

**ANALIZA CLUSTER IN STUDII DE
PROXIMITATE A SOMAJULUI
ÎNREGISTRAT LA NIVELUL
JUDEȚELOR ROMÂNIEI LA
ÎNCEPUTUL CRIZEI ECONOMICE**

Prof. univ. dr. Babucea Ana-Gabriela
Lect. univ. dr. Dănăcică Emanuela-Daniela
Universitatea „Constantin Brâncuși” din
Târgu Jiu

Abstract:

Metoda Cluster clasifică un set de observații în două sau mai multe grupuri de necunoscute ce se exclud reciproc pe baza combinării unor variabile, lucru ce s-a dovedit a fi foarte util. Scopul clasificării este acela de grupare a obiectelor în funcție de asemănările și deosebirile dintre ele, oferind astfel o descriere sintetică sau o diminuare a datelor.

Prin metoda Cluster, vom realiza, în această lucrare, analiza disparităților în județe din România în căutarea numărului de șomeri înregistrați conform celor mai recente date statistice oficiale.

Cuvinte cheie: metoda Cluster, criza economica, statistici, somaj.

1. Metoda Cluster – metodă de analiză multivariată

Pentru a descrie complet un set de date este în general nevoie de un număr mare de variabile. Numărul acesta poate fi redus, fără a pierde din generalitate dacă se apelează la analiza multivariată care constă în esență în prelucrarea simultană a datelor referitoare la mai multe variabile cu scopul de a evidenția eventualele relații ce ar putea să existe între acestea cu scopul generării sau testării de ipoteze statistice. Problema ce se ridică de cele mai multe ori este aceea de a organiza datele observate sau înregistrate astfel încât să aibă semnificație. Una din tehnicile de analiză multivariată prin care avem posibilitatea

**USING CLUSTER ANALYSIS FOR
STUDYING THE PROXIMITY OF
REGISTERED UNEMPLOYMENT AT
THE LEVEL OF COUNTIES IN
ROMANIA AT THE BEGINNING OF
THE ECONOMIC CRISIS**

Prof. PhD Babucea Ana-Gabriela
Lect. PhD Dănăcică Emanuela-Daniela
„Constantin Brâncuși” University of Târgu
Jiu

Abstract:

Cluster analysis classifies a set of observations into two or more mutually exclusive unknown groups based on combination of interval variables and it has proven to be very useful. The classification aim is grouping the objects between their similarities or dissimilarities and so providing a synthetic description or a cut of data.

In this paper we analyze the disparities into the counties of Romania looking the number of registered unemployed according to the latest official statistical data using one technique of clusters analysis.

Key words: Cluster analysis, economic crisis, statistics, unemployment.

1. Cluster analysis – a multivariate analysis method

To fully describe a set of data is generally need a large number of variables. The number can be reduced without loss of generality if choice for multivariate analysis which is essentially the simultaneous processing of data on several variables in order to highlight any relationships that might exist between them in order to generate or test statistical assumptions.

Problem that is raised most often is to organize the observed or recorded data so as to have significance. One of the techniques of multivariate analysis in which we can organize

organizării în grupe omogene a informațiilor cu referire la mai multe variabile este analiza de cluster.

Analiza de cluster (*cluster analysis*) este o metodă de clasificare descriptivă care presupune identificarea într-o mulțime de obiecte a grupurilor similare.

Datele sunt (sau pot fi) organizate ca o matrice (liniile sunt observații, obiecte, coloanele sunt variabile, atribute). În plus, sunt omogene în sensul că are sens calculul distanțelor dintre elemente și există suficient de multe date încât prelucrările statistice elementare nu oferă o imagine satisfăcătoare a structurii datelor. Drept rezultat al clasificării descriptive se obțin grupurile de elemente, clasele identificate.

Termenul de *cluster* se referă la o mulțime de obiecte (elemente) similare între ele și nesimilare obiectelor din alte cluster. În timp ce termenul de *analiză cluster* este utilizat presupune aplicarea unor algoritmi, clasele rezultând ca urmare a unei suite de operații efectuate recursiv sau repetitiv. Analiza de cluster, cunoscută și ca analiză de segmentare sau de taxonomie are ca scop identificarea unui set de grupe omogene prin gruparea elementelor astfel încât să minimizeze variația în cadrul grupei și să maximizeze variația dintre grupe. Este deci o tehnică de analiză uni sau multivariată care cuprinde un număr de algoritmi de clasificare a unor obiecte sau indivizi în grupe omogene.

Problema esențială constă în determinarea (identificarea) clusterelor este cea a specificării proximității (apropierii, similarității) și cum se determină aceasta.

Proximitatea este o noțiune dependentă de problema reală cercetată. Structurile uzuale de date privind obiectele supuse analizei cluster sunt:

- *Matricea de pattern-uri*. Este cazul obiectelor care sunt prezente prin atributele lor – n obiecte și p atribute vor furniza o matrice de tip $n \times p$. Liniile sunt obiecte (*pattern-uri*), coloanele sunt atribute (variabile).

- *Matricea de proximitate*. Elementele $d(i,j)$ reprezintă proximitățile dintre obiectele i și j . Proximitatea poate fi:

in groups of similar information with reference to several variables is cluster analysis.

Cluster analysis is a descriptive classification which involves the identification of many groups of similar objects.

The data are (or can be) organized as a matrix (the lines are comments, articles; the columns are variables, attributes). Furthermore, are homogenous in that sense of distance calculation and there are elements enough statistical data processing that provides a basic no satisfactory picture of the structure data.

As a result of the descriptive classification to obtain groups of elements, classes identified.

The term cluster refers to a variety of objects (elements) similar among themselves and other objects dissimilar clusters while the term is used cluster analysis involves the application of algorithms, resulting in classes as a result of a suite of operations performed recursively or repetitive. Cluster analysis, also known as segmentation analysis or taxonomy is to identify a set of homogeneous groups by grouping items so as to minimize variation within the group and maximize the variation between groups. So, it's a technique of multivariate analysis merged or which contains a number of algorithms for classification of objects or individuals in homogeneous groups.

The priority problem in determining (identifying) clusters is the specification of proximity (proximity, similarity) and how to determine this.

Proximity is a concept related to the real problem investigated. Structures common data objects are subjected to cluster analysis:

- *Matrix of patterns*. Appropriate objects that are present by attribute - n objects and p attributes will provide a matrix of $n \times p$. Lines are objects (patterns), the columns are attributes (variables).

- *Proximity matrix*. Elements $d(i, j)$ is the proximity between objects i and j . Proximity can be:

- *Similarity*, such as correlation coefficient, or

- *Dissociation* (distance,

- *similaritate* (asemănare), cum ar fi coeficientul de corelație, sau

- *disociere* (depărtare, diferențiere), cum ar fi distanța euclidiană.

1.2. Tehnici de clusterizare

Există mai multe tehnici de construire a clusterelor, tehnici bazate pe diverse tipuri de algoritmi:

- *ascendenți* (de agregare, de sinteză) – clasele sunt construite prin agregarea succesivă a elementelor, se obține astfel o ierarhie de partiții, de clase.

- *descendenți* (de divizare) – mulțimea obiectelor este divizată succesiv în submulțimi de obiecte din ce în ce mai omogene; se poate obține și aici o ierarhie a partițiilor.

- *de partiționare* – se pleacă de la o partiție (structură de clase) care se modifică pentru a maximiza omogenitatea fiecărei clase.

Primele două categorii pot fi reunite în clasificarea ierarhică (*hierarchical cluster analysis*). Algoritmul fundamental de clasificare ascendentă ierarhică presupune determinarea perechilor de elemente cele mai apropiate între ele. Prin agregarea lor, se produc noi elemente care se clasifică în același mod până când mulțimea elementelor care se clasifică va avea un singur element. Metoda ierarhică descendentă constă în construirea ierarhiei în ordine inversă: se pleacă de la mulțimea totală a elementelor și la fiecare pas următor, cea mai "mare" (cea mai eterogenă) grupare este divizată în două subgrupări. Algoritmul se oprește atunci când toate grupurile constituite au câte un singur element.

Pentru definirea grupului cel mai eterogen se utilizează diametrul grupului, definit ca *distanța maximă* dintre două elemente din grup. Evident că se pot utiliza și alte metode în acest scop.

Metodele de partiționare din analiza cluster au ca idee esențială aceea că se poate porni de la o partiție oarecare a mulțimii de obiecte și se poate ajunge, prin migrarea obiectelor între clase, la o partiție care îndeplinește un criteriu de optim.

Majoritatea metodelor au drept criteriu

differentiation), such as Euclidian distance.

1.2. Clustering techniques

There are several techniques to construct clusters, techniques based on various types of algorithms:

- *Ascendant algorithms* (aggregation, summary) - classes are built by aggregation of successive elements, to obtain such a hierarchy of partition classes.

- *Descendant algorithms* (the division) - many objects are divided into successively subsets objects continually homogeneous. Can be obtained here as a hierarchy of partitions.

- *Partition algorithms* - to go from one partition (class structure) that is adjusted to maximize the homogeneity of each class.

The first two categories can be brought together in hierarchical classification (hierarchical cluster analysis). Fundamental algorithm ascending hierarchical classification involves determining the pairs of elements close to one another. By aggregating them to produce new elements which are classified in the same way until many elements to be classified will have a single item.

Hierarchical descending method consists in building the hierarchy in reverse order: it goes from the crowd and all the elements at each step following the "high" (most heterogeneous) group are divided into two subgroups. The algorithm stops when all groups have a single item.

Group to define the most heterogeneous group diameter is used, defined as the maximum distance between two elements of the group. Obviously we can use other methods for this purpose.

Partitioning methods of cluster analysis are essential idea that can start from any partition of a set of objects and can be reached by the migration of objects between classes in a partition that satisfies the criterion of optimum.

Most methods have as a criterion for obtaining the optimal partition that minimizes the sum of the squares of errors (the distance from Ward). The error is the distance from an object to its cluster center.

de optim obținerea partiției care minimizează suma pătratelor erorilor (apare la distanța Ward). Eroarea este distanța de la un obiect la centrul clusterului său.

1.3. Distanțe între obiecte

În continuare sunt prezentate metodele uzuale de calculare a distanțelor dintre obiecte (elemente sau grupuri deja constituite). Alegerea unei anumite distanțe modifică grupările care se constituie.

➤ *Metoda celui mai apropiat vecin (nearest neighbor method)*: Distanța dintre două grupuri este distanța minimă dintre două elemente ale grupurilor (distanța dintre cele mai apropiate elemente din clase diferite).

$$d(\pi_1, \pi_2) = \min_{x \in \pi_1, y \in \pi_2} d(x, y) \quad (1)$$

Astfel, un lanț de obiecte apropiate două câte două este atribuit, de regulă, la un singur grup, ceea ce nu produce o grupare acceptabilă.

➤ *Metoda celui mai depărtat vecin (furthest neighbor method)*. Această metodă utilizează calcularea distanței dintre două grupuri drept distanța maximă dintre două elemente ale grupurilor (distanța dintre cele mai depărtate elemente din clase diferite).

$$d(\pi_1, \pi_2) = \max_{x \in \pi_1, y \in \pi_2} d(x, y) \quad (2)$$

Metoda are avantajul că nu aglomerează grupuri legate printr-un lanț. Gruparea obținută corespunde mai bine și grupării intuitive (efectuată de un operator uman).

➤ *Metoda legăturii medii*. Distanța dintre două grupuri este distanța medie dintre perechile de elemente ale grupurilor

$$d(\pi_1, \pi_2) = \frac{1}{n_1 n_2} \sum_{x \in \pi_1} \sum_{y \in \pi_2} d(x, y)$$

➤ *Metoda distanței centrelor (average group linkage)*. Se consideră, ca distanță dintre două grupuri π_1 și π_2 , distanța dintre centrele grupurilor

1.3. Distances between objects

The following are the usual methods of calculating the distance between objects (elements or groups already formed). The choosing a certain distance amended the groups.

➤ *The method of the nearest neighbor (nearest neighbor method)*: The distance between the two groups is the minimum distance between the two groups (the distance between the nearest elements from different classes).

$$d(\pi_1, \pi_2) = \min_{x \in \pi_1, y \in \pi_2} d(x, y) \quad (1)$$

Thus, a chain of objects close second two are assigned, usually to a single group, which does not produce an acceptable group.

➤ *Method of the most distant neighbor (furthest neighbor method)*. This method uses the calculation of distance between two groups as the maximum distance between the two groups (the distance between the far elements of different classes).

$$d(\pi_1, \pi_2) = \max_{x \in \pi_1, y \in \pi_2} d(x, y) \quad (2)$$

The method has the advantage that crowded groups linked by a chain. Grouping obtained corresponds better and intuitive grouping (by a human operator).

➤ *Average link method*. The distance between the two groups is the average distance between pairs of elements of groups

$$d(\pi_1, \pi_2) = \frac{1}{n_1 n_2} \sum_{x \in \pi_1} \sum_{y \in \pi_2} d(x, y)$$

➤ *Center distance method (group average linkage)*. It believes that distance between two groups π_1 and π_2 , the distance between the centers of groups

$$d(\pi_1, \pi_2) = d(G_{\pi_1}, G_{\pi_2}) \quad (4)$$

$$d(\pi_1, \pi_2) = d(G_{\pi_1}, G_{\pi_2}) \quad (4)$$

De remarcat că centrul unui grup este dinamic, fiecare nou element putând produce deplasarea lui. Centrul unui grup format dintr-un singur element este chiar acel element.

➤ *Metoda distanței Ward (Ward's linkage)*. Distanța Ward este bazată pe creșterea “suma de pătrate a erorilor” după contopirea grupurilor într-unul singur. Metoda Ward selectează grupările care minimizează creșterea sumei de pătrate a erorilor.

$$d(\pi_1, \pi_2) = SP(\pi_1 \cup \pi_2) - (SP(\pi_1) + SP(\pi_2))$$

unde $SP(\pi) = \sum_{x \in \pi} |x - \bar{x}_\pi|^2$ (5)

1.4. Arbori de clasificare - Dendrograma

Ca rezultat al algoritmului se obține arborele de clasificare (dendrograma) care este de fapt o sinteză a clasificării. O dendogramă care diferențiază clar clusterelor va avea distanțe mici la ramurile mai îndepărtate ale arborelui și diferențe mari la ramurile apropiate. Când distanțele dintre ramurile îndepărtate sunt mari comparativ cu ramurile apropiate, atunci gruparea nu este chiar eficace, iar dendograma se va interpreta cu prudență. Utilitatea dendogramei poate fi remarcabilă mai ales pentru identificarea acelor obiecte care nu pot fi alăturate nici unui grup fiind excepții ale structurii de grupare și care nu se alătură nici unui grup până la ultimul pas.

1.5. Calitatea clasificării

Deoarece într-o problemă de clusterizare nu se cunoaște numărul de clusterelor *a priori*, evaluarea calității partiției obținute este o etapă foarte importantă. Evaluarea trebuie să ia în considerare atât faptul că, poate, mulțimea inițială nu are o structură bine determinată de clase, cât și faptul că diferite metode conduc la clase diferite.

Procedurile uzuale de evaluare sunt fie vizualizarea dendogramei, fie determinarea unor indicatori de calitate precum coeficienții de divizare (*divisive coefficient*) și de aglomerare (*agglomerative coefficient*) care oferă indicatori

Note that the group is dynamic, every new element of movement may cause. Center of a group consisting of a single item is just that element.

➤ *Distance Ward method (Ward's linkage)*. Ward distance is based on increasing "the amount of square errors" after union groups into one. Ward method selects clusters that minimize the sum of squares increased errors.

$$d(\pi_1, \pi_2) = SP(\pi_1 \cup \pi_2) - (SP(\pi_1) + SP(\pi_2))$$

where $SP(\pi) = \sum_{x \in \pi} |x - \bar{x}_\pi|^2$ (5)

1.4. Classification trees – Dendrogram

As a result of the algorithm is to obtain the classification tree (dendrogram) which is a summary of the classification. A dendrogram has clearly distinguishable clusters which will have smaller distances to more distant branches of the tree and the large differences in the branches close. When distances between remote branches are large compared with the branches close, then the grouping is not effective, and dendogram be interpreted with caution. The utility of the dendogram may be especially notable for identifying those objects that can not be attached to any group being the exception of the structure group and not join any group until the last step.

1.5. Quality of the classification

Because in a matter of clustering not know the number of clusters a priori, to evaluate the quality of the partition obtained is a very important step. Assessment must take account that may crowd the original structure are not well determined by class, and the fact that different methods lead to different classes.

The usual assessment is either viewing dendogram or determination of indicators of quality as factors of division (*Divisive coefficient*) and agglomeration (*agglomerative coefficient*), which provides indicators (average) global:

- *Divisive Coefficient* - DC: For each object, the calculated $d(i)$ as the ratio of the diameter of the last cluster (in the order given

(medii) globali:

- *Divisive Coefficient* - DC: Pentru fiecare obiect i , se calculează $d(i)$ ca fiind raportul dintre diametrul ultimului cluster (în ordinea dată de algoritmul de divizare) la care aparține obiectul înainte de a fi separat ca un singleton și diametrul mulțimii totale de obiecte (clusterul inițial). Atunci:

$$DC = \frac{1}{n} \sum d(i) \quad (6)$$

- *Agglomerative coefficient* - AC este un indice de calitate pentru clasificarea ascendentă: Pentru fiecare obiect i , se calculează $d(i)$ ca fiind raportul dintre disocierea primului cluster (în ordinea dată de algoritmul) la care se atașează obiectul și diametrul mulțimii totale de obiecte (clusterul final).

$$AC = \frac{1}{n} \sum [1 - d(i)] \quad (7)$$

AC tinde să crească o dată cu numărul de obiecte.

2. Analiza comparativă a șomajului la nivelul județelor României aplicând analiza de cluster

Pe fondul crizei economice mondiale și a acutizării efectelor acesteia în România, piața forței de muncă evoluează încet dar sigur într-o direcție puțin favorabilă.

Analiza șomajului este cu atât mai importantă cu cât se pune tot mai insistent problema efectelor crizei economice și în România. Vom încerca aplicând algoritmul *Hierarchical Cluster* cu produsul program SPSS for Windows pentru a analiza efectele similare ale crizei economice la sfârșitul lunii februarie 2009 din perspectiva numărului de șomeri înregistrați.

2.1. Setul de indicatori și datele considerate

Conform datelor oficiale rata șomajului înregistrat în luna februarie a fost de 5,3%, față de 4,9% în luna ianuarie 2009 și 4,2% în aceeași

by the division algorithm) that belonged to the subject before being separated as a Singleton and total diameter set of objects (first cluster). Then:

$$DC = \frac{1}{n} \sum d(i) \quad (6)$$

- *Agglomerative coefficient* - AC is an index of quality for ascending classification: For each object, the calculated $d(i)$ as the ratio of the dissociation first cluster (in order of time algorithm) that is attached to the object diameter and the total set of objects (final cluster).

$$AC = \frac{1}{n} \sum [1 - d(i)] \quad (7)$$

AC tends to increase once the number of objects.

2. Comparative analysis of unemployment in the counties of Romania applying cluster analysis

On the background of global economic crisis and its effects in Romania, the employment market is evolving slowly but surely in a direction least favorable. Analyze of unemployment is especially important as it puts more strongly question the economic crisis in Romania. We will try applying the algorithm Hierarchical Cluster with the program SPSS for Windows to analyze the effects of similar economic crisis at the end of February 2009 from the number of registered unemployed.

2.1. The set of indicators and data considered

According to official registered unemployment rate in February was 5.3% versus 4.9% in January 2009 and 4.2% in the same month of 2008. These data show that since the beginning of the labor market has come close to maximum forecast of the Romanian government to end this year. The Executive prognosticates for 2009 a high unemployment rate of 5.5%.

The total number of unemployed

lună a anului 2008. Aceste date arată că încă de la începutul anului situația pe piața muncii a ajuns aproape de maximele prognozate de Guvernul român până la sfârșitul acestui an. Executivul pronostica pentru 2009 un nivel maxim al ratei șomajului de 5,5%.

Numărul total al șomerilor înregistrați în evidențele agențiilor județene pentru ocuparea forței de muncă a fost de 477.860 de persoane, din care 207.574 femei. Din totalul șomerilor înregistrați la nivel național, 202.470 sunt șomeri indemnizați, iar 275.390 neindemnizați. La agențiile județene ale ANOFM s-au înregistrat în februarie 480 de mii de șomeri, cu aproape 100 de mii în plus față de aceeași lună a anului trecut și cu peste 30 de mii mai mulți decât în prima luna a lui 2009.

Dacă avem în vedere numai indicatorul rata șomajului se constată disparități evidente între județele țării: în județele: Ilfov (1,5%), București (1,7%), Timiș (2,1%), Bihor (3,4%), Cluj (3,7%), Arad (3,9%), Sibiu (4%), Satu Mare (4,1%), Bistrița și Constanța (cate 4,2%), Maramureș (4,3%), Botoșani (4,8%), Giurgiu, Prahova și Suceava (cate 4,9%), Mureș (5%), rata șomajului a fost sub procentul înregistrat la nivel național în timp ce județele cele mai afectate de șomaj au fost Vaslui (11%), Mehedinți (10,3%), Teleorman și Dolj (câte 9%), Alba (8,3%), Galați (8,2%), Gorj și Covasna (câte 8%).

Pentru a realiza o clasificare a județelor după nivelul șomajului înregistrat ca prim efect al crizei economice și pentru a identifica similaritatea între ele s-au folosit datele oficiale publicate în Buletinul Statistic nr. 2/2009 (Tabelul nr. 1) pentru un set de indicatori relevanți: *rata șomajului înregistrat* (raportul dintre numărul șomerilor și populația activă civilă) și alți indicatori expresivi ca *numărul total al șomerilor înregistrați* la ANOFM, *numărul total al șomerilor indemnizați* și *numărul total al șomerilor neindemnizați* și *numărul șomerilor proveniți din sectorul privat*.

2.2. Rezultatele obținute în urma clusterizării

Apelând la analiza cluster se dorește clasificarea județelor în funcție de variabilele considerate în ipoteza că fiecare dintre acestea

registered in the records of county agencies for employment was 477,860 persons, of which 207,574 women. Of the total unemployed registered at the national level, 202,470 are unemployed indemnity and 275,390 no indemnity. At the NAE county agencies were registered in February 480 thousand unemployed, with almost 100 thousand in addition to the same month last year and over 30 thousand more than in the first months of 2009.

If we consider only the unemployment rate is found obvious disparities between the counties of the country: in the counties: Ilfov (1.5%), Bucharest (1.7%), Timiș (2.1%), Bihar (3.4%) Cluj (3.7%), Arad (3.9%), Sibiu (4%), Satu Mare (4.1%), Bistrița and Constanta (by 4.2%), Maramureș (4.3%) , Botoșani (4.8%), Giurgiu, Prahova and Suceava (by 4.9%), Mureș (5%), unemployment rate was below the rate recorded at the national level while the counties most affected by unemployment were Vaslui (11%), Mehedinti (10.3%), and Teleorman Dolj (each 9%), White (8.3%), Galați (8.2%), Gorj and Covasna (8%).

To achieve a classification of the counties where the registered unemployment in the first effect of economic crisis and to identify the similarity between them have used the official data published in Statistical Bulletin no. 2 / 2009 (Table no. 1) for a set of relevant indicators: the unemployment rate recorded (the number of active population unemployed and civil) and other significant indicators that the total number of unemployed in the NAE, the total number of registered unemployed, the total number of indemnity, or no indemnity unemployed and the number of unemployed people from the private sector.

2.2. Results obtained using cluster analysis

Using cluster analysis is desirable classification counties according to the variables considered in the assumption that each of them is based on the geographical situation and the specific activities of certain elements characteristic. He made a hierarchical cluster analysis - using *Hierarchical Cluster Method* to the most distant neighbor (furthest

are în funcție de așezarea geografică și specificul activităților economice anumite elemente caracteristice.

S-a realizat o analiză cluster ierarhică – **Hierarchical Cluster** folosind *Metoda celui mai depărtat vecin (furthest neighbor method - Complete linkage)* cu *Squared euclidian distance* ca metodă de calcul a distanțelor pe cazurile avute la dispoziție cu pachetul de programe SPSS for Windows.

neighbor method - Complete linkage) with squared Euclidian distance as a method of calculating the distance taken on cases with available software package SPSS for Windows.

Numărul total de șomeri înregistrați și rata șomajului pe țară și pe județe la data de 28.02.2009 / The number of registered unemployed and unemployment rate in the country and counties on 28.02.2009

Tabel nr. 1

Nr. ctr.	JUDETUL	Numar total someri	Nr. total som. indemnizati	Nr. total som. neindemnizati	Total someri din sector privat	Rata șomajului FEB (%)
0	1	2	4	5	6	7
1	ALBA	15238	9174	6064	14719	8,3
2	ARAD	8368	4690	3678	7855	3,9
3	ARGES	16152	9304	6848	8750	5,9
4	BACAU	15735	7863	7872	4950	6,7
5	BIHOR	9800	4359	5441	4515	3,4
6	BETRIUA	5469	4045	1424	3343	4,2
7	BOTOSANI	7672	3248	4424	7069	4,8
8	BRASOV	13620	5296	8324	11138	5,5
9	BRAILA	7266	2315	4951	2487	5,3
10	BUCURESTI	18575	5102	13473	7559	1,7
11	BUZAU	12727	3809	8918	10456	6,6
12	CAJAS SEVERIN	9273	3499	5774	8909	7,0
13	CALARASI	6319	1950	4369	2657	5,9
14	CLUJ	12672	5566	7106	12241	3,7
15	CONSTANTA	13288	7507	5781	10635	4,2
16	COVASNA	7564	2628	4936	7163	8,0
17	DAMBOVITA	12141	2880	9261	10034	5,7
18	DOLJ	26272	7886	18386	16135	9,0
19	GALATI	17869	5161	12708	7270	8,2
20	GIURGIU	4452	1452	3000	4099	4,9
21	GORJ	11864	6012	5852	11699	8,0
22	HARGHITA	10619	6368	4251	10417	7,6
23	HUNEDOARA	15881	8071	7810	15202	7,6
24	IALOMITA	7189	3085	4104	6750	6,7
25	IASI	19720	4924	14796	14356	6,3
26	ILFOV	2203	764	1439	1299	1,5
27	MARAMURES	8856	3771	5085	2509	4,3
28	MEHEDINTI	12750	2757	9993	9180	10,3
29	MURES	12462	5055	7407	12071	5,0
30	NEAMT	10725	4353	6372	3061	5,3
31	OLT	11395	5061	6334	9856	6,2
32	PRAHOVA	15471	10059	5412	10368	4,9
33	SATU-MARE	6319	3429	2890	4501	4,1
34	SALAJ	7123	3462	3661	6779	6,7
35	SIBIU	7418	4338	3080	6994	4,0
36	SUCEAVA	12349	6891	5458	11646	4,9
37	TELEORMAN	15507	3690	11817	4913	9,0
38	TIMIS	7237	4625	2612	6757	2,1
39	TULCEA	4884	2399	2485	4549	5,3
40	VASLUI	17726	6336	11390	16799	11,0
41	VALCEA	11480	6802	4678	10153	6,4
42	VRANCEA	8190	2284	5906	4184	5,4
	TOTAL	477860	202.470	275.390	347.707	5,3

Sursa: Buletin statistic lunar. nr. 2/2009, pag. 130 / Source: Monthly statistical bulletin. no. 2 / 2009, p. 130

S-a optat pentru afișarea în output a tuturor combinațiilor din fiecare iterație, distanțe etc. cu opțiunea *Agglomeration schedule*, afișarea distanțelor sau similaritățile dintre elemente cu opțiunea *Proximity matrix* și afișarea apartenenței la clustere în una sau mai multe iterații cu opțiunea *Cluster Membership* – din modulul Statistics.

De asemenea, s-a optat pentru reprezentarea grafică de tip diagramă de aglomerare în formatul *Dendrogram*. În reprezentarea dendrogramei, distanțele dintre elementele care se unesc sunt transformate pe o scală 0 – 25, cu păstrarea raportului distanțelor.

Was chosen for displaying the output of all combinations of each iteration, distances, etc.- *Agglomeration with the schedule*, showing distances or similarities between elements of the matrix and display - *Proximity belonging to cluster in one or more option iterations - Cluster Membership - Statistics from the module*. It was also chosen for the graphical representation of the agglomeration in chart format *Dendogram*. In dendogram representation, distances between items those are processed together on a scale 0 - 25, with keeping the distance.

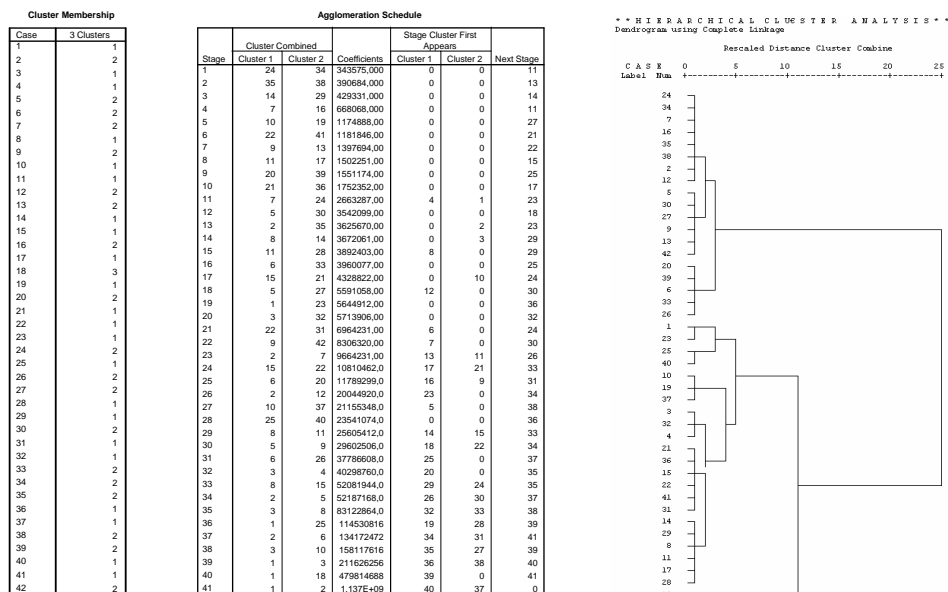


Fig. 1 - Componența clusterelor și Dendograma / Composition of clusters and Dendrogram

3. Concluzii

Analizând dendograma din figura 1 se constată că dacă procesul iterativ începe cu 9 grupe de clase oarecum omogene și județul Dolj. La un nivel de agregare sub 5, grupele se restrâng la 4 plus județul Dolj.

Având în vedere setul de indicatori considerați pentru caracterizarea șomajului la un grad de agregare între 5 și 10 județele României și Municipiul București se pot clasifica în 3 cluster după cum urmează:

- din clusterul 1 fac parte județele: 1- Alba, 3- Argeș, 4 - Bacău, 8 - Brașov, 10 – Municipiul București, 11 - Buzău, 14 - Cluj, 15 - Constanta, 17 - Dâmbovița, 19 - Galați, 21- Gorj, 22 - Harghita, 23 - Hunedoara, 25 - Iași, 28 - Mehedinți, 29 - Mureș, 31 - Olt, 32 - Prahova, 36 - Suceava, 37- Teleorman, 40 – Vaslui și 41 – Vâlcea,

- din clusterul 2 fac parte județele: 2 – Arad, 5 – Bihor, 6 – Bistrita, 7 - Botosani, 9 – Braila, 12 -Caras Severin, 13 – Calarasi, 16 – Covasna, 20 – Giurgiu, 24 – Ialomita, 26 – Ilfov, 27 – Maramures, 30 – Neamt, 33 - Satu-Mare, 34 - Salaj, 35 – Sibiu, 38 – Timis, 39 – Tulcea, 42 – Vrancea;

- județul 18 – Dolj este oarecum diferit de structura întregii colectivități făcând singur parte din cel de-al treilea cluster care aderă la clusterul 2

3. Conclusion

Analyzing the dendrogram in Figure 1 is found that iterative process starts with 9 groups of relatively homogeneous classes and Dolj. At a level of aggregation than 5, groups 4 to restrict the addition Dolj.

Given the set of indicators considered for the characterization of unemployment to a level of aggregation between 5 and 10 counties of Romania and Bucharest City are classified into 3 clusters as follows:

- Cluster no. 1 counties: Alba, 3 - Argeș, 4 - Bacău, 8 - Brasov, 10 - Municipality of Bucharest, 11 - Buzău, 14 - Cluj, 15 - Constanta, 17 - Dâmbovița, 19 - Galati, 21 - Gorj, 22 - Harghita, 23 - Hunedoara, 25 - Iasi, 28 - Mehedinti, 29 - Mureș, 31 - Olt, 32 - Prahova, 36 - Suceava, 37 - Teleorman, 40 - and 41 Vaslui – Vâlcea;

- Cluster no. 2 counties: 2 - Arad, 5 - Bihar, 6 - Bistrita, 7 - Botosani, 9 - Braila, Caras-Severin, 12, 13 - Calarasi, 16 - Covasna, 20 - Giurgiu, 24 - Ialomita, 26 - Ilfov, 27 - Maramures, 30 - round 33 - Satu-Mare, 34 - Salaj, 35 - Sibiu 38 - Timis, 39 - Tulcea, 42 - Vrancea;

- Cluster no. 3: County 18 - Dolj is somewhat different from the whole mass by one part of the third cluster joining the cluster 2 to a

la un grad de agregare între 10 și 15.

Se constată că județul Dolj, prin prisma indicatorilor considerați, diferă de structura considerată, el încadrându-se în una din clasele omogene de abia în ultimul pas al algoritmului. Constatăm, de altfel, numărul cel mai mare de șomeri înregistrați și respectiv de șomeri neindemnizați și o rată a șomajului dintre cele mai mari din țară situație neîntâlnită la niciun alt județ.

Trebuie avut în vedere faptul că metodele analizei cluster vor determina întotdeauna o grupare care poate fi sau nu utilă în clasificarea obiectelor. Dacă grupările fac diferența între variabilele nefolositoare grupării considerate și acele diferențieri sunt utile atunci analiza de cluster este utilă.

4. Bibliografie

1. Babucea, Babucea, A.G., - *The analysis of Romanian Regional Disparities Using Statistical Methods*, International Scientific Conference *Globalization – Social and Economic Impacts '03*, 2003, Ražečke Teplice, Slovacia, p.3-8;
2. Babucea A.G., - *Algoritmi de clasificare utilizând analiza de cluster și SPSS for Windows*, Lucrările sesiunii internaționale "Integrare europeană în contextul globalizării", Pitești, 2003, Ed. Agir, p.409-418;
3. Babucea, A.G., - *Analiza cluster în statistica teritorială*, Lucrările științifice ale simpozionului internațional ROPET 2003, Ed. Universitas, p.13-18;
4. Dănăcică, D.E., - *Cluster Analysis in the Study of Life Quality on the Central and Eastern European Countries*, Zagreb International Review of Economics and Business, ZIREB, Croatia, Vol. X, Nr. 1, 2006, pag. 45-61;
5. Furtună, T.F., - *Algoritmi de clasificare în statistica teritorială*, Revista de statistică, nr. 2/2002, p.72-80;
6. Voineagu, V., Furtună, F., Voineagu, M., Ștefănescu, C., *Analiza factorială a fenomenelor social-economice în profil regional*, Ed. Aramis, 2002.

degree of aggregation between 10 and 15.

It notes that Dolj County, through the indicators considered, differs from the structure as it falling into homogeneous classes until the last step of the algorithm. See, moreover, the largest number of registered unemployed and unemployed no indemnity and unemployment rate of the highest in the country no such situation in any other county.

It must be borne in mind that methods of analysis will determine anytime cluster groups which may or may not be useful in the classification of objects. If groups are the difference between groupings variables considered useless and differentiation that are useful when cluster analysis is useful.

4. Bibliography

1. Babucea, A.G., - *The analysis of Romanian Regional Disparities Using Statistical Methods*, International Scientific Conference *Globalization – Social and Economic Impacts '03*, 2003, Ražečke Teplice, Slovacia, p.3-8;
2. Babucea A.G., - *Algoritmi de clasificare utilizând analiza de cluster și SPSS for Windows*, Lucrările sesiunii internaționale "Integrare europeană în contextul globalizării", Pitești, 2003, Ed. Agir, p.409-418;
3. Babucea, A.G., - *Analiza cluster în statistica teritorială*, Lucrările științifice ale simpozionului internațional ROPET 2003, Ed. Universitas, p.13-18;
4. Dănăcică, D.E., - *Cluster Analysis in the Study of Life Quality on the Central and Eastern European Countries*, Zagreb International Review of Economics and Business, ZIREB, Croatia, Vol. X, Nr. 1, 2006, pag. 45-61;
5. Furtună, T.F., - *Algoritmi de clasificare în statistica teritorială*, Revista de statistică, nr. 2/2002, p.72-80;
6. Voineagu, V., Furtună, F., Voineagu, M., Ștefănescu, C., *Analiza factorială a fenomenelor social-economice în profil regional*, Ed. Aramis, 2002.