ARGUMENTA OECONOMICA
No 1 (48) 2022
ISSN 1233-5835; e-ISSN 2720-5088

*Justyna Brzezińska**

A STUDY ON THE OECD BETTER LIFE INDEX USING MULTIVARIATE STATISTICAL ANALYSIS

One of the more interesting indexes used for the measurement of the quality of life was launched in 2011, and having catered for more than 100,000 people from 180 countries, covers 11 topics dealing with living conditions and quality of life such as: housing, jobs, work-life balance, income, health, safety, education, community, civic engagement, environment and life satisfaction. The OECD Better Life Index, next to the most used GDP index, reveals inequalities between people from different socioeconomic spheres (groups), disparities in well-being and other very important economic phenomena. This paper presents statistical analysis based on Better Life Index data (2017). The author applied multivariate statistical analysis, namely correspondence analysis and classification methods for grouping countries into subgroups of countries with a similar level of well-being and life satisfaction. As a result, the study presents classification and visualization (dendrogram and perceptual map) of countries with a similar level of quality of life in terms of the importance of the 11 characteristics consisting of 24 variables that define well-being. The study shows that multivariate statistical methods can be used as an alternative for the classical measurement proposed by the OECD. All the calculations are conducted using the **R** software (www.r-project.org).*

Keywords: OECD Better Life Index, multivariate statistical analysis, cluster analysis, correspondence analysis

JEL Classification: C38, C39
DOI:10.15611/aoe.2022.1.10

Quote as: Brzezińska, J. (2022). A study on the OECD Better Life Index using multivariate statistical analysis. *Argumenta Oeconomica*, 1(48).

1. INTRODUCTION

In recognition of the limitations of Gross Domestic Product (GDP) as an indicator of social progress, Stiglitz, Sen and Fitoussi in 2009 called for better measurement through a richer mix of statistics as the basis for decisions about policy. In response, the Organisation for Economic Cooperation and Development (OECD) identified the measurement of well-being and societal progress as an area of top priority and

* Department of Economic and Financial Analysis, University of Economics in Katowice, Poland. ORCID: 0000-0002-1311-1020.

produced an extensive data base. The OECD Better Life index shows that there is more to life than the cold numbers of GDP and economic statistics, and it allows to compare well-being across countries, based on 11 topics consisting of 24 variables. The OECD has identified the essentials of this measurement of the areas of material living conditions and quality of life (housing, income, jobs, community, education, environment, civic engagement, health, life satisfaction, safety, work-life balance).

What makes for a good life? While the richness of human experience cannot be captured in numbers alone, it is important that the statistics shaping public policy reflect both material living conditions of people, and the quality of their lives. This includes how life is changing over time, how lives differ across different population groups, and whether today's well-being is achieved at the cost of depleting future resources. This fourth edition of the OECD "How's Life?" aims to meet this need, providing a picture of peoples' well-being in the OECD and partner countries. More than just providing a quantitative analysis of quality of life across the globe, the data in the Better Life Index also reveal stories about the different environments, the struggles and accomplishments of individuals and communities around the world. For instance, in Australia the OECD Better Life Index revealed that nearly 14% of employees work long hours, which can affect other areas of well-being such as health and relationships. This has led to the work-life balance indicated as the greatest concern in Australia, and therefore in order to increase quality of life, focus must be placed on policies that support a more stable work-life balance.

The OECD Better Life Index, created in May 2011 following a decade of work on this issue, is the first attempt to bring together internationally comparable measures of well-being in line with the recommendations of the Commission on the Measurement of Economic Performance and Social Progress (also known as the Stiglitz-Sen-Fitoussi Commission). The recommendations made by this Commission sought to address concerns that standard macroeconomic statistics like GDP have failed to give a true account of peoples' current and future well-being. The OECD Better Life Initiative includes two main elements: "Your Better Life Index" and "How's Life?".

Your Better Life Index (BLI), launched in May 2011, is an interactive tool that allows to compare countries' performances according to their own preferences in terms of what makes for a better life. It was designed by the Berlin-based agency Raureif, in collaboration with Moritz Stefaner. First published on 24 May 2011, it includes 11 dimensions presented on the OECD website. New indicators and dimensions are planned to be added to the Better Life Index in the future. For example, the Better Life Index was criticized for not showing inequalities in a society. Future editions of the index are planned to take inequalities into account by focusing on the well-being achievements of specific groups of the population.

The financial crisis that touched almost all countries around the world have had a deep and long-lasting impact on peoples` lives. Looking at changes in well-being since 2005, it is crucial to know whether there are countries that are similar or not to each other in terms of the group of variable describing the quality of life. This paper presents a statistical analysis of the OECD Better Life Index using multivariate statistical methods: cluster analysis and correspondence analysis using **R** software for computations (https://www.r-project.org/). The results were presented using graphical plots showing countries belonging to the cluster of a particular level of well-being and the Better Life Index, as well as using a perceptual map presenting the correspondence between countries.

2. CLUSTER ANALYSIS

Cluster analysis involves sorting data objects (or items) into exclusive groups based on their similarity. Grouping data is a very important process because it can reveal information about the data such as outliers, dimensionality, or previously latent or hidden relations that cannot be seen from the raw data as such. In cluster analysis there is often no prior specification about the number or nature of the groups to which the objects will be assigned. The grouping is done based solely on similarity measures, and the ideal number of groups is often determined within the clustering algorithm. These characteristics can make cluster analysis difficult, and a wide variety of algorithms have been proposed to produce the best clustering of objects based on a set of observed data. An important class of clustering methods is hierarchical cluster analyses. There are two main types of hierarchical clustering methods, agglomerative and divisive. This paper presents and applies only agglomerative methods. An agglomerative hierarchical method begins with each object as its own cluster. It then successively merges the most similar clusters together until the entire set of data becomes one group. In order to determine which groups should be merged in agglomerative hierarchical clustering, various linkage methods can be used.

Single linkage (Sneath 1957) merges groups based on the minimum distance between two objects in two groups. The distance between clusters R and Q is defined as:

$$d_S(R,Q) = \min_{i\in R,\, j\in Q} d(i,j), \tag{1}$$

where $d(i, j)$ is the distance between the i-th cluster and j-th object.

Complete linkage (McQuitty 1960; Sokal and Sneath 1963) merges groups based on the maximum distance between two objects in two groups which means that the distance between clusters R and Q is defined as:

$$d_C(R,Q) = \max_{i\in R,\, j\in Q} d(i,j). \tag{2}$$

Average linkage (Sokal and Michener 1958) merges groups based on the average distance of all the objects in one group to all the objects in the other. The distance in average linkage is defined as:

$$d_A\left(R,Q\right)=\frac{1}{\left|R\right|\left|Q\right|}\sum_{i\in R,\, j\in Q} d\left(i,j\right), \tag{3}$$

where: $|R|$ and |Q| are the numbers of objects in cluster R and Q, respectively.

Another important hierarchical clustering method is the Ward method (Ward 1963), similar to the linkage methods in that it begins with N clusters, each containing one object, however it differs in that it does not use cluster distances to group objects. Instead, the total within-cluster sum of squares (SSE) is computed to determine the next two groups merged at each step of the algorithm. The error sum of squares (SSE) is defined as:

$$SEE=\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(\mathbf{y}_{ij}-\overline{\mathbf{y}}_i\right)^2, \tag{4}$$

where $\mathbf{y}_{ij}$ is the *j*-th object in the *i*-th cluster and n_i is the number of objects in the *i*-th cluster.

There are also plenty of other methods that allow for the partitioning of multivariate datasets into separate classes, such as the well-known *k*-means (MacQueen 1967) and *k*-medoids (Kaufman and Rousseeuw 1987) algorithm, as well as model-based clustering methods (Banfield and Raftery 1993). However, based on a survey of actual cluster analyses in the scientific literature, Kettenring (2006) indicated that hierarchical clustering was by far the most widely used form of clustering in practice. The great advantage of the presented clustering methods is that is a graphical representation using a dendrogram. In this paper, the author compares the performances of only hierarchical methods for different approaches using **R** software and `hclust` function from the `stats` package.

3. CORRESPONDENCE ANALYSIS

Correspondence analysis is a multivariate statistical method designed for categorical variables in a contingency table. Correspondence analysis transforms a data table into two sets of new variables called factor scores (obtained as linear combinations of, respectively, the rows and columns): one set for the rows and one set for the columns. These factor scores give the best representation of the similarity structure of, respectively, the rows and the columns of the table. In addition, the factor scores can be plotted as maps that optimally display the information in the original table. In these maps, the rows and columns are shown as points whose coordinates are the factor scores and where the dimensions are also called factors, components (by analogy with PCA), or simply dimensions. Interestingly, the factor scores of the rows and the columns have the same variance and, therefore, the rows and columns can be conveniently represented in one single map.

The correspondence analysis is based on the correspondence matrix, defined as the matrix of elements of **N** divided by the grand total of **N**: $\mathbf{P} = \left[\frac{n_{hj}}{n}\right]$. The aim of the correspondence analysis is the geometrical display of two or more categorical variables by showing the categories of the variables as points in a low-dimensional space. The vector of the row and column sums of **P** are denoted by **r** and **c** defined as:

$$\mathbf{r} = \left[\frac{n_{h.}}{n}\right] = \left[p_{h.}\right], \tag{5}$$

$$\mathbf{c} = \left[\frac{n_{.j}}{n}\right] = \left[p_{.j}\right]. \tag{6}$$

The row and column profiles of **P** are defined as the vector of the rows and columns of **P** divided by their respective sums. Row profiles are defined as:

$$\mathbf{R} = \mathbf{D}_{\mathbf{r}}^{-1}\mathbf{P} = \left[\frac{n_{hj}}{n_{h.}}\right] = \left[\frac{p_{hj}}{p_{h.}}\right], \tag{7}$$

and column profiles are defined as:

$$\mathbf{C} = \mathbf{D}_{\mathbf{c}}^{-1}\mathbf{P}^{\mathrm{T}} = \left[\frac{n_{hj}}{n_{.j}}\right] = \left[\frac{p_{hj}}{p_{.j}}\right]. \tag{8}$$

Both the row profiles and column profiles are written in the rows of **R** and **C**, respectively. The row profiles and column profiles are very important elements of correspondence analysis, as they impact on the principal axes. Each row and column profile can be presented as a point in a multidimensional space. Thus a profile will tend to lie closer to vertices for which it has higher values. Each row and column profile has a unique weight associated with it, called mass, which is proportional to the row and column sum in the cross-tabulation. The average row and column profile is then the centroid of the row and column profiles, where each profile is weighted by its mass in the averaging process. The problem of how to graphically present co-occurrence of categories of variables becomes more serious when categorical variables are characterized by a large number of categories. For this purpose, the decomposition of matrix **A** by singular value decomposition should be applied, thanks to which it is possible to determine the coordinates of the categories of variables of interest and to determine the degree of dispersion.

This paper applied correspondence analysis for a graphical presentation of categories that define the Better Life Index by the OECD. The analysis allows to see which categories describing the Better Life Index correspond with countries of the OECD. The author applied the `ca` package using the `ca` function.

4. APPLICATION AND DATA ANALYSIS USING R

In this section of the paper we focus on empirical computations based on the data on well-being. Such dataset is available at the OECD website on Better Life Index (Edition 2017). It consists of two-dimensional matrix of 11 characteristics of well-being and the OECD countries. Well-being as a phenomena is defined via eleven characteristics and a set of the following variables:

1. Housing: housing conditions and spending (X1: dwellings without basic facilities, X2: housing expenditure, X3: rooms per person),
2. Income: household income and financial wealth (X4: household net adjusted disposable income, X5: household net financial wealth),
3. Jobs: earnings, job security and unemployment (X6: labour market insecurity, X7: employment rate, X8: long-term unemployment rate, X9: personal earnings),
4. Community: quality of social support network (X10: quality of support network),
5. Education: education and what one gets out of it (X11: educational attainment, X12: students skills, X13: years in education),
6. Environment: quality of environment (X14: air pollution, X15: water quality),
7. Governance: involvement in democracy (X16: stakeholder, X17: voter turnout),
8. Health (X18: life expectancy, X19: self-reported health),
9. Life Satisfaction: X20: level of happiness,
10. Safety: murder and assault rates (X21: feeling safe walking alone at night, X22: homicide rate),
11. Work-life balance (X23: employed working very long hours, X24: time devoted to leisure and personal care).

The dataset contains information on the OECD countries including Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States.

For cluster analysis the study applied the stats package using `hclust` function. We present the comparative analysis of three different clustering methods and we will provide graphical presentation of results. First, the average linkage method was employed. A dendrogram using this linkage method is presented in Figure 1.

Secondly, the complete linkage method was applied (see Figure 2). Using the Ward method, the following dendrogram was obtained (see Figure 3).

After the comparison of these three methods, one can conclude that there are three clusters containing the same objects, no matter which clustering method was applied. The first cluster consists of two countries: Switzerland and the United States; the second cluster consists of 16 countries: Mexico, Turkey, Norway, Finland, Korea, Spain, Portugal, Chile, the Czech Republic, Hungary, Slovenia, Latvia, Estonia,

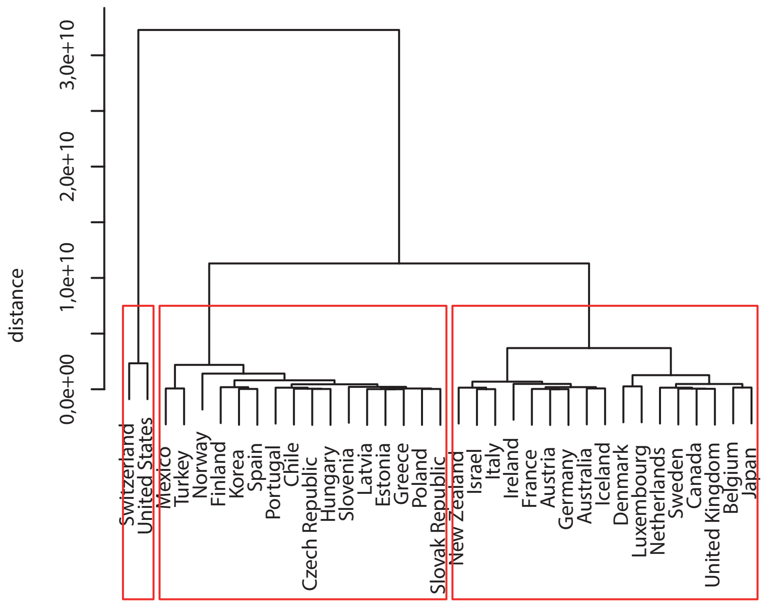

Fig. 1. Dendrogram using the average linkage method

Source: own calculations.

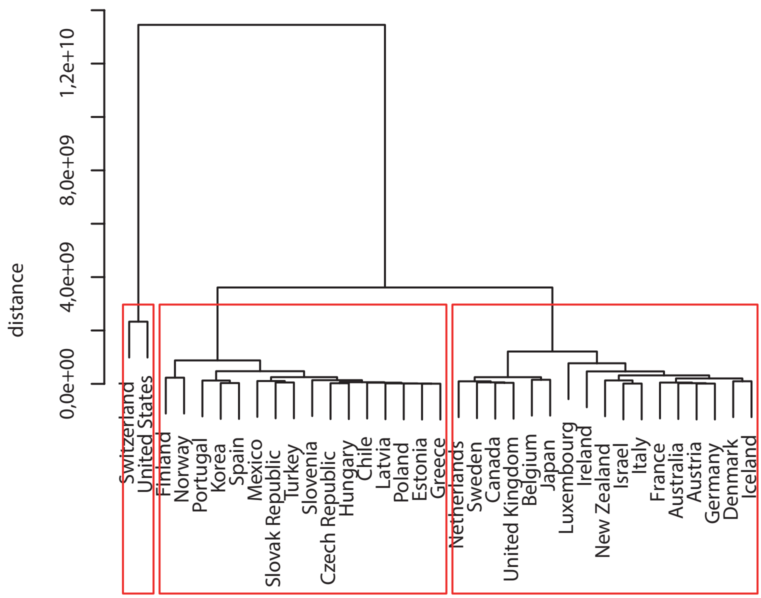

Fig. 2. Dendrogram using the complete linkage method

Source: own calculations.

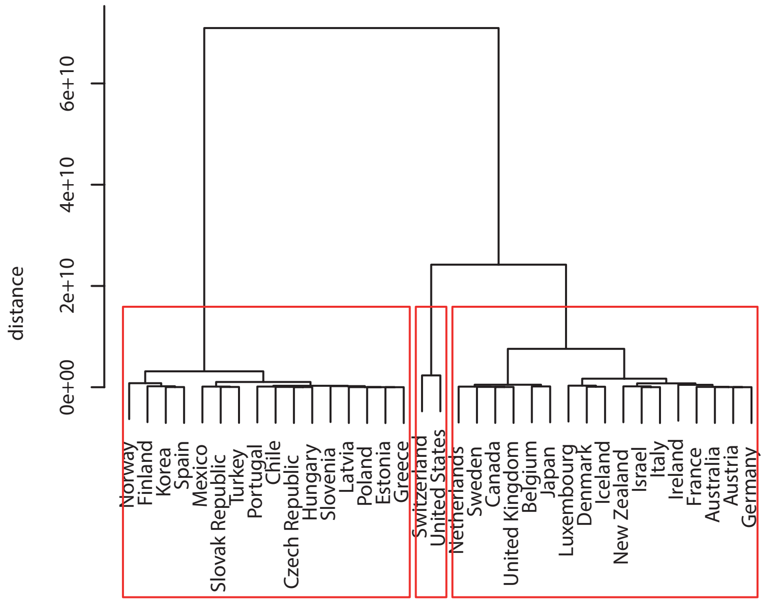

Fig. 3. Dendrogram using the Ward linkage method

Source: own calculations.

Greece, Poland, and the Slovak Republic. The third cluster consists of 17 countries: New Zealand, Israel, Italy, Ireland, France, Austria, Germany, Australia, Iceland, Denmark, Luxembourg, the Netherlands, Sweden, Canada, the United Kingdom, Belgium, and Japan. The only difference is the distance where those objects are linked together with the nearest one.

One can also apply correspondence analysis with a graphical presentation of a perceptual map for all the objects (countries) in the analysis. Correspondence analysis is available in the `ca` package using the `ca` function. After the analysis of the perceptual map using correspondence analysis, one can observe that there were similar clusters of countries that were obtained using cluster analysis. One can see that the United States and Switzerland lie close to each other. Japan, Sweden, the United Kingdom, Belgium, Canada, the Netherlands, Italy, Israel, New Zealand, France, Denmark, Austria, Germany, Australia, Iceland, and Luxembourg are close to one another, creating a cluster of countries that are similar in terms of the Better Life Index. The rest of the countries (16 objects) are in the third cluster, i.e. Norway, Finland, Slovenia, Spain, Korea, the Slovak Republic, Poland, Turkey, Chile, Greece, Estonia, the Czech Republic, Latvia, Portugal, Hungary, and Mexico.

The conducted statistical analysis shows that there are OECD countries that demonstrate a similar level of well-being in terms of the Better Life Index. This means that in countries belonging to one cluster the level of the analysed 11 variables describing the index is very similar. Indeed, one can see there are three groups of countries that show similarities, no matter what analysis was used. This study presented an alternative approach to the raw data as presented by the OECD. The statistical methods that were applied, such as cluster analysis and correspondence analysis, prove that there are similarities that lead to groups of countries belonging to a particular cluster. The author presented statistical analysis using **R** software as this is one of the most powerful and available tools.

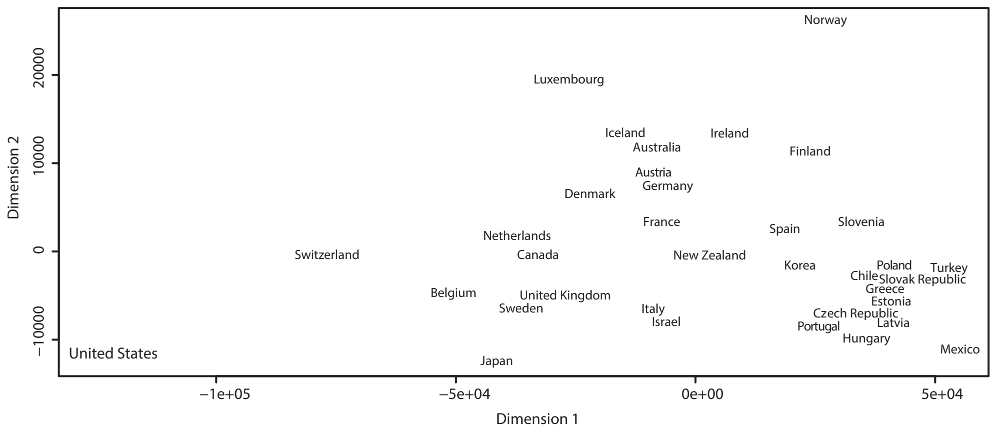

Fig. 4. Perceptual map using correspondence analysis

Source: own calculations.

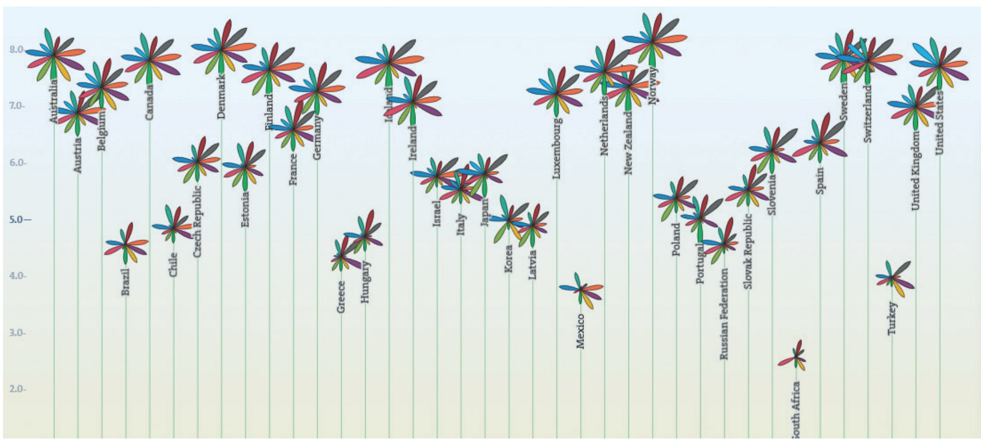

Fig. 5. Graphical presentation of the OECD Better Life Index

Source: OECD (2017).

Looking at the plot provided by the OECD, one can observe that it presents all the countries alphabetically in the form of colourful flower petals corresponding to the 11 characteristics of well-being. The difference in the length of petals represents a higher level of the importance; on the left side there is an axis showing the scale.

Comparing these flowers and their position with the results and graphical plots obtained in multivariate statistical methods allowed to observe certain similarities. The level of the importance for some countries (Figure 5) indicates the cluster the countries belong to in Figures 1 to 3. It is also comparable to the results in the correspondence analysis obtained for all the countries analysed as points in a 2-dimensional plot.

As a result, it can be said that multivariate statistical methods allowed to divide groups of countries into clusters with a similar level of well-being, and at the same time representing a similar level of the importance of the 11 characteristics defining well-being by the OECD.

CONCLUSION

In this paper the author presented the Better Life Index-Edition 2017 prepared by the OECD. This allowed to compare well-being across countries, based on the 11 topics the OECD identified as essential in the areas of material living conditions and the quality of life. The author applied statistical analysis and visualization tools for selecting groups of the OECD countries with a similar quality of life, employing hierarchical analysis with the use of an agglomerative hierarchical method with R software. It was found that using multivariate statistical methods such as cluster analysis and correspondence analysis leads to obtaining a graphical representation of clusters containing countries with a similar level of well-being and the Better Life Index. Several methods were applied, always leading to three clusters. The first cluster, i.e. Switzerland and the United States, the second consists of 16 countries (Mexico, Turkey, Norway, Finland, Korea, Spain, Portugal, Chile, the Czech Republic, Hungary, Slovenia, Latvia, Estonia, Greece, Poland, and the Slovak Republic), while the third cluster contains 17 countries (New Zealand, Israel, Italy, Ireland, France, Austria, Germany, Australia, Iceland, Denmark, Luxembourg, the Netherlands, Sweden, Canada, the United Kingdom, Belgium, and Japan).

As a conclusion one can say that multivariate statistical methods, in particular correspondence analysis and cluster analysis, allowed to group countries into clusters with a similar level of well-being, while representing a similar level of importance of the 11 characteristics (consisting of a set of 24 variables) defining well-being according to the OECD. The plot provided by the OECD in the form of colourful flowers perfectly corresponds to the dendrogram and the perceptual map obtained as a result of the cluster analysis and the correspondence analysis, respectively. This means that multivariate statistical methods can be used as alternative methods for measuring and ranking countries according to their Better Life Index. The R software was used for all the calculations in the paper.

REFERENCES

Banfield, J. D., Raftery, E. A., *Model-based Gaussian and non-Gaussian clustering*, Biometrics, 49, pp. 803-821, 1993.

Markos, A., Menexes, G., Papadimitriou, I., *The CHIC analysis software v1.0*, Classification as a Tool for Research, pp. 409-416, 2010.

Jérôme, K., Antoine, R.,*OECD's 'Better Life Index': Can any country be well ranked?*, Journal of Applied Statistics, 39 (10), pp. 2223-2230, 2012.

Kaufman, L., Rousseeuw, P. J., *Clustering by means of medoids* [in:] Dodge, Y. (ed.) *Statistical Data Analysis Based on the L1-norm and Related Methods*, pp. 405-416. Elsevier/North Holland, New York 1987.

Kettenring, J. R., *The practice of cluster analysis*, Journal of Classification, 23, pp. 3-30, 2006.

MacQueen, J. B., *Some methods for classification and analysis of multivariate observations*, Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, pp. 281-297, 1967.

McQuitty, L. L., *Hierarchical linkage analysis for the isolation of types*, Educational and Psychological Measurement, 20, pp. 55-67, 1960.

R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, http://cran.r-project.org/, 2022.

Sneath, P. H. A., *The application of computers to taxonomy*, Journal of General Microbiology, 17, pp. 201-226, 1957.

Sokal, R. R., Sneath, A. H. P., *Principles of Numerical Taxonomy*. Freeman, San Francisco 1963.

Ward, J. H., Jr., *Hierarchical grouping to optimize an objective function*, Journal of the American Statistical Association, 58, pp. 236-244, 1963.

Received: March 2020, revised: October 2021