Zagreb International Review of Economics & Business, Vol. 9, No. 1, pp. 45-60, 2006

ISSN 1331-5609; UDC: 33+65

Cluster Analysis in the Study of Life Quality on the Central and Eastern European Countries

Daniela-Emanuela Dănăcică*

Abstract: The aim of this research is to study life quality using the cluster analysis methodology. The analysis includes a number of classifing algorithms of elements into relatively homogenous groups and makes a global analysis of statistical units using a high number of characteristics. The required hipotheses are minimal.

JEL Classification: I000

Key words: cluster analysis, life quality, matrix, dendrogram, Ward algorithm

Introduction

The concept of 'life quality' was launched in the 1960s by the developed societies that had as objective the fast growth of welfare.

In the popular sense, it was assimilated to the old concept of happiness. 'Happiness' is an emotional feeling, an ethic individual concept but 'life quality' takes into account the conditions that generate and maintain the development of happiness, having a collective sociological perspective.

Life quality includes the totality of physical, economical, social, cultural, politic, of health, environmental conditions, in which the men live, the contain and the nature of the unfolded activities, the characteristic of the relationship and social process they take part, goods and services they have access, life style, the subjective states of satisfaction or dissatisfaction, etc.

It depends on simultaneously and equilibrated satisfaction of all human requirements as well as the economic and social security, health, life condition, culture, educations, environment, etc.

The aim of this research is to study the life quality in the following Central and Eastern European countries: Albania, Bosnia and Herzegovina, Bulgaria, Croatia,

* Daniela-Emanuela Dănăcică is at the 'Constantin Brâncuşi', University of Tg-Jiu, Romania.

Czech Republic, Estonia, Hungary, Austria, Latvia, Slovakia, Slovenia, Serbia and Montenegro, Turkey, Ukraine, Russia, Belarus. The used methodology is multivariate statistical method cluster analysis.

The term 'cluster analysis', which was used for the first time by Tyron, includes a number of classifing algorithms of elements into relatively homogenous groups. The method makes a global analysis of statistical units using a high number of characteristics. The required hipotheses are minimal. A set of groups that minimises and maximises variation within the groups, can be identified using cluster analysis.

The Variables

The variables which will constitute the basis of the groups setting up were selected having in view theoretical, conceptual and practical reasons, trying to be relevant for the investigate problems and in straight connection with the analysis objectives.

We have also tried here to balance the number of the social-economical demographic variables with the variables of the living level.

Demographic Variables

We use life expectancy at birth (years), fertility rate (birth per women) and infant mortality rate (per 1000 life births) indicators. The data were gathered from the corresponding structure of each country given by World Bank for 2003.

Socio-economical Variables

The used socio-economical indicators are: gross domestic product/capita USD $, public expenditure on education (% of GDP), public expenditure on health (% of GDP), and unemployment rate (% of total labor force), indicators in straight connection with the evaluation of life quality. Data were gathered from the corresponding structure of each country given by UNDP and CIA World Fact book, the reference year being 2003 for GDP and unemployment rate, 1990/2001 for public expenditure on education and 2001 for public expenditure on health.

Variabiles of the Level of Living

The following subjective indicators of the level of living were used: civil and political liberties – measured as mean in between the two indicators: civil liberties and political liberties, measured on a scale from 1 (low level of liberties) to 7 (high level of liberties), with data provided by Freedom House for 2003; corruption measured with values from 0 (highly clean) to 10 (highly corrupt) and which include police corruption, business corruption and political corruption with data for 2003, life satisfaction - most scores are based on responses to the following question 'All things considered how satisfied or dissatisfied with your life-as-a hole now? '1 dissatisfied to 10 satisfied', World Database of Happiness, Happiness in Nations, Rank Report 2004 and objective indicators – literacy (% of total population) with data provided by CIA World Fact book for 2003.

Analysis

The corresponding data for the indicators chosen for this study are presented in Table 1.

In order to eliminate the influence of the unit of measure in determining the distances between countries, I changed the initial data into standardised data using the transformation $a'_{ij} = (a_{ij} - \bar{a}_{ij}) / \sigma$, where a_{ij} represents the data for each indicator, $\bar{a}_{ij}$ represents their mean and σ standard deviation.

The mean and the standard deviation for each variable calculated with SPSS 8.0 are presented in Table 2.

The next step consists in introducing the standardised data in the PC for being processed with the SPSS 8.0.

Using as a measure of distance the Squared Euclidian Distance and for classification the Ward algorithm, we obtain the proper dendrogram presented in Figure 1.

Table 1: Set of 11 Indicators for the Evaluation of Life Quality in Central and Eastern European States

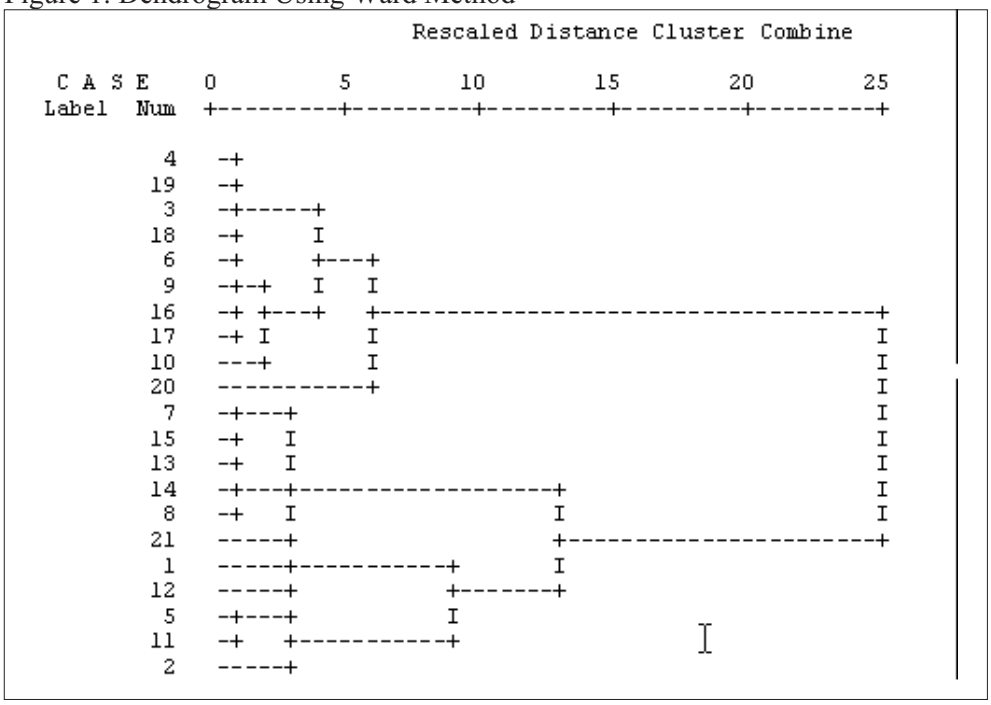
Nr. crt.	Country	Life Expectancy (years)	Fertility rate (birth per women)	Infant mortality rate (per 1000 live births)	PIB/capita US $	Public expenditure on education (% from PIB)	Public expenditure on health (% from PIB)	Literacy (% of total population)	Unemployment (% of total labour force)	Civil liberties (1-7)	Corruption (0-10)	Life satisfaction (1-10)
1	Albania	74.30	2.20	18.00	4900	5.80	2.4	86.5	14.80	2.50	7.50	4.60
2	Bosnia& Hertzegovina	74	1.30	4.00	6511	4.0	2.8	93.0	44.00	2.50	6.70	5.10
3	Latvia	70.70	1.30	10.00	11584	5.9	3.4	99.8	8.80	5.50	6.20	4.80
4	Lithuania	71.90	1.30	8.00	12676	5.2	4.2	99.6	8.00	5.50	5.30	4.90
5	Macedonia	73.60	1.80	10.00	7041	4.1	5.8	96.0	37.70	3.50	7.70	4.90
6	Poland	74.60	1.20	6.00	11984	5.4	4.6	99.8	19.50	5.50	6.40	5.90
7	Romania	70.10	1.30	18.00	7680	3.5	5.2	98.4	6.30	5.00	7.30	4.70
8	Moldova	67.00	1.40	26.00	1926	4.0	2.8	99.1	8.00	4.00	7.60	3.50
9	Slovakia	73.40	1.20	7.00	14525	4.1	5.1	99.7	13.10	5.50	6.30	5.60
10	Slovenia	76.10	1.20	4.00	19597	5.4	6.3	99.7	6.40	5.50	4.10	6.30
11	Serbia &Montenegro	72.80	1.70	12.00	2426	5.4	6.5	93.0	30.00	2.50	6.70	5.10
12	Turkey	68.60	2.40	33.00	7303	3.7	3.3	86.5	9.30	2.50	6.90	5.60
13	Ukraine	68.30	1.20	15.00	6307	4.2	2.9	99.7	3.50	3.00	7.70	3.60
14	Russia	66.70	1.10	18.00	9817	3.1	3.7	99.6	8.30	2.00	7.30	4.40
15	Bulgaria	72.10	1.20	12.30	6255	3	3.9	98.6	12.70	4.50	6.10	4.50
16	Croatia	74.00	1.40	6.00	11195	4.2	7.3	98.5	13.80	4.50	6.30	5.90

17	Czech Republic	75.20	1.20	3.90	16915	4.4	6.7	99.8	10.60	5.50	6.10	6.70
18	Estonia	71.60	1.20	10.00	14427	7.4	4.3	99.8	9.60	5.50	4.50	5.20
19	Hungary	71.70	1.20	8.00	14920	5.1	5.1	99.4	5.90	5.50	5.20	5.50
20	Austria	79.10	1.4	4.50	31265	5.9.	5.5	98.0	4.40	6.00	2.00	7.00
21	Belarus	68.20	1.3	13.00	6844	6.0	4.8	99.6	2.00	1.00	5.80	4.30

Table 2: Mean and Standard Deviation

	Life expectancy	Fertility rate	Infant mortality	PIB/capita	Expenses for education	Expenses for health	Literacy	Unemployment	Civil liberties	Corruption	Life satisfaction
Mean	72.0952	1.4048	11.7476	10757.048	4.75	4.6	97.3381	13.1762	4.1667	6.1762	5.1476
Standard deviation	3.1688	0.3427	7.5651	6623.0954	1.1197	1.4184	4.1385	11.0691	1.5193	1.3881	0.9053

Figure 1: Dendrogram Using Ward Method

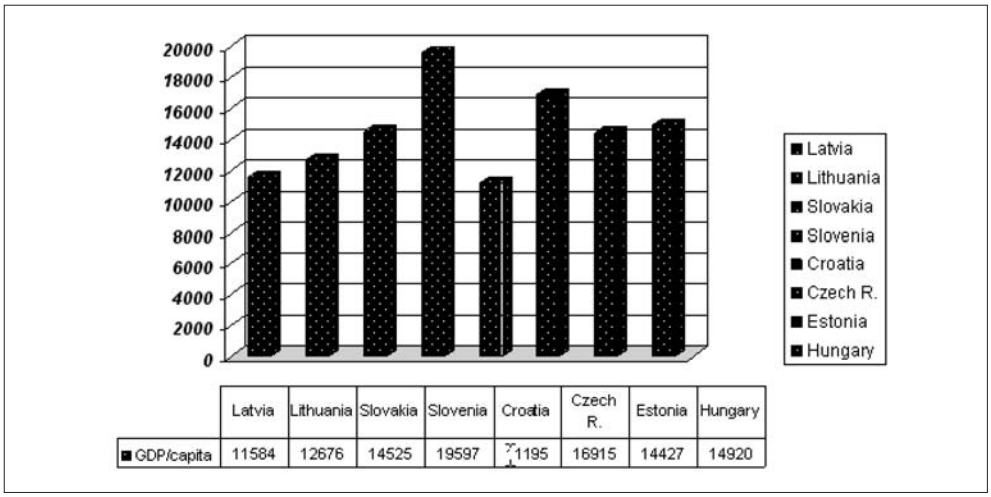


Analysing agglomeration schedule and dendrogram using Ward Method we notice that countries aggregate at low levels in between 0 and 5, but the clusters aggregation is realised at high levels (25). Differences are explained by using the Ward method meaning the variation of each class must be as small as possible while the intercluster variation must as high as possible. On the basis of dendrogram, we can classify the 21 studied countries as following:

- Class 1 is formed of Lithuania, Hungary, Latvia, Estonia, Slovakia, Poland, Croatia, Czech Republic, and Slovenia. The distances between these countries are small and relatively closed (the smallest distance Latvia-Lithuania, generated by the similitude between these two former USSR countries). The countries of this class have a high GDP/capita USD $ (Czech Republic has the maximum value 16 915 and Croatia the minimum value 11 995) and give high percents of GDP for education (Estonia gives the maximum value 7.4 and Slovakia the minimum value 4.1) and for health (Czech Republic gives the maximum value 6.7 and Latvia the minimum one 3.4). These countries present low unemployment rates (Croatia the maximum value 13.80 and Hungary the minimum one 5.90), a high life expectancy at birth, with values in between

70.70 and 74.00 years, an approximately equal fertility rates, the most reduced infant mortality rate among the Central and Eastern European countries (Czech Republic 3.90 - the most reduced rate of geographical zone), a high level of civil and politic liberties (5.50 excepting Croatia with 4.50) and high levels of life satisfaction (Czech Republic with the maximum value 6.70 and Latvia with the minimum value 4.80). Instead they have problems with corruption, recording high levels in between 4.10 (Slovenia) and 6.30 (Croatia). In conformity with World Bank standards 2004, Croatia, Estonia, Latvia, Lithuania, Poland, Czech Republic, Slovakia, Hungary are upper middle-income countries and Slovenia is high-income country.

Figure 2: Cluster 1 GDP per capita (US$)



	Latvia	Lithuania	Slovakia	Slovenia	Croatia	Czech R.	Estonia	Hungary
GDP/capita	11584	12676	14525	19597	11195	16915	14427	14920

- Class 2 is formed of Austria. The isolation of Austria compared to other countries is emphasised by the big distances in comparison with the rest of the countries, being closed only by Slovenia. Austria has a GDP of 31 265 USD/capita, low rate unemployment, 4.40 percent of total labor force, the greatest life expectancy at birth of the analysis countries, 79.10 years, one of the lowest rates of infant mortality, 4.50, the highest level of civil and political liberties, 6 (maximum is 7), the highest level of life satisfaction and the lowest rate of corruption, 2. According to the World Bank Standards 2004, Austria was considered as a high income country.

Figure 3: Cluster 1 - Civil and Political Liberties, Corruption, Life Satisfaction

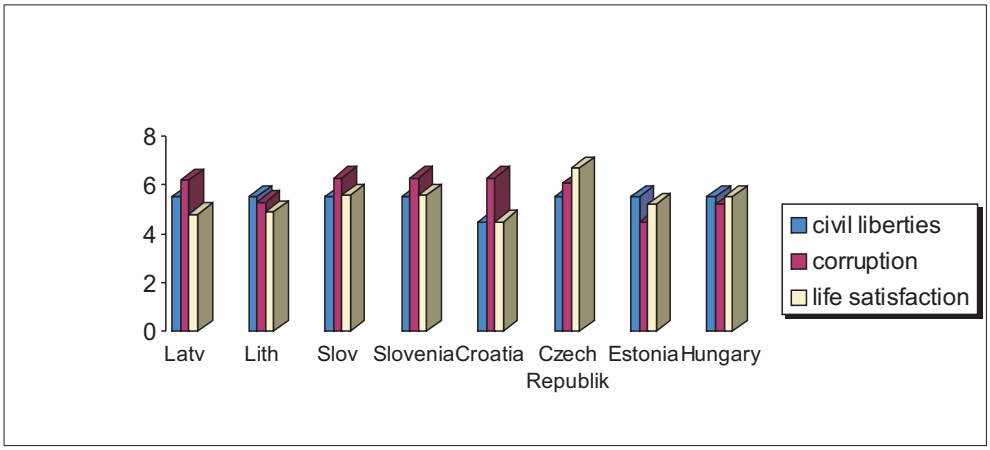
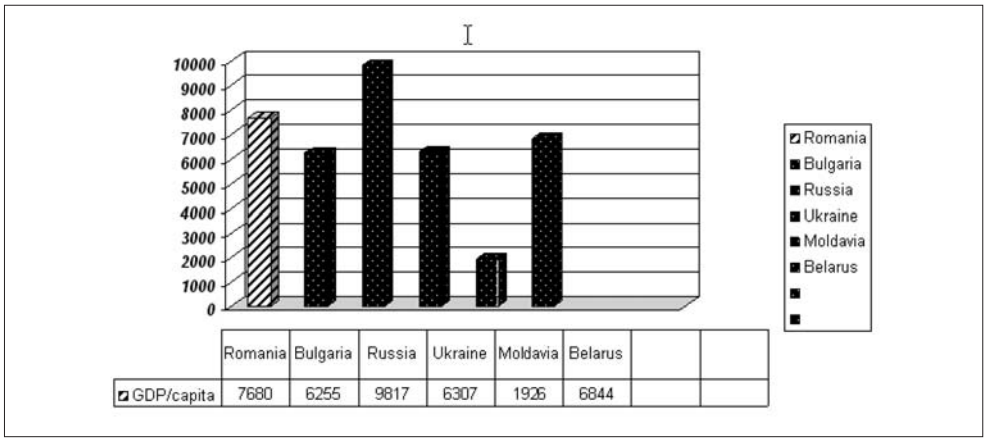


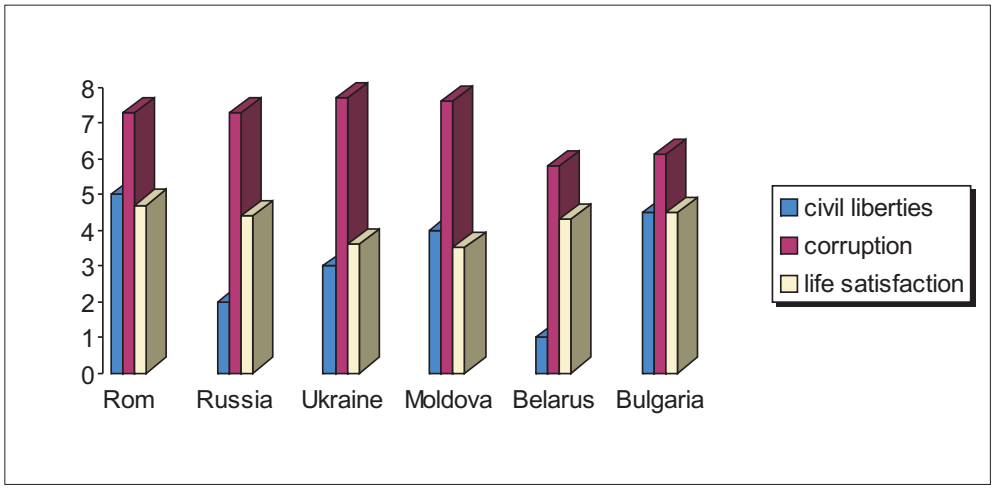
Figure 4: Cluster 3 - GDP per capita (US$)



- Class 3 is formed by Moldavia, Ukraine, Russia, Belarus (former USSR countries) and Romania and Bulgaria that passed through a prolonged transition process with negative results. They are countries with a modest GDP/capita, Russia having the maximum value – 9 817 and Moldavia the minimum one, 1 680 USD/inhabitant. They present low GDP provided for health and education, medium rates of unemployment but low life expectancy at birth in between 72.10 (Bulgaria) and 66.70 (Russia). They also have a reduced rate of fertility (Russia 1.10 the lowest level in the studied countries), the highest rate of infant mortality (26 in Moldavia, 18 Russia, 15 Ukraine), the lowest levels of civil and political liberties (Belarus 1.00, Russia 2.00, Ukraine

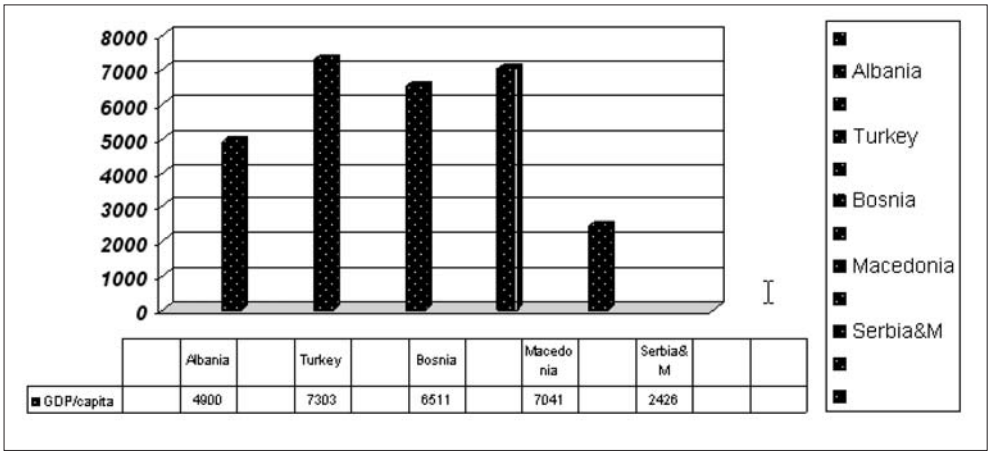
3.00, maximum Romania 5.00), a low life satisfaction and a very high level of corruption (Ukraine 7.70, Moldavia 7.60, Russia 7.30, Romania 7.30). According to World Bank standards these countries are considered as follows: Romania, Ukraine, Russia, and Belarus - lower middle income, Moldavia - low income.

Figure 5: Cluster 3 - Civil and Political Liberties, Corruption, Life Satisfaction



- Class 4 contains Albania, Turkey, Macedonia, Serbia and Montenegro, Bosnia and Herzegovina. They are countries with a very low GDP/capita: Albania, 4 900, Turkey 7 303, Macedonia 7 041, Serbia 2 426, Bosnia 6 511, low percents of GDP earmarked for education and health, high rates of unemployment (Bosnia 44% of total labor force, the highest value of the studied countries, Macedonia 37.70, Serbia 30.00, Albania 14), the highest rate of fertility in the region, due to the prevailing Muslim religion, but the highest infant mortality rates (Turkey 33, the first place among the studied countries, Albania 18, Serbia 12), the lowest level of liberty in the region, 2.50 Turkey, Albania, Bosnia and Herzegovina, Serbia, and 3.50 Macedonia, a low life satisfaction, big problems with corruption, Albania 7.50, Turkey 6.90, Macedonia 7.70, Serbia and Bosnia 6.70. Macedonia, Serbia and Montenegro, Bosnia and Herzegovina are former countries of Yugoslavia that encountered the hell of civil wars. According to World Bank standards 2004 these countries are lower middle income countries.

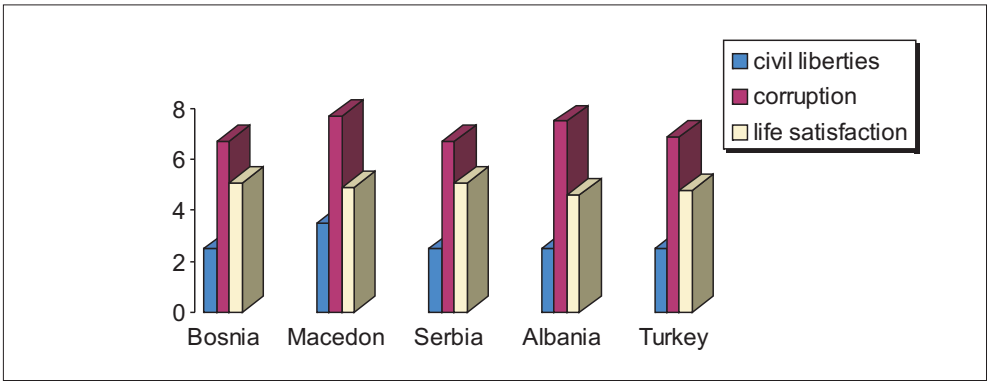
Figure 6: Cluster 4 - GDP per capita (US$)



The map after using Ward algorithm shows as follows:

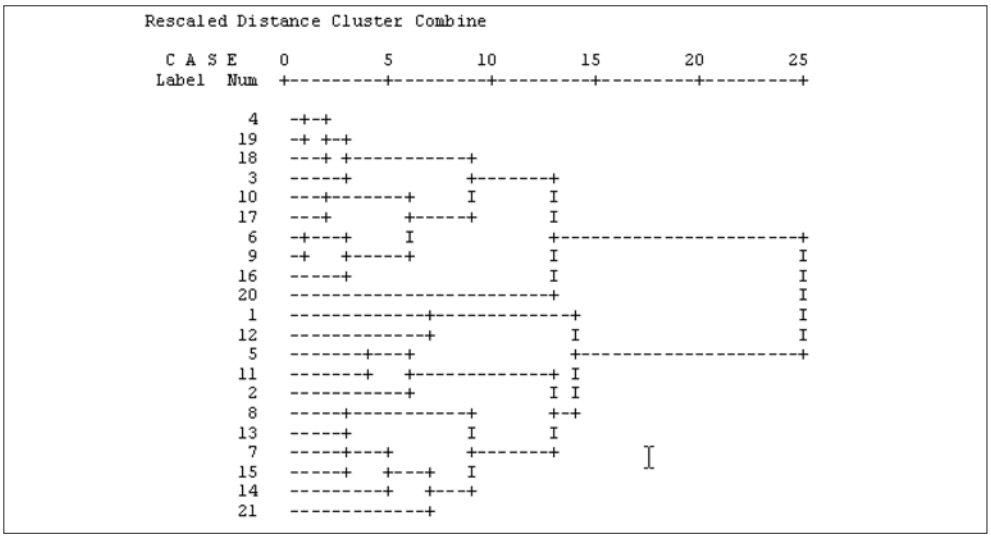
Figure 7: The resulted groups map

Figure 8: Cluster 4 - Civil and Political Liberties, Corruption, Life Satisfaction



Using City Block as a measure of distance and Complete Linkage method, we obtain the dendrogram from figure 9.

Figure 9: Dendrogram Using Complete Linkage



Analysing the resulted dendrogram we obtain the same results, with the difference that the isolation of Austria and of the class formed by Albania and Turkey is obvious.

For the next step we have improved the presented solution of Ward algorithm, using this time K-means of iterative partition method and the centroids of the

above-presented groups as cluster initial centers. We chose the two-cluster solution classifying the analysed countries as following:

Class 1 - good level of life quality - Lithuania, Latvia, Estonia, Hungary, Czech Republic, Croatia, Slovakia, Poland, Slovenia, Austria.

Class 2 - low level of life quality – Romania, Moldavia, Ukraine, Russia, Bulgaria, Belarus, Albania, Bosnia and Herzegovina, Turkey, Serbia and Montenegro, Macedonia.

The medium values of the indicators used in this analysis for each cluster are presented in Table 6.

Table 6: Summary Statistics

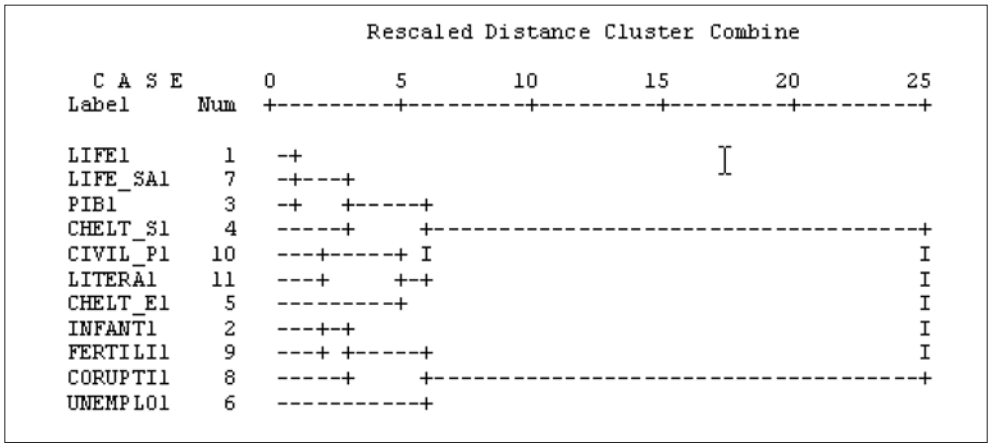
Mean	Cluster 1 10 countries Good quality of life	Cluster 2 11 countries Low quality of life	Mean difference	t	Sig (2-tailed)
Literacy	99.41	95.45	3.9555	2.444	0.024
Civil liberties	5.54	3	2.4500	6.369	0.027
Unemployment	10.01	16.05	-6.0445	-1.269	0.000
Life satisfaction	5.78	4.57	1.2073	4.070	0.000
Corruption	5.24	7.02	-1.7873	-3.818	0.220
Life expectancy	73.83	70.51	3.3118	2.759	0.211
Fertility rate	1.26	1.53	-0.2764	-1.975	0.001
Infant Mortality	6.74	16.30	-9.5600	-3.696	0.003
PIB	15 888.800	6091.81	9796.9818	5.050	0.012
Educational expenses	5.30	4.25	1.0455	2.371	0.063
Health expenses	5.25	4.00	1.2409	2.182	0.002

The differences between the two clusters are obvious. Comparing the two clusters we observe that: literacy is higher with 3.95 percent, civil and political liberties with 2.45, life expectancy at birth with 3.31 years, life satisfaction with 1.20, GDP with 9 796, 98, the percent from GDP for education with 1.05 and for health with 1.24 and

the unemployment rate is lower with 6.04, corruption with 1.78, fertility rate with 0.27 and infant mortality with 9.56. Using the T-test we notice that the differences between the averages of the indicators are statistical significant.

Applying cluster analysis on the variables, using Squared Euclidian Distance and Ward method we obtain the dendrogram presented in Figure 10.

Figure 10: Dendrogram using Ward Method



From the analysis of the dendrogram we can notice that:

- Class 1 includes life expectancy, life satisfaction, GDP and public expenses for health, economic and life level indicators that present a direct and strong enough correlation.
- Class 2 includes civil and political liberties, literacy, public expenses for education, economic and living level indicators that present a direct and strong correlation.
- Class 3 includes infant mortality, fertility rates, corruption level, unemployment rate that are social, demographic and living level indicators.

The correlations matrix between variables is presented in the Table 8.

Looking at the correlation matrix between the chosen variables, we notice the following important correlations:

- Life expectancy is positively correlated with GDP, life satisfaction and civil and political liberties are negatively correlated with corruption and infant mortality;

Table 8: Correlations Matrix

		LIFE1	INFANT1	PIB1	CHELT_S1	CHELT_E1	UNEMPLO1	LIFE_SA1	CORUPTI1	FERTILI1	CIVIL_P1	LITERA1
LIFE1	Pearson Correlation	1.000	-.722	.642	.501	.319	.290	.800	-.583	.000	.511	-.050
INFANT1	Pearson Correlation	-.722	1.000	-.567	-.532	-.333	-.230	-.552	.524	.576	-.501	-.488
PIB1	Pearson Correlation	.642	-.567	1.000	.416	.360	-.342	.787	-.870	-.325	.667	.365
CHELT_S1	Pearson Correlation	.501	-.532	.416	1.000	.087	.042	.639	-.361	-.195	.351	.319
CHELT_E1	Pearson Correlation	.319	-.333	.360	.087	1.000	-.157	.244	-.553	-.008	.230	.067
UNEMPLO1	Pearson Correlation	.290	-.230	-.342	.042	-.157	1.000	.031	.327	.240	-.259	-.382
LIFE_SA1	Pearson Correlation	.800	-.552	.787	.639	.244	.031	1.000	-.663	-.017	.541	.018
CORUPTI1	Pearson Correlation	-.583	.524	-.870	-.361	-.553	.327	-.663	1.000	.291	-.550	-.303
FERTILI1	Pearson Correlation	.000	.576	-.325	-.195	-.008	.240	-.017	.291	1.000	-.424	-.915
CIVIL_P1	Pearson Correlation	.511	-.501	.667	.351	.230	-.259	.541	-.550	-.424	1.000	.513
LITERA1	Pearson Correlation	-.050	-.488	.365	.319	.067	-.382	.018	-.303	-.915	.513	1.000

- Infant mortality rate is positively correlated with fertility, corruption (!) and negatively with life expectancy at birth, GDP, life satisfaction, civil and political liberties and literacy;
- Life satisfaction is positively correlated with life expectancy at birth, GDP, health, civil and political liberties and negatively with infant mortality and corruption;
- Corruption is positively correlated with infant mortality rate and negatively with GDP, public expenses for education, life satisfaction, civil and political liberties;
- Literacy is positively correlated with civil liberties and negatively with infant mortality and fertility;
- Civil and political liberties are positively correlated with GDP, life satisfaction, literacy and negatively with corruption and infant mortality rate.

Conclusion

We used in this paper the cluster analysis for classifying the Central and Eastern European countries depending on the life quality. The procedure suggests four groups of countries using Ward method and two groups of countries after the improvement of Ward method. The most used variables from the study are also indicators of social and economic development and their values increase or decrease if the indicator is positively or negatively correlated with the economic development.

In both cases, with four or two groups, the revealed image is the same. There is a class of countries with a high level of life quality, characterised by a high socio-economic standard, and consequently good life conditions, and a class made of low socio-economic standard countries with low GDP due to the inefficiency of economical politics or to the hell of civil, with big problems regarding corruption, civil and political liberties, life satisfaction, infant mortality and unemployment. As we can see on the European map, the countries with high life quality and implicitly a high social economic life standard are those placed in the middle of the continent that present favorable straight connection with the developed western part of the Europe, while the countries with a low life standard are countries that were under the former USSR influence or with civil and politic problems, having an impact on the social-economic development.

The two final classes resulted from the life quality study are presented in Table 9.

Table 9: Life quality in Central and Eastern European Countries

Nr. Crt.	Good life quality	Low life qualit
1		Albania
2		Bosnia&Hertzegovina
3	Latvia	
4	Lithuania	
5		Macedonia
6	Poland	
7		Romania
8		Moldova
9	Slovakia	
10	Slovenia	
11		Serbia&Montenegro
12		Turkey
13		Ukraine
14		Russia
15		Bulgaria
16	Croatia	
17	Czech Republic	
18	Estonia	
19	Hungary	
20	Austria	
21		Belarus

REFERENCES

http://hdr.undp.org/statistics/data/countries.cfm
http://www.imf.org
http://www.worldbank.org
http://www.worldfactsandfigures.com/gdp_country_desc.php
http://unstats.un.org/unsd/demographic/products/dyb/dyb2.htm
http://www.cia.gov/cia/publications/factbook/
http://www1.eur.nl/fsw/happiness/hap_nat/findingreports/RankReport2004-1c.htm