Comparison of Health Care Results in Public Health Systems of European Countries

Viera Pacáková[1], Pavla Jindrová[2], David Zapletal[3]

[1] University of Pardubice
Faculty of Economics and Administration, Institute of Mathematics and Quantitative Methods
Studentská 84, 532 10 Pardubice, Czech Republic
E-mail: Viera.Pacakova@upce.cz
[2] University of Pardubice
Faculty of Economics and Administration, Institute of Mathematics and Quantitative Methods
Studentská 84, 532 10 Pardubice, Czech Republic
E-mail: Pavla.Jindrova@upce.cz
[3] University of Pardubice
Faculty of Economics and Administration, Institute of Mathematics and Quantitative Methods
Studentská 84, 532 10 Pardubice, Czech Republic
E-mail: David Zapletal@upce.cz

Abstract: *Health of citizens is a precondition for economic prosperity. Ensuring optimal functioning of health systems at limited financial resources of countries and regions is a difficult task. Its solution requires a lot of information that can be obtained only by appropriate analysis of data collected by authorized world, European and national institutions. About the state of public health systems there are collected and published a number of data on the regional, national, EU or OECD countries, continental and the world level. These databases can be used for comparative analyses on health status, risk factors to health, health care resources and utilization, as well as health expenditure and financing. Huge differences in health and healthcare exist between and within countries and regions. The aim of this article is to present the results of application of multivariate comparison statistical methods, regression analysis, factor analysis and cluster analysis, which provide an overview of the health care status and public health systems expenditures, various causal relations and regional differences or similarities. This information is essential to the development of national and international health policies for treatment and financial budget of public health systems.*

Keywords: public health systems, health status, comparison, multivariate methods

JEL codes: I11, I13, I14, C38, F36

1 Introduction

Quality health care system is a priority for the citizens of each country. Citizens' health is a core EU priority. EU health policy complements national policies to ensure that everyone living in the EU has access to quality healthcare. Cancer, heart disease, diabetes, respiratory, mental and other chronic diseases represent great suffering to citizens and represent a huge cost to society and the economy. It is estimated that they will cost the global economy around €22,5 trillion between 2012 and 2030 (EU, 2013).

Huge differences in health and healthcare exist between and within EU countries and regions. The level of disease and the age at which people die are strongly influenced by factors such as employment, income, education and ethnicity, as well as access to healthcare. For example, life expectancy at birth varies by 10 years between EU countries (EU, 2013). Huge differences in health also exist between social groups across the EU and within Member States. People with a lower level of education, a lower occupational class or a lower level of income, people in deprived areas and in poverty, the unemployed, the homeless, the disabled, the mentally or chronically ill, the elderly on low pensions and single parents tend to experience higher levels of disease and premature death.

Health is a precondition for economic prosperity; efficient spending on health can promote growth, so Europe needs smart investments in health. Cost-effective and efficient health expenditure can increase the quantity and the productivity of labor by increasing healthy life expectancy. Ensuring optimal functioning of health systems at limited financial resources of countries and regions is a very difficult task. Its solution requires a lot of information that can be obtained only by appropriate analysis of data collected by authorized world, European and national institutions.

The main aim of the article is to compare the health status and health outcomes in European countries, depending on risk factors, health expenditures and health care resources using appropriate statistical methods.

2 Methodology and Data

About the state of public health there are collected and published a number of data on the regional, national, EU countries, continental and the world level. These databases contain mostly time-space series of reporting aggregate data of diseases or deaths according to various diagnosis and a lot indicators of health care in the public and partly private sector.

The basic source of data is the database of the *World Health Organization* (WHO) for Europe, which provides a selection of core health statistics covering basic demographics, health status, health determinants and risk factors, and health-care resources, utilization and expenditure in the 53 countries of the WHO European Region. The analysis is focused on all EU countries that are completed by Norway and Switzerland, i.e. 30 countries. In accordance with the stated objectives we have chosen these variables (Source: WHO, Eurostat available from http://www.who.int/en/; http://ec.europa.eu/eurostat):

A. Health expenditure

$X1$ Total health expenditure, PPP$ per capita, 2013

$X2$ Gross domestic product (GDP), US$ per capita, 2013

B. Health status

$X3$ Life expectancy at birth (years), 2011

$X4$ Life expectancy at age 65 (years), 2012

$X5$ Probability of dying before age 5 per 1000 live births, 2012

C. The incidence of serious diseases

$X6$ SDR, diseases of circulatory system, all ages, per 100 000, 2012

$X7$ SDR, ischemic heart disease, all ages, per 100 000, 2012

$X8$ SDR, cerebrovascular diseases, all ages, per 100 000, 2012

$X9$ SDR, malignant neoplasms, all ages, per 100 000, 2012

$X10$ SDR, trachea/bronchus/lung cancer, all ages, per 100 000, 2012

$X11$ SDR, diabetes, all ages, per 100 000, 2012

D. Risk factors

$X12$ Age-standardized prevalence of overweight (defined as BMI ≥ 25 kg/m2) in people aged 18 years and over, WHO estimates (%), 2010

$X13$ % of regular daily smokers in the population, age 15+, 2013

$X14$ Pure alcohol consumption, litres per capita, age 15+, 2011

E. Health care resources

$X15$ Hospital beds per 100 000, 2013

$X16$ Physicians per 100 000, 2013

$X17$ Pharmacists (PP) per 100 000, 2013

$X18$ Average length of stay, all hospitals 2013

For analysis the most recent available data were used. The data of some variables have a one or two year's lag but the minimal change for most of these indicators is typical during the years. Application of multivariate statistical methods, such as correlation analysis, factor analysis and cluster analysis, preferably with graphical output, provides an overview of the gravity of the health situation by monitoring indicators, various causal relations and regional similarities or differences (Pacáková and Jindrová, 2014). The selected statistical methods were applied using MS Excel and statistical software packages Statgraphics Centurion XV and Statistica 12.

3 Results and Discussion

Statistical analysis begins with examination of causal relationships between variables. For this we use Spearman rank correlations (Šoltés, 2008) between each pair of variables. These correlation coefficients range between -1 and +1 and measure the strength of the association between the variables. In contrast to the more common Pearson correlations, the Spearman coefficients are computed from the ranks of the data values rather than from the values themselves. Consequently, they are less sensitive to outliers than the Pearson coefficients.

The values of Spearman coefficients provide a number of interesting facts. Variable *X*1-total health expenditure, PPP$ per capita, is strongly positively correlated with variables *X*2 ($r_{1,2}$ = 0,957), *X*3 ($r_{1,3}$ = 0,748), *X*4 ($r_{1,4}$ = 0,786), strongly negatively correlated with variables *X*6 ($r_{1,6}$ = -0,808), *X*8 ($r_{1,8}$=-0,875), moderately potent negatively correlated with variables *X*5 ($r_{1,5}$ = -0,510), *X*9 ($r_{1,9}$ = -0,524) and *X*13 ($r_{1,13}$ = -0,584) and a weakly correlated with variables *X*10, *X*11, *X*12, *X*14 and with all variables *X*15-*X*18 of health care resources. Variables *X*15-*X*18 are poorly correlated with all other variables, only one Spearman rank coefficients slightly exceeds 0,5.

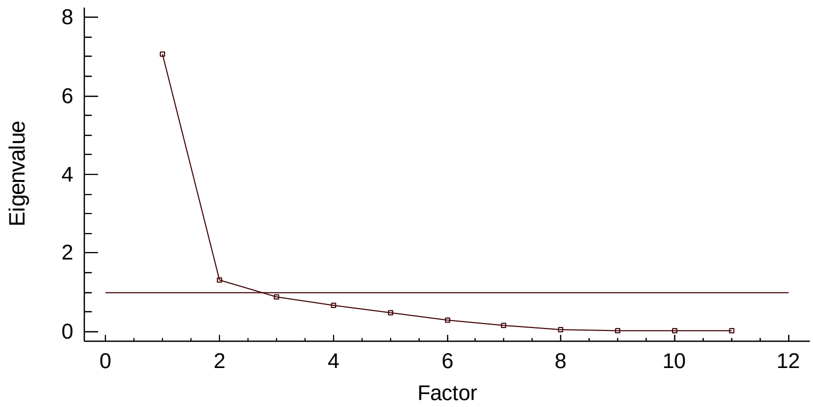
The dependence of the variables *X*3, *X*4, expressed by the Spearman rank correlation is $r_{3,4}$ = 0,915. Both variables *X*3, *X*4 are strongly positive correlated with variables *X*1, *X*2 and strongly negative with variables *X*6 ($r_{3,6}$ = -0,814, $r_{4,6}$ = -0,813), *X*7 ($r_{3,7}$ = -0,738, $r_{4,7}$ = -0,729), *X*8 ($r_{3,8}$ = -0,821, $r_{4,8}$ = -0,815), *X*9 ($r_{3,9}$ = 0,742, $r_{4,9}$ = -0,733) and are a weakly correlated with all other variables.

Based on these results, we tried to apply the factor analysis on 11 variables: *X*1, *X*2, *X*3, *X*4, *X*5, *X*6, *X*7, *X*8, *X*9, *X*13, *X*14.

- **Factor analysis results**

The purpose of factor analysis (Stankovičová and Vojtková, 2007) is to obtain a small number of factors which account for most of the variability in the 11 variables. Factor is a linear combination of the original variables. In this case, two factors have been extracted (Figure 1), since two factors had eigenvalues greater than to 1,0. Together they account for 76,093 % of the variability in the original data. Since we have selected the principal components method, the initial communality estimates have been set to assume that all of the variability in the data is due to common factors.

Figure 1 Scree Plot



Source: Own calculation, output from Statgraphics Centurion XV

The *Scree Plot* (Hebák et al., 2004) can be very helpful in determining the number of factors to extract, because displays the eigenvalues associated with a component or factor in descending order versus the number of the components or factors. We use scree plots to visually assess which factors explain most of the variability in the data.

Factor loadings (Table 1) present the correlation between the original variables and the factors and they are the key to understanding the nature of a particular factor. Rotation is useful method used to rotate the factor loading matrix after it has been extracted. *Varimax rotation* maximizes the variance of the squared loadings in each column (Johnson and Wichern, 2007).

Table 1 Factor Loading Matrix after Varimax Rotation

Variable	Factor 1	Factor 2
*X*1	-0,829504	-0,231396
*X*2	-0,772249	-0,212522
*X*3	-0,807011	-0,538974
*X*4	-0,831973	-0,453525
*X*5	0,868977	-0,088421
*X*6	0,883450	0,382960
*X*7	0,614292	0,599055
*X*8	0,922428	0,186451
*X*9	0,421982	0,698485
*X*13	0,643112	0,098378
*X*14	-0,092937	0,858875

Source: Own calculation, output from Statgraphics Centurion XV

Substantive interpretation of two extracted factors is based on the significant higher loadings. Factor 1 (*F*1), which explains 64,266 % variability of the variability in the data, has 6 significant loadings with positive signs with variables *X*5-*X*13, and significant loadings with negative signs with variables *X*1-*X*4. Therefore, this factor can be interpreted as a *factor of favourable conditions and health care results*. The low value of this factor means high level of health expenditure and life expectancy and a low level of serious diseases incidence. The higher the values of Factor 1 the worse are conditions and health care results. Significant positive correlation with variables *X*7, *X*9 and *X*14 is the reason that we interpret Factor 2 (*F*2) as a *factor of negative consequences and factors of health.* The higher the ischemic heart disease, malignant neoplasms incidence and pure alcohol consumption, the higher is the value of Factor 2 and vice versa.

Table 2 Table of Factor Scores

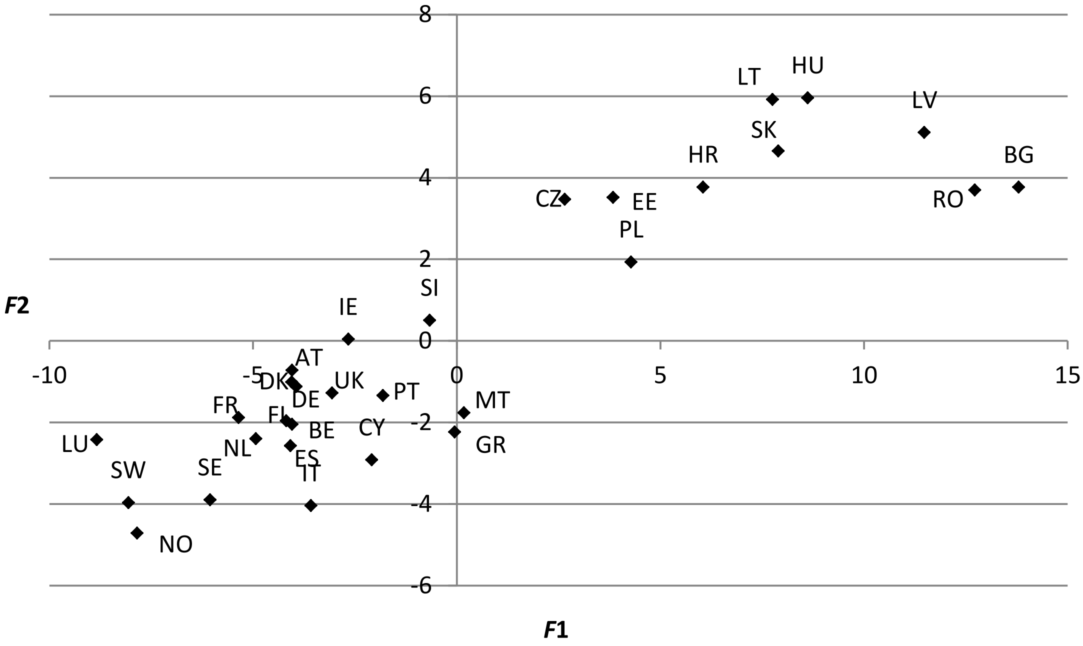
Country	*F*1	*F*2	Country	*F*1	*F*2
Austria - AT	-4,039	-0,721	Latvia - LV	11,483	5,102
Belgium - BE	-4,041	-2,043	Lithuania - LT	7,759	5,912
Bulgaria - BG	13,798	3,761	Luxembourg - LU	-8,834	-2,421
Croatia - HR	6,048	3,767	Malta - MT	0,179	-1,762
Cyprus - CY	-2,083	-2,919	Netherlands - NL	-4,930	-2,401
Czech Republic - CZ	2,654	3,464	Norway - NO	-7,847	-4,714
Denmark - DK	-4,048	-1,010	Poland - PL	4,282	1,932
Estonia - EE	3,844	3,513	Portugal - PT	-1,807	-1,343
Finland - FI	-4,182	-1,959	Romania - RO	12,722	3,698
France - FR	-5,353	-1,885	Slovakia - SK	7,898	4,653
Germany - DE	-3,935	-1,121	Slovenia - SI	-0,667	0,503
Greece - EL	-0,045	-2,234	Spain - ES	-4,077	-2,574
Hungary - HU	8,620	5,953	Sweden - SE	-6,054	-3,898
Ireland - IE	-2,659	0,037	Switzerland - SW	-8,057	-3,965
Italy - IT	-3,575	-4,042	United Kingdom -UK	-3,059	-1,282

Source: Own calculation, output from Statgraphics Centurion XV

Table 2 shows the factor scores for each selected country. A country with a low value of *F*1, together with a low value of *F*2 has (by interpretation of these factors) *favourable conditions and health care* status. On the other hand countries with high values of *F*1 and *F*2 are in bad situation concerning the health status and risk factors.

Graphical display (Figure 2) of selected countries in a two-dimensional coordinate system with axes *F*1 and *F*2 allows us to quickly assess the health situation in each country and allows also compare situation in all selected countries.

Figure 2 Location Selected Countries in the Coordinate System of the Factors *F*1 and *F*2



Source: Own processing according to Table 2

- **Cluster analysis results**

Cluster analysis (Hair et al., 2007) is an analytical technique that can be used to develop meaningful subgroups of object, in our case of countries. The objective is to classify a sample of objects into a small number of mutually exclusive groups based on the similarities among the objects. The clusters are groups of observations with similar characteristics.

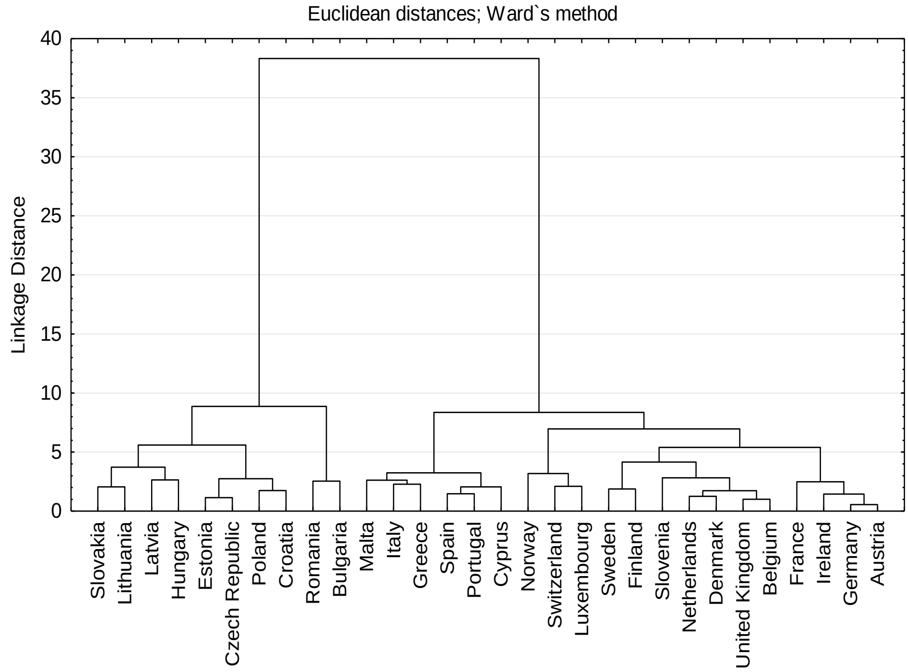
In order to create clusters of observations, it is important to have a measure of "similarity" so that like objects may be joined together. When observations are to be clustered, the closeness is typically measured by the distance between observations in the *p* dimensional space of the variables. We have used *Euclidian distance* for measuring the distance between two items (i.e. countries), represented by x and y

$$d(x,y) = \sqrt{\sum_{i=1}^{p}(x_i - y_i)^2} \tag{1}$$

A number of different algorithms are provided for generating clusters. Some of the algorithms are *agglomerative*, beginning with separate clusters for each observation and then joining clusters together based upon their similarity. To form the clusters, the procedure began with each observation in a separate group. It then combined the two observations which were closest together to form a new group. After re-computing the distance between the groups, the two groups then closest together are combined. This process is repeated until only 1 group remained.

Ward's method which has been used for clustering defines the distance between two clusters in terms of the increase in the sum of squared deviations around the cluster means that would occur if the two clusters were joined. The results of the analysis are displayed in several ways, including a dendrogram. Working from the bottom up, the dendogram shows the sequence of joins that were made between clusters. Lines are drawn connecting the clustered that are joined at each step, while the vertical axis displays the distance between the clusters when they were joined.

Figure 3 The Dendrogram of Cluster Analysis



Source: Own calculation, output from Statistica 12

The results of cluster analysis by 11 variables, the same as in factor analysis, are consistent with the results of factor analysis, as we can see from dendrogram on Figure 3. Cluster, consisting of the old EU countries, complemented by the Norway and Switzerland, has been joined with a cluster of new EU member states on a very large distance. It means that the health situation in these two groups of countries noticeably different and according to the results of factor analysis to the disadvantage of new members of the EU.

- **Multidimensional comparative analysis**

Multidimensional comparative analysis (Sokolowski, 1999) deals with the methods and techniques of comparing multi-feature objects, in our case selected European countries. The objective is establishing a linear ordering among a set of objects in a multidimensional space of features, from the point of view of certain characteristics which cannot be measured in a direct way (the level of socio-economic development, the standard of living, product quality, economic performance, public health situation ...).

At the beginning of the analysis, the type of each variable should be defined. It is necessary to identify whether the "great" values of a variable positively influence the analysed processes (such variables are called stimulants) or whether their "small" values are favourable (these are called destimulants). In comparative analyses of public health by 11 variables the stimulants are X1, X2, X3, X4, destimulants X5 – X9, X13, X14. The variables of the third type, nominants (which have an "optimal" level and deviations either upwards or downwards are undesirable), like X15 – X18) are not suitable for this analysis.

The initial variables employed in composing an aggregate measure are, usually, measured in different units. The aim of normalisation is to bring them to comparability.

Normalisation is performed according to the formulas (Stankovičová and Vojtková, 2007):

$$\text{for stimulants} \rightarrow \quad b_{ij} = \frac{x_{ij}}{x_{\max, j}} \cdot 100 \tag{2}$$

$$\text{for destimulants} \rightarrow \quad b_{ij} = \frac{x_{\min, j}}{x_{ij}} \cdot 100 \tag{3}$$

The aggregate measure of health care level for each country has been calculated as the average of the point b_{ij}, i = 1, 2, ..., 30. According to the formulas (2), (3) obviously applies that the more higher the value of the average score, the higher the level of health care. The rank assigned to the countries by ascending order from 1 to 30 we can see in Table 3.

Table 3 The Results of Multidimensional Comparative Analysis

Rank	Countries	Points	Rank	Countries	Points
1	Luxembourg	79,822	16	Ireland	61,973
2	Switzerland	78,260	17	Cyprus	61,705
3	Norway	78,022	18	Greece	61,668
4	Sweden	74,475	19	Malta	61,334
5	France	72,753	20	Slovenia	58,709
6	Netherlands	70,688	21	Estonia	52,970
7	Finland	70,208	22	Czech Republic	50,519
8	Belgium	68,517	23	Lithuania	50,344
9	Spain	66,894	24	Poland	49,022
10	Denmark	66,462	25	Slovakia	48,699
11	United Kingdom	66,324	26	Romania	46,475
12	Italy	65,737	27	Croatia	46,130
13	Germany	64,227	28	Latvia	44,914
14	Austria	63,858	29	Bulgaria	44,429
15	Portugal	63,194	30	Hungary	42,636

Source: Own calculation

The highest level of health care was observed in Luxembourg, then followed Switzerland, Norway, Sweden, France ... the latest position is occupied by Hungary. The last ten include all former socialist countries.

Table 4 shows the Spearman rank coefficients between each pair of variables *X*1 (total health expenditure), *B*, *C*, *D*, *E*. Variables *B*, *C*, *D*, *E* are aggregate (synthetic) variables which have been computed by formulas (2), (3) from variables *X*3, *X*4, *X*5 of health status – variable *B*, from variables *X*6 – X11 of the incidence of serious diseases – variable *C*, from variables *X*12 – *X*14 of risk factors – variable *D* and from variables *X*15 – *X*18 of health care resources – variable *E*.

Table 4 Spearman Rank Coefficients

	***X*1**	***B***	***C***	***D***	***E***
***X*1**		0,694	0,741	0,517	0,089
B	0,694		0,706	0,377	0,095
C	0,741	0,706		0,456	0,002
D	0,517	0,377	0,456		-0,048
E	0,089	0,095	0,002	-0,048	

Source: Own calculation

As shown in Table 4, variable *E* is not correlated with any of the variables *X*1, *B*, *C*, *D*. We can say that the high values of variables which characterize health care resources (var. *E*) have no effect on health status (var. *B*), on incidence of serious diseases (var. *C*) and on risk factors (var. *D*) in selected European countries. Health care resources are also not correlated with the total health expenditure (var. *X*1), the value of Spearman correlation coefficient is only 0,089. Interesting results also provide Spearman Rank Coefficients of synthetic variable *D* of risk factors with other synthetic variables. It confirms the low impact of risk factors on health status ($r_{B,D}$ = 0,377) and moderate impact of risk factors on the incidence of serious diseases ($r_{C,D}$ = 0,456).

4 Conclusions

The results of statistical analysis confirm the appropriateness of the used methods and the suitability of the chosen variables of health situation in EU countries. The chosen methods enable to extract two common factors instead of the original 11 variables. This allowed obtaining transparent and visual information about the health care situation in the EU countries completed by Norway and Switzerland and the possibility of graphical presentation of results. Cluster analysis and multidimensional comparative analysis supplemented and deepened results of factor analysis. It means that the health situation in the group of the old and the new members of European Union is noticeably different and according to the results of factor analysis to the disadvantage of the new members of EU. The multidimensional comparative analysis provides some surprising results, such insignificant impact of health care resources variables in the health status of the European countries. This suggests ineffective functioning of the public health systems.

References

EU (2013): Public health. Luxembourg: Publications Office of the European Union. Retrieved from: http://ec.europa.eu/health/health_policies/docs/improving_health_for_all_eu_citizens_en.pdf

Eurostat database. Retrieved from: http://ec.europa.eu/eurostat

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., Tatham, R. L. (2007). *Multivariate Data Analysis*. Sixth Ed. New Jersey: Pearson Education.

Hebák, P., Hustopecký, J., Jarošová, E., I. Pecáková, I. (2004). *Vícerozměrné statistické metody* (Multivariate statistical methods). Praha: Informatorium.

Johnson, R. A., Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Sixth Ed. New Jersey: Pearson Prentice Hall.

Pacáková, V., Jindrová, P. (2014) Quantification of Selected Factors of Longevity. In: *Proceedings of the 2014 International Conference on Economics and Applied Statistics* (*EAS '14*), Saint Petersburg, 2014, pp. 218-221.

Sokolowski, A. (1999). *Regional Differences in Living Standards in Eastern European Countries – Changes during the Transition*. Research Support Scheme, Open Society Institute. Retrieved from: http://rss.archives.ceu.hu/archive/00001052/01/52.pdf

Stankovičová, I., Vojtková, M. (2007). *Viacrozmerné štatistické metódy s aplikáciami* (Multivariate statistical methods with applications). Bratislava: IURA Edition.

Šoltés, E. (2008). Regresná a korelačná analýza s aplikáciami (Regression and correlation analysis with applications). Bratislava: Iura Edition.

World Health Organization (WHO) database. Retrieved from: http://www.who.int/en/