**Multiple correspondence analysis for customer segmentation in large retail groups**
By Massari, Manca, Girone

# Multiple correspondence analysis for customer segmentation in large retail groups

Antonella Massari[*a], Fabio Manca[b], and Francesco Girone[a]

[a]*Department of Business and Law Studies, University of Bari "Aldo Moro"*
[b]*Department of Education Science, Psychology, Communication Science, University of Bari "Aldo Moro"*

The aim of the following work is to highlight the importance of categorical methods as applied to marketing strategies. In the past, multivariate statistical techniques were used for quantitative data in marketing decision support systems (MDSS), however, given the introduction of many categorical variables in present day marketing research, the elaboration of these variables requires the use of categorical statistical methods.
In order to carry out this research, we have employed multiple correspondence analysis, which, by means of maps constructed on a limited number of latent dimentions, simplifies the reading of both the intricate relations among the numerous categorical variables observed and their categories. In market research on the purchasing behavior of consumers, these analyses have been used to determine the essential aspects of consumer behavior as a rational basis for adopting opportune marketing strategies.

**keywords:** multiple correspondences analysis, market research.

## 1 Introduction

The primary aim of this study is to underline the efficacy of the Multiple correspondence analysis, which represents an optimal choice when considering marketing strategies, particularly regarding customer segmentation (De Luca, 1998). On the basis of

_____

[*]Corresponding author: antonella.massari@uniba.it

extensive research, we have sought to determine the key purchasing drivers encouraging consumers to choose the Mongolfiera Ipercoop shopping centers in the Apulia region, Italy, identifying spending ratios, and analysing the product characteristics of the purchases considered.

To attain these aims, we have employed multiple correspondence analysis which, by means of maps constructed on a limited number of latent variables, simplifies the reading of both the intricate relations among the numerous categorical variables observed and their categories. In market research on the purchasing behavior of consumers, this analysis has been used to determine the essential aspects of consumer behavior as a rational basis for adopting appropriate marketing strategies.

## 2  Sample group

The sample group is represented by 1.000 Ipercoop customers, who responded to questionnaires distributed at the entrance of the mall, which were then collected at the exit. The interviews were carried out at the exits of 6 Mongolfiera Ipercoop malls in Apulia (Bari Santa Caterina, Bari Japigia, Molfetta, Taranto, Brindisi and Lecce).

In order to preserve the salient characteristics of the observed population, random criteria were used in the selection of statistical units, employing the constant systematic sampling technique upon customer entry to the malls in question. The representative samples were collected in the periods of weekdays, public holidays and pre-holidays in the months of May and June 2015.

The given size of the sample (1.000) is sufficient to guarantee an a priori error estimation of 3.5% with a confidence level of 95%.

The anagraphical variables analysed in the sample considered gender, age, and profession, 58.8% of those interviewed being female and 41.2% male. 26.0% of the interviewees belong to the age group of up to 30 years old, 25.5% belong to the over-60 group, with 24.5% falling between the limits of 51 and 60, and 24.0% in the 31-50 age group. The largest categories, according to professional status, are those of pensioners and the self-employed, registering the same percentage of 16.4% each, followed by housewives (13.1%), freelance professionals (12.0%), employees (10.0%), unemployed (9.1%), students (8.1%), executives (7.7%), and other (7.2%).

Further variables under consideration include purchasing average expenditure, the characteristics of the products purchased and the relevant purchasing drivers. These variables allow us carry out an in-depth examination of consumer purchasing behaviour through an appropriate utilization of multiple correspondence analysis. Considering the average expenditure of each purchase: 47.5% of the customers spent between 30 and 59 euros, 24.3% between 60 and 89 euros, 15.5% until 29 euros, with 12.7% spending 90 euros and over.

Regarding the purchasing drivers for our sample group: 28,0% of purchases were driven by discounts, 25.9% by price, 22.2% by brand, with a very small percentage driven by other product characteristics such as nutritional value, product origin, whether or not the product was organic, in season, or new and/or improved, and finally, packaging.

The key drivers leading our consumers to opt for Ipercoop malls were subdivided as follows: 40.8% were motivated by the satisfaction of getting good value for money, 13.2% to follow the current trend, 12,0% to be a member of a pre-established group, 10.8% desired freedom of choice, and 10,0% self-gratification. Percentages were low for the motivational drivers: the desire to be alone, to communicate, and to participate in social interaction.

## 3 Multiple correspondence analysis

Multiple correspondence analysis is the extension of correspondence analysis to more than two variables, making it possible to study the structure of relationships existing in a set of categorical variables. Its primary function is to interpret the associations among the categories of those variables by means of their projection in a space with a reduced number of dimensions (almost always two or three).

Any survey leads to a double-entry table, the *observations table*, in which the correspondence of the different statistical units, generally shown in lines, and the different variables observed, generally shown in columns, are reported the categories indicated by the respondents.

The observations table is transformed into a wider table, also double-entry, known as the complete disjunctive table, where, in correspondence with each statistical unit in the lines, and with each category of the observed variables in the columns, the value 1 is attributed if the said category has been indicated by the statistical unit, and the value 0 if it has not been indicated. The data matrix of the complete disjunctive table is generally indicated by **Z**.

The matrix product of the transposed **Z'** of the matrix **Z** for the same matrix **Z** is a symmetrical square matrix in which both the lines and columns indicate the categories of all the variables. This matrix. know as the *Burt matrix* and indicated by **B**, contains all the contingency tables that can be obtained by intersecting the categories of the starting variables two-by-two, including the diagonals, of each variable with itself.

Multiple correspondence analysis, which can be carried out by utilizing either a dissociative table or the Burt matrix, leads to the determination of a number of latent dimensions, orthogonal to each other, in number equal to the rank of the data matrix employed. The total variability of the table is denominated as *total variance* or *total inertia* and is equal to the sum of the inertias of the dimensions. The passage from the categories of the original variables to the dimensions makes it possible to obtain a lower number of orthogonal variables. The successive dimensions express decreasing quotas of the total inertia, so that only a few are necessary, in almost all cases 2 or 3, to concisely and effectively represent the most relevant quotas of the relationships among the categories of the observed variables. Projecting the categories on the axes of the dimensions produces simple representations, facilitating interpretation, and more specifically, by limiting attention to two dimensions, it is easy to consider the proximity or distance between the different categories of the variables as association or disassociation.

Multiple correspondence analysis involves complex processing that can be carried out

using specific softwares such as the Categories module of SPSS (*Statistical Package for Social Sciences*) as utilized in this study (Clausen, 1998).

## 4 Preliminary results of the analysis

Multiple correspondence analysis was applied to the following Q = 6 variables: gender, age, profession, "*How much do you spend on average every time you make purchases at Ipercoop?*", "*Which of the following characteristics do you consider at the moment of purchasing a product?*", "*Which of the following motivations lead you to go to a Ipercoop or a malls?*" corresponding to a total of $K = 36$ categories. It is of note that the first three variables allow us to determine the characteristics of the respondents, while the last three concern their purchasing behavior (Di Franco, 2013).

The complete dissociative table of the survey data has allowed to obtain $K - Q = 30$ singular values therefore the same number of eigenvalues $\lambda_l, l = 1, 2, \ldots, 30,$. This analysis has been conducted considering the eigenvalues higher than their average, calculating the corrected quotas of inertia of the corresponding dimensions by utilizing the Benzecri criteria, where the variance quota of dimension $l$ is:

$$\tau_l = \frac{Q\lambda_l}{K - Q} \tag{1}$$

and the average of eigenvalues is:

$$\bar{\lambda} = \frac{1}{Q} \tag{2}$$

Due to the high dimensionality of the clouds, the variance quotas of the first dimensions are usually quite low. In order to better appreciate the importance of the first dimensions, Benzecri (1979, p.378 and 1992, p.412) proposed to use *modified or corrected variance quotas.*

For: $l = 1, 2, \ldots, l_{max}$ so that: $\lambda_l > \bar{\lambda}$,
we calculate:

1) the pseudoeigenvalues:

$$\lambda_l' = \left(\frac{Q}{Q-1}\right)^2 (\lambda_l - \bar{\lambda})^2; \tag{3}$$

2) The sum:

$$S = \sum_{l=1}^{l_{max}} \lambda_l' \tag{4}$$

therefore: for $l < l_{max}$ the corrected quotas are:

$$\tau_l' = \frac{\lambda_l'}{S} \tag{5}$$

(Le Roux and Rouanet, 2010, p.41).

Table 1 shows Cronbach's indexes, the eigenvalues and the corrected quotas of inertia for the first three dimensions. The values of the Cronbach indexes highlight the validity of the dimensions considered. In particular, the first two dimensions explain almost two-thirds and the first three more than three-quarters of the total inertia. Therefore, in what follows, we will limit ourselves to consider the results of the analyses for the first two dimensions. These results are presented with the measure of discrimination for each variable.

Table 1: Synthetic results of the correspondence analyses of the survey of Apulian Ipercoop customers in 2015

| Dimension | Cronbach's Alpha | Eigenvalue | Explained variance | |
| | | | Correct inertia (%) | Cumulated correct inertia (%) |
| --- | --- | --- | --- | --- |
| 1st | 0.628 | 0.350 | 0.347 | 0.347 |
| 2nd | 0.580 | 0.320 | 0.249 | 0.596 |
| 3rd | 0.483 | 0.280 | 0.131 | 0.727 |

The measurement of the discrimination of the variables adds an important contribution to the interpretation of the analyses. The greater the value of discrimination associated with a variable, the better that dimension is able to discriminate between the categories of that variable.

Table 2 presents the measures of discrimination. In our case, the professional status has a high discrimination value of 0.771 with the first dimension and a high value of 0.602 for the second, hence the categories are well discriminated by the first two dimensions. The professional status is followed by that of the motivations to go to Ipercoop with a discrimination value of 0.711 for the first dimension and 0.382 for the second. The other variables exhibit lesser discrimination values.

Table 2: Measures of discrimination of the variables

| Variables | Dimension 1st | Dimension 2nd | Vector lenght |
|---|---|---|---|
| Gender | 0.124 | 0.035 | 0.128 |
| Age | 0.130 | 0.522 | 0.538 |
| Professional status | 0.771 | 0.602 | 0.978 |
| How much do you spend on average every time you make purchases at Ipercoop? | 0.149 | 0.153 | 0.214 |
| Which of the following characteristics do you analyse at the moment of purchasing a product? | 0.215 | 0.236 | 0.319 |
| Which of the following motivations lead you to go to a Ipercoop? | 0.711 | 0.382 | 0.807 |

The vector length of the variables is associated with the high capacity of vector discrimination. The graph in Figure 1 clearly confirm the results shown in the previous table, namely, that the variables "profession" and "motivations", which exhibiting considerable vector length, are closely linked with the two dimensions, differently from the gender variable which assumes a notably shorter vector than the other vector streams.

## 5 Results of multiple correspondence analysis on single variable

To present clearly and to interpret effectively the results of the multiple correspondence analysis applied to the survey of customers of the mall Ipercoop, in this section we consider the results obtained from the said analysis concerning single variables, in next section we consider the aspects regarding the relations between each personal variable of the customers and their buying behaviour.

We begin by examining the variables describing interviewee characteristics.

Table 3 shows the results for the multiple correspondence analysis for the gender variable. Firstly, the prevalence of females with respect to males (6 compared to 4 out of 10). On the plane of the two dimensions considered (Figure 2), the points of the two genders are close together and are close to the barycenter. As a result, despite their masses, the two genders are only slightly discriminated by the two dimensions, a fact which also emerges from the limited values of the discrimination measures, regarding both the first dimension and even more so in the case of the second. The contribution of the gender variable to overall inertia and its two categories is very low. The squared cosines indicate the scarce quality of the coordinates of the gender points projected on the axes of the dimensions.
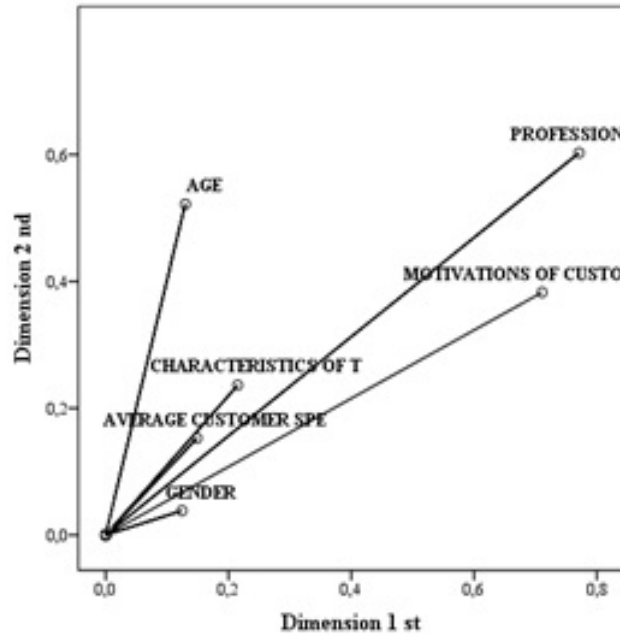
Figure 1: Measure of variable discrimination

Table 3: Multiple correspondence analysis results for the gender variable and its categories

| Gender | Percentage frequency | Coordinates | | Contribution % | | Squared cosines | | Measures of discrimination | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1st | 2nd | 1st | 2nd | 1st | 2nd | 1st | 2nd |
| Female | 58.8 | 0.295 | -0.164 | 2.5 | 0.8 | 0.124 | 0.037 | 0.051 | 0.015 |
| Male | 41.2 | -0.421 | 0.234 | 3.6 | 1.1 | 0.124 | 0.037 | 0.073 | 0.020 |
| Total | 100.0 | | | 6.1 | 1.9 | | | 0.124 | 0.035 |

Table 4 shows the results of the multiple correspondence analysis for the age variable. Firstly, it should be noted that the four age groups considered here register as having approximately the same weight. On the plane of the two dimensions considered (Figure 3), the points of the two categories of the under 30s and the 60s and older are distant from each other and from the barycenter. In addition, the younger group presents positive coordinates on both dimensions, while the older age group presents negative coordinates. As such, the two extreme age groups are markedly discriminate, as highlighted by the discrimination measures. The points for the central age groups present coordinates of a certain interest in only one dimension and almost none in the other, resulting in almost

negligible discriminating values. The contribution to overall inertia for the age variable is considerable (27.9%); followed by the contribution to the second dimension of the oldest age group (18.4%). The quality of that category on the second dimension also appears considerable as shown by the relative squared cosine. In synthesis, the plane of the two dimensions clearly highlights the contrast between the young and the old age groups, as well as the importance on the second dimension of the 60s and over age group from all points of view: coordinates, contribution, quality and discriminating value.

Table 4: Correspondence analysis indicators for the age variable and its categories

| Age classes (years) | Percentage frequency | Coordinates | | Contribution % | | Squared cosines | | Measures of discrimination | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1st | 2nd | 1st | 2nd | 1st | 2nd | 1st | 2nd |
| Under 30s | 26.0 | 0.456 | 0.506 | 2.7 | 3.3 | 0.073 | 0.086 | 0.054 | 0.067 |
| 30 - 49 | 24.0 | 0.179 | 0.132 | 0.4 | 2.1 | 0.010 | 0.005 | 0.008 | 0.004 |
| 50 - 59 | 24.5 | -0.140 | 0.584 | 0.2 | 4.1 | 0.006 | 0.114 | 0.005 | 0.084 |
| 60s and over | 25.5 | -0.499 | -1.201 | 3.1 | 18.4 | 0.085 | 0.491 | 0.063 | 0.368 |
| Total | 100.0 | | | 6.4 | 27.9 | | | 0.130 | 0.522 |

Table 5 shows the results of the analysis of the professional status variable. This variable has nine categories, each of which has a different weight: this goes from the 7.7% of executives, to the 16.4% of the self-employed and pensioners, with the incidence of housewives (13.1%) and employees (10.0%) also registering as relevant. The categories furthest from the barycenter are housewives, pensioners and executives, with the other categories clustering closer to the barycenter. On the plane of the two dimensions considered (Figure 4), the housewives and pensioners categories are those with the highest measures of discrimination, particularly on the first and second dimensions, respectively. This is not the case for the executives, for whom the frequency is lower. In synthesis, the first dimension contrasts housewives and pensioners, the second dimension contrasts those two categories and the executives. The contribution of the professional status variable to the overall inertia is quite considerable on both dimensions; 38.5% and 30.1% respectively, with the housewives contributing particularly in the first dimension (24.9%), and pensioners in the second dimension (15.5%). Regarding the quality of the category coordinates, here too, housewives emerge in the first dimension with pensioners in the second.

Table 5: Correspondence analysis indicators for the professional status variable and its categories

| Professional status | Percentage frequency | Coordinates | | Contribution % | | Squared cosines | | Measures of discrimination | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1st | 2nd | 1st | 2nd | 1st | 2nd | 1st | 2nd |
| Executive | 7.7 | -0.538 | 1.231 | 1.1 | 5.8 | 0.024 | 0.131 | 0.022 | 0.117 |
| Freelance professional | 12.0 | -0.415 | 0.505 | 1.0 | 1.5 | 0.023 | 0.033 | 0.021 | 0.031 |
| Self-employed | 16.4 | -0.088 | 0.209 | 0.1 | 0.4 | 0.002 | 0.009 | 0.001 | 0.007 |
| Employee | 10.0 | -0.333 | 0.293 | 0.6 | 0.4 | 0.012 | 0.010 | 0.011 | 0.009 |
| Housewife | 13.1 | 1.950 | -0.769 | 24.9 | 3.9 | 0.573 | 0.088 | 0.498 | 0.077 |
| Pensioner | 16.4 | -0.979 | -1.374 | 7.9 | 15.5 | 0.188 | 0.369 | 0.157 | 0.310 |
| Student | 8.1 | 0.720 | 0.453 | 2.1 | 0.8 | 0.046 | 0.017 | 0.042 | 0.017 |
| Unemployed | 9.1 | 0.202 | 0.591 | 0.2 | 1.6 | 0.004 | 0.034 | 0.004 | 0.032 |
| Other | 7.2 | -0.454 | -0.211 | 0.7 | 0.2 | 0.016 | 0.003 | 0.015 | 0.003 |
| Total | 100.0 | | | 38.5 | 30.1 | | | 0.771 | 0.602 |

We now pass to the variables expressing the interviewees' behavior.
Table 6 shows the results of the multiple correspondence analysis for the average expenditure variable. The distribution of the interviewees according to their average expenditure is unimodal, with slightly less than half of the frequencies concentrated in the central group (47.5%), close to another quarter in the 60-90 euro group, about a sixth in the lowest expenditure group and lesser frequencies for the highest expenditure groups.
On the plane of the two dimensions considered (Figure 5), the points of the two categories furthest from the barycenter are those of the highest and lowest spending classes, while the intermediate spending classes are closer to the barycenter. It is to be noted that on the first dimension the youngest age class is close to the lowest spending class and is the only one with both positive coordinates, so that dimension contrasts the lowest spending class with all the others. On the second dimension the highest coordinates are those of the two highest spending classes. There is little discrimination by the two dimensions in the case of all the spending classes, including the central ones despite their high frequency, as is shown by the modest discrimination measures for both dimensions. Both the average spending variable and its categories make a very limited contribution to the overall inertia. All the squared cosines indicate a scarce quality of the coordinates of the category points projected on the axes.

Table 6: Correspondence analysis indicators for the average spending variable and its categories

| Average spending (€) | Percentage frequency | Coordinates | | Contribution % | | Squared cosines | | Measures of discrimination | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1st | 2nd | 1st | 2nd | 1st | 2nd | 1st | 2nd |
| Less than 30 | 15.5 | 0.852 | 0.135 | 5.6 | 0.1 | 0.133 | 0.003 | 0.112 | 0.003 |
| 30 - 59 | 47.5 | -0.054 | -0.230 | 0.1 | 1.3 | 0.003 | 0.048 | 0.001 | 0.025 |
| 60 - 89 | 24.3 | -0.303 | -0.123 | 1.1 | 0.2 | 0.030 | 0.005 | 0.022 | 0.004 |
| 90 - 119 | 6.0 | -0.447 | 0.626 | 0.6 | 1.2 | 0.013 | 0.025 | 0.012 | 0.024 |
| 120 and above | 6.7 | -0.084 | 1.206 | 0.0 | 4.9 | 0.001 | 0.107 | 0.000 | 0.097 |
| Total | 100 | | | 7.4 | 7.6 | | | 0.149 | 0.153 |

Table 7 shows the results of the multiple correspondence analysis for the variable characteristics of the product to buy. Of the nine categories of this variable, according to the interviewees' replies, only three have a considerable frequency: discounts and special offers (28.0%), price (25.9%) and brand (22.2%). On the plane of the two dimensions considered (Figure 6), the points of the product characteristic categories indicated by the interviewees that are furthest from the barycenter are, in order: organic nature and typicality, nutritional value, packaging and seasonality. These are also the least frequent categories. On the contrary, the characteristics of the products most frequently indicated by the interviewees are close to the barycenter. As a result the three most frequently indicated categories, despite their mass, are not greatly discriminated by the two dimensions, a fact which emerges also from the low measures of discrimination for both dimensions. The product characteristics variable and its categories make a very limited contribution to the overall inertia. The squared cosines indicate also a scarce quality of the coordinates of the category points projected on the axes of the two dimensions.

Table 7: Correspondence analysis indicators for the product characteristics variable and its categories

| Characteristics of the product to buy | Percentage frequency | Coordinates | | Contribution % | | Squared cosines | | Measures of discrimination | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1st | 2nd | 1st | 2nd | 1st | 2nd | 1st | 2nd |
| Organic nature and typicality | 2.5 | -0.161 | 1.617 | 0.0 | 3.3 | 0.001 | 0.070 | 0.001 | 0.065 |
| Packaging | 3.6 | -0.044 | 1.262 | 0.0 | 2.9 | 0.000 | 0.058 | 0.000 | 0.057 |
| Brand | 22.2 | -0.275 | 0.225 | 0.8 | 0.6 | 0.022 | 0.016 | 0.017 | 0.011 |
| Novelty | 5.3 | 0.476 | 0.538 | 0.6 | 0.8 | 0.013 | 0.015 | 0.012 | 0.015 |
| Price | 25.9 | -0.262 | -0.180 | 0.9 | 0.4 | 0.024 | 0.012 | 0.018 | 0.008 |
| Source / origin | 2.4 | 0.646 | -0.251 | 0.5 | 0.1 | 0.010 | 0.002 | 0.010 | 0.002 |
| Discounts / offers | 28.0 | -0.107 | -0.358 | 0.2 | 1.8 | 0.004 | 0.050 | 0.003 | 0.036 |
| Seasonality | 5.9 | 1.067 | 0.417 | 3.4 | 0.5 | 0.071 | 0.011 | 0.067 | 0.010 |
| Nutritional value | 4.2 | 1.444 | -0.859 | 4.4 | 1.5 | 0.091 | 0.033 | 0.088 | 0.031 |
| Total | 100.0 | | | 10.8 | 11.8 | | | 0.215 | 0.236 |

Table 8 shows the results of the multiple correspondence analyses for the variable "motivation for shopping at the malls". The frequencies for the seven categories are quite differentiated, those most often indicated by the interviewees are: to get a bargain (40.8%), to follow fashion (13.2%), to spend time with friends (12.0%); the other categories are indicated much less frequently. On the plane of the two dimensions considered (Figure 7) the points of the seven categories are on average distant from the barycenter, those furthest away are: to follow fashion, communication and social interaction. The motivations variable makes quite a considerable contribution to the overall inertia:

35.5% on the first dimension and 19.1% on the second. The follow fashion category makes a very high contribution to the first dimension and is strongly discriminated on that dimension presenting a value of excellent quality.

Table 8: Correspondence analysis indicators for the motivations for shopping at the malls variable and its categories

| Motivations for shopping at the malls | Percentage frequency | Coordinates | | Contribution % | | Squared cosines | | Measures of discrimination | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1st | 2nd | 1st | 2nd | 1st | 2nd | 1st | 2nd |
| Self-gratification | 10.0 | -0.421 | 0.674 | 0.9 | 2.3 | 0.020 | 0.052 | 0.018 | 0.045 |
| Communication and social interaction | 5.2 | 0.421 | 1.092 | 0.5 | 3.1 | 0.010 | 0.066 | 0.009 | 0.062 |
| Freedom of choice | 10.8 | -0.213 | 0.654 | 0.2 | 2.3 | 0.005 | 0.051 | 0.005 | 0.046 |
| To get a bargain | 40.8 | -0.469 | -0.560 | 4.5 | 6.4 | 0.152 | 0.218 | 0.090 | 0.128 |
| To be alone | 8.0 | -0.565 | 0.550 | 1.3 | 1.2 | 0.028 | 0.026 | 0.026 | 0.024 |
| To follow fashion | 13.2 | 2.065 | -0.566 | 28.1 | 2.1 | 0.649 | 0.048 | 0.563 | 0.042 |
| To spend time with friends | 12.0 | 0.061 | 0.533 | 0.0 | 1.7 | 0.000 | 0.038 | 0.000 | 0.034 |
| Total | 100.0 | | | 35.5 | 19.1 | | | 0.711 | 0.382 |
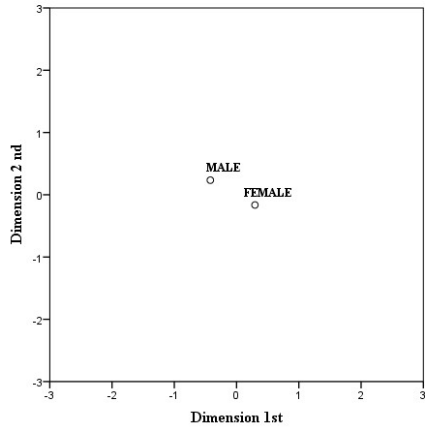
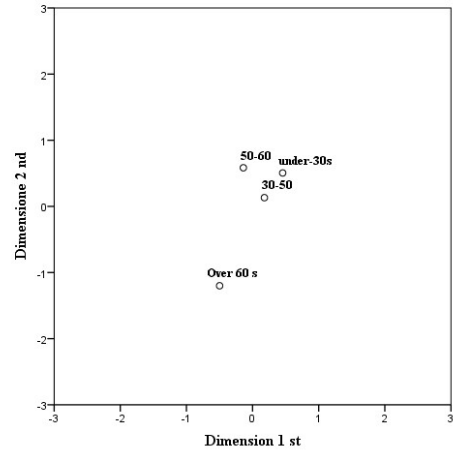Figure 2: Coordinates of the gender variable categories on the plane of the two dimensions



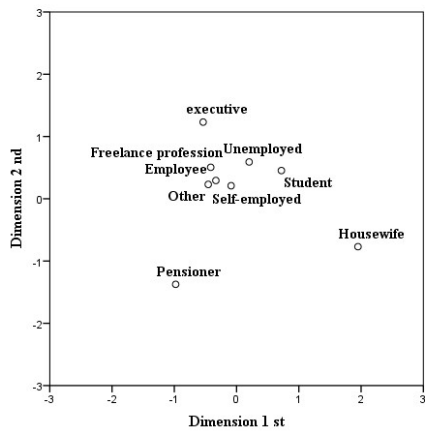Figure 3: Coordinates of the age variable categories on the plane of the two dimensions



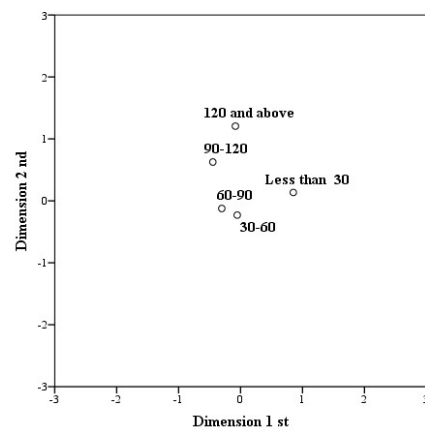Figure 4: Coordinates of the professional stats variable categories on the plane of the two dimensions



Figure 5: Coordinates of the average spending variable categories on the plane of the two dimensions
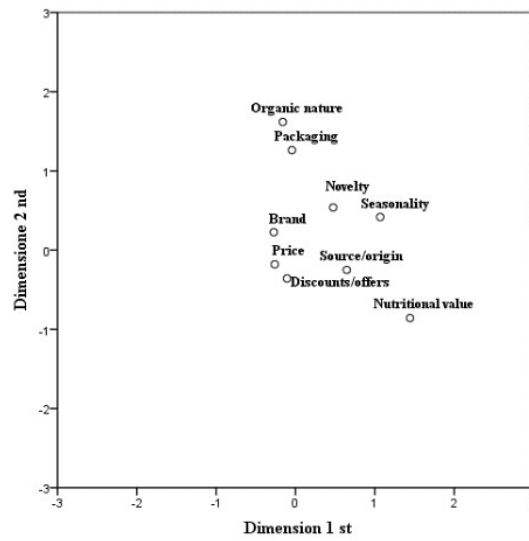
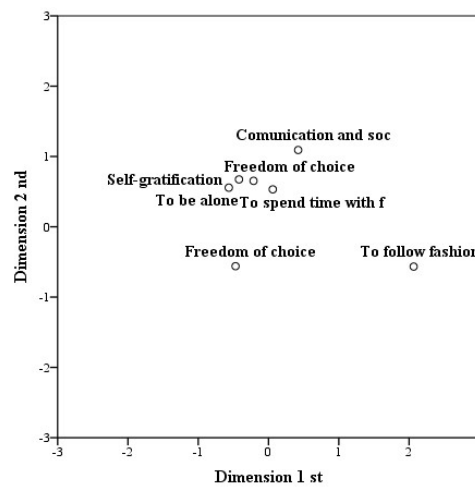Figure 6: Coordinates of the product characteristics variable categories on the plane of the two dimensions



Figure 7: Coordinates of the motivations for shopping at the malls variable categories on the plane of the two dimensions

# 6 Results of multiple correspondence analysis on the complete set of variables

Figure 8 shows the 36 categories of the 6 variables in relation to the two dimensions considered; this figure is the synthesis of Figures 2-7 which showed the categories of the single variables. Through the closeness or distance between the categories, the figure makes it possible to identify the associations and disassociations between the same categories. We will restrict our attention to the associations between the categories regarding the characteristics of the interviewees (gender, age and profession) and those of their behavior (average expenditure, product characteristics and motivations for shopping at the malls). We will draw attention to the most substantial and considerable associations or disassociations.

We begin with gender. The women seem to pay particular attention to the brands and prices of the products, their expenditure is medium to medium-high, they seem to pay less attention to the nutritional value of those products, and infrequently go to the malls with the idea of following fashion. The men pay particular attention to the origin of the products, and a marked lack of attention to the organic aspects and typicality of those products, their average expenditure is prevalently medium-low and, like the women, they do not go to a shopping center with the idea of following a trend.

Passing to the age groups, we consider, in particular, young people (under 30s) and the oldest age group (60 and over). The young people generally concentrate in the lowest expenditure class, they seem to pay particular attention to the products' origin and little attention to their organic nature and typicality, nor do they go to the shopping center with the idea of following fashion. The oldest age group, who are to be found in the lower left square of the graph, appear to pay little attention to almost all the characteristics of the products and have no particular motivations for going to the shopping center. Their average expenditure is generally in the intermediate classes, and they pay a low level of attention to promotional offers and the idea of following fashion.

Finally, we consider professional status which, as we have already seen, represents the most discriminating variable, housewives, who are positioned in the lower right square of the graph, appear to pay little attention to almost all the characteristics of the products and do not appear to have particular motivations for going to the malls. Their average expenditure is generally in the lowest groups. They pay minimal attention to the nutritional aspects of the products they purchase. Their visits to the malls seem to be a routine task without any specific motivations.

The executives are concentrated in the highest spending bracket and pay particular attention to packaging and the organic and specific aspects of their purchases. They seem to have diverse motives for going to the shopping center, with minimal attention to personal gratification and free choice.

The unemployed do not generally correspond to any single expenditure bracket, they favour novelty in products, and pay particularly little attention to nutritional value.

Employees predominantly belong to the middle to high expenditure class, they pay great attention to brand and the price of the products they buy and very little attention to

nutritional value, they go to the shopping center motivated above all by the possibility of choice that it offers.

The freelance professionals tend towards the medium to high average expenditure class, they pay particular attention to the brand and little attention to the nutritional value of the products, and they have diverse motivations for going to the malls: to be alone, gratification and freedom of choice.

Like the oldest age group, also the pensioners, found in the lower left square of the graph, appear to pay little attention to almost none the characteristics of the products and are quite indifferent to the motivations for going to the shopping center. Their average expenditure tends towards the medium to low group. However, the pensioners are less inattentive to promotional offers and are more motivated by the search for bargains at the malls.

The students tend towards the lowest average expenditure group, they seem to pay attention to seasonality and the novelty of products, and their motivations for going to the malls are social in that it is an opportunity to spend time with friends.
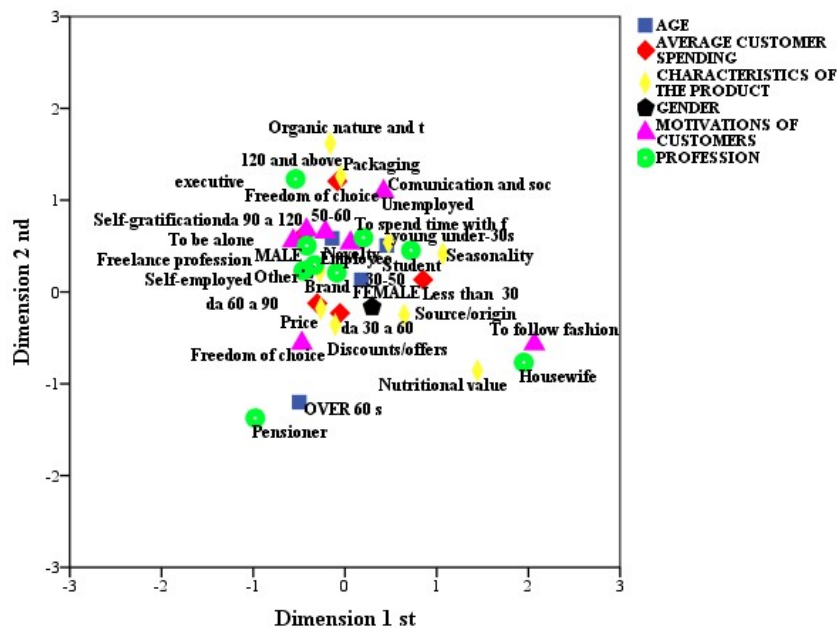


Figure 8: Coordinates of the categories of all the variables on the first two dimensions

# 7 Conclusion

Multiple correspondence analysis allows a number of different categorical variables to be analysed simultaneously. The passage from numerous original variables to a few latent

variables, orthogonal to each other, facilitates immediate comprehension and interpretation of the results of the analyses. The elaboration of data using modern statistical software allows results of great value to be obtained, in a graphic form that is easily read and interpreted, and above all, increases the strategical efficacy of marketing decisions. In this work we applied the multiple correspondence analysis to the data from an extensive survey carried out at the malls of Apulia. It has been possible to segment the customers according to how much they spend, the characteristics of the products they consider buying and their motivations for going to the shopping center (Zani and Cerioli, 2007).

The questionnaire is composed of two sections, one relating to personal data and the other relating to the objectives of the survey, the decision to limit the set of active variables to one of two groups, that could help identify a stronger structure, it was not considered because of the limited number of available variables.

# References

Benzecri, J. P. (1979). Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire, addendum et erratum à [bin. mult.]. *Les cahiers de l'analyse des données*, 4(3):377–378.

Benzécri, J.-P. (1992). *Correspondence analysis handbook*. Marcel Dekker.

Clausen, S. E. (1998). *Applied correspondence analysis: An introduction*, volume 121. Sage.

De Luca, A. (1998). *Marketing bancario e metodi statistici applicati*, volume 3. FrancoAngeli.

Di Franco, G. (2013). *Corrispondenze multiple e altre tecniche multivariate per variabili categoriali*, volume 15. FrancoAngeli.

Le Roux, B. and Rouanet, H. (2010). *Multiple correspondence analysis*, volume 163. Sage.

Zani, S. and Cerioli, A. (2007). *Analisi dei dati e data mining per le decisioni aziendali*. Giuffrè Editore.