International Journal of Business and Social Research
Issue 04, Volume 10, 2014

Comparison of Turkey and European Union Countries' Health Indicators by Using Fuzzy Clustering Analysis

Nesrin Alptekin[1]

ABSTRACT

In this study, it is aimed to classify of 27 European Union countries and Turkey with the healthcare indicators by using fuzzy clustering analysis. This study also investigates the position of Turkey compared to the European Union countries in terms of healthcare statistics. Fuzzy clustering analysis has been applied to the data obtained from 2012 World Health Report. Based on the Fuzzy clustering analysis, the countries were classified into two different groups. Turkey is placed in the same cluster as Bulgaria, Cyprus, Estonia, Hungary, Latvia, Lithuania, Poland, Romania and Slovakia.

Keywords: EU Countries, fuzzy clustering analysis, health indicators.
JEL Codes: C38, I10, I15.
Available Online: 30th October 2014.
MIR Centre for Socio-Economic Research, USA.

1.0 INTRODUCTION

The main purpose of public health policies is to maintain and improve the level of the nation's health. Therefore, it is important to identify the factors contributingnation's health status.It is difficult to measure a nation's health directly because it is created by some factors such as economic, social and environmental.There is a reciprocal casual relationship between the level of economic development and thenation's health.Developments in theeconomy have a positive effects on healthcare indicators.Countries become more industrialized, they are able to allocate more resources to health services. According to statistics from the World Health Organization, Turkey's life expentancy was 63 for men and 67 for women in 1990 and increased to 72 for men and 77 for women in 2009. The ratio of total healtcare expenditures were 4.9% of Turkey's GDP in 2000, and increased to 6.7% in 2009. On the other hand, in European Union countries, life expentancy was 68 for men and 75 for women in 1990

[1]E-mail: nesrinesen@anadolu.edu.tr

and increased to 72 for men and 77 for women in 2009.The ratio of total healtcare expenditures were 8.0% of EU region's GDP in 2000, and increased to 9.3% in 2009.

Comparing healthcare indicators of Turkey and European Union countries is important because thehealth sector which is gaining importance day by day is an important indicator of a country's socio-economic development level. A large body of literature has compared different countries or different country groups by using various statistical classification techniques such as cluster analysis, discriminant analysis, factor analysis and multidimensional scaling analysis (Sen et al. 1991; Gerdtham et al. 1992; Schieber et al. 1994;Anderson and Hussey 2001; Kısa et al. 2002; Gauld et al. 2006; Nixon and Ulmann 2006; Kısa et al. 2007; Ersöz 2008; Lorcu and Bolat 2012; Girginer 2013).

2.0 FUZZY CLUSTERING ANALYSIS

Clustering analysis is a statistical technique that can be used to organize data into clusters based on similarities among the individual data items. The clusters developed by clustering analysis denote a high level of homogeneity within each cluster and high level of heterogeneity between clusters. Clustering analysis does not rely on assumptions common to other multivariate statistical analysis methods, such as underlying statistical distribution of data, the number of clusters or the cluster structures.

The conventional (hard) clustering methods are based on classical set theory and restrict that each point of the data set belongs to exactly one cluster. Hard clustering means partitioning the data into a specified number of mutually exclusive subsets.

In fuzzy clustering, the data points can belong to more than one cluster, and associated with each of the points are membership grades which indicate the degree to which the data points belong to the different clusters(Wolfram Mathematica, 2014).

Fuzzy set theory proposed by Zadeh in (1965) gave an idea of uncertainty of belonging which was described by a membership function. Fuzzy clustering analysis is an appropriate method in case units are not separated from each other significantly. In fuzzy clustering, each object is 'spread over' various clusters and the degree of belonging of an object to different clusters is quantified by means of membership coefficients, which range from 0 to 1, with the stipulationthat the sum of their values is one. This is called a fuzzification of the cluster configuration. It has the advantage that it does not force every object into a specific cluster. It has the disadvantage that there is much more information to be interpreted (NCSS Statistical Software, 2014).

The particular technique in fuzzy clustering used in this paper is called Fuzzy c-means (hereafter, FCM) clustering method developed by Dunn(1973) and improved by Bezdek(1981).

Let $\{x_1, x_2, ..., x_N\}$be a set of *N* data objects represented by *n*_dimensional feature vectors $x_k = [x_{1k}, x_2k, ..., x_{nk}]^T \in \Re^n$. A set of N feature vectors is then represented as a $n \times N$ data m

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nN} \end{bmatrix} \quad (1)$$

A fuzzy clustering algorithm partitions the data *X* into *K* fuzzy clusters, forming a fuzzy partition in *X*. A fuzzy partition can be conveniently represented as a matrix *U*, whose elements $u_{ik} \in [0,1]$ represent the membership degree of x_k in cluster *i*. Hence, the *i*th row of *U* contains values of the *i*th membership functionin the fuzzy partition.

The FCM algorithm used in this paper is developed by Kaufman and Rousseeuw(1990). This algorithm seeks to minimize the following objective function:

$$\sum_{k=1}^{K}\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}u_{ik}^{2}u_{jk}^{2}d_{ij}}{2\sum_{j=1}^{N}u_{jk}^{2}} \qquad (2)$$

subject to the following constraints,

$$u_{ik}\geq 0, \sum_{j}u_{ij}=1, \text{ for } i=1,2,\ldots,N;\ k=1,2,\ldots,K \qquad (3)$$

in which u_{ik} describes the unknown membership of the object *i* in cluster *k*, u_{jk} describes the unknown membership of the object *j* in cluster *k* and d_{ij} is dissimilarity between objects *i* and *j*. The dissimilarity coefficient or distance, $\|x_i - x_j\|$, between two objects , x_i and x_j , is defined as the Euclidean distance:

$$\|x_i - x_j\| = \sqrt{\sum_{k=1}^{p}\left(x_{ik}-x_{jk}\right)^{2}} \qquad (4)$$

One of the most important indicators found in the data is *Dunn's partition* coefficient. The amount of fuzziness in a solution is measured by Dunn's partition coefficient, which is defined as the sum of squares of all the membership coefficients divided by the number of objects and may be further normalized as in the following formula:

$$F_k(U)=\frac{K\sum_{i=1}^{N}\sum_{k=1}^{K}\frac{u_{ik}^{2}}{N}-1}{K-1} \qquad (5)$$

The normalized Dunn's coefficient, $F_k(U)$, ranges from 0 to 1. The value of $F_k(U)$ close to 1(hard cluster) indicates no fuzziness in the data whilst it close to 0 indicates complete fuzziness.

Another partition coefficient, given in Kaufman and Rousseeuw (1990), is defined as

$$D(U)=\frac{1}{n}\sum_{k=1}^{K}\sum_{i=1}^{N}\left(h_{ik}-m_{ik}\right)^{2} \qquad (6)$$

with normalized version

$$D_k(U)=\frac{K\cdot D(U)-1}{K-1} \qquad (7)$$

in which *q* is the cluster for which u_{ik} is maximal. For the optimum number of clusters, $F_k(U)$ and $D_k(U)$ together give a good indication. It should be chosen*K* so that $F_k(U)$ is large and $D_k(U)$ is small.

Average silhouette width can be used to measure how well a cluster or the whole data set is classified. Silhouette width may be obtained as

$$s(i)=\frac{b(i)-a(i)}{\max\left[a(i),b(i)\right]},\ -1\leq s(i)\leq 1 \qquad (8)$$

where *a(i)* denotes the intra-dissimilarity and *b(i)* denotes the smallest inter-dissimilarity.

The average silhouette coefficient of a cluster is calculated as the average of the *s(i)* for all objects in that cluster, and is thus an indicator of how well a cluster is classified (Artis and Zhang 2002).

3.0 DATA AND FINDINGS

In this study, the data on 7 different variables compiled from 2012 World Health Organization Report is used in Table 1. The healthcare indicators used as variables are:
X_1: Life expectancy at birth(years),
X_2: Healhty life expectancy at birth(years),
X_3: Mortality rate in children under age 5(per 1000 live births),
X_4: Adult mortality rate in ages between 15-69 (years),
X_5: Total expenditure on health as % of GDP,
X_6: Per capita total expenditure on health ($),
X_7: General government expenditure on health as % of total expenditure on health.

Table 1: Health indicators of EU countries and Turkey

Life expectancy and mortality					*Health Expenditure*		
Countries	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Austria	80	4	76	72	11,0	4.288	77,7
Belgium	80	4	82	72	10,8	3.948	75,1
Bulgaria	74	13	146	66	7,2	995	55,4
Cyprus	81	4	61	70	6,1	1.874	41,5
Czech Republic	77	4	10	70	8,0	2.107	84,0
Denmark	79	4	86	72	11,5	4.345	85,0
Estonia	75	5	156	66	6,7	1.338	78,4
Finland	80	3	90	72	9,0	3.226	74,7
France	81	4	86	73	11,9	3.969	77,9
Germany	80	4	76	73	11,7	4.219	76,9
Greece	80	4	75	72	10,6	3.054	61,7
Hungary	74	6	164	66	7,6	1.510	69,7
Ireland	80	4	77	73	9,4	3.761	75,0
Italy	82	4	59	74	9,4	3.071	77,9
Latvia	72	10	195	64	6,6	1.066	61,6
Lithuania	73	7	185	63	7,5	1.292	73,4
Luxembourg	81	3	76	73	7,9	6.592	84,0
Malta	80	6	60	72	8,5	2.141	64,8
Netherlands	81	4	66	73	12,0	4.881	79,0
Poland	76	6	137	67	7,4	1.391	72,3
Portugal	79	4	89	71	10,7	2.690	67,8
Romania	73	14	155	65	5,6	818	79,0
Slovakia	75	8	129	67	9,1	2.084	65,7
Slovenia	79	3	93	71	9,3	2.551	73,4
Spain	82	5	69	74	9,6	3.067	73,6
Sweden	81	3	61	74	10,0	3.722	81,5
United Kingdom	80	5	77	72	9,8	3.438	84,1
Turkey	75	13	104	66	6,7	957	75,1

Source: WHO 2012 Report

Variables used in this study are in different measure units and different numerical sizes. Preventing from weighting the variables more or less from the others, besides the raw dataand standardized data have been also used. Standardization of the data, (x-μ)/σ formula is used. To compare the findings, the values obtained from standard and raw data are used.

Average silhouette coefficients (ASC), the normalized Dunn's coefficients $F_k(U)$ and another normalized partition coefficients $D_k(U)$ obtained from healthcare indicators by using fuzzy clustering analysis of 28 countries are shown in Table 2.

Table 2: Fuzzy clustering analysis results of the countries' health indicators

Number of Clusters(k)	Standard Data			Raw Data		
	ASC	$F_k(U)$	$D_k(U)$	ASC	$F_k(U)$	$D_k(U)$
2	0,527	0,300	0,128	0,605	0,480	0,083
3	0,307	0,136	0,385	0,511	0,462	0,141
4	0,208	0,084	0,502	0,523	0,482	0,138
5	0,158	0,080	0,604	0,524	0,440	0,154

According to the findings in Table 2, the appropriate number of clusters is 2 both for standard and raw data. Two clusters maximize the Average silhouette coefficient and the normalized Dunn's coefficients ($F_k(U)$) while minimize another normalized partition coefficients ($D_k(U)$). There are 20 countries in cluster 1 for standard data and 17 of them are also in cluster 1 for raw data. Czech Republic, Malta and Slovenia take part in cluster 1 for standard data while they take part in cluster 2 for raw data. Bulgaria, Cyprus, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia and Turkey are the countries which take part in cluster 2(Table 3).

Statistics of correctly classified percentages for 2 clusters is obtained by using discriminant analysis. Based on the discriminant analysis, 100,0% of original grouped cases correctly classified for standardized data and 96,4% of original grouped cases correctly classified for raw data (Table 4).

Table 3: Countries in clusters

Countries	Standard Data	Raw Data
Austria	1	1
Belgium	1	1
Bulgaria	2	2
Cyprus	2	2
Czech Republic	1	2
Denmark	1	1
Estonia	2	2
Finland	1	1
France	1	1
Germany	1	1
Greece	1	1
Hungary	2	2
Ireland	1	1
Italy	1	1
Latvia	2	2
Lithuania	2	2
Luxembourg	1	1
Malta	1	2
Netherlands	1	1
Poland	2	2
Portugal	1	1
Romania	2	2
Slovakia	2	2
Slovenia	1	2
Spain	1	1
Sweden	1	1
United Kingdom	1	1
Turkey	2	2

Table 4: Correct classification rates for 2 clusters with discriminat analysis

Statistics		Predicted Group Membership			
		Standard Data		Raw Data	
		1st Cluster	2nd Cluster	1st Cluster	2nd Cluster
Count	1st Cluster	18	0	15	0

	2nd Cluster	0	10	1	12
%	1st Cluster	100%	0%	100%	0%
	2nd Cluster	0%	100%	7.7%	92.3%

100,0% of original grouped cases correctly classified for standartized data,96,4% of original grouped cases correctly classified for raw data.

4.0 CONCLUSION

In this paper, it is compared 27 European Union countries and Turkey with fuzzy clustering analysis by using the healthcare indicators given with seven variables. This study also investigates the position of Turkey compared to the European countries in terms of healthcare statistics.

Fuzzy clustering analysis results indicate that the classification of 28 countries in 2 clusters both in standard and raw data. Except Czech Republic, Malta and Slovenia, the other countries belong to the same group in both data types. The countries in both groups are different in terms of Adult mortality rate ages between 15-60 years, Healthy life expentancy at birth and Per capita total expenditure on health.

Finally, it is concluded that Turkey needs to close the gap between EU countries by reducing the mortality rates between ages 15-60 years and by increasing healthy life expentancy at birth and per capita total expenditure on health.

REFERENCES

Anderson, G., &Hussey, P. S. (2001). Comparing health system performance in OECD Countries. Health Affairs 20(3), 219-232.

Artis, M.J. &Zhang, W.(2002). Membership of EMU:A Fuzzy Clustering Analysis of Alternative Criteria. Journal of Economic Integration, 17(1), 54-79.

Bandemer, H. &Gottwald, S.(1995), Fuzzy Sets Fuzzy Logic Fuzzy Methods with Applicaitons, West Sussex, John Wiley and Sons Inc., England.

Bezdek, J. C. (1981), Pattern Recognition with Fuzzy Objective Function Algorithm, New York, Plenum Press, USA.

Bezdek, J.C., Ehrlıch, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. Computers and Geosciences, 10, 191–203.

Dunn, J.C.(1973). A fuzzy relative of the isodata process and its use in detecting compact, wellseperated clusters. Journal of Cybernetics, 3(3), 32–57.

Dunn, J. C. (1976), Indices of Partition Fuzziness and the Detection of Clusters in Large Data Sets, Fuzzy Automata and Decision Processes, Elsevier, NY.

Ersöz, F.(2008). Analysis of health levels and expenditures of Turkey and OECD Countries. Journal of Statisticians: Statistics and Actuarial Sciences, 2, 95-104.

Gauld, R., Ikegami, N., Barr, M. D., Chiang, T. L., Gould, D., & Kwon, S. (2006). Advanced Asia's health systems in comparison. Health Policy 79(3), 325-336.

Gerdtham UG, Sogaard J., Andersson F., & Jonsson B. (1992). An econometric analysis of health care expenditure: a cross-section of OECD countries. Journal of Health Economics 11,63–84.

Girginer, N.(2013). A Comparison of the Healthy Indicators of Turkey and The European Union Members Countries Using Multidimensional Scaling Analysis and Cluster Analysis. İktisat, İşletme ve Finans, 28(323), 55-72.

Hedayaty, M.K. &Rezazadeh, J.(2012). Fuzzy C- Means Clustering for the Analysis of Investors' Behavior in Tehran Stock Exchange (TSE). Science Series Data Report, 4(1), 56-73.

Kaufman, L. &Rousseeuw, P.J.(1990), Finding Groups and Data: An Introduction to Cluster Analysis, John Wiley and Sons Inc., NY.

Kaymak, U. &Setnes, M. (2000). Extended fuzzy clustering algorithms. ERIM report series Research in Management, Erasmus Research Institute of Management, Erasmus University Rotterdam, Netherlands.http://repub.eur.nl/res/pub/57/erimrs20001123094510.pdf.

Kısa, A., Kavuncubaşi, S., & Ersoy, K. (2002). Is the Turkish health care system ready to be a part of the European Union? Journal of Medical Systems 26(2), 89-95.

Kısa, A., Younis, M. Z., & Kısa S. (2007). A comparative analysis of the European Union's and Turkey's health status: How health-care services might affect Turkey's accession to the EU. Public Health Reports 122(5), 693-701.

Klir, George J., ST. Clair, Ute H. & Yuan, B.(1997), Fuzzy Set Theory Foundationas and Applications, Prentice Hall Inc., NJ.

Lorcu, F., &Bolat Acar, B. (2012). Comparison Member and Candidate Countries to the European Union by Means of Main Health Indicators. China-USA Business Review, 11(4), 556-563.

NCSS Statistical Software, Chapter 448, Fuzzy Clustering(2014, August 25). Retrieved From the NCSS Documentation website: http://ncss.wpengine.netdna-cdn.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Fuzzy_Clustering.pdf

Nixon, J. &Ulmann, P. (2006). The relationship between healthcare expenditure and health outcomes: evidence and caveats for a casual link. European Journal of Health Economics 7, 7-18.

Pal, R.N. & Bezdek, J.C. (1995). On Cluster Validity for the Fuzzy c-Means Model. IEEE Transactions on Fuzzy Systems, 3(3), 370-379.

Rousseeuw, P. J.(1987). Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. Journal of Computational and Applied Mathematics, 20(1), 53–65.

Schieber, G., Poullier, J. J. P., &Greenwald, L. M. (1994). Health system performance in OECD Countries: 1980-1992. Health Affairs, 13(4), 100-112.

Şen, F., Akkaya, Ç., Cyrns, M., Czock, H., Efendioglu, U., &Yarar, F. (1991). The comparison analysis of Turkey's socio-economic structure with new EU members. İktisat İşletme ve Finans 6(58), 32-47.

Trauwaert, E. (1988). On the meaning of dunn's partition coefficient for Fuzzy Clusters. Fuzzy Sets and Systems, 25, 217-242.

Wolfram Mathematica Documentation Center (2014, September 5). Fuzzy Clustering. Retrieved From the Wolfram Mathematica Documentation Center website: http://reference.wolfram.com/applications/fuzzylogic/Manual/12.html

WHO (2010-2012), World Health Statistics.

Zadeh, L.A.(1965). Fuzzy sets. Information and Control, 8(3), 338-353.