ANNALS OF THE UNIVERSITY OF ORADEA
Fascicle of Management and Technological Engineering
ISSUE #2, AUGUST 2016, http://www.imtuoradea.ro/auo.fmte/

THE ANALYSIS OF THE MAIN CHARACTERISTICS CONSIDERED FOR THE GROUPING EU MEMBER STATES ACCORDING TO GROSS INLAND ENERGY CONSUMPTION

Catalin POPESCU,
Petroleum-Gas University from Ploiesti, cpopescu@upg-ploiesti.ro

***Abstract*—** Gross inland energy consumption represents the quantity of energy necessary to satisfy inland consumption of the geographical entity under consideration. It describes the total energy needs of a country. According to some Eurostat reports from 2013 and 2014, over the last two decades, gross inland energy consumption in the EU28, as can be noted, has changed pretty much: in 1990 cumulated 1670 million tonnes of oil equivalent (Mtoe), rose to a peak of 1830 Mtoe in 2006 and then decreased to 1680 Mtoe in 2012. These contradictory values required a comprehensive analysis and therefore present paper describes the main characteristics related to the grouping EU Member States according to gross inland energy consumption in order to discover some common patterns among EU member states.

***Keywords*—** ANOVA test, clusters, Gross inland energy consumption, total renewable parameters

I. Introduction

Gross inland energy consumption, sometimes abbreviated as gross inland consumption, is the total energy demand of a country or region. Gross inland energy consumption covers:

1) *consumption by the energy sector itself;*
2) *distribution and transformation losses;*
3) *final energy consumption by end users;*
4) *some 'statistical differences'.*

The difference between *gross inland energy consumption* and *gross (energy) consumption* is that in gross energy consumption the transformation output (electricity or heat produced from other energy sources) is included.

Therefore, gross energy consumption is a product-specific consumption and does not reflect the demand for primary energy. In the analysis, the main table used is related to the Gross inland consumption (GIC).

To calculate the (GIC) values the relation is: Final Consumption + (Transformation input + Consumption of the energy branch + Distribution losses) - (Transformation output + Exchanges and transfers, returns).

According to [1], the gross inland consumption of each EU Member State depends, to a large degree, on the structure of its energy system, the availability of natural resources for primary energy production, and the structure and development of each economy; this is true not only for conventional fuels and nuclear power, but also for renewable energy sources.

On the other hand, energy intensity is measured as the ratio between gross inland consumption of energy and gross domestic product (GDP); this indicator is a key indicator for measuring progress under the Europe 2020 strategy [2] for smart, sustainable and inclusive growth. If an economy becomes more efficient in its use of energy and its GDP remains constant, then the ratio for this indicator should fall.

Gross inland consumption of energy is calculated as the sum of gross inland consumption of five energy types: coal, electricity, oil, natural gas and renewable energy sources. In fact, *energy intensity* measures the energy consumption of an economy and its energy efficiency. Also, *the GDP figures* are taken at constant prices to avoid the impact of inflation.

It should be noted that the economic structure of an economy plays an important role in determining energy intensity, as service-based economies will display relatively low energy intensities, while economies with heavy industries may have a considerable proportion of their economic activity within industrial sectors, thus leading to higher energy intensity.

Also, due to [1] there are many factors that impact on energy use for transport, for example, overall economic growth, the efficiency of individual transport modes, the take-up of alternative fuels, advances in transport technology and fuel, and lifestyle choices.

Within the paper is carrying out a cluster analysis as a statistical method for grouping a set of data objects into clusters. Is necessary choosing an adequate clustering method because a good clustering method produces high quality clusters with high intra-class similarity (similar to the objects in the same cluster) and low inter-class similarity (dissimilar to the objects in other clusters). Also the paper contained an approach related with hierarchical clustering that has as result the design of a dendrogram. Also as part of the methodology in order to test the statistical significance of the variables inclusion in clusters is used ANOVA. The analyses of variances technique in One-way Analysis of Variance (ANOVA) determine whether the mean of a variable differs significantly between groups.

II. METHODOLOGY

Departing from the main objective of this article, namely the analysis of the main characteristics taken into account when grouping EU-28 member states into clusters according to the Gross inland energy consumption, out of the eight categories indicated in the Energy Balance 2014 (as is reachable on the website http://ec.europa.eu/eurostat/web/energy/data/energy-balances), four indicators have been selected: Gross inland solid fuels energy consumption, Gross inland oil energy consumption, Gross inland gas energy consumption, and Gross inland total renewables energy consumption.

After the first analysis of the data, taking into account the fact that data on all four indicators haven't been identified for Cyprus and Malta as well as the fact that, from the point of view of Gross inland energy consumption, their weight within the EU-28 is quite reduced, the survey was narrowed down to 26 states.

In order to group the states according to the values of the indicators considered, a cluster analysis [3] and specifically a hierarchical cluster analysis was carried out.

Starting from the matrix $X = \{x_{ij}\}_{i=\overline{1,n},\, j=\overline{1,m}}$ where n=26 (the number of countries), and m=4 (the number of variables) the Z-sore transformation was applied to each of its elements as follows:

$$y_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \qquad (1)$$

where $\bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}$ and $\sigma_j = \sqrt{\frac{\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2}{n-1}}$.

It resulted in the matrix:

$$Y = \left\langle y_{ij} \right\rangle_{i=\overline{1,26},\, j=\overline{1,4}} \qquad (2)$$

with normal values (the mean is 0 and the dispersion coefficient is 1) of the initial values of the 4 variables.

In order to determine the proximity matrix (W) the Chebychev distance has been selected

$$w_{ik} = \max_i \left| y_{ik} - y_{ij} \right|,\ j = \overline{1,24}, k = \overline{1,4},\ j \neq i, k \neq i \qquad (3)$$

In order to determine the distance between clusters, the average linkage within groups' method was used.

If $\Phi = \{r_i\}_{i=\overline{1,N_\Phi}}$ and $\Psi = \{p_i\}_{i=\overline{1,N_\Psi}}$ represents the two clusters, then the distance between them is:

$$D(\Phi,\Psi) = \frac{1}{(N_\Phi + N_\Psi)(N_\Phi + N_\Psi - 1)} \cdot \sum_{R,P \in \Phi \cup \Psi} d(R,P) \qquad (4)$$

For data processing and analysis the SPSS conformable to [4] and [5] as well as the Excel [6] software program (namely the Data Analysis module) were used.

In order to test the statistical significance of the variables' inclusion in clusters, the ANOVA was used, hence the hypotheses of the test were:

H_{0_1}: the analysed variable is not significant in relation to the inclusion in the cluster (the variance among groups does not differ significantly from the variance within the groups)

H_{1_1}: the analysed variable is significant in relation to the inclusion in the cluster (the variance among groups is significantly greater than the variance within the groups).

In order to test the statistical significance of the mean values pertaining to the analysed variables, the Confidence interval for means, for a Confidence level of 95% and 90%, respectively, was determined within the clusters. The tested hypotheses are:

H_{0_2}: the mean of the analysed variable is not statistically significant in relation to the inclusion in the cluster (the confidence interval for the variable mean includes the 0 value; the limits of the confidence interval are of different signs).

H_{1_2}: the mean of the analysed variable is statistically significant in relation to the inclusion in the cluster (the confidence interval for the variable mean does not include the 0 value; the limits of the confidence interval have the same sign).

III. RESULTS AND DISCUSSIONS

The first significant clusters were obtained by halting the generation of 14 clusters. In order to test the statistical significance of the inclusion of the four cluster grouping variables, the variational analysis was used. The results obtained from the application of ANOVA are rendered in Table I.

Considering that for the chosen significance threshold, for all the four variables *Sign.F*=0.00<α=0.05 (Table I) the null hypothesis (H0) is rejected whereas the alternative hypothesis is accepted, therefore all the variables are significant in relation to the inclusion in the cluster.

This fact is emphasized by the values of the F-statistic which, in all four cases, has far greater F-statistic values than the critical value $F_{0.05;11;14} \cong 2.39$.

The resulting dendrogram is illustrated in Figure 1.

TABLE I
THE ANOVA TABLE FOR TESTING THE STATISTICAL SIGNIFICANCE OF THE SOLID FUELS, OIL (TOTAL), GAS AND TOTAL RENEWABLES INDICATORS CORRESPONDING TO THE 12 CLUSTER GROUPING

		Sum Of Squares	Df	Mean Square	F	SIG.
Solid fuels	Between Groups	7758724438	11	705338585.3	144.11	0.00
	Within Groups	68522610	14	4894472.2		
	Total	7827247048	25			
Oil (total)	Between Groups	19096228728	11	1736020793.5	89.47	0.00
	Within Groups	271649714	14	19403551.0		
	Total	19367878442	25			
Gas	Between Groups	8689093260	11	789917569.2	100.07	0.00
	Within Groups	110513310	14	7893807.9		
	Total	8799606571	25			
Total Renewables	Between Groups	2031027356	11	184638850.6	305.51	0.00
	Within Groups	8461179	14	604370.0		
	Total	2039488536	25			

Source: devised by the author

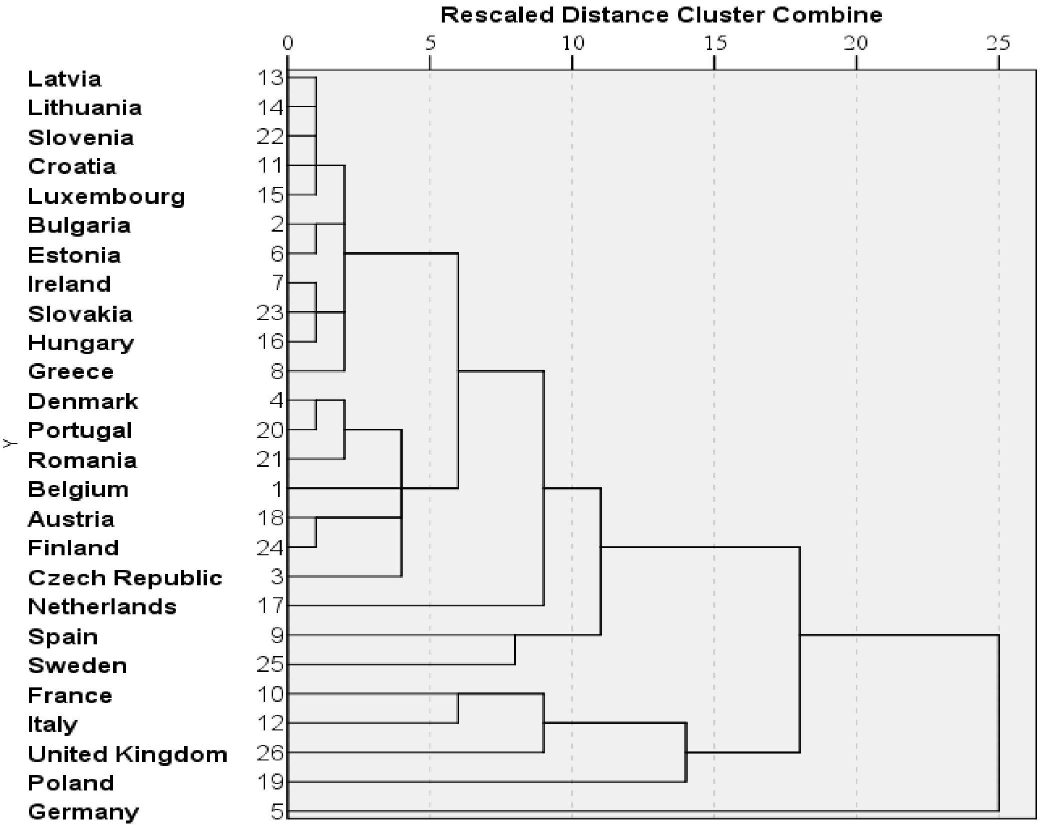

Fig.1. Dendrogram using Average Linkage (Within groups) for the Solid fuels, Oil (total), Gas indicators.

Source: devised by the author by means of the SPSS

The first three clusters include 17 states. The composition and main characteristics of the mean values of the variables subjected to analysis in each cluster are presented in Table II.

In cluster C1, the means of the Solid fuels, Oil and Total Renewables variables for a Confidence Level of 95% (α=0.05) are statistically significant (the null hypothesis H_{0_2} is rejected whereas the alternative hypothesis H_{1_2} is accepted) and for the Gas variable, the null hypothesis (H0_2) can be rejected only for the Confidence Level of 90% (α=0.01).

In cluster C1, the means of the Solid fuels, Oil and Total Renewables variables for a Confidence Level of 95% (α=0.05) are statistically significant (the null hypothesis H_{0_2} is rejected whereas the alternative hypothesis H_{1_2} is accepted) and for the Gas variable, the null hypothesis (H0_2) can be rejected only for the Confidence Level of 90% (α=0.01).

TABLE II
MEAN, MINIMUM AND MAXIMUM VALUES, CONFIDENCE INTERVALS FOR THE SOLID FUELS, OIL (TOTAL), GAS AND TOTAL RENEWABLES PARAMETERS WITHIN THE C1, C2 AND C3 CLUSTERS

Cluster	Included states
C1	Belgium, Denmark, Portugal, Romania
C2	Bulgaria, Estonia, Ireland, Greece, Croatia, Latvia, Lithuania, Luxembourg, Hungary, Slovenia, Slovakia
C3	Austria, Finland

Characteristics

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min.	Max.
						Lower Bound	Upper Bound		
Solid fuels	C1	4	3525.0	1527.8	763.9	1093.9*	5956.1*	2386	5746
	C2	11	2478.8	2459.7	741.6	826.4*	4131.3*	53	6687
	C3	2	3760.5	1009.0	713.5	-744.5**	8265.5**	3047	4474
Oil (total)	C1	4	12149.0	7543.7	3771.9	145.2*	24152.8*	6589	23249
	C2	11	4080.0	3134.5	945.1	1974.2*	6185.8*	1100	12034
	C3	2	10740.0	1634.8	1156.0	3441.0**	18039.0**	9584	11896
Gas	C1	4	7066.5	4716.9	2358.5	1517.0**	12616.0**	2830	12599
	C2	11	2399.9	1898.6	572.4	1124.4*	3675.4*	436	6982
	C3	2	4480.5	2778.2	1964.5	-7923.4**	16884.4**	2516	6445
Total Renewables	C1	4	4860.5	1222.1	611.1	2915.8*	6805.2*	3357	6124
	C2	11	1427.5	625.4	188.6	1007.3*	1847.6*	190	2446
	C3	2	9970.0	261.6	185.0	7619.4*	12320.6*	9785	10155

Source: devised by the author

In what is follows, the values and calculus refer to a representative notion for energy sector which is called *tonne of oil equivalent (toe)*. This is a unit of energy defined as the amount of energy released by burning one tonne of crude oil. It is approximately 42 gigajoules. For multiples of the tonne of oil equivalent is usually used kilotonne of oil equivalent - ktoe (10^3 tonne).

The C1 Cluster is characterized by a very high mean value of Gross inland oil energy consumption (12194 ktoe), followed by Gas (7066.5 ktoe) and Solid fuels (3525 ktoe), whereas the Total Renewables consumption amounted to 4860.5 ktoe which accounts for 17.61% of the Total Gross inland energy consumption for all the four categories subjected to analysis. From the point of view of the consumption volume, in 2014, maximum values were recorded in Romania for Solid fuels (5746 ktoe) and Total Renewables (6124 ktoe), in Belgium the maximum consumption was recorded for Oil (23249 ktoe) and Gas (12599 ktoe) as well as a minimum consumption of Total Renewables, whereas, within this cluster, the minimum consumption of Solid fuels (2386 ktoe), Oil (6589 ktoe) and Gas (2830 ktoe) were recorded in Belgium.

In the C2 cluster, the means of all the four variables are statistically significant (the null hypothesis H_{0_2} is rejected whereas the alternative hypothesis H_{1_2} is accepted) for a Confidence Level of 95% (α=0.05). This cluster is characterized by very low mean values of all the four variables subjected to analysis.

Thus, the average energy consumption is lower in C2 than in C1 in the case of Solid fuels (1.42 times lower), for Gross inland gas consumption (2.94 times lower), as well as for Gross inland oil consumption (2.98 times lower).

The most significant differences between cluster C2 and the other two clusters are recorded in the case of Gross inland total renewables consumption for which, in cluster C2, the average consumption is 3.41 times lower than in C1 and 6.98 times lower than in C3.

The lowest consumption values for Solid fuels and Total Renewables have been recorded in Luxembourg (53 ktoe and 190 ktoe, respectively), and for Oil and Gas consumption in Estonia (1100 ktoe and 436 ktoe, respectively). On the other hand, in Greece, the highest values have been recorded for Gross inland solid fuel energy consumption (6687 ktoe) and for oil energy consumption (12034 ktoe).

In the C3 cluster, the means of the Oil and Total Renewables variables are statistically significant for a Confidence Level of 90% and of 95%, respectively. The other two variables (namely Solid fuel and Gas) are not statistically significant (the null hypothesis H_{0_2} is accepted and the alternative hypothesis H_{1_2} is also accepted) therefore they cannot be taken into consideration.

In cluster C3, in comparison with the other two clusters, high mean values were recorded for Gross inland total renewables energy consumption (2.05 times higher than in cluster C1 and 6.98 times higher than in cluster C2), even though both states had a significant consumption of such type of energy, specifically 9785 ktoe in Finland and 10155 ktoe in Austria. Regarding Gross inland oil energy consumption, the average production is 2.63 times more consistent than in C2, but by 11.6% lower than in C1.

TABLE III

MEAN, MINIMUM AND MAXIMUM VALUES, CONFIDENCE INTERVALS FOR THE SOLID FUELS, OIL (TOTAL), GAS AND TOTAL RENEWABLES PARAMETERS WITHIN THE A1, A2 AND A3 CLUSTERS, IN 2006

Cluster	Included states
C4	Check Republic, Spain, Netherlands, Sweden
C5	France, Italy, Great Britain
C6	Poland
C7	Germany

Characteristics									
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min.	Max.
						Lower Bound	Upper Bound		
Solid fuels	C4	4	9619.0	5758.9	2879.5	455.3*	18782.7*	2103	15876
	C5	3	17432.0	10994.8	6347.8	-1103.7**	35967.7**	9290	29939
	C6	1	49238.0					49238	49238
	C7	1	79617.0					79617	79617
Oil (total)	C4	4	25584.8	18728.2	9364.1	3550.9**	47618.5**	9072	49066
	C5	3	67231.0	10775.6	6221.3	40462.8*	93999.2*	55825	77240
	C6	1	22373.0					22373	22373
	C7	1	108415.0					108415	108415
Gas	C4	4	14927.0	13565.8	6782.9	-1033.2**	30887.2**	801	29058
	C5	3	47695.7	13841.2	7991.2	13312.1*	82079.2*	32597	59784
	C6	1	13405.0					13405	13405
	C7	1	64057.0					64057	64057
Total Renewables	C4	4	10526.0	8095.5	4047.8	1001.6**	20059.4**	3400	17768
	C5	3	19978.7	7295.2	4211.9	1856.5*	38100.8*	12107	26512
	C6	1	8591.0					8591	8591
	C7	1	35406.0					35406	35406

*Confidence Level 95% (α=0.05)
**Confidence Level 90% (α=0.01)
Source: devised by the author

In the case of the other nine states which have not been included into the C1-C3 clusters, according to the Gross inland energy consumption from the four categories of analysed resources, their distribution on clusters is illustrated in Table III.

IV. CONCLUSIONS

The first conclusion that can be drawn from Table 3 as well as from Figure 2 is that in the C4-C7 clusters, the values of consumption corresponding to the four categories of resources are significantly higher than in the C1-C3 clusters.

In the C4 cluster, the mean values for Gross inland solid fuels energy consumption are statistically significant for a Confidence level of 95% and 90%, respectively, as well as for Gross inland oil energy consumption and Total Renewables.

On the other hand, for Gross inland gas energy consumption the mean value is not statistically significant. Within this cluster, the minimum values were recorded in Sweden for Gross inland solid fuels energy consumption (2103 ktoe) and Gross inland gas energy consumption (801 ktoe), in the Check Republic for Gross inland oil energy consumption (9072 ktoe) and in the Netherlands for Gross inland total renewables energy consumption (3400 ktoe), whereas the maximum values were recorded in Spain for Gross inland total renewables energy consumption (17768 ktoe) and Gross inland oil energy consumption (49066 ktoe), in the Check Republic for Gross inland solid fuels energy consumption (15876 ktoe) and in the Netherlands for Gross inland gas energy consumption (29058 ktoe).

For cluster C5, with the exception of Gross inland solid fuels energy consumption whose mean is not statistically significant, the means of the other three variables are statistically significant for a Confidence level of 95%.

Within this cluster, the mean values of Gross inland energy consumption were recorded in France for Gross inland solid fuels energy consumption (9290 ktoe) and Gross inland gas energy consumption (32597 ktoe), in Italy for Gross inland oil energy consumption (55825 ktoe) and in Great Britain for Gross inland solid total renewables energy consumption (12107 ktoe), whereas the maximum values were recorded in Great Britain for Gross inland solid fuels energy consumption (29939 ktoe) and Gross inland gas energy consumption (59784 ktoe), in France for Gross oil fuels energy consumption (77240 ktoe) and in Italy for Gross inland solid total renewables energy consumption (26512 ktoe).

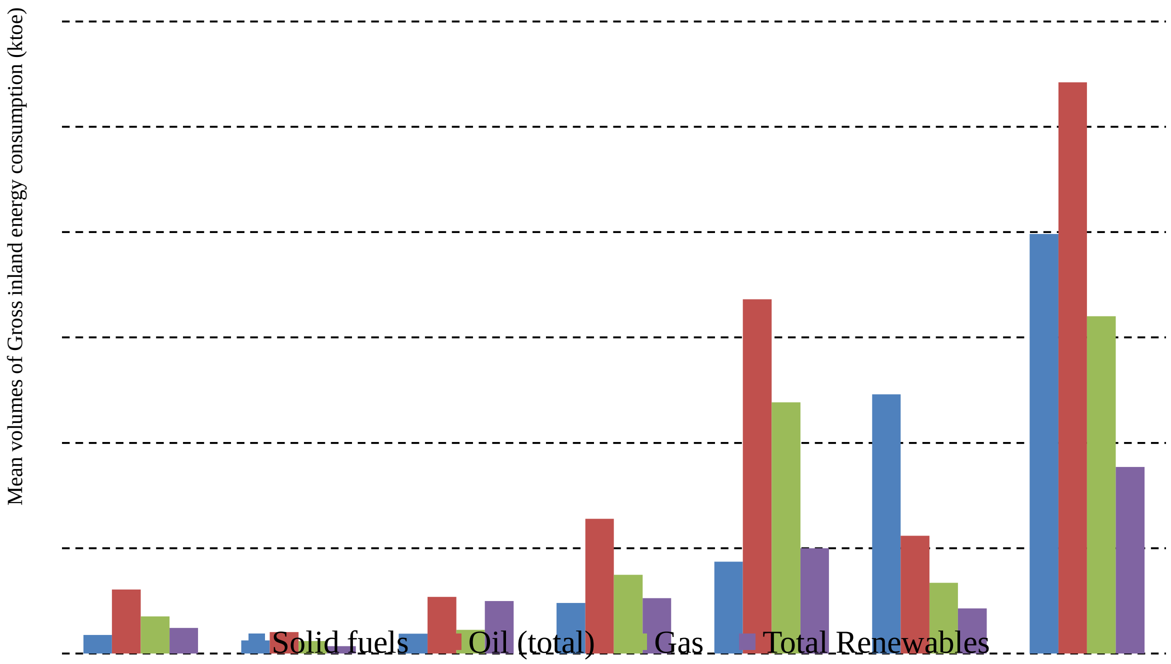

Fig. 2. A comparative presentation of the mean volumes of Gross inland energy consumption (ktoe) for Solid fuel, Oil, Gas and Total renewable

Each of the last two clusters, namely C6 and C7, consist of only one state, Poland and Germany, respectively. In these two countries, with regard to the Gross inland energy consumption from the four categories of resources subjected to analysis, there have been significantly different values as compared to the others, hence their inclusion in the other clusters (Poland in C4 and Germany in C5) would lead to the loss of statistical significance of the mean values of the variables that characterize these clusters both for a Confidence level of 95% as well as for a Confidence level of 90%.

REFERENCES

[1] http://ec.europa.eu/eurostat/statistics-explained/index.php/Consumption_of_energy, data extracted on May 2015
[2] http://ec.europa.eu/europe2020/index_en.htm, data extracted on January 2016
[3] T. Rotaru, (coordonator), G. Bădescu, I. Culic, E. Mezel, C. Mureşan, "*Statistical methods applied in social sciences*" ("*Metode statistice aplicate în ştiinţele sociale*"), Ed. Polirom, Iaşi, 2006, pp.272-303.
[4] D. Horwitt, D. Cramer, "*Introduction to SPSS in Psychology*", Third Edition, Pearson Education Limited, 2005.
[5] M. Popa, "*Statistics for psychology. SPSS Theory and Applications*" ("*Statistică pentru psihologie. Teorie şi aplicaţii SPSS*"), Editura Polirom, Iaşi, 2008.
[6] C. Oprea, M. Zaharia, "*Elements of data analysis and modeling using Excel*" ("*Elemente de analiza datelor şi modelare utilizând Excel*"), Editura Universitară, Bucharest, 2011.