QUANTITATIVE METHODS IN ECONOMICS
Vol. XIV, No. 2, 2013, pp. 210 – 219

THE USE OF CORRESPONDENCE ANALYSIS IN THE EVALUATION OF THE ROLE OF FIBROUS AND MEDICINAL PLANTS IN PLANT PRODUCTION IN FARMS

Agnieszka Sompolska–Rzechuła
Department of Mathematics Applications in Economy
West Pomeranian University of Technology in Szczecin
e-mail: asompolska@zut.edu.pl
Grzegorz Spychalski
Department of Economic Sciences
Koszalin University of Technology
e-mail: grzegorz.spychalski@tu.koszalin.pl

Abstract: The paper presents the usage of multidimensional correspondence analysis to estimation the role of fibre and medicinal plants in farms plant production in the context of the determinants influencing the choice of these plants for crop rotation scheme. The source of data were the survey questionnaires collected among the farmers running agricultural activity in the period 2011-2012 in the Wielkopolska region. The correspondence analysis enabled to indicate relations a selected categorical variable and such categories as age, sex, education, number of person per household, farm area and farm income in 2010 and 2011.

Keywords: fibrous and medicinal plants, survey questionnaire, correspondence analysis

INTRODUCTION

The paper aims at the analysis of the role of fibrous and medicinal plants in the plant production at Polish farms in terms of the factors determining the choice of these plants for crop rotation schemes. The source of data for the analysis were survey questionnaires carried out at farms that actively run agricultural activity in Wielkopolska region in years 2011 and 2012.

In order to complete the research task multidimensional correspondence analysis was employed to assess relations between specific categories of variables.

The correspondence analysis allows for precise indication of simultaneous occurrence of two or more variables, which were measured in a nominal scale. Its advantage is the possibility of graphical presentation of concurrent occurrence of specific categorical variables. Despite its advantages, correspondence analysis is used relatively rarely in research and when used it is mostly applied in economic and social sciences.

METHODS

The correspondence analysis is a specialized tool for exploring data that presents the associations between the variables and objects, most often in a graphical form. It enables not only to analyze quantitative data but also the data measured in nominal and ordinal scales and does not have any requirements as for the number of the set of objects. The objective of using this method is to gain knowledge from data sets by analyzing correlation between specific variants of the observed variables.

The following procedure was employed for studying correlations between categorical variables regarding cultivation of fibrous and medicinal plants and the remaining categorical variables[1]:

- Determination of the Burt matrix – this method of recording data is most commonly used for correspondence analysis. As a result a symmetrical block matrix is obtained, where apart from the main diagonal contingency tables are prepared that represent two different variables and contain a number of objects with specific categories of these two variables. The diagonal matrices are placed on the main diagonal, where the non-zero values indicate the number of occurrences of a certain categorical variable,
- determination of the actual space of correlation of the *K* categorical variable, according to the following formula:

$$K = \sum_{q=1}^{Q} \left(J_q - 1 \right) \tag{1}$$

where: J_q - number of the categorical variable q $\left(q = 1,2,\ldots,Q\right)$, Q – number of variables,

- verification to what degree the values of the property space of lower dimension explains total inertia. The Greenacre's criterion was applied, which says that significant inertia are those main inertia of value higher than $1/Q$,
- modification of Eigen values according to the formula below:

[1] A detailed description of the method can be found in: [Stanimir 2005, Gatnar, Walesiak 2006, Machowska-Szewczyk, Sompolska-Rzechuła 2010, Sompolska-Rzechuła 2010].

$$\tilde{\lambda}_k = \left(\frac{Q}{Q-1} \right)^2 \cdot \left(\sqrt{\lambda_k} - \frac{1}{Q} \right)^2 \quad (2)$$

where: Q - number of variables, λ_k - k eigenvalue,

- application of the Ward's method for classification of categorical variables. To present graphically the correlation of variables in the dimension higher than 3 selected classification methods can be used. Categories of all analyzed variables must be defined as objects, where the variables are the values of projection coefficient for each category. Classification methods are also useful when the number of all the variants of variables is high and the distribution of the points in a figure does not allow for unambiguous definition of the classes. This study made use of one of the most common agglomerative classification of the results of correspondence analysis i.e. Ward's method [Ward 1963; Gordon 1999; Ostasiewicz 1998],
- graphical presentation of the categorical variable associations in two- or three-dimensional spaces.

CHARACTERISTICS OF THE RESEARCH MATERIAL

The role of fibrous and medicinal plants in plant production on farms, assessed with correspondence analysis, was determined on the basis of questionnaire answered by farmers active in agricultural activity in Wielkopolska region of Poland in years 2010 and 2011. Presently, agricultural producers choose the crop structure by analyzing natural and economic conditions. They apply the principle of income diversification.

In this respect farmers introduce industrial plants i.e. the so called specialty crops, which do not supply food products and have special role in several production-consumption [Wojdyła 2006].

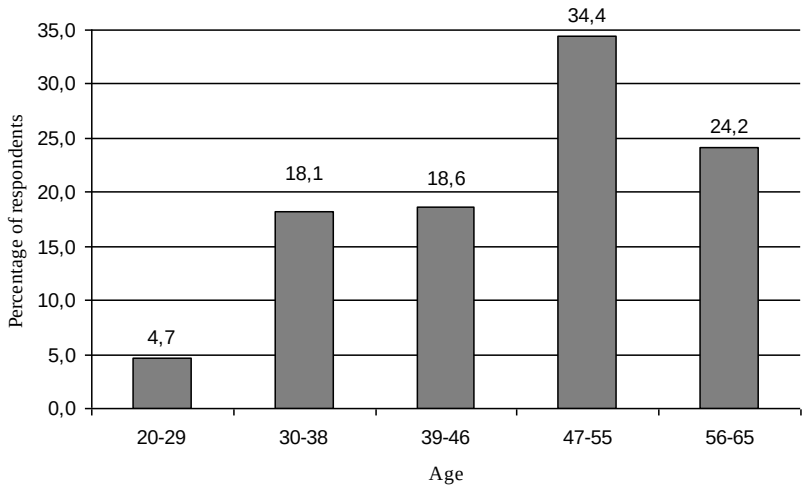
This group of crops includes fibrous and medicinal plants that are used for textile, protective, construction and medical applications. In Polish agriculture there is centuries long tradition of cultivation of flax, hemp and herbs, therefore the producers' skills are good and guarantee sufficient quality of the produced raw materials.

The survey questionnaire comprised two parts: 14 questions concerned the cultivation of fibrous and medicinal plants and 6 questions were demographic.

The total of 224 farmers responded to the questionnaire, provided their acceptance to take part in the survey.

Among the questioned farmers, the biggest group were people between 47-55 years old (34,4%), while the smallest group were the people between 20-29 years old (4,7%). The age structure of the respondents is presented in Figure 1.

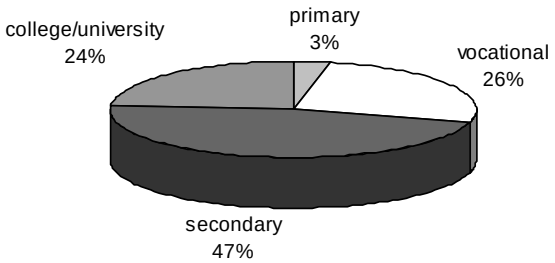
Figure 1. Age structure of the respondents



Source: own study

Among respondents, the largest group was individuals with high school education, while the smallest with primary school education (see Figure 2).

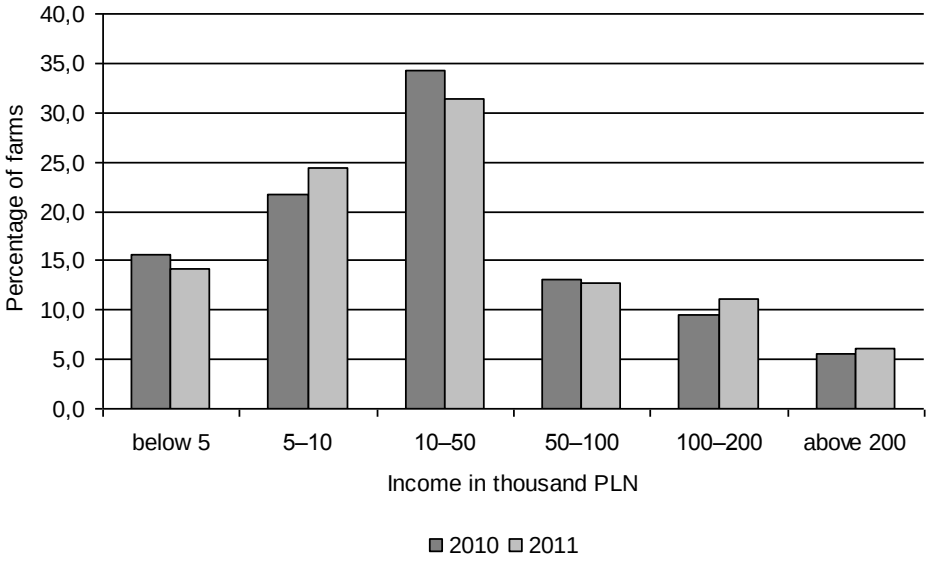
Figure 2. Education structure of the respondents



Source: own study

The farms studied in the questionnaire varied greatly in terms of their area with the smallest of 0,18 ha and the largest of 470 ha. The farms of 20 ha were the dominating group (15 farms). Similar situation was observed in case of the area of arable land, where the variation coefficient was more than 131%. The questioned farmers were also asked about the income levels in years 2010-2011. In both periods the most farms reached the income between 10-50 thousand PLN, in 2010 such income was reported for 34,3% of farms, while in 2011 – 31,5%. The income structure of farms in years 2010 and 2011 is presented in Figure 3.

Figure 3. The farm structure in terms of income levels in years 2010 and 2011



Source: own study

The technical part of the questionnaire concerning the role of cultivation of fibrous and medicinal plants in the plant production of farms included questions on cultivation of these plants, their position in the sowing structure, conditions for crop rotation schemes, knowledge on textile industry and herb processing industry in Poland, Polish research organizations active in this field and the expectations for the academic world.

As much as 18% of the questioned farmers declared that they grow fibrous (flax or hemp) or medicinal plants. At only 8% of the farms fibrous plants were cultivated, of which 9% of the farmers believed fibrous plants improve the soil structure and 5% that growing these plants is profitable. A similar situation was observed for medicinal plants. The highest number of farmers (42%) thought low popularity of fibrous plants is due to absence of that tradition for growing these plants. The same reason was given by 38% of questioned farmers regarding

absence of medicinal plants in the sowing structure of their farms. If they were to introduce a new plant, the highest group of farmers (51%) would opt for growing herbal plants, 33% - for flax, and 16% for fibrous hemp. As much as 56% of the respondents believe that cultivation of specialty crops can increase the farm income, 23% that it improves the soil quality, 16% respondents think it will facilitate sales of the yields. Half of the questioned farmers expect that academic world will provide advice while 43% that the science will provide tried and tested technology.

The survey questions were accompanied with the following sets of features and categories:

- cultivation of fibrous and/or medicinal plants: Y (yes), N (no),
- sex: F (female), M (male),
- age: A (20-38 years old), B (39-55 years old), C (more than 55 years old),
- education level: A (high school or college/university), B (vocational or primary school education),
- number of people in the household: A (1-3 individuals), B (3-6 individuals), C (more than 6 people),
- farm area: A (below 50 ha), B (50-100 ha), C (above 100 ha),
- area of arable land: A (below 50 ha), B (50-100 ha), C (above 100 ha),
- income per farm in 2010: A (below 50 thousand PLN), B (above 50 thousand PLN),
- income per farm in 2011: A (below 50 thousand PLN), B (above 50 thousand PLN).

Before applying multidimensional correspondence analysis we tested whether there is dependence between those features, what was confirmed in the results.

RESULTS

In case of studying the associations between categories of dependent variable i.e. cultivation of fibrous and/or medicinal plants and the other categories of variables associated with it a Burt matrix was obtained of the dimension 22×22. The dimension of actual answer correlation space was 13.

The next step involved verification how the eigenvalues of the space with lower dimension explains the total inertia. The results are presented in Table 1 that includes: eigenvalues λ_k, singular values γ_k, share of main inertia in total inertia (in percentage of λ_k / λ) and share of eigenvalues from *K* dimension in total inertia (accumulative percentage τ_K).

Table 1. Eigen and singular values and the degree of explanation of the total inertia

Number of K dimensions	Values		Percentage	
	Eigen λ_k	singular γ_k	inertia λ_k / λ	accumulated τ_K
1	0,3124	0,5589	21,6291	21,6291
2	0,2040	0,4516	14,1216	35,7507
3	0,1534	0,3917	10,6219	46,3726
4	0,1408	0,3753	9,7492	56,1218
5	0,1317	0,3629	9,1197	65,2415
6	0,1125	0,3354	7,7886	73,0302
7	0,1048	0,3237	7,2534	80,2836
8	0,0992	0,3149	6,8658	87,1494
9	0,0863	0,2938	5,9752	93,1246
10	0,0583	0,2414	4,0356	97,1602
11	0,0273	0,1652	1,8883	99,0485
12	0,0117	0,1084	0,8131	99,8615
13	0,0020	0,0447	0,1385	100,0000
	λ =1,4444			

Source: own study

According to Greenacre's criterion, the best dimension of projection of category variables is the one where eigenvalues fulfil the condition: $\lambda_k > 1/Q$. In the analyzed case, this value is 0,1111 for $Q = 9$. The data presented in Table 1 indicate that these are inertia for the R^6 dimension and total dimension in this case is 1,4444. Modification of eigenvalues was made according to the Greenacres's criterion. The values of modified eigenvalues and singular values and the degree of explanation of the total inertia are presented in Table 2.

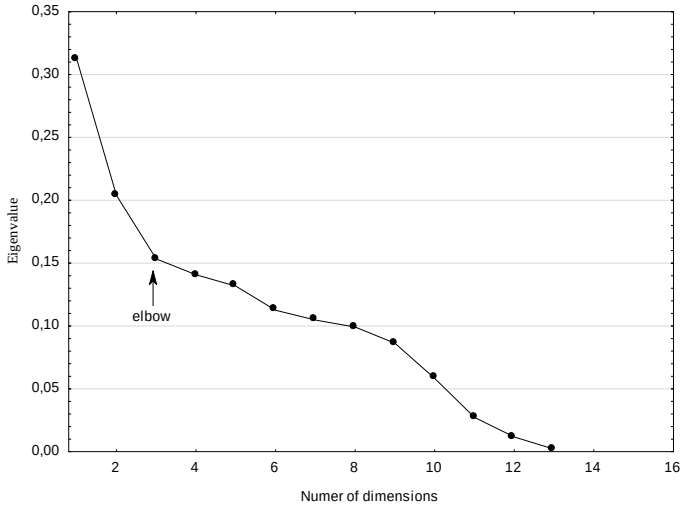
Table 2. Modified Eigen and singular values and the degree of explanation of the total inertia

Number of K dimensions	Values		Percentage	
	Eigen $\tilde{\lambda}_k$	singular $\tilde{\gamma}_k$	inertia $\tilde{\lambda}_k / \tilde{\lambda}$	accumulated $\tilde{\tau}_K$
1	0,2538	0,5038	34,6530	34,6530
2	0,1468	0,3831	20,0361	54,6892
3	0,0996	0,3157	13,6033	68,2925
4	0,0883	0,2972	12,0563	80,3487
5	0,0803	0,2833	10,9581	91,3068
6	0,0637	0,2523	8,6932	100,0000
	$\tilde{\lambda}$ =0,7325			

Source: own study

Additionally, a graph was prepared that shows eigenvalues (see Figure 4).

Figure 4. The graph presenting Eigen values

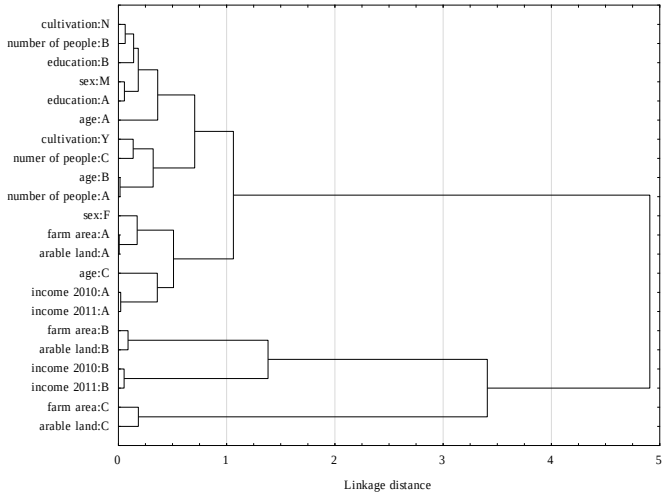


Source: own study

One of the methods of determining the number of eigenvalues, which indicate coordinates significant for projection with low dimension is the so called „elbow" criterion. A place is searched on the graph with all non-zero specific/proper values in descending order where a slight drop in these values is observed (a number of the eigenvalue is indicated, where the 'bend' is visible) [Stanimir 2005]. Figure 4 shows that 'the elbow' occurs for $k = 3$. Therefore, the correlation analysis between categorical variables will take place in three-dimensional space, which explains almost 68,3% of total inertia.

The correlations between categorical variables are presented in a dendrograph obtained with the Ward's method (Figure 5).

Figure 5. The dendrograph presenting the division of categorical variables according to the Ward's method



Source: own study

The use of correspondence analysis allowed for separating two groups of associations between categorical variables. Two groups were distinguished because of two categories of dependent variable i.e. cultivation of fibrous and/or medicinal plants. This variable is of dychotomic nature and takes two values Yes or No. The category 'non-cultivation of fibrous and/or medicinal plants' is mostly associated with the B category of the variable 'number of people per farm' i.e. between 3 and 6 people per farm. Moreover, absence of these crops is declared by the surveyed farmers with vocational and primary school education but also with high school and university education and men aged 20-38 years.

Cultivation of fibrous and/or medicinal plants is characteristic for the farms with more than 6 people in the household and also with 1-3 people. The age range of the farmers that grow fibrous and/or medicinal plants is 39-55 years. Weaker correlation is visible between the cultivation and the following categories of variables: income per farm both in 2010 and 2011, farm area and the area of arable land above 50 ha. These categories constitute a separate class not associated with the categorical variable of cultivation of fibrous and/or medicinal plants.

REFERENCES

Gatnar E., Walesiak M. (2006) Metody statystycznej analizy wielowymiarowej w badaniach marketingowych, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, pp. 285-315.

Gordon A. D. (1999) Classification, Chapman & Hall/CRC, Boca Raton.

Machowska-Szewczyk M. Sompolska-Rzechuła A. (2010) Analiza korespondencji w badaniu postaw osób dokonujących zakupów przez Internet, Ekonometria 29, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 141, Wrocław, pp. 9-21.

Ostasiewicz W. (red.) (1998) Statystyczne metody analizy danych, Akademia Ekonomiczna we Wrocławiu, Wrocław, pp. 95-97.

Panek T. (2009) Statystyczne metody wielowymiarowej analizy porównawczej, Szkoła Główna Handlowa w Warszawie, pp. 22.

Sompolska–Rzechuła A. (2010) Zastosowanie metody analizy zgodności w badaniu jakości życia kobiet, Wiadomości Statystyczne, nr 1, pp. 53-64.

Stanimir A. (2005) Analiza korespondencji jako narzędzie do badania zjawisk ekonomicznych, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław

Ward J. H. (1963) Hierarchical grouping to optimize an objective function, Journal of the American Statistical Association, 58, pp. 236-244.

Wojdyła T. (2006) Rośliny przemysłowe w przemyśle spożywczym oraz metody analiz stosowanych w ich przetwórstwie, Wydawnictwo Akademii Techniczno-Rolniczej, Bydgoszcz.