HEALTH AND ECONOMIC STATUS OF THE POPULATION OF THE CZECH REPUBLIC IN THE LAST 20 YEARS

Kornélia Cséfalvaiová – Elena Makhalova – Michaela Antovová

Abstract

Since the final decade of the 20th century Czech society has gone through significant changes that are noticeable mainly in economic, health and social area. Due to better health care Czech citizens live longer and spend more years in good health. On the one hand, health status of the Czech population has improved substantially. On the other hand, proportion of patients suffering from cancer, obesity, diabetes, heart attack, depression and anxiety is increasing. The aim of this paper is to point out the connection between health and economic status of the population of the Czech Republic using cluster analysis for selected economic and health indicators during the last twenty years. Social climate and national economy may significantly affect human health and health status and for this reason it is necessary to monitor these determinants simultaneously. In this context, it is essential to emphasize the importance of employment and job opportunities that positively contribute to the life quality and living standards of residents. On the other hand, decelerating economic growth and high unemployment have negative effects on social and health care costs. This paper examines the development of unemployment in a long-term view in the Czech Republic and focuses on typical causes of diseases and death.

Key words: Health status, Economic status, Unemployment, Cluster analysis

JEL Code: I130, J100

Introduction

Health status of individuals, people or society is significantly influenced by a number of factors, including economic factors, environmental factors, social and social factors, lifestyle and standard of living, health and quality of health care (Malina, 2005). Czech Republic has undergone considerable economic, political and social changes over the last 20 years. After 1989, according to extensive investments in medical and diagnostic technologies, the health status of inhabitants of the Czech Republic has improved significantly, especially in terms of decrease in cardiovascular diseases (heart attack, stroke). However, the appearance of cancer is increasing recently.

This increase is caused by lifestyle or environment. Due to better health care people live longer, and there is a greater probability of cancer in higher ages. According to forecasts, number of oncology diseases will triplicate by 2030 and cancer will become the most common cause of death instead of cardiovascular diseases. The most endangered group are people aged 50 years and over (Petruželka, 2014).

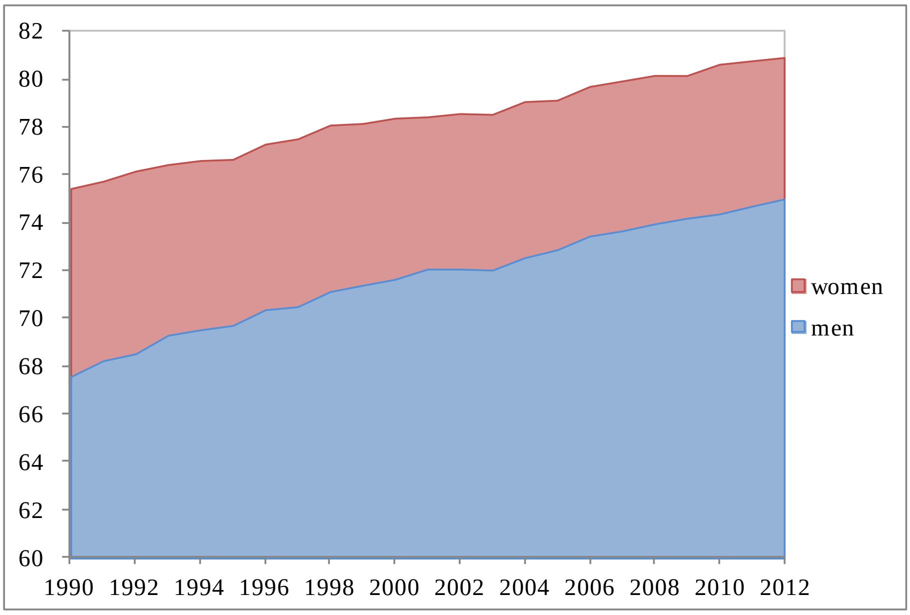
1 Development of selected health and economic indicators in the Czech Republic

Health status of the Czech population has significantly changed in various stages of development and monitoring causes of morbidity and mortality. After 1940, improving health status of the population and decreasing infant mortality was influenced by the decline of infectious diseases due to better hygiene care and the introduction of compulsory vaccination. During the 1960s of the 20th century number of malignant neoplasms and circulatory system diseases increased. Current health status of the population of the Czech Republic is most threatened by cancer. Cancer endangers younger people as well as people in older age groups.

1.1 Life expectancy at birth

Evolution of life expectancy at birth has a growing tendency in the period 1990–2012 (see figure 1). In the future, we expect a further increase of this indicator simultaneously for both men and women.

Figure 1: Life expectancy at birth for men and women, 1990-2012



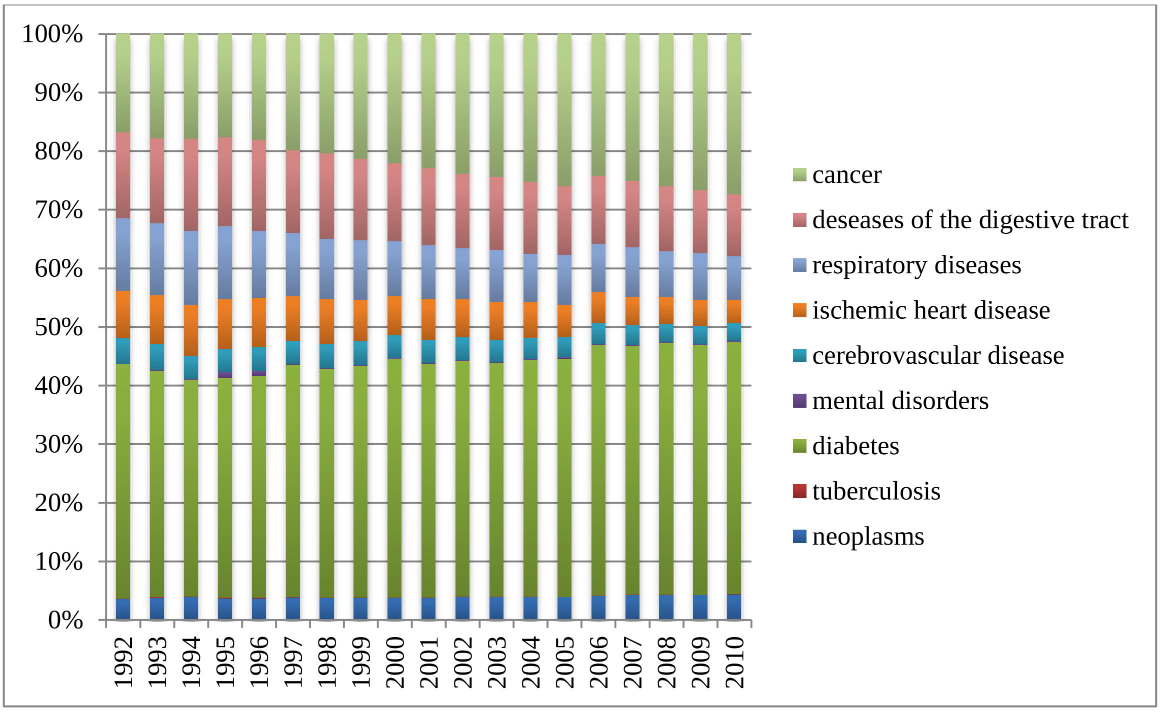
Source: Czech Statistical Office, own construction

In the evolution of life expectancy (e0) we have not experienced any significant fluctuations in the period under review, life expectancy has a linear upward trend. This indicator is higher for women than for men (e0 = 80.9 for women and e0 = 75.0 for men in 2012). Between 1990 and 2012 difference of life expectancy for women is 5.5 years, whereas in the case of men, the difference is 7.4 years. Human life prolonging and decrease of mortality in older age groups is a process typical for all developed countries.

1.2 Development prevalence rates of selected diseases in the Czech Republic

From figure 2 it is seen that the highest proportion occupies diabetes and cancer. The increasing incidence of diabetes is caused by changes in physical activity of the population. The movement of children and adults is much lower than before.

Figure 2: Prevalence rates of selected diseases in the Czech Republic in 1992–2010



Source: World Health Organisation, Institute of Health Information and Statistics, Czech Statistical Office, own construction

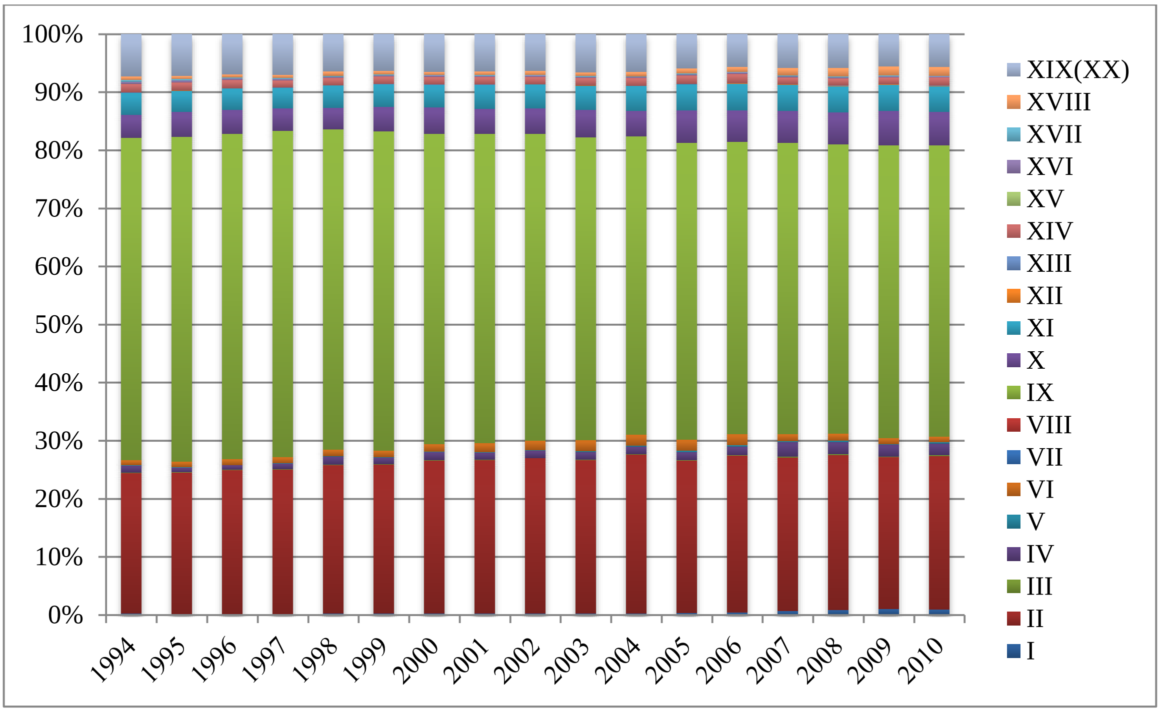
Employment is usually in front of computer (without physical activity), it has increase the accessibility and grows excessive consumption of unhealthy food. Thus driven lifestyle leads to obesity and diabetes. During the years 1992–2010, the incidence of cancer is almost 2,7multiplied. Significant influence on this has improvement of diagnosis. Previously, they misdiagnose and subsequent deaths were classified as natural death due to high age, even if it was a cancer. Diseases of the digestive, respiratory, cerebrovaskular and ischemic heart disease in 1992-2010 exhibit a constant evolution without a significant increase or decrease.

Neoplasms increased nearly 1,8times compared to 1992. Mental disorders and tuberculosis are least represented. Tuberculosis was widespread disease in the between the war period. The disappearance of tuberculosis contribute the development of antituberculosis, improving access to medical care and compulsory vaccination in the late 50s (Postgraduální medicína, 2014; Holčík, 2010).

1.3 Development of mortality rates by cause of death in the Czech Republic

For the analysis of mortality rates in 1990–2010, the 10th revision of International Classification of Diseases was used, which tracks the causes of death from 1994 to the present. Throughout the period the most common deaths are circulatory system diseases. Very often people die on neoplasms, injury, poisoning and diseases of the respiratory, digestive system. Group of injury and poisoning includes a wide range of causes of death: injuries at work, home, leisure, road accidents, suicides and poisonings.

Figure 3: Mortality rates by cause of death in the Czech Republic in the years 1994–2010 (10th revision)



Source: Czech Statistical Office, own construction

2 Fuzzy cluster analysis and evaluation of fuzzy clustering results

This chapter presents the fuzzy clustering method for economic data and estimation of the results of clustering. First of all, let's see what fuzzy clustering is.

Fuzzy c-means (FCM) is a data clustering technique where each data point belongs to a cluster to some degree that is specified by a membership grade. This technique was originally introduced by Jim Bezdek in 1981 as an improvement on earlier clustering methods. It provides a method that shows how to group data points that populate some multidimensional space into a specific number of different clusters. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2, 1 \le m < \infty \tag{1}$$

where m is any real number greater than 1, u_{ij} s the degree of membership of x_i in the cluster j, x_i is the i-th of d-dimensional measured data, c_j is the d-dimension center of the cluster, and $\|*\|$ is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right)^{\frac{2}{m-1}}}, \; c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \tag{2}$$

Estimate the suitable number of clusters will found with the follow indexes help. The basic validity index associated with fuzzy clustering is Dunn's coefficient defined by:

$$PC = \frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{2} \tag{3}$$

Here $u_{ij} \in \langle 0,1 \rangle$ are membership degrees of objects to clusters. This index assumes only the compactness measurement for each cluster and for the data structure. In according with the theory of fuzzy sets, the sum of memberships every object to all clusters is 1. It can be seen, that the value of membership degrees is decreasing with increasing number of clusters. Hence, raising the membership's value in square we obtain smaller values. It must be concluded, that with increasing number of clusters the value of this coefficient is decreasing.

We see that the Silhouette Coefficient (*SC*) always defines the compactness and separation measures for the data structure and provides representation of how well each object lies within its cluster. Let's see what the index is.

For an individual point *i* this coefficient is based on two measures: the average distance between *i* and every points in its cluster is calculated and min average distance between *i* and points in other clusters is calculated than. The total average Silhouette coefficient represends how tightly grouped all the data in the cluster are and may vary up from -1 to 1 inclusive:

$$SC = \frac{1}{N}\sum_{i=1}^{N}\frac{\min\left(A_{ij}, j \in C_{-i}\right) - A_{ic_i}}{\max(\min\left(A_{ij}, j \in C_{-i}\right), A_{ic_i})} \tag{4}$$

Where C_{-i} denotes cluster labels that do not include case i as a member, while c_i is the cluster label which includes case i, if $max(min(A_{ij},j \in C_{-i})A_{iCi})$ equals 0, the Silhouette of case i is not used in the average operations. Than the Silhouette coefficient is closer to 1, it means the best clustering result. When a variability in clusters is small, SC the usually determined the number of clusters correctly.

The next modification method combines two components using the harmonic mean. Those two elements have the different nature: one of these components is based on fuzzy clustering theory and other one is based on hard clustering theory.

With the help of this combination we can reduce disadvantages of both components. The first element here is Dunn's coefficient. It is calculated according to the formula (1) with values from the interval *[1/k,1]*. Let us consider extreme situations: completely fuzzy clustering: all $u_{ih}=1/k=>PC=1$ hard clustering: for one u_{ih} : $u_{ih}=1$ and for all others: $u_{ih}=0=>PC=1$. The second element is the difference between 1 and the ratio of the distance (C). The ratio of the distance is defined by the next formula:

$$C = \frac{\sum d_{i,\min}}{d_{1,\min}} \tag{5}$$

Here $\sum d_{n,min}$ is the minimal sum on the Euclidean distances between points in the case of n clusters; and d1,min is the minimal sum on the Euclidean distances between points in the case of one cluster, when the dataset is one cluster, it means before clustering. The sum of dn,min should be minimal for the best clustering, it means the minimal value of C should achieve the minimum for the best clustering.

Since the aggregate function we combine two parts, one of them is Dunn's coefficient (the maximum value for the best clustering), and the second one should also strive to maximum, that's why we introduce N (which achieves the maximum value for the best clustering):

$$N = 1 - C \tag{6}$$

And now we must to solve the optimization problem. An optimization problem can be represented in the following way:

$$E = \frac{2}{\left(\frac{1}{PC} + \frac{1}{N}\right)} \to \min \tag{9}$$

This function tends to its minimum for the best clustering, because the inverse values of the indexes of *PC* and *N* receive its maximum for the best clustering.

An optimization problem consists of minimizing a real function by systematically choosing input values from within an allowed set and computing the value of the function. The allowed set we can describe as:

$$E = \begin{cases} \frac{1}{PC} + \frac{1}{N} \neq 0; \\ PC \neq 0; \\ N \neq 0; \\ N \neq 1, \\ PC \in [0,1], N \in (0,1). \end{cases} \tag{7}$$

In the case, if $N=0$ we have a dataset like one cluster.

Applications in actual data:

The data set

Data are typically observations of some physical process. Each observation consists of n measured variables, grouped into n n-dimensional column vector $z_k=[z_{1k},\ldots,z_{nk}]^T$, $z_k \in \Re^n$. A set of N observations is denoted by $z=\{z_{k|k=1,2,\ldots,N}\}$ and represented as n x N matrix (where the z_{n1}-z_{n15} the attributes, and z_{1N}-z_{20N} the annual variation of the attribute in relative values, from 1993 till 2012).

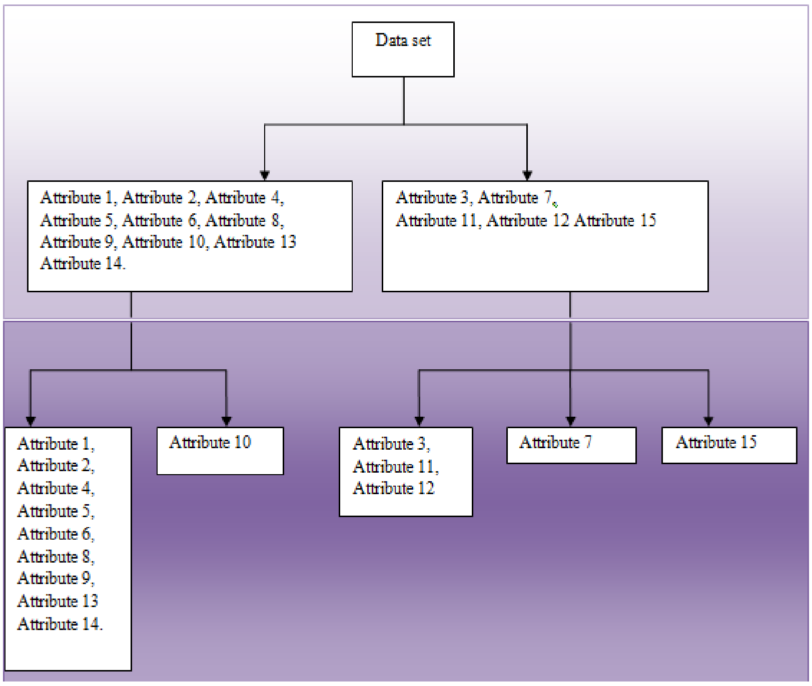
Table 1: Significance of the attributes

Number of attributes	Significance
z_{n1}	1 GDP per capita
z_{n2}	Exchange rate CZK to EUR
z_{n3}	Purchasing power parity
z_{n4}	Exchange rate CZK to USD
z_{n5}	The inflation rate
z_{n6}	State debt
z_{n7}	The share of gross value added (public sector)
z_{n8}	Population aged 15 years and over
z_{n9}	Work force (= economically active)
z_{n10}	The unemployment rate of persons aged 15-64 years
z_{n11}	The share of "High Technology Products" of the total foreign trade
z_{n12}	The share of "High Technology Products" of the total foreign trade
z_{n13}	The average wage
z_{n14}	The average monthly amount of retirement income
z_{n15}	Ratio of the average pension to average wage

Source: Czech Statistical Office, own calculation

We provided clustering for 5 cases: organized data into groups with 2,3,4,5 and 6 clusters. And evaluated received clusters on the basic of the Dunn's coefficient, Silhouette coefficient and Modificated coefficient *(E)*. Results are shown in table 1 and figure 1. The highest value of the Dunn's and Silhouette coefficients means the best clustering. The lowest value of the Modificated coefficient *(E)* means the best clustering. The Dunn's coefficient achieved its highest value for 2 clusters, whereas Silhouette showed the best number of clusters is 5. The lowest value of the Modificated coefficient *(E)* is for 2 clusters. Let's see on the data structure for cases 2 and 5 clusters.

Figure 4: Results of clustering (case with 2 and 5 clusters)



Source: own calculation

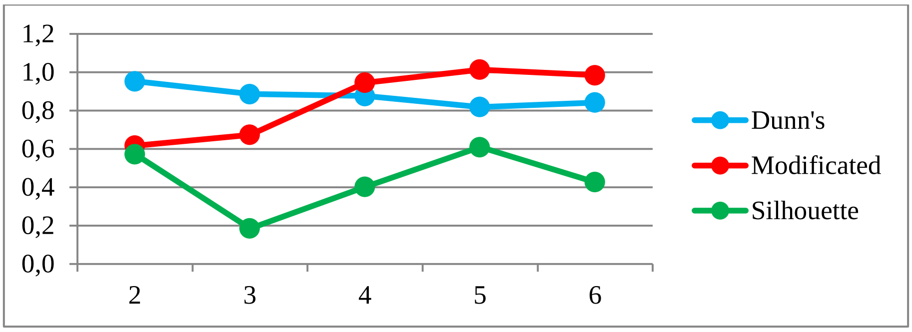
Table 2: Value of the coefficients

Name of the coefficient / Number of clusters	Silhouette coefficient	Dunn's coefficient	Modificated coefficient *(E)*
2	0,5726	**0,9525**	**0,6161**
3	0,1847	0,8857	0,6730
4	0,4015	0,8759	0,9444
5	**0,6086**	0,8177	0,9829
6	0,4264	0,8414	0,9837

Source: own calculation

This type of clustering is problematic, given the fact that the structure of the data is very heterogeneous and there is no clearly marked cluster group (can see on the figure 5 and 6). In addition on the figures 6 and 7 shown that some of the data is very distant from the others, which is why Silhouette coefficient refers them as separate clusters (grouping the noise points to separates clustering). Interestingly, after removing the attributes identified in separate clusters (attribute 10, 7.15) Silhouette coefficient takes a maximum value for the two clusters, without changing the structure of the initial clusters (clusters from the first experiment). Therefore, these attributes can be considered noise.

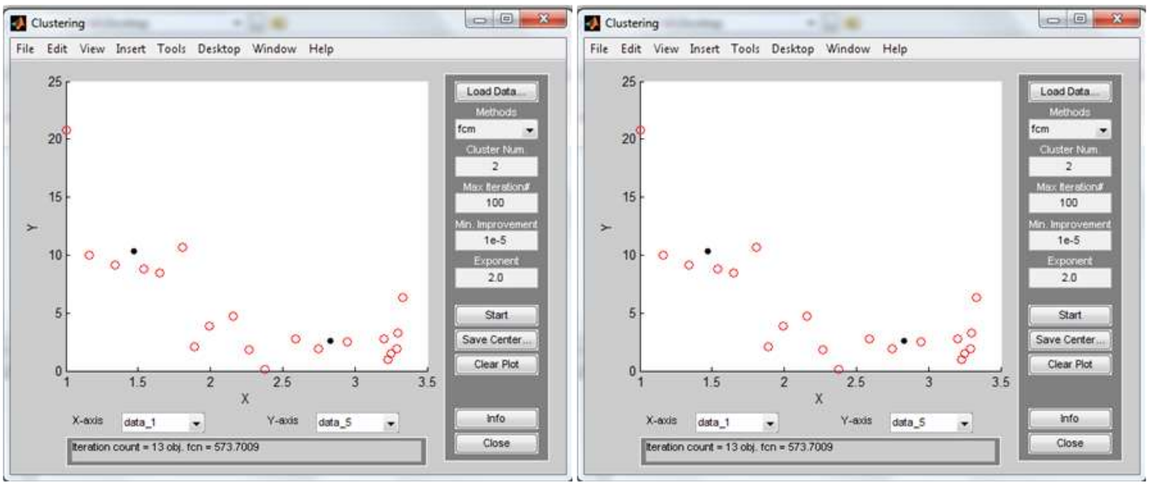
Figure 5: The value of the coefficient



Source: own calculation

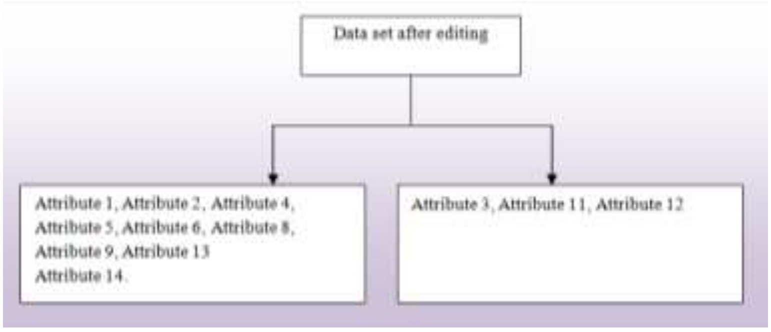
As we can see, the clustering for 2 groups showed the best data structure. The data inside each cluster are as similar as possible and the data between 2 clusters are as different as possible. After data editing (deleting the noise points) all three indexes show the same results: the best number of clusters is 2.

Figure 6: Results of clustering (case with 2 clusters)



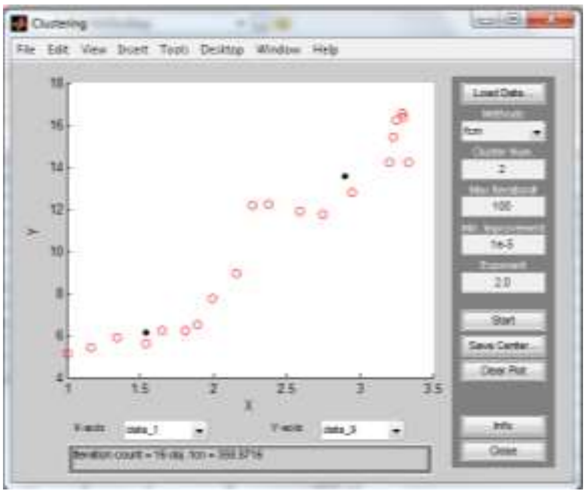
Source: Matlab, own calculation

Figure 7: Results of clustering (case with 2 clusters, after editing)



Source: own calculation

Figure 8: Results of clustering after editing (case with 2 clusters)



Source: Matlab, own calculation

Living standards may affect the outputs of the market economy, which includes people and their increasing life expectancy at birth (Šimpach; Langhamrová, 2012).

Conclusion

Presented paper analyzed the health and economic status in the Czech Republic in the last 20 years. From the obtained results it is visible that life expectancy at birth is increasing and this trend results in population ageing. Nowadays, the health status of the Czech population is improving due to progressed health care. Nevertheless, more and more people suffer from cancer, diabetes and circulatory system disorders. What is more, the unemployment rate and the amount of earnings in the Czech regions affect the entire economic process and the life of the population and that is why it is important to analyse these indicators (Löster; Langhamrová, 2012).

In this paper we used the cluster analysis for the classification of objects into groups. Clustering methods may be compared with each other or optimal number of clusters can be determined (Löster; Pavelka, 2013). Coefficients show that the best number of clusters for this data set is equal to 2. Presented dataset included some noise points that were identified and excluded through the analysis.

Acknowledgment

This article was supported by the Internal Grant Agency of University of Economics, Prague No. 68/2014 under the title "Economic and health connections of population ageing."

References

Česká republika od roku 1989 v číslech. (2014, April 15). Retrieved May 1, 2014, from http://www.czso.cz/csu/redakce.nsf/i/cr_od_roku_1989#05

Česká republika od roku 1989 v číslech: Tab.13.01 Vybrané ukazatele ze zdravotnictví. (2013, January 1). Retrieved January 1, 2014, from http://www.czso.cz/csu/redakce.nsf/i/cr_od_roku_1989#01

Demografická příručka 2012: 8-5b Míry úmrtnosti podle tříd příčin smrti v letech 1994–2012 (10. revize Mezinárodní klasifikace). (2013, January 1). Retrieved January 1, 2014, from http://www.czso.cz/csu/2013edicniplan.nsf/t/8E0017980B/$File/4032130805b.pdf

European health for all database (HFA-DB). (2013, January 1). Retrieved January 1, 2014, from http://data.euro.who.int/hfadb/

Fiala, T., & Langhamrová, J. (2013). Development of economic and social dependency and population ageing. Politická ekonomie, 61, 338-355.

Holčík, J. (2010, January 1). Přednáška o vývoji úmrtí v České republice v letech 1919–2006. Lecture conducted from Prague.

Löster, T. (2012). Kritéria pro hodnocení výsledků shlukování se známým zařazením do skupin založená na konfuzní matici, Forum Statisticum Slovacum, sv. 8, № 7 p. 85—89.

Löster, T., & Langhamrová, J. (2012). *Disparities between regions of the Czech republic for non-business aspects of labour market* .In Loster Tomas, Pavelka Tomas (Eds.), *6th International Days of Statistics and Economics* (pp. 689-702). ISBN 978-80-86175-86-7.

Löster, T., & Pavelka, T. (2013). *Evaluating of the Results of Clustering in Practical Economic Tasks*. In Loster Tomas, Pavelka Tomas (Eds.), *The 7th International Days of Statistics and Economics* (pp. 804-818). ISBN 978-80-86175-87-4.

Malina, A., & Drbal, C. (2005). Postgraduální vzdělávání v oblasti veřejného zdravotnictví v České republice. Retrieved May 2, 2014, from http://usm.lf1.cuni.cz/download/uvod_do_vz.pdf

Murthy, V., & Okunade, A. (2014, April 13). Population health status and economic growth in Chinese provinces: some policy implications. Retrieved from http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=Refine&qid=5&SID=Y1XP5MeVdEuKrPd6cuJ&page=2&doc=18#

Petruželka, L. (2014, April 23). Počet nádorových onemocnění se ztrojnásobí. *Lidové noviny*, p. 15.

Postgraduální medicína. (2014, January 1). Retrieved January 1, 2014, from http://www.postgradmed.cz

ŠIMPACH, O., & LANGHAMROVÁ, J. (2012). Czech Household Computer Facilities as a Reliable Variable in a Life Expectancy Forecast Model up to the Year 2060. In: IDIMT-2012. Linz: Trauner, p. 143–152. ISBN 978-3-99033-022-7.

White, K., Mondesir, F., Bates, L., & Glymour, M. (2014). Diabetes risk, diagnosis and control: Do psychosocial factors predict hemoglobin A1C defined outcomes or accuracy of self-reports? Retrieved from http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=2&SID=Y1XP5MeVdEuKrPd6cuJ&page=1&doc=5

Zdravotnická ročenka České republiky. (2013, January 1). Retrieved January 1, 2014, from http://www.uzis.cz/katalog/rocenky/zdravotnicka-rocenka-ceske-republiky

Contact

Ing. Kornélia Cséfalvaiová

University of Economics, Prague

W. Churchill Sq. 4

130 67 Prague 3

xcsek00@vse.cz

Mgr. Elena Makhalova
University of Economics, Prague
W. Churchill Sq. 4
130 67 Prague 3
xmake900@vse.cz

Ing. Michaela Antovová
University of Economics, Prague
W. Churchill Sq. 4
130 67 Prague 3
xantm11@vse.cz