciavolino E., Redd R., Evrinomy A., Falcone M., Fini V., Kadianaki I., Kullasepp K., Mannarini T., Matsopoulos A., Mossi P. *et al*. (2017). Views of context. an instrument for the analysis of the cultural milieu. A first validation study. *Electronic Journal of Applied Statistical Analysis*, **10**(2), 599-628.

D'Ambra A. (2017). Cumulative correspondence analysis using orthogonal polynomials. *Communications in Statistics-Theory and Methods*, **46**(6), 2942-2954.

D'Ambra L., Lauro C. (1989). Non symmetrical analysis of three-way contingency tables. In R. Coppi and S. Bolasco (Eds.), *Multiway data analysis*, (pp. 301-315), North-Holland, Amsterdam.

D'Ambra L., Beh E.J., Amenta P. (2005). Catanova for two-way contingency tables with ordinal variables using orthogonal polynomials. *Communications in Statistics Theory and Methods*, **34**(8), 1755-1769.

Davy P., Rayner J., Beh E. *et al*. (2003). Generalised correlations and simpsons paradox. *Current Research in Modelling, Data Mining and Quantitative Techniques*, 63-73.

Dennis M.L., Scott C.K., Funk R., Foss M.A. (2005). The duration and correlatesvof addiction and treatment careers. *Journal of substance abuse treatment*, **28**(2), S51-S62.

Emerson P.L. (1968). Numerical construction of orthogonal polynomials from a general recurrence formula. *Biometrics*, 695-701.

Fisher R.A. (1940). The precision of discriminant functions. *Annals of Human Genetics*, **10**(1), 422-429.

Goodman L.A. (1981). Association models and canonical correlation in the analysis of cross-classifications having ordered categories. *Journal of the American Statistical Association*, **76**(374), 320-334.

Goodman L.A., Kruskal W.H. (1954). Measures of association for cross-classification. *Journal of the American Statistical Association*, **49**(268), 732-764.

ISTAT (2015). Relazione annuale al Parlamento sullo stato delle tossicodipendenze in Italia. Dipartimento Politiche Antidroga.

Kamalja K.K., Khangar N.V. (2017). Multiple correspondence analysis and its applications. *Electronic Journal of Applied Statistical Analysis*, **10**(2), 432-462.

Light R.J., Margolin B.H. (1971). An analysis of variance for categorical data. *Journal of the American Statistical Association*, **66**(335), 534-544.

Lombardo R., Beh E.J. (2016). Variants of simple correspondence analysis. *R Journal*, **8**(2), 167-184.

Lombardo R., Beh E.J., D'Ambra L. (2007). Non-symmetric correspondence analysis with ordinal variables using orthogonal polynomials. *Computational Statistics and Data Analysis*, **52**(1), 566-577.

Lombardo R., Beh E.J., D'Ambra A. (2011) Studying the dependence between ordinal-nominal categorical variables via orthogonal polynomials. *Journal of Applied Statistics*, **38**(10), 2119-2132.

ODP (2011). Primo rapporto sulle dipendenze patologiche. Sanit Puglia.

Onukogu I.B. (1985). Reasoning by analogy from anova to catanova. *Biometrical Journal*, **27**(8), 839-849.

Reinarman C. (2005). Addiction as accomplishment: The discursive construction of disease. *Addiction Research and Theory*, **13**(4), 307-320.

Singh B. (1996). On catanova method for analysis of two-way classified nominal data. *Sankhya: The Indian Journal of Statistics, Series B*, **58**(3), 379-388.

Takane Y., Shibayama T. (1991). Principal component analysis with external information on both subjects and variables. *Psychometrika*, **56**(1), 97-120.