Statistical Analysis of Health Inequalities in European Countries

Viera Pacáková[1,*], *Pavla Jindrová*[1], and *Lucie Kopecká*[1]

[1]University of Pardubice, Faculty of Economics and Administration, Institute of Mathematics and Quantitative Methods, Studentská 95, 532 10 Pardubice, Czech Republic

Abstract. Efficiently functioning health systems are a prerequisite for high-quality health care and healthy life expectancy. Health care management at all levels requires a lot of information that can be obtained only by relevant analyses of health data. There are collected and regularly updated on-line published a large number of databases and enormous number of indicators about health status, health expenditures and health systems functioning at regional, national, EU member countries, OECD countries and on the world level. Paradoxically, the extent of these data sets is the reason why without at least basic statistical analysis the level of provided information is minimal. Advanced statistical methods aimed at reducing the dimension and quantification of causal relationships can provide significant information added value. The objective of this article is to analyse causal relationships between health status, health expenditures and sources of health care in selected European countries and to identify determinants of health inequalities in European countries by applying multidimensional statistical methods.

1 Introduction

Quality health care system is a priority for citizens of each country and a precondition for economic prosperity. Significant differences in health status exist between European countries and regions. Health inequalities exist along many demographic or social dimensions, including sex, age, geographic area and socio-economic status.

The Europe 2020 strategy, which aims to deliver smart, sustainable and inclusive growth with high levels of employment, productivity and social cohesion, is the main vehicle for achieving this. Europe 2020 sets targets against which the process will be measured and emphasises that a major effort is needed to reduce health inequalities to ensure that everybody can benefit from economic growth. [1]

Actions to improve health are an important part of two of the seven flagship initiatives that contribute to implementing Europe 2020. Achieving the Europe 2020 targets, particularly the target of reducing by 20 million the number of people in or at risk of poverty and social exclusion, will contribute substantially to creating a more equitable distribution of health. [2]

Over the last century, average health status improved in Europe. However, these gains are not evenly distributed across countries or across social groups within the same country. Health inequities can be observed in higher and lower income countries across the European Region. Despite improvements of health status in European countries, important questions about how successful countries are in achieving the Europe 2020 targets on different dimensions of health system performance remain. Answering these questions is by no mean an easy task. The aim of this article is to help shed light on how well countries do in promoting the health of their population and on several dimensions of health system performance. Application of advanced multidimensional statistical method on a selected set of indicators of health and health system functioning in selected European countries could summarize some of the relative strengths and weaknesses and can be useful to identify possible priority areas for actions.

According to the above mentioned we have used correlation, factor and cluster analysis [3] on a selected set of health and social indicators from the OECD Health Statistics [4] and OECD Social Statistics databases [5], so selected countries are the European countries that are the members of OECD. For analysis were used the most recent data available.

2 Data and methods

2.1 Selected indicators

In accordance with the objectives of analysis we have selected 19 indicators (Table 1). Indicators (variables) H1 to H7 together characterize the state of health, E1 to E3 the state of healthcare expenditure, C1 to C5 healthcare resources and indicators S1 to S4 the social determinant of health of the inhabitants in the selected countries.

* Corresponding author: viera.pacakova@upce.cz

Table 1. Indicators (2016 or latest available year).

Health status	
Health status	
H1	Life expectancy at birth
H2	Healthy life expectancy
H3	Healthy life years at age 65
H4	Mortality Circulatory syastem (standardized death rates per 100 000 inhabitants)
H5	Mortality Cancer (standardized death rates per 100 000 inhabitants)
H6	Amenable deaths of residents (standardized death rates per 100 000 inhabitants)
H7	Perceived health status good/very good health status, total 15+ (% of population)
Health expenditures	
E1	Health expenditure as a share of GDP
E2	Health expenditure per capita (US PPP)
E3	Expenditures Long-term care (% of total expenditure on health care)
Health care	
C1	Employment in health and social work as a share of total employment
C2	Doctors (total) per 1000 inhabitants
C3	Nurses (total) per 1000 inhabitants
C4	MRI units (per million population)
C5	Computed tomography scanners (per 1 000 000 population)
Social determinants	
S1	Poverty rate (relative threshold)
S2	Living in working households
S3	Disposable income (US dolar at PPP rates)
S4	Unemployment rate (% of labour force aged 15+)

Source: OECD Health Statistics 2017

2.2 Selected multidimensional methods

Factor analysis is a statistical approach that can be used to analyse interrelationships among a large number of variables and to explain these variables in terms of their common underlying factors. The general purpose of factor analytic techniques is to find a way of condensing (summarizing) the information contained in a number of original variables into a smaller set of new composite factors with a minimum loss of information. Numerous variations of the general factor model are available. The two most frequently employed approaches are principal component analysis and common factor analysis. The component model is used when the objective is to summarize most of the original information (variance) in a minimum number of factors. The Scree Plot 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 factors. [6], [7]

An important concept in factor analysis is the rotation of factors. In practice, the objective of all methods of rotation is to simplify the rows and columns of the factor matrix to facilitate interpretation. The Varimax criterion centres on simplifying the columns of the factor matrix. With the Varimax rotation approach, there tend to be some high loadings (i.e., close to -1 or +1) and some loadings near 0 in each column of the matrix. The Factor Loadings show the correlation between the original variables and the factors and they are the key to understanding the nature of a particular factor. The Factor Scores in output of Factor analysis procedure display the values of the rotated factor scores for each of n cases, in our analysis for each of 25

European countries. Factor score show where each country falls with respect to the extracted factors. [6], [7]

Cluster Analysis procedure is designed to group observations (countries) into clusters based upon similarities between them. A number of different algorithms is provided for generating clusters and are described in detail in many statistical publications, for example in [3]. We have used the agglomerative algorithm, beginning with separate clusters for each observation or variable 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 one group remained. The results of the analysis are displayed in a dendrogram.

The distance between two observations we calculate by Euclidean distance, defined as

$$D_E(x_i, x_j) = \sqrt{\sum_{i=1}^{m}(x_{il} - x_{jl})^2} \qquad (1)$$

and distance between two clusters by Ward's method. Ward's method 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 dendrogram 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. [7], [8], [9]

2 The results of multidimensional methods

The results of the correlation analysis in graphic form show the correlation coefficients between each pair of indicators and their clusters. The results indicate a strong positive dependence of health indicators on E1-E3 healthcare expenditure, employment in health and social work (C1), as well as the number of nurses per 1000 inhabitants (C3 indicator), moderate dependence on the number of physicians and technical C4 and C5 sources and strong negative dependence on social determinants S1-S4, see Figure 1.
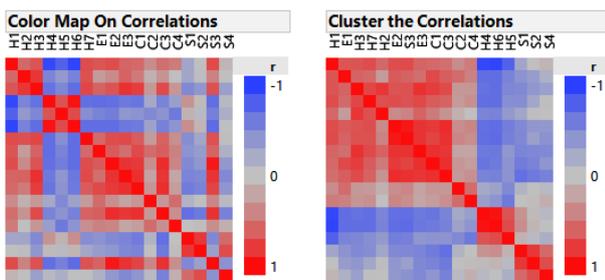

Figure 1. Correlation maps of selected health indicators.
Source: OECD Health Statistics 2017, self-processed in SAS JMP

By application of factor analysis we try to obtain a small number of common factors which account for most of the variability in the original variables. To assess the suitability of indicators for the factor analysis, we applied the Kaiser-Meyer-Olkin measure.

The KMO = 0.7544399 show suitability of the source variables for factor analysis. The first three factors were used according to eigenvalues which are higher than 1, see Figure 2.

Factor loadings which present the correlation between the original variables and the factors and they are the key to understanding the nature of a particular factor. After varimax rotation we obtained factor loadings shown in Table 2. Rotation is performed in order to simplify the explanation and naming of the factors. Based on those factor loadings, we found out that the 1st factor has strong positive correlation with the indicators of health status and health expenditures, the 2nd factor demonstrated rather moderate positive correlation with the indicators of Employment in health and social work and Disposable income and strong negative correlation with other social indicators, the 3rd factor showed strong positive correlation with the indicators of personal and technical resources. The high values of each factor mean a high level of the observed reality.

Based of above mentioned we have named three common factor as:
- F1 – Factor of health status and health expenditures,
- F2 – Factor of social determinants of health,
- F3 – Factor of personal and technical resources of healthcare.

Table 3 shows the factor scores for each monitored country. The Factor Scores displays the values of the rotated factor for each country. Graphical display of countries in a two-dimensional coordinate system with the axes of the selected factors allows us to quickly assess the observed situation in each country and also compare the situation in different countries.
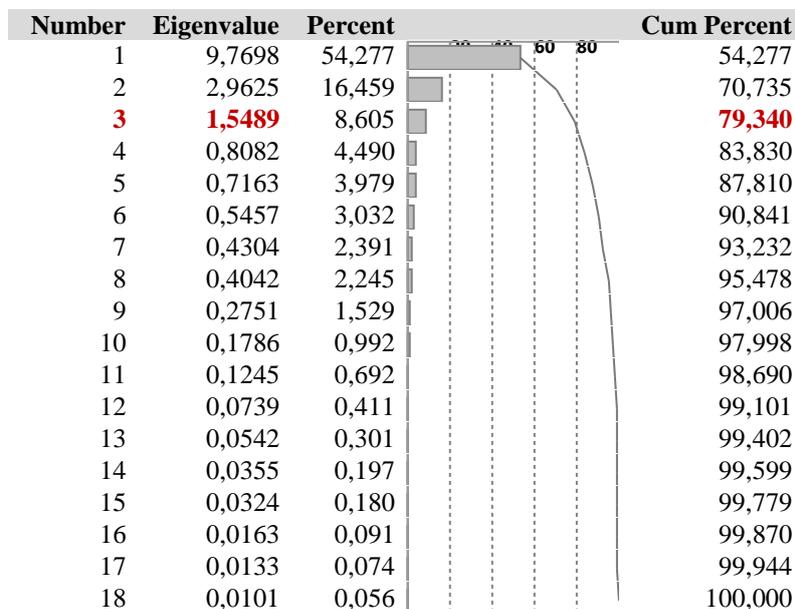
Number	Eigenvalue	Percent		Cum Percent
1	9,7698	54,277		54,277
2	2,9625	16,459		70,735
3	**1,5489**	**8,605**		**79,340**
4	0,8082	4,490		83,830
5	0,7163	3,979		87,810
6	0,5457	3,032		90,841
7	0,4304	2,391		93,232
8	0,4042	2,245		95,478
9	0,2751	1,529		97,006
10	0,1786	0,992		97,998
11	0,1245	0,692		98,690
12	0,0739	0,411		99,101
13	0,0542	0,301		99,402
14	0,0355	0,197		99,599
15	0,0324	0,180		99,779
16	0,0163	0,091		99,870
17	0,0133	0,074		99,944
18	0,0101	0,056		100,000

Figure 2. Eigenvalues and percent of explained variability.
Source: OECD Health Statistics 2017, self-processed in SAS JMP

Table 2. Factor loadings Matrix After Varimax Rotation.

Indicator	Factor 1	Factor 2	Factor 3
H1	**0,906528**	-0,00934	0,318014
H2	**0,78895**	0,052413	-0,200819
H3	**0,768442**	0,426931	0,124728
H4	**-0,911259**	-0,05192	-0,199143
H5	**-0,633472**	0,093032	-0,296436
H6	**-0,900476**	-0,07838	-0,253168
H7	**0,807485**	0,25077	0,106453
E1	**0,660973**	0,342799	0,463807
E2	**0,659004**	0,529455	0,330547
E3	**0,677245**	0,520442	0,07212
C1	0,514384	**0,634132**	0,192238
C2	0,152075	0,116312	**0,751755**
C3	0,481221	0,306835	**0,752156**
C4	0,219703	-0,03421	**0,848875**
C5	-0,022708	-0,02492	**0,841822**
S1	-0,272452	**-0,79072**	0,045056
S2	0,00262	**-0,90275**	0,064685
S3	**0,665835**	**0,511544**	0,317976
S4	0,116944	**-0,87776**	0,065708

Source: self-processed in Statistica 12

In the coordinate system of the factors F1 and F2 three groups of countries were created, one with high values of both factors, including all the old EU countries, the second with low values of both factors, including the new EU countries and the third with the middle level of the first and the low to medium level of the second factor, see Figure 3.
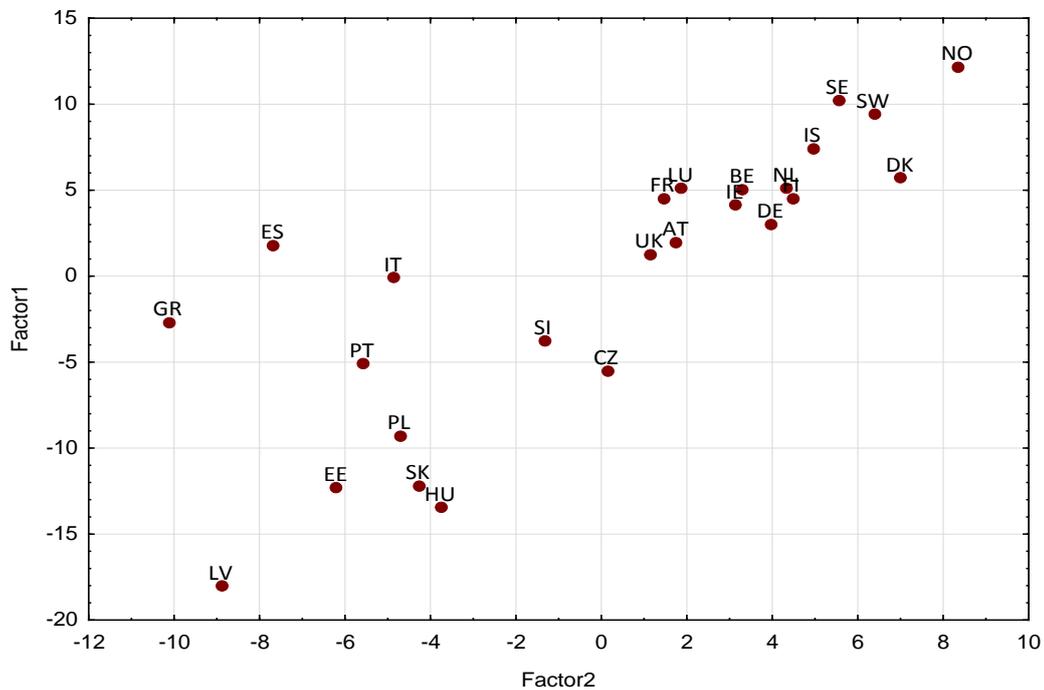

Figure 3. Location of countries in the coordinate system of the factors F1, F2.
Source: self-processed in Excel

Table 3. Table of Factor Scores.

Country	Code	Factor1	Factor2	Factor3
Austria	AT	2,00378	1,74353	4,46643
Belgium	BE	5,08323	3,29357	0,57092
Czech Republic	CZ	-5,45320	0,11487	-3,71813
Denmark	DK	5,73055	6,99177	3,07196
Estonia	EE	-12,25560	-6,22476	-4,45906
Finland	FI	4,55566	4,45395	3,07897
France	FR	4,55871	1,43485	0,59611
Germany	DE	3,03376	3,95455	5,77128
Greece	GR	-2,62676	-10,14940	1,14920
Hungary	HU	-13,41090	-3,78231	-8,10042
Iceland	IS	7,46812	4,94746	3,91766
Ireland	IE	4,16867	3,13029	-1,21098
Italy	IT	-0,05871	-4,87130	2,99287
Latvia	LV	-17,99340	-8,90168	-4,54024
Luxembourg	LU	5,15976	1,83866	-0,01253
Netherlands	NL	5,14330	4,32427	0,07244
Norway	NO	12,21040	8,34577	4,50509
Poland	PL	-9,24465	-4,73269	-6,37747
Portugal	PT	-5,06248	-5,59551	-0,95386
Slovakia	SK	-12,12520	-4,27758	-5,29907
Slovenia	SI	-3,72020	-1,35464	-3,78259
Spain	ES	1,85508	-7,69628	0,83205
Sweden	SE	10,24510	5,53130	3,11245
Switzerland	SW	9,44452	6,36282	7,31856
United Kingdom	UK	1,29049	1,11854	-3,00166

Source: Self-processed in Statistica 12

Figure 4, showing the location of Europe's monitored countries in the coordinate systems of factors F1 and F3, indicates that F3 - Factor of personnel and technical resources of health care has a positive effect on F1 - Health and healthcare expenditure. Three clusters of monitored countries were created. One group consists of states of Northern and Western Europe with high values of both factors, the group with the lowest level of both factors again forming the same five countries as in Figure 2 and third, the most numerous group of countries with medium level of both factors, again belonging the Czech Republic.

Figure 5 shows the dependence of Factor F3 - Factor of Personal and Technical Resources of Health Care and Factor F2 - Factor of Social Determinants of Health. Again, there is a direct dependence of these two factors and, similarly to Fig. 3. We can observe the specific situation of the groups of countries of Greece, Spain, Italy and Portugal where, even at a low level of social

determinants of health, the medium to high level of personnel and technical resources of health care. Unfortunately, the group of five new EU Member States with the lowest level of both factors is the same as in Fig. 3 and Fig. 4.

The factor analysis based on principal component method resulted in 3 mutually independent factors, each representing one dimension of health situation. These factors are appropriate for the cluster analysis. Dendrogram and parallel plots represent the results in the visual form, see Figure 6.

According to the legend, the red colour presents the high, so desired value of each factor, and the size of the values is indicated by the intensity of the colour. Low factor values are analogously shown in blue.

The colour map in the 1st column refers to the 1st factor of the health status and health expenditures, the 2nd column of colour map represents the social determinants of health and the 3rd column represents the

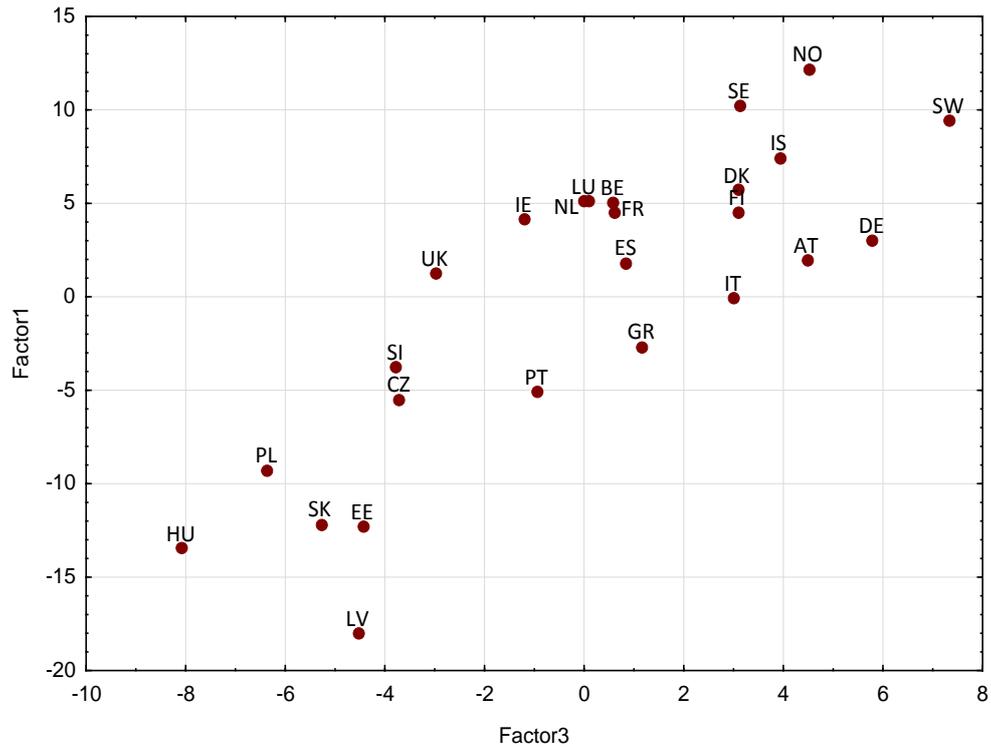* Corresponding author: viera.pacakova@upce.cz

Figure 4. Location of European countries in the coordinate system of the factors F3, F1.
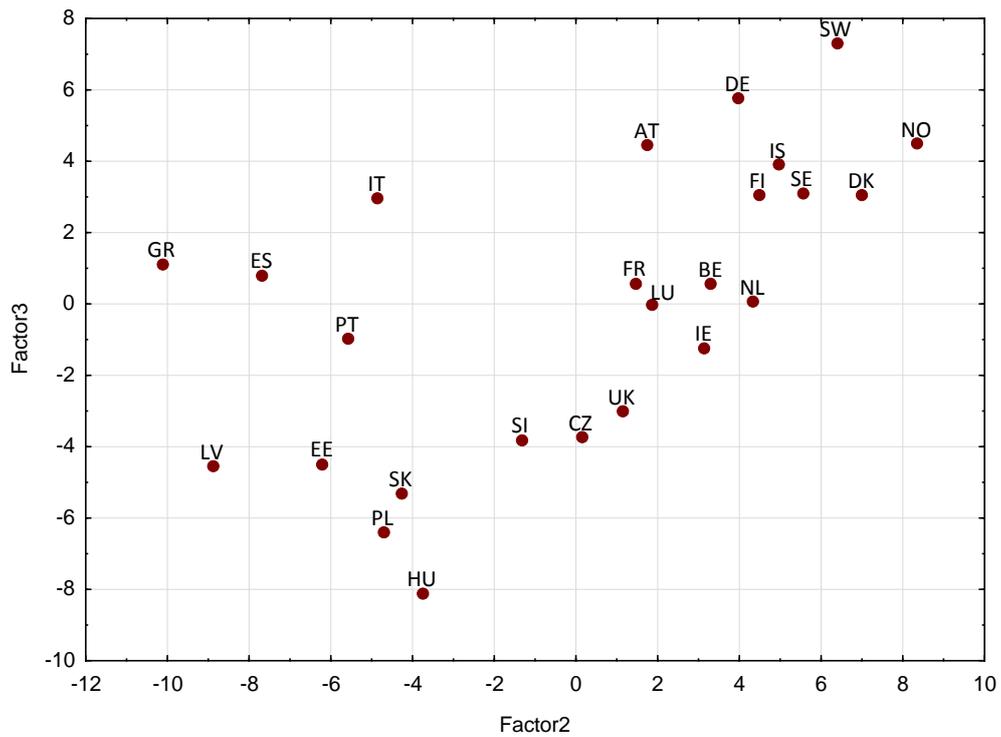*Source: self-processed in Excel*

Figure 5. Location of European countries in the coordinate system of the factors F2, F3.
Source: self-processed in Excel

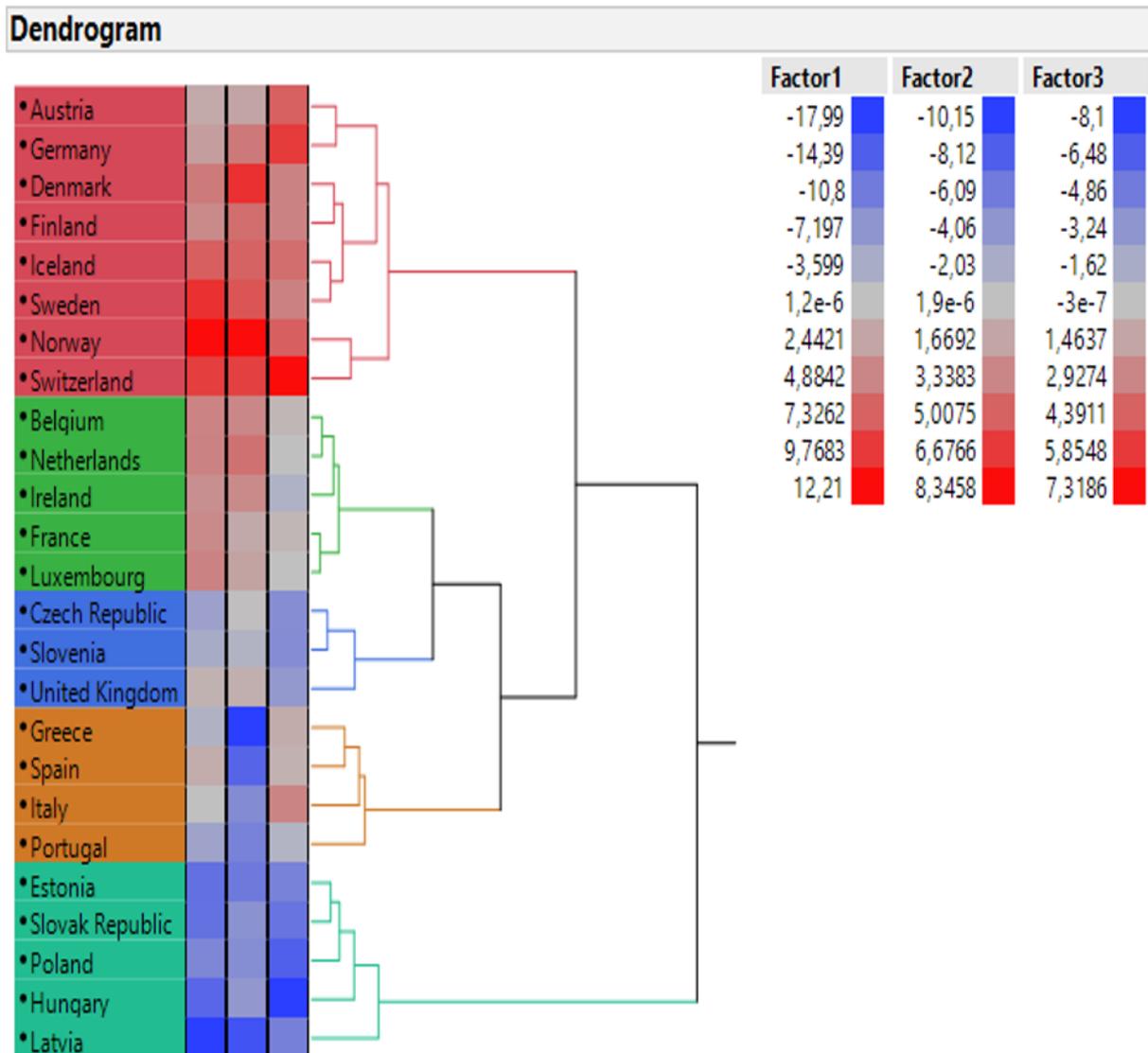* Corresponding author: viera.pacakova@upce.cz

Dendrogram

Figure 6. Dendrogram of the European countries clusters according to extracted factors.
Source: self-processed in SAS JMP

3rd factor of personal and technical resources of healthcare.

According to the dendrogram of Ward's method we have considered 5 different clusters. The first red cluster includes 8 countries with high values of all three factors. In the five countries of the green cluster there are the slightly lower values of the first two factors and the significantly lower value of the third factor. In the blue cluster of 3 countries there are again degraded values of all three factors compared with the previous two clusters. The brown cluster of four countries is characterized by poor social determinant of health. To the last cluster belong the countries with the lowest level according to all three factors.

Analogous interpretation of five clusters from the dendrogram in Figure 5 also provide parallel graphs in Figure 6.

Conclusion

The results of the selected multidimensional methods have confirmed the usefulness of their use to reduce the dimension of large-scale data sets of health indicators in Europe, assessing health inequalities and identifying some of its determinants.

The correlation analysis provided quantification of causal relationships of health indicators, health care expenditures, personnel and technical resources, and social determinants of health care. The application of factor analysis allowed to replace the 19 original indices with three common factors explaining almost 80% of variables of the original variables. Identifying these factors using factor loads has made it possible to assess the impact of social determinants and personnel and technical resources on health status as well as health inequalities in monitored countries caused by these factors. Charts 1 to 3 confirm that despite the efforts
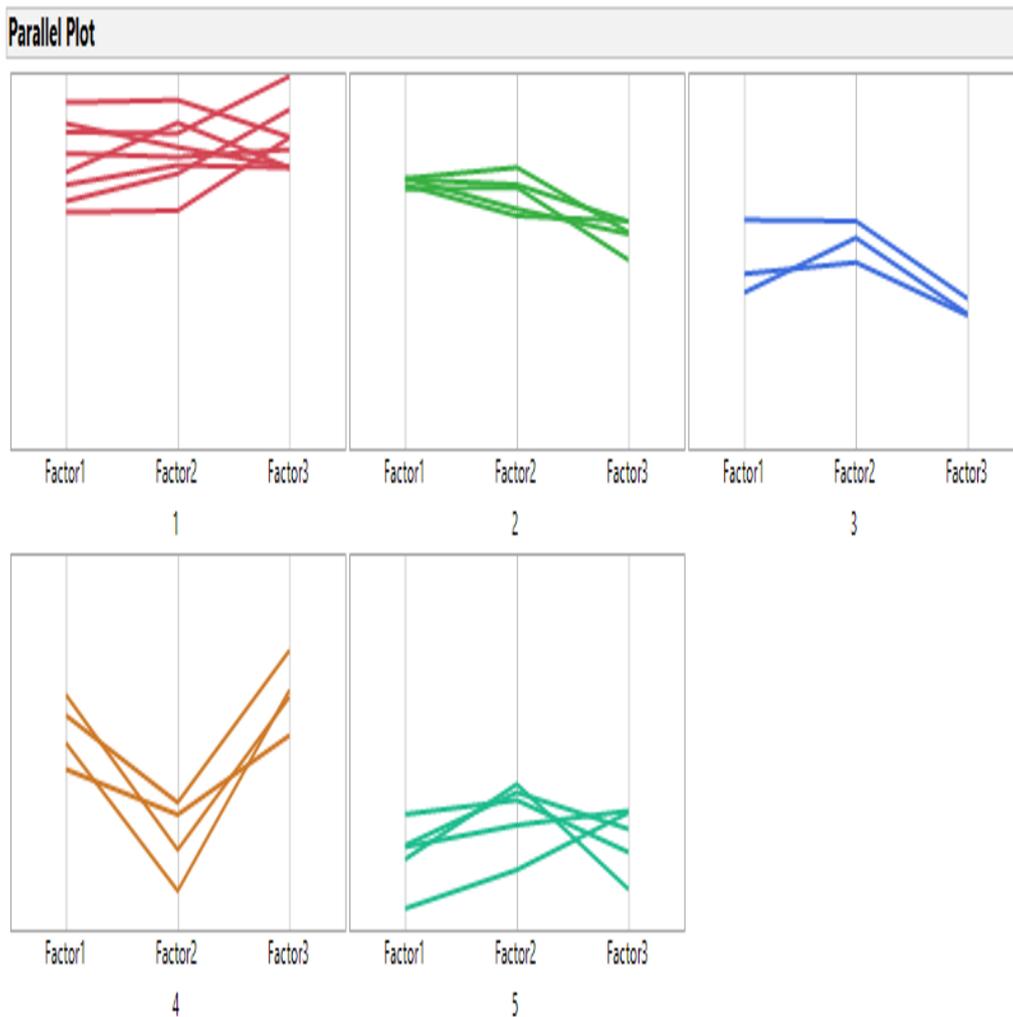
* Corresponding author: viera.pacakova@upce.cz

Figure 6. Parallel plots of the European countries clusters according to extracted factors
Source: self-processed in SAS JMP

and actions of the European Commission, these inequalities are significant. This also confirms the results of cluster analysis that are consistent with the results of factor analysis.

The use of appropriate statistical software packages and Excel spreadsheet allows the publication of the results of advanced statistical methods in clear visual form, understandable also to people without thorough knowledge of these methods.

Acknowledgement

References

1. European Commission. *Report on health inequalities in the European Union*. Brussels (2013). Retrieved from: http://ec.europa.eu/ health/health_policies/docs/improving_health_for_al l_eu_citizens_en.pdf
2. European Commission. *Improving health for all EU citizens. Luxembourg*. Publications Office of the European Union (2013). Retrieved from: https://ec.europa.eu/health//sites/health/files/health_ policies/docs/improving_health_for_all_eu_citizens _en.pdf
3. J. F. Hair, W. C. Black, B. J. Babin, R. E. Anderson, R. L. Tatham, *Multivariate Data Analysis*. Sixth Edition. New Jersey: Pearson Education, Inc., (2007).
4. OECD Health Statistics. *OECD*, Paris. (2018). Retrieved from: http://www.oecd.org/els/health-systems/health-data.htm
5. OECD. *Society at a Glance 2016: OECD Social Indicators*, OECD Publishing, Paris. (2016). Retrieved from: http://dx.doi.org/10.1787/9789264261488-en
6. V. Pacáková, P. Jindrová, D. Zapletal. Comparison of Health Care Results in Public health Systems of European Countries. *In: European Financial Systems 2016. Proceedings of the 13th International*

* Corresponding author: viera.pacakova@upce.cz

Scientific Conference, Brno: Masaryk University, pp. 534-541, (2016).

7. V. Pacáková,, L. Kopecká, Inequalities in Health Status Depending on Socio-economic Situation in the European Countries. *E+M Economics and Management, XXI*(2), pp. 4-20, (2018).

8. J. Křupka, R. Provazníková, M. Kašparová, J. Langer, Standards of Living Modeling based on Objective Indicators with Using Cluster Analysis. *In: International Journal of Mathematical Models and Methods in Applied Sciences,* 8(7), pp. 782-792, (2013).

9. E. Šoltés, T. Šoltésová. Comparison of Income Poverty and Social Exclusion in EU in 2008 and 2016. In : *The 12th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-economic Phenomena*, Poland: Zakopane (2018).