ANALELE ŞTIINŢIFICE ALE UNIVERSITĂŢII „ALEXANDRU IOAN CUZA" DIN IAŞI
Tomul LVI Ştiinţe Economice 2009

THE EVALUATION OF THE REGIONAL PROFILE OF THE ECONOMIC DEVELOPMENT IN ROMANIA

Elisabeta JABA[*], Alina Măriuca IONESCU[**], Corneliu IAŢU[***],
Christiana Brigitte BALAN[****]

Abstract

The main purpose of the paper is to identify the regional profile induced both by the existing resources and the level of development. The paper provides reliable information for regional development policies taking into consideration local resources.

The research uses the multivariate statistical analysis of the macroeconomic statistical data after 1990. The results of the research pointed out the following conclusions:

- *local resources are used to a smaller extent at the regional level;*
- *the correlation between resources and the level of economic development underlines the necessity of adopting a development policy that would better use the present resources.*

Key words: economic development, regional, profile, Romania, multivariate analysis
JEL classifications: O18, R12, R58

1. Introduction

Regional economic development is a key issue at present at European and worldwide level. The main problems that arise are about the attractiveness of economic activities and about the economic consolidation (there are regions which are confronted with the problems of conversion, transition etc.) (Varga and Schalk 2004, 977-989). Building a regional economic profile implies taking into account a range of indicators not only economic but also socio-economic (Chih-Kai, 2008, 21-31). An important role falls to the demand for services, the key point being the accessibility. The importance of the construction of the regional profile is vital for the diagnosis of the intervention of local authorities (Goschin et al., 2008, 80-105).

The characterization of the regional economic development aims to highlight the specificity of the counties and their development prospects, to observe the disparities between the

[*] **Elisabeta JABA** (ejaba@uaic.ro) , PhD, Professor, "Al. I. Cuza" University of Iasi, Faculty of Economics and Business Administration.

[**] **Alina Măriuca IONESCU** (alina.ionescu@yahoo.com), PhD, "Al. I. Cuza" University of Iasi, Faculty of Economics and Business Administration.

[***] **Corneliu IAŢU** (corneliu_iatu@yahoo.fr) , PhD , "Al. I. Cuza" University of Iasi, Faculty of Geography.

[****] **Christiana Brigitte BALAN** (christiana.balan@uaic.ro), PhD, Lecturer, "Al. I. Cuza" University of Iasi, Faculty of Economics and Business Administration.

concerned areas and to provide elements for the allocation of resources and for making the right decisions to diminish these disparities (Coulombe and Tremblay, 2009, 5–18, Arntz and Wilke, 2009, 43–61). The statistical reliability and the relevance of the indicators are key elements to achieve a characterization in accordance with the economic realities of a territory (Serban et al. 2008, 80-106).

This study aims to achieve the identification of a regional profile of socio-economic development of Romania by the means of multivariate statistical methods, to synthesize the many facets of this concept.

2. Variables and method

Initially, there were used 25 variables extracted from the database of the National Institute of Statistics of Romania. These variables are presented in Table 1.

The statistical methods used in the paper are: principal components analysis (PCA) – for the preliminary analysis of the data; cluster analysis – to identify homogenous clusters of Romanian counties according to economic development characteristics; discriminant analysis – to validate the solution obtained with cluster analysis.

Data are recorded at county level, the reference being 2005. Data source is the Statistical Yearbook of Romania 2006. Statistical data processing was conducted using SPSS software.

3. Results

3.1. Preliminary data analysis using PCA

Principal components analysis (PCA) is justified by data set dimension (25 characteristics for the 42 counties), all the 25 variables being quantitative continuous. Using PCA the dimensionality of data is reduced by creating principal components from the original variables (Schott, 2006, 827-843).

In the context of this study, principal components analysis is used in order to explore the original data set and to select the appropriate variables used to identify a regional profile of economic development in Romania.

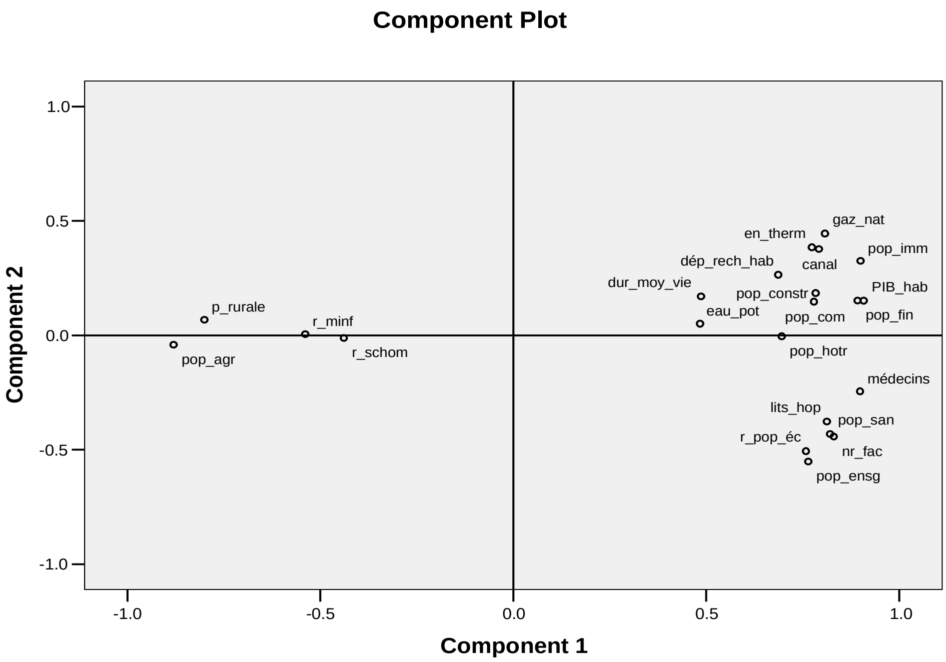
In order to verify the adequacy of data for a factorial analysis, the Barlett's test of sphericity (to test the null hypothesis that the variables in the correlation matrix of the population are uncorrelated), and the indicator MSA (Measure of Sampling Adequacy) of Kaiser-Meyer-Olkin (to evaluate in which degree each variable may be predicted by all the other variables) were used.

The results obtained by data processing with SPSS are presented in Table 2. The significance level associated to Barlett's test of sphericity, $Sig = 0.000$, is smaller than 0.05 (conventional value), which means the null hypothesis of variables' uncorrelation is rejected. Therefore one can conclude that the considered variables are adequate for a PCA. The value of the indicator MSA of KMO (0.798), greater than 0.5 and very closed to 0.8, also indicate the suitability of the considered data for factor analysis (Richarme, 2001).

Another indicator of the adequacy of variables for the considered analysis is the anti-image correlation matrix. Each value of the main diagonal of the matrix shows the measure of sampling adequacy (MSA) for the respective item. In our example the following variables: percentage of the population occupied in industry, abandon rate in primary and secondary education and index of net using the tourists accommodation capacity in function had the values of MSA under 0.5. These variables will be excluded from further analysis because the results indicate they are variables that seem to not be correlated with the structure of the other variables.

The extraction communalities, that are estimates of the variance in each variable accounted for by the components in the factor solution, may also suggest unsuitable variables. In the context of this study, the variables abandon rate in primary and secondary education and index of net use of tourist accommodation capacity in function have values of these estimations under 0.5, and shouldn't be kept in further analysis as they don't fit well with the factor solution.

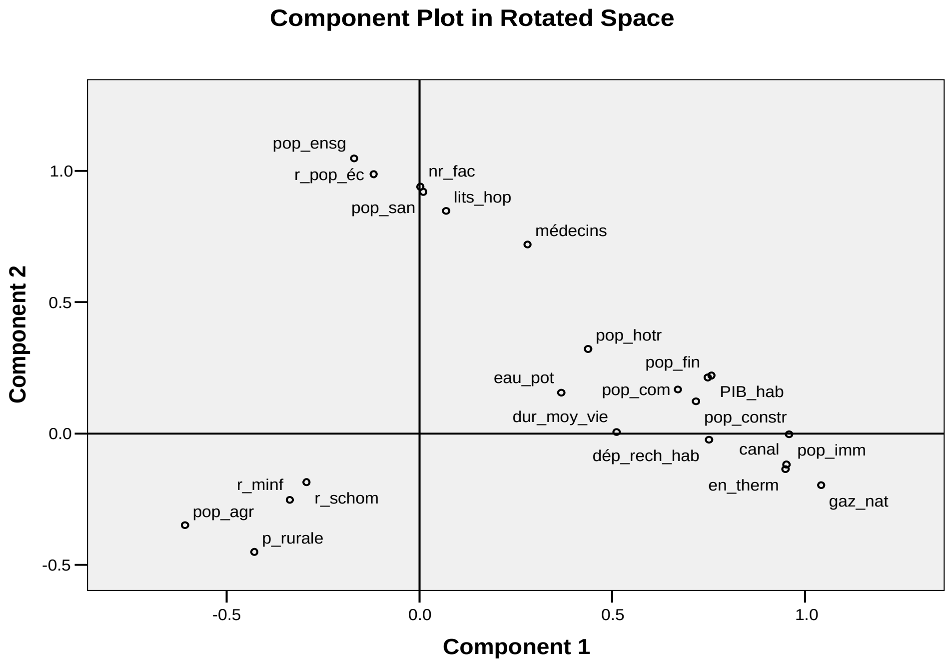
The elimination of the 3 variables (which are not correlated with the structure of the others) from the analysis has resulted in an increase in the measure of sampling adequacy Kaiser-Meyer-Olkin from 0.767 to 0.846, the retained variables being more appropriate for a factor analysis. The explanatory power of the principal components has also improved up to 84.139%, the variance explained by the first two axes increasing from 84.139% to 67.208% of the total variance.



Source: Output obtained in SPSS with PCA

Figure 1: Variables' position on the first two factorial axes

In the graphical representation of the variables' positions on the first two factorial axes (figure 1), one can notice that the first axis opposes, on the one hand, variables that describe the percentage of rural population and the percentage of population occupied in agriculture, and on the other hand, the variables that express the development of the infrastructure and the economic results (Jaba et al., 2007, 1-22).



Source: Output obtained in SPSS with PCA

Figure 2: Variables' position on the first two factorial axes after the rotation of the axes

As some variables present correlation coefficients with the factorial axes that have comparable values on both axes, for a better interpretation of PCA results, it was also generated the rotated solution using an oblique rotation with Direct Oblim method, available in SPSS software.

After axes' rotation, the correlation of the variables with the two axes is better and the principal components are more easily observed: the potential of development (first axis) and resources' quality (second axis).

The analysis of the factorial maps obtained before and after axes' rotation (Figure 2) shows that there are differences in regional profile as regards the economic development and available resources of the counties.

Graphical representation of the counties in the plane of the first two factorial axes (Figure 3) highlights the existence of an outlier (Bucharest). Since the capital of the country presents very different characteristics of economic development compared to other administrative-territorial units, it requires an individual analysis of these features, and it is not included in further analysis.

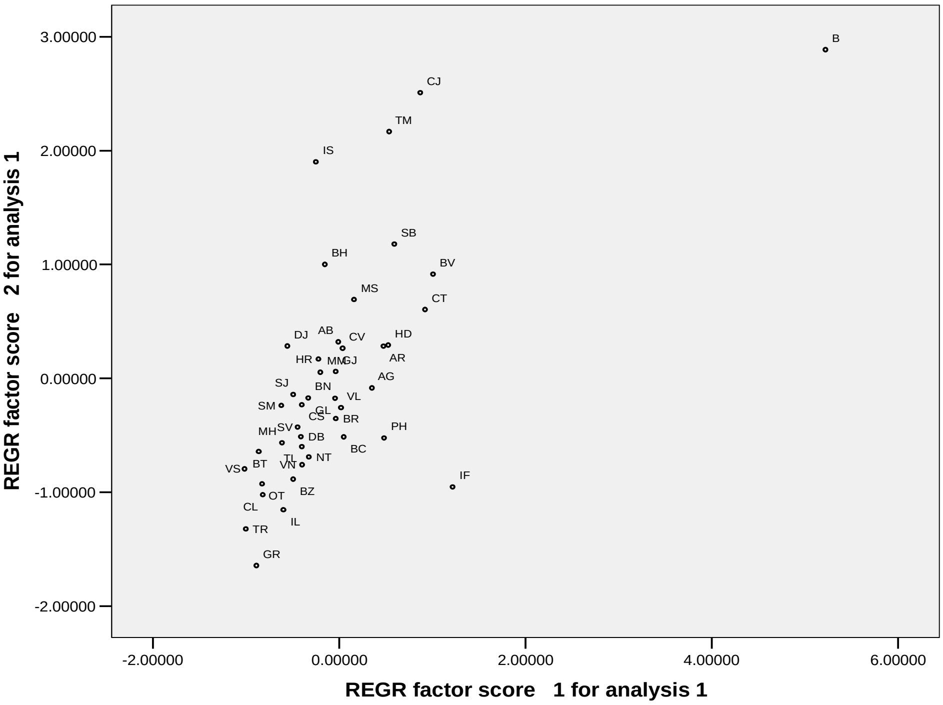


Figure 3: Graphical representation of the counties on the first two factorial axes after the axes' rotation

3.2. Results of the Cluster Analysis

Cluster Analysis is used to identify homogenous groups of counties according to their economic development.

This analysis allows presenting graphically the regional profile of the economic development (Del Campo, et al, 2008, 600-612) by identifying the homogenous clusters of counties according to existing resources and the development level with the aim to optimize the decisions of economic policy.

Due to the fact that the size of the studied population is quite small (n=41 counties after eliminating the outliers), the hierarchical classification method was applied and the squared Euclidian distance measure, frequently used as dissimilarity measure for interval data, was used.

After applying the methods of hierarchical classification available in SPSS, it was noticed that the following methods Within-groups linkage, Complete linkage (Furthest neighbor), and Ward's method clustered most clearly the counties according to the considered variables and resulted in most compact and balanced clusters (Jaba et al., 2008, 123-136).

In order to establish the optimum number of clusters, there is not pre-determined criteria, but useful information on this issue can be drown from the dendrogram and the coefficient agglomeration schedule that show the way in which the counties are combined at each stage of the analysis.

By analyzing independently the dendrogram and the coefficients agglomeration schedule for the three methods, there were identified three possible solutions, each solution grouping the counties in 5 clusters (the optimal solution is presented more detailed in Section 6).

3.3. Results of the Discriminant Analysis (DA)

Discriminant Analysis allows identifying and describing the significant differences among the counties groups.

Discriminant Analysis (Vaughn and Wang, 2008, 315–340) is used in order to find out the solution for which one gets a combination of predictor variables that provide the best discrimination between the clusters of counties.

In our study, the discriminant variables (predictor variables) are considered the 22 independent variables selected by PCA and the grouping variable, the variable that is subject to classification, is considered the cluster membership obtained by Cluster Analysis.

The significant differences between the groups are identified by the discriminant functions, linear combinations of the uncorrelated predictor variables: $\mathrm{D} = \mathrm{b}_1\mathrm{X}_1 + \mathrm{b}_2\mathrm{X}_2 + ... + \mathrm{b}_\mathrm{p}\mathrm{X}_\mathrm{p} + c$ where D=discriminant function; Xj=the vector of discriminating variables; $j = \overline{1, p}$; bj=discriminant coefficients; c=constant.

The use of discriminant analysis implies the following assumptions: the predictor variables have normal multivariate distributions (the normality of the multivariate distributions), the variances are equal among groups (homoscedasticity) and the predictors are not perfectly correlated (lack of multicollinearity).

For testing the predictor variables normality in SPSS, there was used the Kolmogorov–Smirnov test, the examples in the literature being quite numerous (D'Alimonte and Cornford, 2008, 613-620, Solomonoff, 2008, 238-240), and the Levene test for testing the variances homogeneity.

The results of the tests generally show the validation of the assumptions with little exception for the normality and homogeneity assumptions. Discriminant analysis is relatively robust, even when normality and homogeneity assumptions are violated (Lachenbruch, 1975). According to this statement, the discriminant analysis may be applied without influencing the conclusions drown based on its results.

Table 3 shows the percentage of counties correctly classified by the discriminant analysis for each of the 3 clusters solutions. Thus, in our study, for all the 3 solutions, the discriminant function correctly classifies 100% of the total cases, that is, all the 41 counties. A case is correctly classified if it is assigned, by its classification score computed for the discrimination function, to the group which it really belongs to.

The results of the original classification offer over-optimistic estimations. The cross validation may solve this issue as each case in the analysis is classified by the functions derived from all cases other than that case.

The cross validation is a method used for the assessment of the classification rules by estimating the error rate (Lachenbruch and Mickey, 1968, 1-10).

The results of the cross validation highlights that the Complete linkage method correctly classifies the highest number of cases (78 %, that is, 32 of 41 counties) generating the smallest error rate (22%). Consequently, this method is the optimal solution for the counties grouping according to the analyzed variables.

4. Clusters of counties

4.1. Classification of Romanian counties by groups

The dendrogram presented in figure 4 shows clearly the grouping of counties in 5 main clusters:

- cluster 3 groups the most developed counties and with a very important R&D activity (Timis - TM, Cluj - CJ, Iasi - IS);
- in cluster 4 one can identify the counties with a developed infrastructure and public utilities, well-known for their industrial role during the communist regime (Brasov - BV, Sibiu - SB, Hunedoara HD, Constanta - CT);
- cluster 5 consists of only one county Ilfov – IF, that is positively and deeply influenced by the neighboring position to the capital city Bucharest;
- cluster 1 is composed by the counties with a moderate and contrasting economic development (Bihor - BH, Mures - MS, Arad - AR, Tulcea - TL, Maramures - MM, Braila - BR, Bacau - BC, Covasna - CV, Harghita - HR, Caras Severin - CS, Alba - AB, Galati - GL, Gorj - GJ, Prahova - PH, Valcea - VL, Arges - AG);
- cluster 2 groups the less economic developed counties with an important agricultural activity (Calarasi - CL, Olt - OT, Botosani - BT, Vaslui - VS, Buzau - BZ, Mehedinti - MH, Ialomita - IL, Giurgiu - GR, Teleorman - TR, Neamt - NT, Suceava - SV, Dambovita - DB, Bistrita Nasaud - BN, Vrancea - VN, Satu Mare - SM, Salaj - SJ, Dolj - DJ).

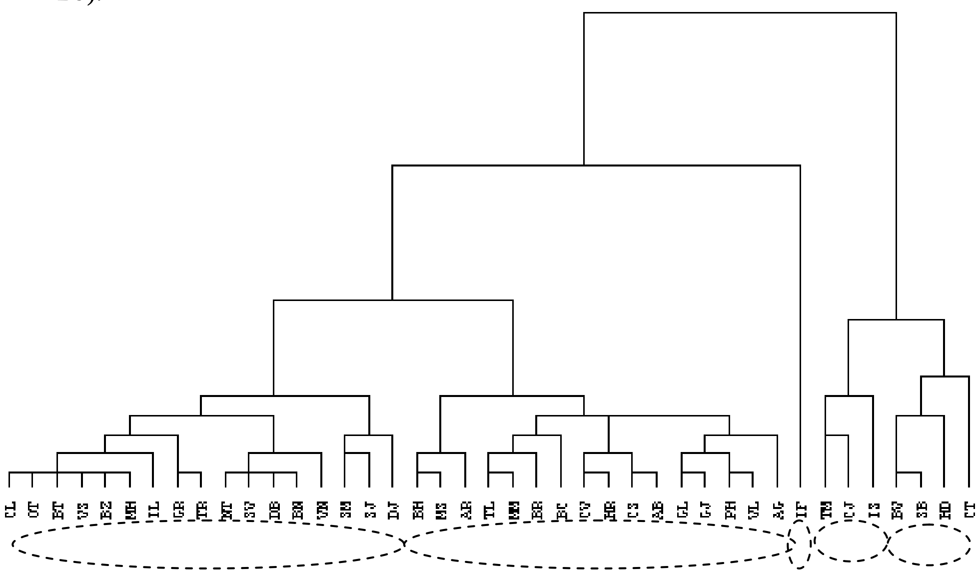


Figure 4: Counties' grouping in 5 clusters

The clusters of counties are highlighted in Figure 5. The superposition of the counties clusters (Figure 5) on the factorial map of PCA (Figure 3) offers some characteristics of the obtained clusters.

The counties in cluster 3 are clearly different of the other counties by their strong correlation with the two factorial axes, having the highest level of economic development. This group of counties has the highest coordinates on the access to health services and education axis which can be explained by the cluster composition: the three counties have a strong tradition in education (especially higher education), culture, health (health services) and, in the same time, an important demographic weight.

Another cluster where the counties are marked by a high level of development is cluster 4. The counties from this cluster have the highest positive contributions on the first factorial axis, these contributions being explained by a good quality of the infrastructure and high values of the economic indicators.

The counties of cluster 1 are represented round the origin of the factorial axes, meaning a moderate level of development. These counties are characterized by important within-counties disparities with strong industrialized core area opposing to less developed ones.

In the second cluster, the counties are displayed in the third dial of the factorial axes plane with the highest negative coordinates on the two factorial axes, showing the lowest development level. The activities of the primary sector are predominant.

The graphical representation of the Figure 5 shows only one atypical county, Ilfov: the highest positive coordinate on the first factorial axis, similarly to the most developed counties, and a negative coordinate on the other axis, similarly to the less developed counties. An important role plays the neighboring of the Ilfov County, as it is near the capital city of Bucharest. This influence is very strong and it explains the paradox in its positioning on the factorial axes.

4.2. The analysis of the counties and clusters positioning on the factorial map

This analysis allows identifying, by the 3σ rule, the most developed and the less developed counties according to each of the factorial axis. One must look for the counties that are situated outside the intervals: $\bar{x} \pm \sigma$, $\bar{x} \pm 2\sigma$ and, respectively, $\bar{x} \pm 2\sigma$ corresponding to the two axes and marked on the graph by stippled lines ($\bar{x} = 0, \sigma = 1$) (Jaba, 2007; Dühr, 2005, 1167-1182).

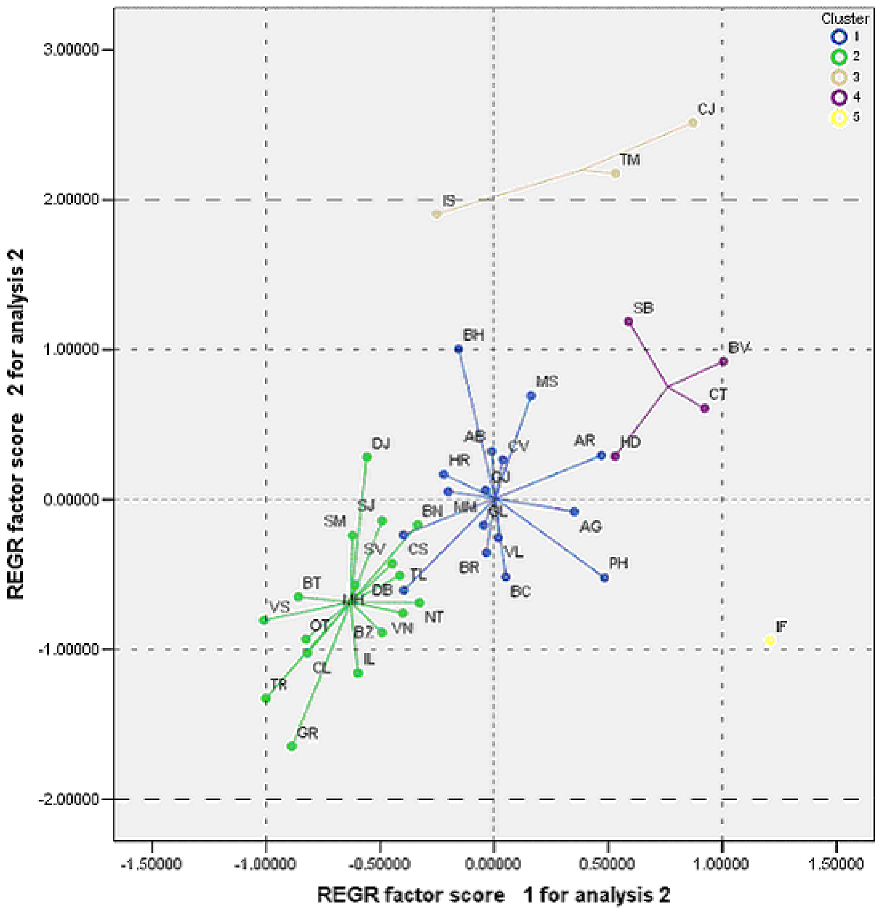


Figure 5: Representation of the clusters of counties on the factorial map

The axes of economic results and the infrastructure of public utilities identify the following counties as the most developed ones: Brasov (BV) and Ilfov (IF). They are followed, closed to the upper bound of the interval $\bar{x} \pm \sigma$, by Constanta (CT) and Cluj (CJ). The lowest economic results and access to public utilities networks are specific to Vaslui (VS) and Teleorman (TR), counties with a predominant agricultural activity.

The most developed counties considering the health and education infrastructure are Bihor (BH), Sibiu (SB), Iasi (IS), Timis (TM) and Cluj (CJ) that are situated at the right of the $\bar{x} \pm \sigma$ interval. The less developed are Calarasi (CL), Ialomita (IL), Tulcea (TL) and Giurgiu (GR) counties situated at the left of the $\bar{x} \pm \sigma$ interval on the second factorial axis. The backwardness of the latter counties is explained both by lack of diversity of economic activities (CL, IL, GR) and by isolation and difficult access (TL).

The map presented in Figure 5 also highlights the counties situated at the bounds of the clusters, any change in their development characteristics making possible the moving towards a neighbor cluster. It is the case of Caras Severin (CS) and Tulcea (TL) counties that are situated in cluster 1 but they strongly resemble to the less development counties of cluster 2. On the contrary, the Bistrita-Nasaud (BN) County from cluster 2 has similar characteristics to the counties in the first cluster.

The Hunedoara County is assimilated to the most developed counties group (important activities in ironworking and mining industries in decline after 1990) though it is situated nearly the group of counties with a moderate development. Sibiu has the highest chances to reach the most developed counties cluster, being the closest to these ones.

Analyzing the regional distribution of counties clusters for the optimal solution, it was noticed that it reproduces in a great extent the geographical distribution, grouping the neighbor counties.

4.3. Geographical distribution of Romanian counties clusters

In a certain way, the solution reproduces the geographical map of counties (Figure 6), and it may prove useful when applying policies for small areas or founding the policies for larger areas such as regions that may consider the regional distribution of the aimed counties.

Considering only the counties when taking decisions on development is a mistake. It should also be taken into consideration the important development disparities among the counties. The disadvantaged areas policy applied in Romania after 1990 tried to solve these issues, but it wasn't very successful.

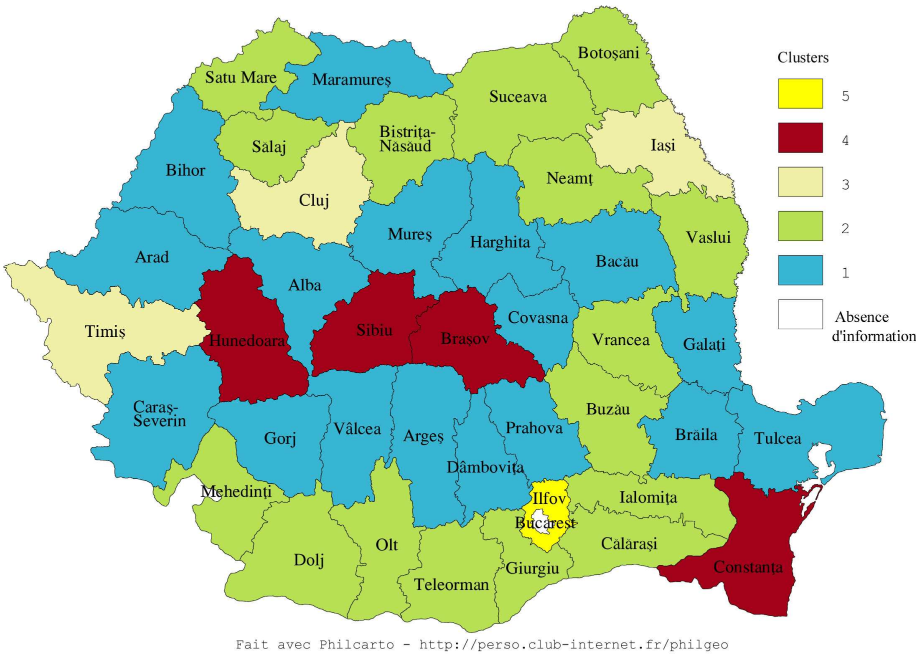


Figure 6: Regional distribution of counties clusters

For a more detailed presentation of the clusters regarding their development level and the available resources, there were determined the mean level of the standardized variables for each cluster. The specific profiles of the clusters are the following:

Cluster 1

- the values are closed to the mean for most of the indicators (Table 4);
- the counties in this cluster are characterized by a moderate access to health and education services, the funds allocated to R&D activities per capita are insignificant;
- the access to utilities is closed to the mean, but slightly higher than the mean for all the utilities, as compared to the other clusters that have an important weight of counties with access to some of the utilities but lack in the same time other utilities;
- the socio-economic indicators (GDP/capita, unemployment rate, rural population rate) are also closed to the mean;
- the population employed in agriculture is more important than for the other clusters, but there is not a particular activity that make a difference from the other counties groups;

Cluster 2

- the highest level of mortality rate, the lowest level of life expectancy rate and the lowest level of access to health services;
- the access to education has a very low level;
- the infrastructure is less developed, having the lowest level of access to natural gas and public canalization equipment networks;
- the counties have a low level of economic development (GDP/capita, minimal investments in R&D activity);
- the urbanization level is small, the unemployment rate is high, the employment rate in agriculture is the highest and the employment rate in other economic activities is very small;

Cluster 3

- in contrast with cluster 2, the counties of this cluster have a high development level, a high level of GDP/capita, important investments in R&D activities, a low unemployment rate, a high urbanization rate, the best health and education infrastructure, the highest life expectancy rate and the lowest rate of child mortality rate;
- the access to utilities is higher than the mean, but the natural gas and public canalization equipment networks are scarcely spread;
- the highest employment rate in health and education services (in chief towns of the counties one can find the most important universities in the country);
- a high employment rate in real estate and construction activities;
- a low employment rate in agriculture.

Cluster 4

- the counties in this cluster are also developed (high GDP/capita, important expenses with R&D activities, good access to all the utilities and a well developed heating energy distribution network);
- a good access to health and education services but in a lower extent that in the counties from the previous cluster;
- the lowest rate of rural population and the lowest employment rate in agriculture;

- the highest level for the employment rate in hotels and restaurants activities (tourist activities very well developed), constructions activity;
- a high employment rate in trade and real estate activities;
- the highest level of the unemployment rate.

Cluster 5

- this cluster consists of only one county, Ilfov (IF) County, that is an atypical case due to the proximity to the capital city;
- the proximity to Bucharest explains mostly the main characteristics of this county, basically rural: the highest GDP/capita, the largest funds granted to R&D activities (important research institutions headquarters), the lowest unemployment rate, the highest rate of employment in trade (the largest warehouses in the country) and real estate activities (the real estate boom is kept up by the capital inhabitants that prefer this area to the crowded city);
- the infrastructure is a paradox: even if the Ilfov County has the most developed natural gas and public canalization networks, it has the lowest access to heating energy and drinking water supply equipment.
- Bucharest has the highest number and the most important medical and educational centers. The Ilfov County population also benefits of these services and therefore, the county has not a well developed health infrastructure by itself. Other characteristics of thus county are the highest life expectancy rate and the lowest rate of child mortality.

5. Conclusions

The clusters obtained are partially homogenous on the inside, but they are very different as it concerns the counties' characteristics and their development level.

The 5 clusters solution gives the possibility to identify the main paths for the foundation of development policies, strategies and programs:

- the development of the natural gas and public canalization networks for cluster 3 that consists of the most developed counties (Timis – TM, Cluj – CJ, Iasi – IS);
- the main issue for cluster 4 is the high unemployment rate, though the counties in this cluster are well developed. Consequently, the measures to be taken should improve the entrepreneurship in order to generate new jobs for a long time and all along the year, and not seasonal jobs. In the Hunedoara County (HD), the high unemployment rate is due not only to the tourist seasonality but also to the recessions of the mining sector after 1990 (carbon extraction). The new jobs would offer a professional alternative to the miners forced to quit the mining industry.
- the improvement of the Ilfov County development level (cluster 5) implies the development of the heating energy and drinking water supply networks, and also the development of health and education networks;
- specific measures to improve the indicators for the cluster of counties moderately developed (cluster 1);
- the allocation of important resources and the implementation of radical programs for the cluster of the less developed counties (cluster 2). These measures should aim firstly the improvement of life quality by a better access to medical services (hiring

medical stuff in the disadvantaged areas, founding the construction of new medical buildings and the extension of the existent ones), and the development of the infrastructure. Another solution aims to encourage the companies willing to invest in the disadvantaged areas, the main objective being the improvement of the economic activities in these counties.

References

Arntz M. and Wilke R. A. “Unemployment Duration in Germany: Individual and Regional Determinants of Local Job Finding, Migration and Subsidized Employment”, *Regional Studies* 43.1, 2009.

Chih-Kai C. “Construct model of knowledge-based economy indicators”, *Transformations in Business and Economics* 7(2), 2008.

Coulombe S. and Tremblay J.F. Migration and Skills Disparities across the Canadian Provinces, Regional Studies 43.1, 2009.

D'Alimonte D. and Cornford D. “Outlier detection with partial information: application to emergency mapping”, *Stochastic Environmental Research and Risk Assessment* 22(5), 2008.

Del Campo C., Monteiro C. M. F. and Soares J. O. “The European regional policy and the socio-economic diversity of European regions: A multivariate analysis”, *European Journal of Operational Research* 187(2), 2008.

Dühr S. “Spatial policies for regional sustainable development – a comparison of graphic and textual representations in regional plans in England and Germany”, *Regional Studie*s 39, 2005.

Goschin Z., Constantin D.L., Roman M. and Ileanu B., “The current state and dynamics of regional disparities in Romania”, *Romanian Journal of Regional Science* 2(2), 2008.

Jaba E. “The “3 sigma” rule used for the identification of the regional disparities”, *Yearbook of the “Gheorghe Zane” Institute of Economic Researches – Jassy* 16, 2007.

Jaba E., Iatu C. and Pintilescu C., “Dynamique de la structure économique en Roumanie et l’impact sur le marché“, *Révue d’Economie Régionale et Urbaine.Grandes questions urbaines et régionales* 4(2), 2007.

Jaba E., Balan C., Roman M., Viorică D. and Roman M. Employment rate prognosis on the basis of the development environment trend displayed by years-clusters, Economic Computation and Economic Cybernetics Studies and Research 42(3-4), 2008.

Lachenbruch P. A. and Mickey M. R., “Estimation of error rates in discriminant analysis”, *Technometrics* 10, 1968.

Lachenbruch P. A., *Discriminant analysis*. Hafner Publishing, New York, 1975.

Schott J. R. A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix, Journal of Multivariate Analysis 97(4), 2006.

Serban D., Mitrut C., Cristache S.E. et al. Intercultural and inter-confessional relations in a Romanian countryside, Journal for the Study of Religions and Ideologies 7(20) , 2008.

Solomonoff R. J. “The probability of "undefined" (non-converging) output in generating the universal probability distribution”, *Information Processing Letters* 106(6), 2008.

Varga A. and Schalk H. Knowledge Spillovers, Agglomeration and Macroeconomic Growth: An Empirical Approach, Regional Studies 38(8), 2004.

Vaughn B. K. and Wang Q. Classification Based on Tree-Structured Allocation Rules, Journal of Experimental Education 76(3), 2008.