THE CORRESPONDENCE ANALYSIS APPLIED ON THE AGRICULTURE SECTOR OF EUROPEAN UNION COUNTRIES USING STATISTICAL ANALYSIS SYSTEM

Ana-Maria Mihaela Iordache[1]
Mihai-Tiberiu Iordache[2]
Ionela-Cătălina Tudorache[3]

ABSTRACT

The main goal of analysis is represented by studying the simultaneous correspondences of lines and columns of a contingency table in order to highlight the connections and the correspondences between the sets of variables. There are two basic ways to achieve the correspondence analysis. First is the analysis of relationships between the two variables whose observation we find a contingency table and the second is the analysis of relationships between a set of variables (types of responses of subjects) and another group of qualitative variables with more ways to respond. In our analysis we use eight variables from agriculture measured on European Union countries, then we applied correspondence analysis on the data set and we showed how the countries group by the quantities of information brought by each indicator.

Keywords: correspondence analysis, inertia, mass, cluster analysis

1. INTRODUCTION

The correspondence analysis is a form of interdependence analysis by which it can be identified the correlation between the variables. This type of analysis is used for a matrix of variables that can be measured on the metric scale.

Correspondence analysis is an exploratory data analytic technique designed to analyze simple two-way and multi-way tables containing some measure of correspondence between the rows and columns. As opposed to traditional hypothesis testing designed to verify a priori hypotheses about relations between variables, exploratory data analysis is used to identify systematic relations between variables when there are not (or rather incomplete) a priori expectations as to the nature of those relations.

Correspondence analysis is also a (multivariate) descriptive data analysis technique. Even the most commonly used statistics for simplification of data may not be adequate for description or understanding of the data. Simplification of data provides useful information about the data, but that should not be at the expense of valuable information. Correspondence analysis remarkably simplifies complex data and provides a detailed

[1]Assitant Lecturer, Phd, School of Computer Science for Business Management, Romanian-American University, Bucharest
[2] Engineer, IMSAT, Bucharest
[3] Student, Romanian-American University, Bucharest

description of practically every bit of information in the data, yielding a simple, yet exhaustive analysis. [5]

A distinct advantage of correspondence analysis over other methods yielding joint graphical displays is that it produces two dual displays whose row and column geometries have similar interpretations, facilitating analysis and detection of relationships. In other multivariate approaches to graphical data representation, this duality is not present.

In a two-way contingency table, the observed association of two traits is summarized by the cell frequencies, and a typical inferential aspect is the study of whether certain levels of one characteristic are associated with some levels of another. Correspondence analysis is a geometric technique for displaying the rows and columns of a two-way contingency table as points in a low-dimensional space, such that the positions of the row and column points are consistent with their associations in the table. The goal is to have a global view of the data that is useful for interpretation. [7]

In a nutshell, correspondence analysis may be defined as a special case of principal components analysis of the rows and columns of a table, especially applicable to a cross-tabulation. However both analyses are used under different circumstances. Principal components analysis is used for tables consisting of continuous measurement, whereas correspondence analysis is applied to contingency tables (*i.e.* cross-tabulations). Its primary goal is to transform a table of numerical information into a graphical display, in which each row and each column is depicted as a point. [5]

The usual procedure for analyzing a cross-tabulation is to determine the probability of global association between rows and columns. The significance of association is tested by the Chi-square test, but this test provides no information as to which are the significant individual associations between row-column pairs of the data matrix. Correspondence analysis shows how the variables are related, not just that a relationship exists. [5]

2. DATA DESCRIPTION

The case study we propose is focused on the agriculture output of the European Union countries. From this point of view, we take into account the 2011's data given by eurostat.com.

The variables used are: cereals including seeds, industrial crops, forage plants, vegetables and horticultural products, potatoes, crop output, animals and animal products; these are denominated in millions of euros, the prices of year 2005, and production value at basic prices (Table 1). From these variables, all are measured on a quantitative scale.

Table 1. The indicators used in the application

Indicator	The full name of the indicator
I1	cereals (including seeds)
I2	industrial crops
I3	fodder plants
I4	vegetables and horticultural products
I5	potatoes
I6	crop output
I7	animals
I8	animal products

The variable that describes the production of 2011 cereals and seeds can be associated with letters A, B and C, where A represents values between 0 and 800, B values between 800 and 1500 and C the rest of the values. The interpretation of this classification is that the C class is the most productive class, while the A class is formed by countries that have the lowest production of cereals and seeds.

Similar to this classification, the variable that represents the industrial crops can be clustered into three classes such as: low (values lower than 100), medium (between 100 and 1000) and high (over 1000). Also, the variable that indicates the forage plants can be represented by inefficient (low than 100), medium (between 100 and 1000) and efficient (over 1000) (Table 2).

Table 2. The values for the eight indicators used in application

	i1	i2	i3	i4	i5	i6	i7	i8
Belgium	412.6288	169.5038	535.4384	1068.3707	281.3786	2764.589	2889.0882	1028.5508
Bulgaria	818.8636	627.7695	158.6724	73.5638	21.0759	1795.6154	399.8602	398.0899
Czech_Republic	992.9728	548.8937	296.8152	165.7950	69.6178	2161.6584	645.5808	702.908
Denmark	1610.6868	283.2277	456.2134	602.9430	96.0245	3136.3782	3143.4119	2163.1676
Germany	6568.5105	2164.4084	8993.5788	3974.2490	2083.9041	25495.285	11827.526	10452.449
Estonia	101.7430	57.3716	32.1979	32.5214	39.8885	267.6125	114.9711	179.6353
Ireland	412.9007	0.0000	945.9967	210.4345	87.5291	1792.123	2820.597	1958.9735
Greece	906.3880	669.4132	606.7470	1417.2459	265.6558	5841.0286	1184.2242	1227.6459
Spain	3834.7560	980.3310	2225.5361	5938.3790	449.4629	21676.741	10119.507	3434.3737
France	9398.6105	4132.4728	4970.6703	4782.1935	1249.5184	35421.622	14292.561	8754.2525
Italy	4467.1776	740.4087	1593.6452	7492.8721	577.5025	24570.114	8757.2154	5452.7829
Cyprus	8.8523	0.4709	6.0014	85.3044	29.1511	287.6839	164.7357	117.6916
Latvia	156.1806	59.3309	64.2285	29.9067	36.5332	353.9019	99.5096	183.3702
Lithuania	596.0140	220.8341	145.0991	68.8748	54.1309	1141.8278	298.8684	412.3927
Luxembourg	21.7954	5.2114	57.3723	4.1864	3.5328	115.361	65.0556	79.2353
Hungary	2189.5385	799.2138	153.1232	615.0281	99.6316	4312.7997	1528.976	721.5466
Malta	0.0000	0.0000	3.5948	27.5919	6.6564	44.2546	37.0942	21.5088
Netherlands	268.7154	343.6750	574.1679	7750.4851	1057.3031	11054.028	4866.109	4627.9933
Austria	812.1383	297.3968	529.0283	459.3787	71.8289	3093.1012	1712.7395	1187.4255
Poland	4101.1337	1366.8590	1102.5194	1761.1310	814.5381	10298.997	4985.066	3816.815
Portugal	223.4773	51.5397	280.9075	890.0185	123.5468	2897.9759	1677.2925	790.3142
Romania	3321.4688	842.9727	1428.3049	1549.7039	1113.6458	9287.6643	1347.8813	1504.2803
Slovenia	82.9742	22.7171	168.0061	56.0734	20.1836	559.8679	281.5215	194.0854
Slovakia	491.3602	197.0592	63.5493	74.7333	16.9625	891.6431	269.5462	292.3848
Finland	482.0740	83.0789	164.4458	476.3071	132.1874	1427.84	748.0076	1339.4447
Sweden	635.7105	156.3253	784.1503	313.6648	144.2268	2091.8429	935.9871	1265.2118
United_Kingdom	3890.0120	2285.8358	396.3183	2703.7489	902.1065	10994.216	9911.0052	5397.913
Norway	301.7992	7.2660	478.2247	313.8146	69.0342	1249.4262	1236.2508	1230.6686
Switzerland	236.8082	161.6743	647.5787	850.7200	106.6406	2651.3744	1467.95	1390.6486
Croatia	538.7446	155.5408	169.0407	239.7703	26.8232	1469.7878	544.3601	349.2497

After describing the variables took into account with descriptive statistics, we saw that the coefficient of variation is higher than 30%, which means that the mean is unrepresentative from statistical point of view, and there is a high heterogeneity between the countries. Therefore, it is necessary to see if there are outliers between countries.

Statistically, it is shown that almost 99.87% of values are between mean-3*standard deviation and mean+3*standard deviation, for a normal standard distribution. The values that are outside this interval are considered to be outliers, and should be eliminated from the analysis. For the countries we analyze, France and Germany are outliers for some indicators. Because of the fact that we analyze the agriculture area of European Union we cannot eliminate countries that are considered to be very important for the analysis.

3. THE CORRESPONDENCE ANALYSIS ALGORITHM

A variant of Euclidean distance, called the weighted Euclidean distance, is used to measure and thereby depict the distances between profile points. Here, the weighting refers to differential weighting of the dimensions of the space and not to the weighting of the profiles.

Distance between two rows i and i' is given by

$$d^2(i,i') = \sum_{j=1}^{J} \frac{1}{N_{+j}} \left(\frac{N_{ij}}{N_{i+}} - \frac{N_{i'j}}{N_{i'+}} \right)^2$$

In a symmetric fashion, the distance between two columns j and j' is given by

$$d^2(j,j') = \sum_{i=1}^{I} \frac{1}{N_{i+}} \left(\frac{N_{ij}}{N_{+j}} - \frac{N_{ij'}}{N_{+j'}} \right)^2$$

The distance thus obtained is called the *Chi-square distance*. The Chi-square distance differs from the usual Euclidean distance in that each square is weighted by the inverse of the frequency corresponding to each term (Table 3).

The division of each squared term by the expected frequency is "variance – standardizing" and compensates for the larger variance in high frequencies and the smaller variance in low frequencies. If no such standardization were performed, the differences between larger proportions would tend to be large and thus dominate the distance calculation, while the differences between the smaller proportions would tend to be swamped. The weighting factors are used to equalize these differences.

Essentially, the reason for choosing the Chi-square distance is that it satisfies the principle of distributional equivalence, expressed as follows:

- If two rows i and i' of I of $N(I, J)$ are proportioned and if they are replaced by only one, which is the sum, column-by-column, then the distances between columns are not changed in $N(J)$.

- If two columns j and j' of J of $N(I, J)$ are proportioned and if they are replaced by only one, which is the sum, row-by-row, then the distances between rows are not changed in $N(I)$.

Inertia is a term borrowed from the "moment of inertia" in mechanics. A physical object has a center of gravity (or centroid). Every particle of the object has a certain mass m and a certain distance d from the centroid. The moment of inertia of the object is the quantity md^2 summed over all the particles that constitute the object. Moment of inertia = $\sum md^2$

This concept has an analogy in correspondence analysis. There is a cloud of profile points with masses adding up to 1. These points have a centroid and a distance (Chi-square distance) between profile points. Each profile point contributes to the inertia of the whole cloud.

The criterion used for dimensionality reduction implies that the inertia of a cloud in the optimal subspace is maximum, but that would still be less than that in the true space. What is lost in this process is the knowledge of how far and in which direction the profiles lie off this subspace. What is gained is a view of the profiles, which otherwise would not be possible. The ratio of inertia inside the subspace to the total inertia gives a measure of the accuracy of representation of a cloud in the subspace.

Correspondence analysis determines the principal axes of inertia and for each axis the corresponding eigenvalue, which is the same as the inertia of the cloud in the direction of the axis. The first factorial axis is the line in the direction of which the inertia of the cloud is a maximum. The second factorial axis is, among all the lines that are perpendicular to the first factorial axis, the one in whose direction the inertia of the cloud is a maximum. The third factorial axis is, among all the lines that are perpendicular to both the first and second factorial axes, the line in whose direction the inertia of the cloud is a maximum, and so on. The optimal subspace is a subspace spanned by the principal axes. The inertia of a profile along a principal axis is called the *Principal Inertia.*

The inertia of a profile point can be computed by the following formula (Table 4).

For the i^{th} row profile,

Inertia = $r_i \sum_j \frac{(r_{ij} - \bar{r}_j)^2}{\bar{r}_j}$, where r_{ij} is the ratio n_{w}/n_{i+} and $\bar{r}_j$ is $n_{.j}/n$.

The row analysis of a matrix consists in situating the row profiles in a multidimensional space and finding the low- dimensional subspace, which comes closest to the profile points. The row profiles are projected onto such a subspace for interpretation of the inter-profile positions. Similarly, the analysis of column profiles involves situating the column profiles in a multidimensional space and finding the low-dimensional subspace, which comes closest to the profile points.

The row and column analyses are very connected. If a row analysis is performed, the column analysis is also ipso facto performed, and vice versa. The two analyses are equivalent in the sense that each has the same total inertia, the same dimensionality and the same decomposition of inertia into principal inertias along principal axes.

Since the sums of the frequencies across the columns must be equal to the row totals, and the sums across the rows equal to the column totals, there are in a sense only independent entries in each row, and independent entries in each column of the contingency table. Thus, the maximum number of eigenvalues that can be extracted from a two- way table is equal to the minimum of the number of columns minus 1, and the number of rows minus 1. If we choose to extract the maximum number of dimensions that can be extracted, then we can reproduce exactly all the information contained in the table.

4. APPLYING CORRESPONDENCE ANALYSIS USING STATISTICAL ANALYSIS SYSTEM

The procedure used in Statistical Analysis System for implementing the algorithm described above is:

```
ods graphics on;
proc corresp data=a;
var i1 i2 i3 i4 i5 i6 i7 i8;
id tara;
run;
```

This procedure runs only if the data set is imported and named a. The variables taken into analysis are named, as shown, from i1 to i8, while de id of the country is the row named "tara" (country). Using the instruction id named "tara" is important in making graphics. In this way, in the representation of countries in the plan defined by two calculated dimensions, the observations are labeled by the name of the country.

Table 3. Correspondence analysis indicators for the entire data table

Singular Value	Principal Inertia	Chi-Square	Percent	Cumulative Percent
0.23071	0.05323	25848.8	40.47	40.47
0.18732	0.03509	17039.8	26.68	67.14
0.15001	0.0225	10928.1	17.11	84.25
0.10662	0.01137	5520.2	8.64	92.89
0.0654	0.00428	2077.2	3.25	96.14
0.06223	0.00387	1880.7	2.94	99.09
0.03461	0.0012	581.8	0.91	100

From Table 3 results that only two components account for more than a half of the information in the cloud of points, the indicator Cumulative Percent is set to 67.14. Further, applying the correspondence analysis for lines and columns is enough to keep the two-dimensional representation of points.

Table 4. Summary statistics for row coordinates

Country	Row Coordinates		Summary Statistics for the Row Points		
	Dim1	Dim2	Quality	Mass	Inertia
Belgium	-0.1984	0.2291	0.5434	0.0188	0.0242
Bulgaria	0.5188	-0.4094	0.7826	0.0088	0.0375
Czech_Republic	0.3967	-0.2096	0.8353	0.0115	0.0211
Denmark	0.1098	0.2	0.3191	0.0237	0.0293
Germany	0.1515	0.2114	0.636	0.1474	0.1191
Estonia	0.221	0.0552	0.2607	0.0017	0.0026
Ireland	0.0719	0.6649	0.9161	0.0169	0.0629
Greece	-0.066	-0.2022	0.5007	0.025	0.0171
Spain	-0.1622	-0.0892	0.5192	0.1002	0.0503
France	0.1311	-0.0739	0.7996	0.1709	0.0368
Italy	-0.2206	-0.1312	0.8323	0.1105	0.0665
Cyprus	-0.3434	0.1249	0.6207	0.0014	0.0024
Latvia	0.3159	-0.0124	0.5675	0.002	0.0027
Lithuania	0.4064	-0.1959	0.8416	0.0061	0.0111
Luxembourg	0.2296	0.5063	0.8077	0.0007	0.0021
Hungary	0.2791	-0.3735	0.8531	0.0215	0.0416
Malta	-0.5266	0.1978	0.8633	0.0003	0.0008
Netherlands	-0.6256	-0.0044	0.8934	0.0629	0.2095
Austria	0.0919	0.089	0.5704	0.0168	0.0037
Poland	0.1464	-0.0611	0.5346	0.0582	0.0208
Portugal	-0.2825	0.0746	0.7679	0.0143	0.0121
Romania	0.1752	-0.2477	0.4401	0.042	0.0668
Slovenia	0.0977	0.2559	0.5764	0.0029	0.0028
Slovakia	0.392	-0.2797	0.8035	0.0047	0.0104
Finland	-0.0625	0.2223	0.2294	0.01	0.0177
Sweden	0.1745	0.2845	0.702	0.013	0.0157
United_Kingdom	0.0381	0.0352	0.0206	0.0751	0.0747
Norway	-0.0079	0.5095	0.9047	0.0101	0.022
Switzerland	-0.1504	0.2524	0.8109	0.0155	0.0125

Croatia	0.1475	-0.1585	0.7842	0.0072	0.0033

Another way of looking at correspondence analysis is to consider it as a method for decomposing the overall inertia by identifying a small number of dimensions in which the deviations from the expected values can be represented. This is similar to the goal of factor analysis, where the total variance is decomposed, so as to arrive at a lower - dimensional representation of variables that allows one to reconstruct most of the variance/covariance matrix of variables (Table 5).

Table 5. Partial contributions to inertia and squared cosines for the row points

Country	Partial Contributions to Inertia for the Row Points		Squared Cosines for the Row Points	
	Dim1	Dim2	Dim1	Dim2
Belgium	0.0139	0.0282	0.2329	0.3104
Bulgaria	0.0447	0.0422	0.4822	0.3004
Czech Republic	0.034	0.0144	0.653	0.1823
Denmark	0.0054	0.027	0.0739	0.2452
Germany	0.0635	0.1877	0.2157	0.4203
Estonia	0.0016	0.0001	0.2454	0.0153
Ireland	0.0016	0.2135	0.0106	0.9055
Greece	0.002	0.0291	0.0483	0.4525
Spain	0.0495	0.0227	0.3987	0.1205
France	0.0552	0.0266	0.6067	0.1929
Italy	0.101	0.0542	0.6148	0.2175
Cyprus	0.0032	0.0006	0.5482	0.0725
Latvia	0.0038	0	0.5667	0.0009
Lithuania	0.0188	0.0066	0.6829	0.1587
Luxembourg	0.0007	0.0053	0.1378	0.6699
Hungary	0.0314	0.0853	0.3057	0.5474
Malta	0.0015	0.0003	0.7566	0.1067
Netherlands	0.4625	0	0.8934	0
Austria	0.0027	0.0038	0.2944	0.276
Poland	0.0234	0.0062	0.4553	0.0793
Portugal	0.0214	0.0023	0.718	0.05
Romania	0.0242	0.0734	0.1467	0.2934
Slovenia	0.0005	0.0053	0.0733	0.5031
Slovakia	0.0137	0.0105	0.5324	0.2711

Finland	0.0007	0.0141	0.0168	0.2126
Sweden	0.0075	0.03	0.1919	0.5101
United Kingdom	0.0021	0.0027	0.0111	0.0095
Norway	0	0.0745	0.0002	0.9044
Switzerland	0.0066	0.0281	0.2125	0.5984
Croatia	0.0029	0.0051	0.364	0.4202

From the information given by index of the coordinates that contribute most to inertia for the row points, it results two groups of countries:

- Group I: Bulgaria, Czech Republic, Estonia, Spain, France, Italy, Cyprus, Latvia, Lithuania, Malta, Netherlands, Poland, Portugal, Slovakia;
- Group II: Belgium, Denmark, Germany, Ireland, Greece, Luxembourg, Hungary, Austria, Romania, Slovenia, Finland, Sweden, United Kingdom, Norway, Switzerland, Croatia.

We will continue to apply the same algorithm of correspondence analysis on the columns too. In Table 6 were calculated the specific indicators, such as inertia, mass, quality. It was observed that the cloud of points can be represented in the plan, based on two factorial axes which bring a maximum of information (Table 7).

Table 6. Summary statistics for columns correspondences

Indicators	Columns Coordinates		Summary Statistics for the Columns Points		
	Dim1	Dim2	Quality	Mass	Inertia
i1	0.3555	-0.1937	0.8815	0.0986	0.1393
i2	0.4282	-0.285	0.6725	0.0359	0.1074
i3	0.269	0.3787	0.5515	0.0577	0.1717
i4	-0.5705	-0.1083	0.9616	0.0907	0.2417
i5	-0.0041	0.0551	0.0103	0.0207	0.0468
i6	-0.0176	-0.1198	0.6089	0.3895	0.0713
i7	-0.0452	0.1671	0.3431	0.182	0.1208
i8	0.0075	0.2592	0.6331	0.1249	0.1009

Table 7. Partial contributions to inertia and squared cosines for the column points

Indicator	Partial Contributions to Inertia for the Columns Points		Squared Cosines for the Columns Points	
	Dim1	Dim2	Dim1	Dim2
i1	0.2341	0.1054	0.6798	0.2017

i2	0.1236	0.0831	0.466	0.2065
i3	0.0785	0.236	0.185	0.3666
i4	0.5544	0.0303	0.9281	0.0335
i5	0	0.0018	0.0001	0.0102
i6	0.0023	0.1593	0.0128	0.5961
i7	0.007	0.1448	0.0234	0.3197
i8	0.0001	0.2393	0.0005	0.6326

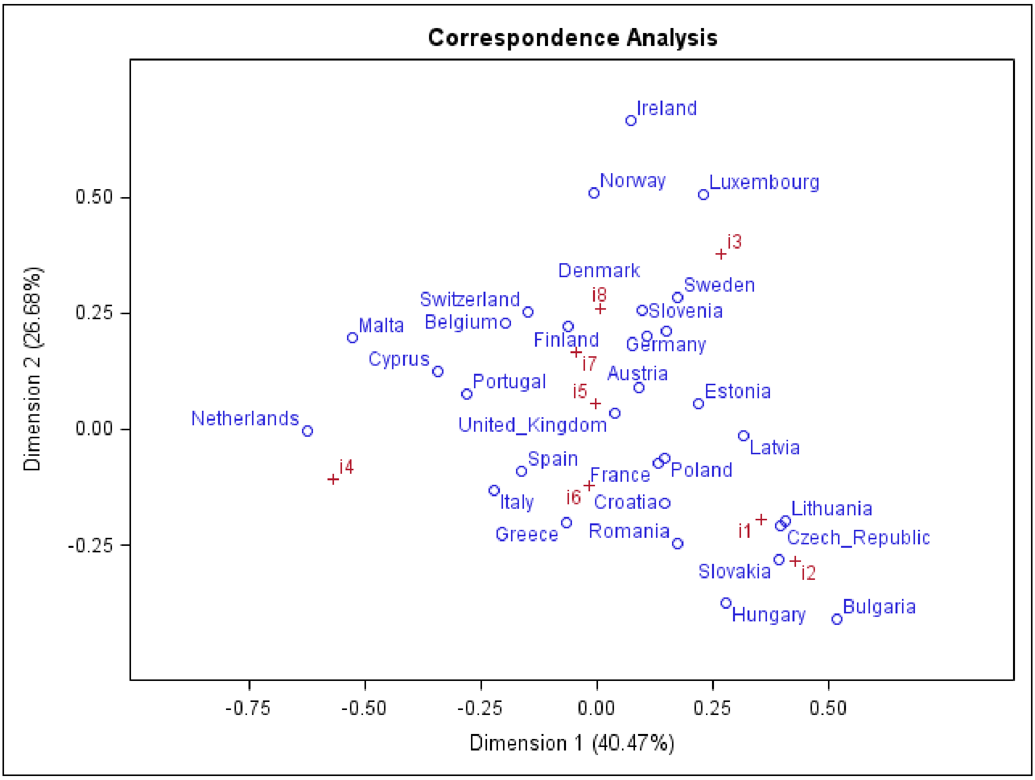
From the information given by index of the coordinates that contribute most to inertia for the column points, the indicators are grouped in two parts:

- Group I: I1 (cereals), I2 (industrial crops), I4 (vegetables and horticultural products).
- Group II: I3 (fodder plants), I5 (potatoes), I6 (crop output), I7 (animals), I8 (animal products).

As in principal components analysis, the results of correspondence analysis are presented on graphs that represent the configurations of points in projection planes, formed by the first principal axes taken two at a time. It is customary to summarize the row and column coordinates in a single plot. However, it is important to remember that in such plots, one can only interpret the distances between row points, and the distances between column points, but not the distances between row points and column points. However, it is legitimate to interpret the relative positions of one point of one set with respect to all the points of the other set

The joint display of row and column points shows the relation between a point from one set and all points of another set, not between individual points between each set. Except in special cases, it is extremely dangerous to interpret the proximity of two points corresponding to different sets of points.

Outlier points plague correspondence analysis. Occasionally, a row or column profile is rare in its set of points that it has a minor role in the determination of the higher order axes. This situation can be discerned easily by considering the point’s contribution to the axes. When a point has a large contribution, at a large principal coordinate at a major principal axis, it is called an outlier. Outlier points should be treated as supplementary variables.



Graph 1. The representation of index and countries in factors plan

The graph above shows the representation of the variables and the observations in the plan determined by dimension 1 and dimension 2. We can determine the way that countries are grouped along the indicators. Identifying 4 major groups, we can say that Netherlands has a high value for indicator 4 (which is vegetables and horticultural products); Bulgaria, Slovakia, Czech Republic and Lithuania have a big production of cereals (including seeds) and industrial crops; Ireland, Norway and Luxembourg are the main producers of forage plants, while the rest of the countries are grouped along the rest of the indicators. It is interesting to observe that the more a country is closer to an indicator (on graph), the more that country is a better producer (or the main producer from European Union) of that product. The points, which do not contribute essentially to the inertia of each axis, are virtually identical to the average profile. Points of a cloud (or set) situated away from the origin, but close to each other have similar profiles.

5. CONCLUSION

In conclusion, we can interpret the correspondence analysis this way: the country that has the lowest distance to an indicator is the main producer of that good. For example, Greece is the main producer for crop output, while Slovakia is the main producer for industrial crops. It is customary to summarize the row and column coordinates in a single plot. However, it is important to remember that in such plots, one can only interpret the distances between row points, and the distances between column points, but not the distances between row points and column points. A point makes a high contribution to the inertia of a principal axis in two ways –when it has a large distance from the barycenter, even if it has a small mass, or when it has a large mass, but a small distance. But, some may say that this conclusion may be seen from the data set chosen. This is true, only that, by the correspondence analysis, we can determine the way that countries are grouping along the indicators chosen, a fact that is not visible in the data set.

BIBLIOGRAPHY

[1] C. Boboc, Analiza statistica multidimensionala. Aplicatii in cadrul studiului produselor si serviciilor, Ed. Meteor Press, 2007

[2] E. Ţiţan, A. Bacescu-Carbunaru, S. Ghiţă, Bazele Statisticii, Ed. Meteor Press, 2002

[3] L. Spircu, M. Calciu, T. Spircu, Analiza datelor de marketing, Ed. All, 1999

[4]http://www.stiucum.com/marketing/cercetarea-de-marketing/Analiza-datelor-de-marketing44747.php

[5] http://forrest.psych.unc.edu/research/vista-frames/pdf/chap11.pdf

[6] http://stefansinclair.name/correspondence-analysis/

[7] http://www.mathematica-journal.com/2010/09/an-introduction-to-correspondence-analysis/

[8] http://www.statsoft.com/textbook/correspondence-analysis/

[9] http://www.upg-bulletin-se.ro/archive/2010-2/11.%20Carbureanu_Ionita.pdf

[10] http://www.tqmp.org/Content/vol07-1/p005/p005.pdf

[11] http://statmath.wu.ac.at/courses/CAandRelMeth/caipA.pdf

[12] http://ro.scribd.com/doc/14480322/Analiza-datelor-Aplicaii-economice