IJCEAS

International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XI, Issue: 2, Year: 2021, pp. 454-468

Analysis of Covid-19 Pandemic Data According to the Countries Including in the Human Development Index by Using Clustering Analysis K-Average Method*

Zafer KANBEROĞLU[1]
Çetin GÖRÜR[2]
İbrahim YILDIRIMÇAKAR[3]
Mustafa TÜRKMENOĞLU[4]

Received: 11.05.2021, Accepted: 30.12.2021
DOI Number: 10.5281/zenodo. 5831774

Abstract

One of the important disasters of the 21st century is the Covid-19 pandemic that started in China in December 2019 and spread all over the world in a short time. The Covid-19 pandemic, which started as a health problem, continues to deeply affect all structures of the countries with its economic and social dimensions. Although there is a parallelism between the development level of the country and its success in combating the pandemic, differentiation is observed from time to time. Within the scope of this study, Covid-19 pandemic data will be obtained by country and will be analyzed using the Cluster analysis K-Average Method. The sample of the study will be forty countries in total, including ten countries in eachcategory, in different developmentcategories in the United Nations human development index. In the first stage of the cluster analysis to be used in the study, the variables of total cases, daily cases, total deaths and daily deaths belonging to countries; In these cond phase, total cases per million, daily cases per million, total deaths per million and deaths per million will be used. It will be analyzed whether the variables used in the said stages make a meaningful clustering. In this context, it will be revealed whether countries are in the same cluster according to their level of development in the fight against the Covid-19 pandemic.

---

*This study is the revised version of the oral presentation presented at ICE-TEA 2021.

[1]Prof. Dr. Zafer KANBEROGLU Department of Economics, Faculty of Economics and Administrative Sciences, Van Yuzuncu Yil University, Van, Turkey e-mail: zkanberoglu@yyu.edu.tr, orcid: 0000-0002-4440-4133

[2] PhD student Çetin GÖRÜR Institute of Social Sciences, Inonu University, Van, Turkey e-mail: gorurcetin@hotmail.com, orcid: 0000-0002-9556-5068

[3] PhD student İbrahim YILDIRIMÇAKAR Institute of Social Sciences, Van Yuzuncu Yil University, Van, Turkey, e-mail: ibrahimvanli.2156@gmail.com, orcid: 0000-0002-1933-9798

[4] Research assistant Mustafa TÜRKMENOĞLU, Department of Economics, Faculty of Economics and Administrative Sciences, Van Yuzuncu Yil University, Van, Turkey e-mail: mustafaturkmenoglu@yyu.edu.tr, orcid: 0000-0001-8556-6959

Key words: K-Means Cluster method, Covid-19, ANOVA.

JEL Code: G22, C12, I18

1.Introduction

Humankind beings have been exposed to many economic, social and environmental disasters since the day they existed. One of the most important disasters of the recent period is the pandemic disease known as Covid-19. The epidemic has significantly affected all countries of the world since December 2019. In the pandemic, which is accepted to have started in China and affected the whole world in a short time, around 164 million cases and around 3.4 million deaths have been recorded so far.

With the Covid-19 pandemic, the health systems of countries have also been questioned. It has also been observed that countries, which are highly trusted with all their technological infrastructure, have not been able to show the same trust in the health sector. There are many examples of this in the world. In this context, the examination of similar aspects and differences of the country with other countries in the health sector according to a certain classification is important for the health of the societies.

The health sectors and social health of countries are important both for their own societies and for other societies in the world. This situation, which was seen in the previous pandemic diseases of the world (Black Death/Plague in Europe between 1347-1351, Cholera in India between 1817-1824, Flu in Spain between 1918-1920, Aids in the USA in 1981) once again revealed how vital public health is with the covid 19 outbreak.

The similarities and classification of pandemic diseases, which are closely related to public health, on the basis of countries, are considered important in terms of countries knowing their situation in combating the pandemic and seeking ways to get rid of this situation. In this study, the indicators and values of the pandemic in forty countries in four categories in the human development index were analyzed using the Clustering analysis method.

Events in nature are complex situations that arise as a result of the individual or combined interactions of many interrelated variables or factors. In this complex structure, it may not be easy to determine the variables that affect the relevant event, to solve their relations with each other and to explain their formation mechanisms in a simple and comprehensible way (İşleyen, 2021). This is because, at a certain time and with a certain number of data, it is possible to present the event within this structure in an accurate and reliable way only by using correct statistical methods (Demir et al., 2021).

As a result of the literature review, we think that the study will contribute to the literature. Since Covid is a virus that affects the world, it is a subject with a wide field of study. The aim of this study is to analyze whether the measures taken by countries to protect against this virus have similar results with this study. In this respect, the study differs in the literature.

2.Literature Review

Theoretical and empirical studies on the examination of Covid-19 pandemic data on the basis of countries are available in the literature. Demircioğlu and Eşiyok (2020), one of these studies, analyzed the Covid-19 outbreak data in line with the health indicators of the countries. The K-means method was used in the samples of 36 OECD and EU countries. Countries clustered according to their similarities in terms of health indicators and The locations of 36 countries have been evaluated relative to each other.

Tekin (2020) analyzed the effects of the Covid-19 pandemic on countries using the hierarchical clustering and Ward's method of cluster analysis. In this study on OECD countries, the pandemic significantly affected every country from every continent on the other hand, it has been observed that some countries pursue a more successful process by preventing the spread of the pandemic of the effect of the reflex they exhibited and at the same time preventing the increase in the number of deaths with effective treatment methods and the structure of health systems. And as a result of this situation, it has been determined that the impact of the pandemic on some economic and financial indicators of the countries is limited.

Sığırlı et al. (2006) analyzed the health level criteria in the samples of 25 countries for the period 1998-2004 with a multidimensional sampling analysis. It is aimed to reveal the similarities and differences between the positions of the countries in terms of health indicators. It has been determined that Turkey differs from other countries other than Slovakia, Hungary and the Czech Republic in terms of health indicators in the first dimension, health expenditures and the share of national income from national income in the second dimension.

Ersöz (2009) examined the similarities of countries with the help of clustering and discriminat analysis in the comparison of selected health indicators in the sample of OECD member countries for 2004. As a result of the cluster analysis, it was determined that Turkey is in the same cluster with Poland, Slovakia, Czech Republic, Hungary, Mexico, Republic of Korea in the stepwise clustering method among OECD countries. In the k-means method, which is the non-progressive clustering method, it has been determined that Turkey is in the same cluster, Portugal, Poland, Slovakia, Hungary, Czech Republic, Mexico and Republic of Korea. In the Medoid clustering method, it was determined that Turkey is in the same cluster with Mexico.

Alptekin and Yeşilaydın (2015) analyzed the health indicators of 34 OECD member countries using fuzzy cluster analysis method. The most suitable number of clusters is five after the analysis; three in the first cluster, nine in the second division, nine in the third division, six in the fourth cluster and it has been determined that there are seven countries in the fifth cluster. In addition, it has been determined that Turkey is in the same cluster with Estonia, Hungary, Mexico, Poland and Chile.

Mut and Akyürek (2017) examined OECD countries with a cluster analysis according to health indicators. In the study, the indicators effective in the clustering of countries were examined and the differences between clusters were revealed. According to the results of the analysis, it has been determined that Turkey, Mexico and Chile are in the same cluster.

Kartal et al (2020), In their study, in which they examined the data of the Covid-19 pandemic disease in the world and in Turkey using the Cluster analysis method, the changing situation of COVID-19 on a global and national scale was evaluated. According to the results of the research, the risk situations and possible similarities of the countries in the same cluster have been determined more carefully, regardless of their geographical location and in which group they are according to time. Thus, it was emphasized that the necessary policy recommendations should be determined quickly and intervened according to the Covid-19 disease states of the countries.

Küçükefe (2020) health economy exchange due to Covid 19 was investigated by comparing GDP declines and deaths per million people in the case of OECD and China. Empirical data revealed that countries with the highest death rates experienced the greatest economic downturns. Clustering analysis found that countries are divided into three parts in terms of current account balance, GDP growth, and the number of deaths per million people.

Neuburger and Egger (2020) examined the relationship between perception of COVID-19, travel risk perception and travel behaviour among travellers in the DACH region (Germany, Austria, Switzerland). Cluster analysis was performed and defined three unique clusters in both periods with distinctive characteristics. The results revealed a significant increase in risk perception of COVID-19, travel risk perception and travel behaviour over a short period of time.

Verelst et al (2020), in their study using Cluster analysis on the impact of the Covid-19 pandemic disease on the European health system, they found that the Covid-19 disease puts a significant pressure on the European health system and this situation was associated with country-specific Covid-19 deaths, active covid-19 cases and health system capacity. According to the results of the said study, they stated that Covid-19 will soon exceed the existing health capacity of European countries and health institutions may be insufficient in this situation. In the study, it was also emphasized that Italy, Spain, France and the Netherlands were the countries where the pressure of the Covid-19 pandemic disease was experienced the most.

3. Methodology

In the study, the corona virus case and death data of the countries were analyzed in the SPSS 22 program using Cluster Analysis. Non-Hierarchical Clustering Method and K-Means Technique were used in the analysis. Covid-19 data of the countries were obtained from the official website of Github.com for the analysis, the number of deaths and cases on 01 May 2020 - 05 May 2021 (370 days) and the total number of cases and deaths per million were discussed. The analyzed countries, according to the 2020 Human Development Index report; Very High Human Development, High Human Development, Medium Human Development and Low Human Development categories. Ten countries selected from each category were included in the analysis. The variables belonging to these countries are; Total Cases are Daily Cases, Total Deaths, Daily Deaths, Total Cases Per Millions, Cases Per Millions, Total Deaths Per Millions, and Daily Deaths Per Millions. Variables were considered as two different groups and analyzed in two stages. First stage; while the variables of Total Cases, Daily Cases, Total Deaths and Daily Deaths are included, the second stage includes the variables Total Cases per Million, Cases per Million Daily Cases, Total Deaths per Millions and Deaths per Millions. The methods used in the study are detailed as headings below.

Cluster Analysis

Although classical methods used to statistically evaluate a large number of data obtained as a result of the analysis provide important information for each variable, they are insufficient to provide real information about the existence of a relationship between two or more different characteristics and do not allow the grouping of samples with homogeneous structure (Demir et al., 2016). Cluster analysis, which is one of the multivariate statistical techniques, is used to classify data with unknown number of groups and ungrouped data according to their similarities. Cluster analysis is a technique that allows data to be collected in discrete clusters in terms of their similarity to each other according to units or variables. Cluster analysis is similar to discriminant analysis in that it aims to collect similar individuals in the same groups, and with factor analysis because it aims to collect similar variables in the same groups, and it has data reduction features (Çakmak, 1999).

The assumption of normality of the data, which is important in other multivariate statistical analyzes, is not very important in cluster analysis, and the normality of distance values is considered sufficient (Tatlıdil, 2002). Clustering is done by looking at the similarity (closeness) or distance measure of two

International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XI, Issue: 2, Year: 2021, pp. 454-468

observations or two variables according to the purpose determined as explained above.

The main assumptions of the cluster analysis are that the data matrices do not divide the predictive and criterion variables into sub-matrices before the analysis and the data is partially homogeneous and partially heterogeneous (Atamer, 1992).

Similarity and Distance Measures Used in Clustering Analysis

In clustering analysis, various distance/similarity measures are used to calculate the distances between each other in dividing n units into clusters according to p variables. Distance or similarity measures vary according to the units of measure of the variables in the data matrix. If the variables are obtained on a proportional or intermittent scale, distance or correlation type measures are used, if the values obtained by counting are the chi-square distance measure or Phi square distance measure, if the values obtained according to binary observations, similarity or difference measures such as Euclidean, square Euclidean are used (Özdamar, 2004). :283). In clustering analysis, while the units are grouped, their proximity to each other is calculated according to the distance criteria. The distances of n units in the data matrix with respect to p number of variables are expressed with the D matrix.

Table 1. Distance Measures Used in Clustering

Euclidean Distance	$d(x_i - y_i) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$
Squared Euclidean Distance	$d(x_i - y_i) = \sum_{i=1}^{n}(x_i - y_i)^2$
Cıty-Block Distance	$d(x_i - y_i) = \sum_{i=1}^{n} \lvert x_i - y_i \rvert$
Chebychey Distance	$d(x_i - y_i) Max_i \lvert x_i - y_i \rvert = (\sum_{i=1}^{n} \lvert x_i - y_i{}^k \rvert)^{1/k}$
Minkowski Distance	$d(x_i - y_i) = (\sum_{i=1}^{n} \lvert x_i - y_i{}^k \rvert)^{1/k}, k \geq 1$
Mahalanobis Distance	$d(x_i - y_i) = (x_i - y_i)' S^{-1} (x_i - y_i)$
Canberra Distance	$d(x_i - y_i) = \sum_{i=1}^{m} \lvert x_i - y_i \rvert / \lvert x_i - y_i \rvert$

Hotelling T^2 Distance	$d(x_i - x_j) = T^2 = \frac{n_1 n_2}{n}(\bar{x}_i - \bar{x}_j)' S^{-1}(\bar{x}_i - \bar{x}_j)$
Biserial Correlation Measure	$I_b = \frac{\bar{x}_p - \bar{x}_q}{S_t}\frac{p.q}{y}$
Pearson Correlation Measure	$r = \frac{\sum x_i y_i - \frac{(\sum x_i \sum y_i)}{n}}{\sqrt{(\sum x_i^2 - \frac{(\sum x_i)^2}{n})}\sqrt{(\sum y_i^2 - \frac{(\sum y_i)^2}{n})}}$
Spearman Rank Correlation Coefficient	$p = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$

Cluster Analysis Methods

Clustering techniques; By using the distance matrix, it allows to create homogeneous and heterogeneous groups between objects or variables. Many algorithms have been proposed for cluster analysis. However, in the literature, these algorithms are grouped under two headings: Hierarchical clustering techniques, Non-hierarchical clustering techniques (Ketchen and Shook, 1996: 444). The common goal of both techniques is to maximize the differences between clusters and the similarities within clusters. That is, while the homogeneity within the cluster is increased, the homogeneity between the clusters is decreased. Although which technique to use depends on the number of clusters, it is much more useful to use both techniques together. Thus, it is possible to compare both the results and which of the two techniques gives more appropriate results (Akın, 2008:8).

Non-Hierarchical Clustering Method

Non-hierarchical clustering techniques are designed to collect units rather than variables in K clusters. The number of clusters (K) can be given as a specific value or determined as a part (part) of the clustering technique. Because the distance (similarity) matrix does not have to be determined, and the underlying data does not have to be stored throughout the computer's operation. Non-hierarchical techniques can be applied to larger datasets than hierarchical techniques (Johnson and Wichern, 1988). Non-hierarchical techniques start either from a fraction of individuals in groups or from a set of source points that form the core of clusters. Two of the most used non-hierarchical clustering techniques are the K-means technique and the maximum likelihood technique. In our study, information about the K-means technique is given.

K-Averages Method

International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XI, Issue: 2, Year: 2021, pp. 454-468

Although the term k-means was first defined and used by Macquenn in 1967, the idea of applying this method dates back to 1956. Over time, it has become one of the most popular cluster analysis algorithms. It has become one of the first methods that comes to mind in the concept of non-hierarchical clustering analysis and has taken its place in almost all clustering analysis package software.

Euclidean distance; according to the calculation as $\left\|x_{ij}-v_j\right\|^2, \quad 1 \le j \le k$, individuals are classified to the closest cluster. Here x_{ij} is the *j*th object in the *i*th cluster; v_j refers to the center of the *j*th cluster. The distance measure is calculated as follows.

$$j(x;v)=\sum_{j=1}^{k}\sum_{x\in c_j}\left\|x_{ij}-v_j\right\|^2$$

The purpose of the k-means method is to maximize in-group homogeneity and heterogeneity between groups, as in other clustering analysis methods (Bulut, 2018). It works by dividing the data into a user-specified number of clusters and then iteratively reassigns the observations to the clusters until some numerical criteria are met. The k-means criterion sets a goal of minimizing the observation distance in clusters and maximizing the distance between clusters. The k-means method is so widely used that it is even used by some researchers to refer only to non-hierarchical cluster analysis (Hair et al., 2014).

4. Findings

In the study, analyzes were performed using the K-Averages technique in the Non-Hierarchical Clustering Method used in Clustering analysis. Analyzes were obtained using three clusters and 10 iterations. Below, the distance between the cluster centers, the clusters in which the countries are located, the ANOVA test to analyze whether the variables show significant clustering, the averages of the variables in the clusters and the numerical values in the clusters are listed as tables.

Table 2 shows how the variables are distributed in the three clusters considered for both stages.

Table 2. Clustering Centers

Cluster		1	3	2
1.Stage	Total Cases	167677,53	1360938,85	1231973,82
	Daily Cases	17996,58	14222,19	37654,51
	Total Deaths	72161,59	41096,35	87511,06
	Daily Deaths	463,53	275,36	621,04
2: Stage	Total Cases Per Mil.	2260,69	17238,39	1964,11
	Daily Cases Per Mil.	319,12	435,94	299,52
	Total Deaths Per Mil.	190,61	421,95	131,26
	Daily Deaths Per Mil.	2,63	3,79	2,34

Table 2 shows the density that occurs in the clusters. When the table is examined for the first stage, the cluster order in the total case is 2, 1, and 3. The cluster ranking in the daily case, total death and daily deaths is 3, 1 and 2.For the second stage, the density order in the clusters differs from the first stage due to the fact that the population is a significant factor.When the table is examined for the second stage,it is observed that density clusters for all variables are in the form of 2, 1 and 3. Table 3 below shows the distances between the clusters formed for both stages.

Table 3. Distance Between Cluster Centers

	Cluster	1	2
1. Stage	2	8691910,239	
	3	26658181,935	17966349,629
2. Stage	2	21979,977	
	3	61895,568	39915,828

When Table 3 is examined, it is observed that the longest distance for the 1st phase is between Cluster 1 and Cluster 3, and the shortest distance is between Cluster 2 and Cluster 1. It is observed that the longest distance for the second phase is between Cluster 1 and Cluster 3, and the shortest distance is between

International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XI, Issue: 2, Year: 2021, pp. 454-468

Cluster 1 and Cluster 2. Table 4 shows the distribution of countries by clusters for both stages, and how many days in which cluster the countries took place between 01 May 2020 - 05 May 2021 (370 days).

Table 4. Clustering of Countries

Cl usters	**1. Stage**	**2. Stage**
Cluster 1	Afghanistan (1-370), Algeria (1-370), Azerbaijan (1-370), Bangladesh (1-370), Brazil (1-150), Chad (1-370), China (1-370), Cuba (1-270),Djibouti (1-370), Egypt (1-370), France (1-337), Gambia (1-370), Germany (1-370), Ghana (1-370), Guyana (1-370), India (1-33), Iran (1-370), Iraq (1-370), Italy (1-370), Kenya (1-370), Kyrgyzstan (1-370), Madagascar (1-370), Mexico (1-370), Nepal (1-370), New Zealand (1-370), Nigeria (1-370), Norway (1-370), Pakistan (1-370), Russia (1-359), Senegal (1-370), Singapore (1-370), Sri Lanka (1-370), Sudan (1-370), Tajikistan (1-370), Thailand (1-370), Turkey (1- 289), Uganda (1-370), UK (1-370), USA (1-97), Yemen (1-370)	Afghanistan (1-370), Algeria (1-370), Azerbaijan (1-219), Bangladesh (1-370), Brazil (1-98), Chad (1-370), China (1-370), Cuba (1-270),Djibouti (1-370), Egypt (1-370), France (1-169), Gambia (1-370), Germany (1-217), Ghana (1-370), Guyana (1-343), India (1-360), Iran (1-144), Iraq (1-215), Italy (1-189), Kenya (1-370), Kyrgyzstan (1-344), Madagascar (1-370), Mexico (1-269), Nepal (1-370), New Zealand (1-370), Nigeria (1-370), Norway (1-309), Pakistan (1-370), Russia (1-203), Senegal (1-370), Singapore (1-370), Sri Lanka (1-370), Sudan (1-370), Tajikistan (1-370), Thailand (1-370), Turkey (1- 193), Uganda (1-370), UK (1-180), USA (1-91), Yemen (1-370)
Cluster 2	Brazil (151-270), France (338-370), India (34-258), Russia (360-370), Turkey (290-310), USA (197-270)	Azerbaijan (220-370), Brazil (99-283), France (200-267), Germany (218-370), Guyana (344-370), India (361-370), Iran (145-370), Iraq (216-370), Italy (190-291), Kyrgyzstan (345-370), Mexico (270-370), Norway (310-370), Russia(204-370), Turkey(194-298), UK(181-256), USA (92-165),
Cluster 3	Brazil (271-370), India (259-370), Turkey (311-370), USA(271-370),	Brazil (284-370), France (268-370), Italy (292-370), Turkey(299-370), UK(257-370), USA(166-370),

When table 4 is examined for the first stage, it is seen that the number of countries is in cluster 1 the most. Cluster 1 is followed by cluster 2 and 3, respectively. In addition, clusters 2 and 3 consist of Brazil, India, Turkey and the USA, where case and death rates are high. According to the human development index, it has been observed that countries with low human development index are in cluster 1 in all 370 days considered in general. Likewise, both the health policies they followed during the pandemic and countries such as Singapore and New Zealand, which have a very high human development index, are included in this cluster.Among the main reasons of this situation, it can be said that countries with low human development index do not act transparently in reporting cases and individuals infected with the virus are not followed up.In its last statement, WHO stated that India, Brazil and the USA were the center of the pandemic. Turkey, on the other hand, has been among these countries with the update of the case sharing and the increase in the number of cases recently. In Table 2, it is obtained that the cluster centers for the first stage are generally cluster 2 and 3. This situation is supported in table 4 and the countries included in the WHO statement have been distributed in this cluster.

When the table for the second stage is examined, it is seen that the number of countries is in cluster 1 the most. Cluster 1 is followed by cluster 2 and 3, respectively. The main reason that clusters differ from the first stage is population. In Table 2, it is seen that cluster centers are listed as cluster 2, 1 and 3 according to the second stage. According to this situation, although the number of countries in cluster 1 is high in table 4, the main reason for the central being in cluster 2 is the numerical low number of cases in cluster 1 and population density. The main reason for this can be shown as the lack of transparent case number sharing, as mentioned above. Likewise, the countries designated by WHO as above the epidemic center showed a large distribution in clusters 2 and 3.

Table 5 contains the statistical information about the variables as a result of using the K-Means method for both stages. At the same time, whether the variables are significant in the cluster distribution is included in the table.

Table 5. ANOVA

		Cluster		Error		F	Sig.
		Mean Square	df	Mean Square	df		
Stage 1	Total Cases	71223343296442800,000	2	1334197826338,679	14739	53382,896	000
	Daily Cases	1909131275525,637	2	296649334,393	14739	6435,650	000
	Total Deaths	23452574217444,797	2	1166281754,518	14739	20108,841	000
	Daily Deaths	448646693,336	2	88213,793	14739	5085,902	000
Stage 2	Total Cases Per Mil.	1305484215221,293	2	24043458,100	14755	54296,857	000
	Daily Cases Per Mil.	37395968,855	2	6390,043	14755	5852,225	000
	Total Deaths Per Mil.	713134831,472	2	37596,942	14755	18967,895	000
	Daily Deaths Per Mil.	14384,751	2	4,251	14755	3383,980	000

Table 5 gives statistical information about the variables generated using the K-Means method for both stages. The ANOVA table allows us to see whether the variables show a significant distribution in the clustering. Considering the table, it is observed that all variables in clustering show a significant distribution (Sig.≤0.05). At the same time, looking at the F values for the first stage, it is seen that the most effective variable in clustering is the total state. The least influential variables are daily mortality and daily case variables, respectively. Looking at the F values for the second stage, it is seen that the most effective variable in clustering is a total case per million. The least influential variables are deaths per million and daily cases per million, respectively. The reason for this is to maximize the difference between clusters with cluster analysis. After all, the distribution of observations in clusters is not a random distribution. Table 6 shows the number of rows in the clusters for both phases.

Table 6. Numerical Values in the Sets

	Clusters	Number of Rows	Total
Stage 1	Cluster 1 Cluster 2 Cluster 3	13983,000 657,000 160,000	14800,000
Stage 2	Cluster 1 Cluster 2 Cluster 3	13519,000 1778,000 561,000	14800,000

When Table 6 is examined, it is observed that the number of rows for the first stage is mostly in Cluster 1. The cluster with the least number of rows is the third cluster. In cluster two, there are 160 rows in total. In the second stage, it is observed that the number of rows is mostly in Cluster 1. The cluster with the least number of rows is the third cluster. There are a total of 1778 lines in cluster two.

5. Conclusions

The Covid -19 pandemic started in China in December 2019 and has spread all over the world since March 2020. Identifying the similarities and differences of the effects of the epidemic between countries has become important in this sense, and it has become important to classify countries in the light of pandemic data. In this classification, it has been tried to determine whether there is a harmony between the socio-economic development of the countries and the cluster in the pandemic process.

The sample of the study consists of a total of forty countries, ten countries from each category in four different categories in the Human Development Index. The econometric findings of the study were evaluated in two stages, in the first stage, density centers of the clustering were determined for all variables (total cases, daily cases, total deaths, daily deaths) and the density; it was determined as Cluster 3, Cluster 1, Cluster 4 and Cluster 2. In the second stage, the order was made according to density due to the fact that the population was a significant factor, and it was determined that this order was Cluster 2, Cluster 4, Cluster 1 and Cluster 3.

Considering the number of days in which cluster the countries under consideration are included in the determined dates, it is seen that the first number of countries is in Cluster 1. Cluster 1 is followed by Cluster 3, Cluster 4 and Cluster 2, respectively. Only USA takes place in cluster 2 for the last 73 days. According to the human development index, it is seen that countries with low human development index are generally in Cluster 1 for all 370 days. Likewise, it was determined that both the health policies they followed during the epidemic and Singapore and New Zealand, which had a very high human development index, were included in this cluster. Among the main reasons for this situation are that countries with low human development index do not act transparently in reporting cases and the individuals infected with the virus are not followed up. In Table 1, it is obtained that the cluster centers for the first stage are generally Cluster 3 (Brazil, France, Germany, India, Italy, Mexico, Russia, Turkey, UK, USA). This situation is supported in Table 3 and The countries included in the statement of the World Health Organization were distributed in this cluster. Turkey took place in Cluster 3 due to the recent case sharing update.

For the second stage, it is seen that the number of countries is the most in Cluster 1. Cluster 1 is followed by Cluster 2, Cluster 4 and Cluster 3, respectively. The main reason that clusters differ from the first stage is population. According to the second stage in Table 1, the result is that the cluster centers are in Cluster 2. According to this situation, the main reason for the center being in Cluster 2 (Azerbaijan, Brazil, France, Germany, Guyana, Iran, Iraq, Italy, Kyrgyzstan, Mexico, Norway, Russia, Singapore, Turkey, UK, USA) in Table 3 is the numerical incidence of cases in Cluster 1. The main reason for this can be shown as the lack of transparent case number sharing, as stated above.

Considering the statistical information about the variables as a result of using the K-Means method, it was seen that the variables were significant in the cluster distribution. It is seen that the most effective variable in clustering is the Total Case. The least effective variables are the Daily Death and Total Death variables, respectively. Looking at the F values for the second stage, it is seen that the most effective variable in clustering is Total Cases in a Million. The least effective variables are Death per Million and Case per Million, respectively.

Finally, it has been determined that there is no harmony between the ranking of the countries in the Human Development Index and the group in which the Covid 19 outbreak takes place. For example, countries with different development levels such as Djibouti, Turkey, England, Sudan, Pakistan and New Zealand have been in the same cluster for a long time. Of course, although this situation may have reasons such as transparent data sharing and deliberately low number of cases, the result of this analysis does not overlook the result of the human development index not being able to reveal the development measures sufficiently.

Conflict of interest

The authors declare no conflicts of interest.

International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XI, Issue: 2, Year: 2021, pp. 454-468

REFERENCES

Akın, Y.K.(2008), Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi, Yayınlanmamış Doktora Tezi, İstanbul: Marmara Üniversitesi, Sosyal Bilimler Enstiüsü.

Alptekin, N. and Yeşilaydın, G.(2015), Classifying OECD Countries According to Health Indicators Using Fuzzy Clustering Analysis, Journal Business Research, 7(4), 137- 155.

Atamer, B.(1992), Kümeleme Analizi (Cluster Analysis) ve Kümeleme Analizinin İlaç Sektörüne Uygulanması, Yayınlanmamış Yüksek lisans Tezi, İstanbul.

Blashfield, R.K. andAldenferder, M.S.(1978), The Literature on Cluster Analysis, MultivariateBehavioral Research, 13, 271–295.

Çakmak, Z.(1999), Kümeleme Analizinde Geçerlilik Problemi ve Kümeleme Sonuçlarının Değerlendirilmesi", Dumlupınar Üniversitesi Sosyal Bilimler Dergisi, 3, 187-205.

Demir, Y., Esenbuğa, N., and Bilgin, Ö. C. (2016). İvesi Koyunlarının Et Kalitesini Değerlendirmede Temel Bileşenler Analizinin (PCA) Kullanılması. - Süleyman Demirel University Journal of Natural and Applied Sciences, 20 (3), pp. 536-541.

Demir, Y., Keskin, S., and Çavuşoğlu, Ş. (2021). Doğrusal Olmayan Temel Bileşenler Analizinin Tanıtımı ve Uygulanabilirliği. -KSU J. Agric Nat, 24 (2), pp. 442-450.

Demircioğlu, M. and Eşiyok, S.(2020), Classifying Countries With Clustering Analysis In Struggling Covid-19 Pandemic, Istanbul Commerce University Journal of Social Sciences, 37, 369-389.

Ersöz, F.(2009), Comparing of Cluster and Discriminant Analysis in Selected Health Indicators, Journal of Medical Sciences, 29 (6), 1650-1659.

Forgy, E.W. (1965), Cluster Analysis of Multivariate Data: Efficiency vs Interpretability of Classifications, Biometrics 21, 768–780.

Forgy, E.W.(1965), Cluster Analysis of Multivariate Data: Efficiency vs Interpretability of Classifications. Biometrics, 21, 768-780.

Grabmeier, J., and Rudolph, A.(2002), Techniques of Cluster Algorithms in Data Mining, Data Mining and Knowledge Discovery, 6, 303-360.

Hair, J. F. Black, W. C.,Babin, B. J., Anderson, R. E. and Tatham, R. L.(2014), Multivariate Data Analysis. USA: Pearson .

Han J. and Kamber M.(2000), Data Mining Concepts and Techniques, M. K: Publishes.

Hartigan, J.(1975), Clustering Algorithms, John WileyandSons Ltd., USA.

İşleyen, Ş., (2021). Clustering Analysis of Employment Sectors According to OECD Countries Using the KAverage Method. International Journal of Contemporary Economics and Administrative Sciences ISSN: 1925 – 4423 Volume: XI, Issue: 1, Year: 2021, pp. 093-105

Johnson, R. A., and Wichern D. W.(1988), AppliedMultivariateStatistical Analysis: (2nd Ed.) PrenticeHall, EnglewoodCliffs, New Jersey.

Junhagen, S.(2000), From Complexity to Simplicity: on the Application of Three

TecniquesforMultivariate Data Analysis, MPP WorkingPaper, 15.

Kalaycı, Ş.(2005), SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri, Asil Yayıncılık,Ankara.

Kartal, E., Balaban, M. E.,and Bayraktar, B. (2021). Küresel Covıd-19 Salgınının Dünyada ve Türkiye'de Değişen Durumu Ve Kümeleme Analizi. Journal Of Istanbul Faculty Of Medicine, 84(1), 9-19.

Küçükefe, B.(2020), Clusterıng Macroeconomıc Impact Of Covid-19 In OECD Countries and China, Journal Of Research in Economics, Politics & Finance 5, 280-291.

MacQueen, J.(1967), Some Methods for Classification and Analysis of Multivariate Observations., Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, 1, 281-297.

MacQueen, J.(1967), Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, 1 (1967) 281-297.

Mut, S. and Akyürek, Ç. E.(2017), Classifying OECD Countries According to Health Indicators Using Clustering Analysis, International Journal of Academic Value Studies 3-12, 411-422.

Neuburger, L. and Egger, R(2020)., Travel risk perception and travel behaviour during the Covid-19 pandemic 2020:a case study of the DACH region. Current Issues in Tourism, https://doi.org/10.1080/13683500.2020.1803807

Özdamar, K.(1999). Paket Programları İle İstatiksel Veri Analizi, Kaan Kitabevi, Eskişehir.

Özdamar, K.(2004), Paket Programlar İle İstatistiksel Veri Analizi 2, Eskişehir: Kaan Kitabevi.

Selanik, M.(2007), Türk Tarımının Avrupa Birliği İçindeki Yerinin Kümeleme Analizi İle Belirlenmesi,. Sosyal Bilimler Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi,Ankara.

Sığırlı, D., Ediz, B., Cangür, Ş., Ercan, İ. and Kan, İ.(2006), Examination of Turkey's and European Union Member States' Health Indicators with Multidimensionel Scaling, Inonu University Journal of Medicine Faculty, 13-2, 81-85.

Tatlıdil, H., (2002). Uygulamalı Çok Değişkenli İstatistiksel Analiz, H.Ü. Fen Fakültesi İstatistik Bölümü,Ankara

Tekin, B. (2020), Classıfıcatıon Of Countrıes In The Context Of Covıd-19, Health And Fınancıal Indıcators In The Perıod Of Covıd-19 Pandemıc: Hıerarchıcal Cluster Analysıs. Research Of Financial Economic And Social Studies, 5(2), 336-349.

Ünler, A.(2006), Sezgisel Yöntemlerle K-Harmonik Ortalama Veri Kümeleme En İyilemesi, Ankara SBE Yayınlanmamış Doktora Tezi .

Verelst, F., Kuylen, E., and Beutels, P. (2020). Indications for healthcare surge capacity in European countries facing an exponential increase in coronavirus disease (COVID-19) cases, March 2020. Eurosurveillance, 25(13), 2000323.