www.turkishstudies.net/turkishstudies

Turkish Studies

eISSN: 1308-2140

INTERNATIONAL BALKAN UNIVERSITY
Sponsored by IBU

Research Article / Araştırma Makalesi

Examining the Data of Covid-19 Virus with Factor Analysis by Country

Ülkelere Göre Covid-19 Virüsü Verilerinin Faktör Analizi ile İncelenmesi

Yıldırım Demir*

Abstract: Factor analysis is a multivariate statistical method that aims to obtain a small number of new variables by combining many interrelated variables. This study aims to determine which countries are associated with each other in terms of COVID-19 disease and which country is under which factor with a data set consisting of daily case numbers. Covid-19 data covering 16 countries for 201 days obtained from Github was used as the material in the study. Factor analysis was performed by applying Principal components and Varimax rotation. With the KMO test, it was determined that the data set was suitable for factor analysis and with the principal component approach, sixteen original variables were reduced to four factors with a loss of approximately 24%. After the Varimax rotation process, it was determined which variable (country) under which factor. According to the results, the USA (35393.14) is the first and China (47.64) is the last (sixteenth) in terms of the number of daily cases. As a result, it can be said that countries with similar characteristics in terms of the number of daily cases are under the same factor and therefore, the factors affecting the virus or the methods of combating the virus in these countries are close to each other. In general, countries in the fourth factor are more successful in combating the virus, and other countries can achieve more successful results by using similar methods.

Structured Abstract: This study aims to determine which countries are associated with each other in terms of COVID-19 disease and which country is under which factor with a data set consisting of daily case numbers. Factor analysis is a multivariate statistical method that explains a structure described by p number of variables associated with each other with k number of new variables (k <p) independent from each other. Although there is no relationship between the new variables found by factor analysis, there is a very high relationship between the variables that make up each factor (Özdamar, 2010: 223-266). The main purpose of factor analysis is to identify a small number of common factors that can best describe the complex and very difficult (even impossible) relationships between multiple variables. Thus, by analyzing the higher dimensional variables, lower-dimensional variables are obtained that enable the relationships to be interpreted more easily with a certain loss of information (Hanlıoğlu, 2011: 27).

Daily Coronavirus case data of 16 countries were used as material. Following this purpose; Germany, the United States, Australia, Brazil, China, South Korea, France, India, Iran, Spain, Italy, Japan, Canada, Singapore and Turkey's 201 days (including dates from 12 March to 28 September 2020) Covid-19 data was obtained from the Github website (Github, 2020). COVID-19, an acute respiratory disease caused

*Dr. Öğr. Üyesi, Van Yüzüncü Yıl Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü
Asst. Prof. Dr, Van Yüzüncü Yıl University, Faculty of Economics and Administrative Sciences, Department of Econometrics
ORCID 0000-0002-6350-8122
ydemir@yyu.edu.tr

Cite as/ Atıf: Demir, Y. (2021). Examining the data of Covid-19 virus with factor analysis by country. *Turkish Studies, 16*(2), 613-623. https://dx.doi.org/10.7827/TurkishStudies.47637
Received/Geliş: 16 November/Kasım 2020 *Checked by plagiarism software*
Accepted/Kabul: 25 April/Nisan 2021 **Published/Yayın:** 30 April/Nisan 2021
CC BY-NC 4.0

by a newly emerging virus of zoonotic origin, affects the respiratory system and is transmitted from person to person through droplets (ECDC, 2020; T.R. Ministry of Health, 2020). The most frequently reported clinical symptoms in laboratory-approved cases; fever (88%), dry cough (68%), fatigue (38%), sputum (33%), shortness of breath (19%), sore throat (19%), headache (14%) and muscle ache (myalgia or arthralgia) (15%) (WHO, 2020). Diarrhea (4%) and vomiting (5%) are less common symptoms. The virus affects the lung and, if left untreated, causes diseases such as severe acute respiratory failure syndrome. Covid-19 causes the death of thousands of people every day around the world, as well as causing huge economic losses every day.

Coronavirus, which has turned into a global problem, continues its effect strongly despite all global measures. In order to determine the global impact of this virus that causes COVID-19 disease and to combat it, research should be done in many different scientific fields about this virus and its effects. Because this virus causes serious problems not only in the health sector but in all sectors, especially in the economy. For this reason, countries are struggling with the virus both nationally and globally. However, due to both economic and other reasons, there are differences between countries' methods of combating the virus. Thus, the spread of the virus also varies from country to country.

In the analysis result, since the KMO value was found to be 0.726, it was determined that the data set fit for factor analysis was good and thus, the sample size used was sufficient.

According to the factor analysis results, it is observed that the country with the highest daily average number of cases is the USA (35393.14), and the country with the lowest daily average number of cases is China (47.64). Turkey (1564.34) ranks 8th among 16 countries in terms of the average daily number of cases. Thus, the top three countries with the highest number of cases are listed as USA (35393.14), India (30222.15) and Brazil (23543.66) and approximately 85% of the total number of cases occurring in these 3 countries. Furthermore, the highest relationship was observed between Germany and Italy (0.85), while the lowest relationship was observed between Germany and Iran and Germany and Japan (0.00).

In the determining the most appropriate factor number, factors that can explain at least 67% of the total variance (explained variance 2/3) are taken into consideration. In addition, in general, the number of factors is determined by taking into account the factors with eigenvalues greater than 1. Thus, the original data set consisting of sixteen variables in total was reduced to four factors with a loss of approximately 24% in the total variance. After applying Varimax rotation, countries with similar characteristics in terms of daily case numbers were included under the same factor and these factors were listed as F1, F2, F3 and F4.

As a result, it can be said that the factors affecting the virus are close to each other in countries under the same factor or that countries' methods of combating the virus are close to each other. It is thought that the policies implemented by the countries in the fourth factor are more successful and therefore other countries can achieve more successful results by using similar methods of struggle.

Keywords: Coronavirus, COVID-19, factor analysis, statistical analysis, statistics, varimax

Öz: Faktör analizi, birbiri ile ilişkili çok sayıda değişkenin bir araya gelmesiyle daha az sayıda yeni değişkelerin elde edilmesini amaçlayan çok değişkenli bir istatistik yöntemdir. Bu çalışmanın amacı, günlük vaka sayılarından oluşan bir veri seti kullanılarak COVİD-19 vakka sayısı bakımından hangi ülkelerin birbiri ile ilişkili olduğunu ve hangi ülkelerinin aynı faktör altında bulunduğunu belirlemektir. Çalışmada materyal olarak Github internet adresinden elde edilen, 16 ülke ve 201 günlük Covid-19 verisi kullanılmıştır. Temel bileşenler ve Varimax döndürme işlemi uygulanarak faktör analizi yapılmıştır. KMO testi istatistiği ile veri setinin faktör analizi için uygun olduğu belirlenmiş ve temel bileşenler yaklaşımıyla on altı orijinal değişken yaklaşık %24 kayıpla dört faktöre indirgenmiştir. Varimax döndürme işlemi sonrası hangi değişkenin (ülke) hangi faktör altına yer aldığı belirlenmiştir. Şayet gerçek veriler manüple edilmediyse, analiz sonuçlarına göre günlük vaka sayısı bakımından ABD (35393.14) ilk ve Çin (47.64) ise son (on altıncı) sırada bulunmaktadır. Sonuç olarak, günlük vaka sayısı bakımından benzer karakteristik özellikler gösteren ülkelerin aynı faktör altında yer aldığı ve bu nedenle, bu ülkelerde virüsü etkileyen faktörlerin veya ülkelerin virüs ile mücadele yöntemlerinin birbirine yakın olduğu söylenebilir. Genel olarak dördüncü faktörde bulunan ülkelerin virüs ile mücadelede daha başarılı oldukları ve diğer ülkeler de benzer yöntemler kullanarak daha başarılı sonuçlar elde edebilirler.

Anahtar Kelimeler: Coronavirus, COVID-19, faktör analizi, istatistiksel analiz, istatistik, varimax

Introduction

In some cases, it is of great importance to understand and interpret the phenomena that occur in the natural cycle. Because the most important way to deal with the problems that arise with these phenomena is often possible by determining the factors that affect them. However, generally, the number of these factors is too many and their effect levels are different from each other. Thus it may not be easy to understand and interpret them. In this case, the data set can be easily interpreted by reducing it to fewer dimensions with dimension reduction methods. Factor analysis, one of the multivariate statistical methods, can be used for dimension reduction.

Factor analysis is a multivariate statistical method that explains a structure described by p number of variables associated with each other with k number of new variables ($k < p$) independent from each other. Although there is no relationship between the new variables found by factor analysis, there is a very high relationship between the variables that consisting each factor (Özdamar, 2010: 223-266). Factor analysis is used in many different fields such as behavioral sciences, medicine and economics, especially social sciences. In this context, a multi-dimensional, complex and very difficult to interpret (Covid-19) data set is analyzed with factor analysis and reduced to smaller sizes and thus can be interpreted more easily.

Coronavirus was first seen in the 1960s, the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) appeared in 2003 and the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) in 2012, caused great losses. In 2019, a new type of Coronavirus (2019-nCoV) emerged and was named SARS-CoV-2 due to its similarity to SARS-CoV. Coronaviruses are positive polarity, enveloped, single-stranded RNA viruses; among these, the SARS-CoV-2 virus causes the disease defined as COVID-19 (Akbıyık & Avşar, 2020; Gorbalenya et al., 2020).

COVID-19, an acute respiratory disease caused by a newly emerging virus of zoonotic origin, affects the respiratory system and is transmitted from person to person through droplets (ECDC, 2020; T.R. Ministry of Health, 2020). The most frequently reported clinical symptoms in laboratory-approved cases; fever (88%), dry cough (68%), fatigue (38%), sputum (33%), shortness of breath (19%), sore throat (19%), headache (14%) and muscle ache (myalgia or arthralgia) (15%) (WHO, 2020). Diarrhea (4%) and vomiting (5%) are less common symptoms. The virus affects the lung and, if late intervention, causes diseases such as severe acute respiratory failure syndrome. Covid-19 causes the death of thousands of people every day around the world, as well as causing huge economic losses every day.

This virus, which emerged in Wuhan, China's Hubei province, in December 2019, globalized in a short time, affecting almost the whole world and started to adversely affect people's daily lives. Thus, there were daily increases in the number of countries affected by the outbreak, the number of confirmed cases, and the deaths associated with the disease. Global data are updated and published daily by organizations such as the World Health Organization (WHO), the European Center for Disease Prevention and Control (ECDC), and the US Center for Disease Control and Prevention (CDC) (WHO, 2019). In addition, each country records its number of cases under international standards.

It was announced by the World Health Organization that COVID-19 created a Public Health Emergencies of International Importance (PHEIC) on January 30, 2020. On March 11, 2020, it was declared as a global epidemic, considering the global spread rate and severity (WHO, 2019). However, despite all global precautions, its spread cannot be prevented and it creates problems for the whole world.

The aim of this study is to determine which countries are associated with each other in terms of COVID-19 disease and which country is under which factor with a data set consisting of the daily number of COVID-19 disease cases of 16 countries.

Material and Method

In the study, daily Coronavirus case data of 16 countries were used as material. Following this purpose; Germany, the United States, Australia, Brazil, China, South Korea, France, India, Iran, Spain, Italy, Japan, Canada, Singapore and Turkey's 201 days (including dates from 12 March to 28 September 2020) Covid-19 data was obtained from the Github website (Github, 2020). Factor analysis was used as a method and analyzes were carried out with the IBM SPSS Statistics 25 statistical package program.

Factor Analysis

Factor analysis aims to explain the original data set with maximum covariance and few factors; it is a multivariate statistical method that enables the observed variables to be interpreted in a more understandable, easy and summary manner (Büyüköztürk, 2007; Özen, 2013). The main purpose of factor analysis is to identify a small number of common factors that can best describe the complex and very difficult (even impossible) relationships between multiple variables. Thus, by analyzing the higher dimensional variables, lower-dimensional variables are obtained that enable the relationships to be interpreted more easily with a certain loss of information (Hanlıoğlu, 2011: 27).

Factor analysis is used not only for dimension reduction, but also for determining the factors that explain the relationships between variables, revealing hidden dimensions that cannot be observed among variables, ensuring the validity of assumptions of various statistical models, and determining variables associated with discriminant and regression analysis (Atlaş, 2002; Karagöz et al., 2010).

Some assumptions should be fulfilled to apply factor analysis. These assumptions; variables can be listed as being quantitative, being in an interval or ratio scale, observations showing the normal distribution and are independent of each other.

Factor analysis basically consists of 5 stages. These are, examining whether the data has a factorable structure and meeting the necessary assumptions, obtaining factor loadings matrix using one of the factor extraction methods, determining the number of factors to be considered with methods such as eigenvalue, slope plot, making factor rotation to make the factors easier to interpret, and interpreting the results as a whole (Karagöz et al., 2010).

Standardized data matrix (Z_{pn}) is used in factor analysis and this data matrix is obtained from the raw data matrix (X_{pn}), which shows p properties of n individuals or objects (Tatlıdil, 1992:141-170; Patır, 2009). Thus, the factor analysis model is defined as a linear model showing the relationship between common factors $f_1, f_2, \dots, f_k$ and standardized variables (z_i). In general, the factor analysis model is shown with equation (1).

$$z_i = a_{i1}f_1 + a_{i2}f_2 + \dots + a_{ik}f_k + b_i u_i \qquad i = 1, 2, \dots, p \qquad (1)$$

Where, a_{ik} shows the weight of the i^{th} variable on the k^{th} factor, while u_i is the residual factor and b_i is the residual factor coefficient. Obtaining the *pxk*-dimensional load matrix ($A = a_{ij}$) with factor load value a is the main purpose of this method (Tatlıdil, 1992: 141-170).

The coefficient explaining the relationship of variables (items) with factors is called "factor coefficient" or "factor loading". Loadings are expected to be high in the factor including the variables. If variables that are highly correlated with a factor in a data set, it can be expected that these variables together measure a concept-structure-factor. Among the variables that define the factor, those with a factor load less than 0.3 are not taken into account. Because the square of this value is less than 0.09 (9%), this indicates that less than 10% of the variance is explained by the factor, and the remaining 90% is under the other variables and/or error variance (Comrey & Lee, 2009). In general, loadings between 0.30 and 0.59 are expressed as medium, and loadings greater

than 0.59, regardless of the sign, are expressed as high degree and these are considered in variable extraction. Factor loadings, which are correlations between original variables and factors, can also be examined in terms of statistical significance. However, even if there is a low relationship, an increase in the sample may increase the probability of a significant relationship (Büyüköztürk, 2007).

Based on the assumption that common factors are unrelated to each other and the residual factor, the variance of z_i variables is calculated with equation (2).

$$Var(z_i) = a_{i1}^2 + a_{i2}^2 + \cdots + a_{ik}^2 + b_i^2$$
$$= h_i^2 + b_i^2 = 1 \quad (2)$$

Where, h_i^2 shows the common factor variance of the *i*th variable and b_i^2 shows the error variance (the variance belonging to the part that the common factors cannot explain). The variance that can be explained by common factors in factor analysis is called common factor variance or common variance.

After finding the *pxk*-dimensional factor load matrix indicated by A, common variance values for each variable are calculated with equation (2). The fact that all of the common variance values are close to 1, the factorization of the correlation matrix is good, and some of them are less than 0.50 indicates that factorization is not good. However, one of the criteria used in determining the most appropriate factor number is the ratio value in equation (3).

$$\sum_{i=1}^{p} \frac{h_i^2}{p} \geq \frac{2}{3} \quad (3)$$

Thus, factors that can explain at least 67% of the total variance (explained variance 2/3) are taken into consideration. In addition, in general, the number of factors is determined by taking into account the factors with eigenvalues greater than 1. The eigenvalue of each factor is equal to the sum of the variances calculated for that factor. Thus, it can be computed as follow;

$$\lambda_j = a_{1j}^2 + a_{2j}^2 + \cdots + a_{pj}^2 \qquad j = 1, 2, \ldots, k \quad (4)$$

Also, it can be expressed as follow (Cureton & D'Agostino, 2009)

$$\sum_{j=1}^{k} \lambda_j = \sum_{i=1}^{p} h_i^2 \qquad j = 1, 2, \ldots, k \qquad i = 1, 2, \ldots, p \quad (5)$$

and variance explanation rate is calculated with equation (6).

$$\%v_j = \frac{\lambda_j}{p} * 100 \qquad j = 1, 2, \ldots, k \quad (6)$$

Where, λ_j shows the eigenvalue of the *j*'th factor, v_j shows the variance explanation ratio of the *j*'th factor and *p* the total number of variables (Child, 1975). Factor loadings (a_{ij}), with e_{ij} being the eigenvector element for the *j*th eigenvalue, can also be found by multiplying the square root of each eigenvalue and the vector elements associated with that eigenvalue (Özdamar, 2010: 223-266).

$$a_{ij} = \sqrt{\lambda_j} e_{ij} \quad (7)$$

Factoring or finding factors can be defined as the process of revealing new variables (concepts) or obtaining functional definitions of concepts by using factor loadings of variables

(items). What is expected from a good factorization; the dimensions of the variables are reduced, new variables (factors) are independent of each other, and conceptual significance is ensured (Büyüköztürk, 2007).

Factors should be rotated to ensure conceptual significance. The transformation of the obtained factors into new factors to better interpret (conceptual significance) is expressed as factor rotation. In the factor rotation process, while the factor loadings are made orthogonal by rotating the axes at a suitable angle, the mathematical properties of the solution remain the same. However, as the loading of the variables on one factor increases with the rotation of the axes, the loadings on other factors decrease, thus variables that are highly correlated with the factors can be found and interpreted easily. Rotation techniques applied for better interpretation of the factors are examined in two groups as vertical (Varimax, Orthomax ... etc.) and oblique (Quartimin, Oblimin ...) (Özdamar, 2010: 223-266).

To determine whether the data set is suitable for factor analysis, the Kaiser-Meyer-Olkin (KMO) test, Bartlett test and one of the methods of creating the correlation matrix are used. The most widely used among these is the KMO method (Akgül & Çevik, 2003). The KMO scale compares the observed correlation coefficients with the partial correlation coefficients (Patır, 2009; Mulaik, 2010: 241).

$$KMO = \sum\sum_{i\neq j}^{n} r_{ij}^2 \Big/ \left(\sum\sum_{i\neq j}^{n} r_{ij}^2 + \sum\sum_{i\neq j}^{n} a_{ij}^2 \right)$$

Where, r_{ij}^2 shows the square of the *i, j*th observed correlation between *i* and *j* variables and a_{ij}^2 shows the partial correlation coefficient between variables *i* and *j*.

When the partial correlation coefficient approaches 0, the KMO value approaches 1. KMO value close to or greater than 0.90 indicate that the fit is perfect, values close to 0.8 indicate good fit, values close to 0.7 indicate that the fit is medium, values close to 0.6 indicate that the fit is bad, and values close to or less than 0.5 indicate that the fit is very bad. Thus, how suitable the data set is for factor analysis depends on how close the KMO value is to 1 (Mulaik, 2010: 241).

Varimax aims to maximize the sum of the squared variances of the loadings in the columns of the factor matrix. This produces high or near zero loadings on each column that is a factor. One of the critical features of Varimax is that it uses squares of loadings rather than the actual loadings of the procedure.

Results and Discussion

In the study, a data set consisting of daily (201 days) COVID-19 cases from 16 countries was used. Descriptive statistics of the data set are given in Table 1.

Table 1: Descriptive statistics for COVID-19 cases by country

Country	Mean	Standard Deviation
USA	35393.14	17291.56
Germany	1413.11	1416.19
Austria	133.97	160.61
Brazil	23543.66	18274.35
China	47.64	53.60
France	2670.57	3334.21
S. Korea	79.13	85.04
India	30222.15	32284.99

UK	2160.96	1660.06
Iran	2181.12	743.87
Spain	3548.29	5034.26
Italy	1491.15	1576.38
Japan	405.79	428.40
Canada	761.35	518.23
Singapore	286.24	275.36
Turkey	1564.34	999.40

In Table 1, it is observed that the country with the highest daily average number of cases is the USA (35393.14), and the country with the lowest daily average number of cases is China (47.64). Turkey (1564.34) ranks 8th among the 16 countries in terms of average daily number of cases. Thus, the first three countries with the highest number of cases are listed as USA (35393.14), India (30222.15) and Brazil (23543.66) and approximately 85% of the total number of cases occurring in these 3 countries.

Table 2 shows the results of KMO and Bartlett tests used to determine whether the data set is suitable for factor analysis. The relationship between the variables used in factor analysis should be high. As the relationship between variables decreases, the reliability of factor analysis results decreases (Durak et al., 2017).

Table 2. KMO and Bartlett test results

Kaiser-Meyer-Olkin Sampling Adequacy		0.726
Bartlett's Test of Sphericity	Approximate Chi-Square	2768.013
	Degree of Freedom	120.000
	p	0.001

Bartlett test given in Table 2 shows that the relationship between variables is high and the data come from multiple normal distributions. This value does not change unless the data has an extremely abnormal structure. For this reason, KMO is considered for compliance. Since the KMO value was found 0.726, it was determined that the data set fit for factor analysis was good and thus, the sample size used was sufficient.

Table 3. Correaltion coefficients between the countries for COVID-19 cases

	Germany	Austria	Brazil	China	France	S. Korea	India	UK	Iran	Spain	Italy	Japan	Canada	Singapore	Turkey
USA	-0.25	0.49	0.73	0.39	0.11	0.09	0.55	-0.21	0.36	-0.03	-0.50	0.59	-0.21	-0.13	-0.07
Germany		0.14	-0.46	0.22	0.32	0.09	-0.19	0.54	0.00	0.29	0.85	0.00	0.34	-0.20	0.54
Austria			0.33	0.73	-0.10	0.10	0.22	-0.35	0.16	0.11	0.03	0.54	-0.36	-0.22	-0.18
Brazil				0.18	0.09	0.27	0.62	-0.48	0.43	-0.10	-0.67	0.48	-0.49	-0.24	-0.38
China					-0.08	-0.05	0.11	-0.18	0.12	0.10	0.13	0.46	-0.20	-0.11	0.05
France						0.39	0.66	0.51	0.30	0.47	0.24	0.18	0.17	-0.43	0.15
S. Korea							0.55	-0.17	0.09	0.37	0.05	0.33	-0.32	-0.50	-0.13
India								-0.06	0.35	0.43	-0.33	0.58	-0.25	-0.47	-0.21
UK									-0.07	0.20	0.50	-0.08	0.81	0.24	0.66
Iran										0.10	-0.26	0.14	-0.18	-0.32	-0.14
Spain											0.25	0.12	-0.01	-0.37	0.14

Italy	-0.17	0.30	-0.14	0.43
Japan		-0.15	-0.18	0.06
Canada			0.53	0.61
Singapore				0.23

According to Table 3, the relationship between the USA and Brazil, Germany and Italy and Australia and China with a correlation value higher than 0.7 is very high. Also, it can be noted that the relationship between countries with a correlation value of 0.11 or less (underlined) is nonsignificant, and the relationship between countries with a value higher than 0.11 is important. The highest relationship was observed between Germany and Italy (0.85), while the lowest one was observed between Germany and Iran, Germany and Japan (0.00).

In Figure 1, the scree plot showing the points between eigenvalue and component axes is given. Each of these points indicates a factor and when determining the factor structure, this diagram is used as well as Table 4. Thus, how many of the factors obtained as a result of factor analysis will be taken into account is determined using either Table 4 or Figure 1.

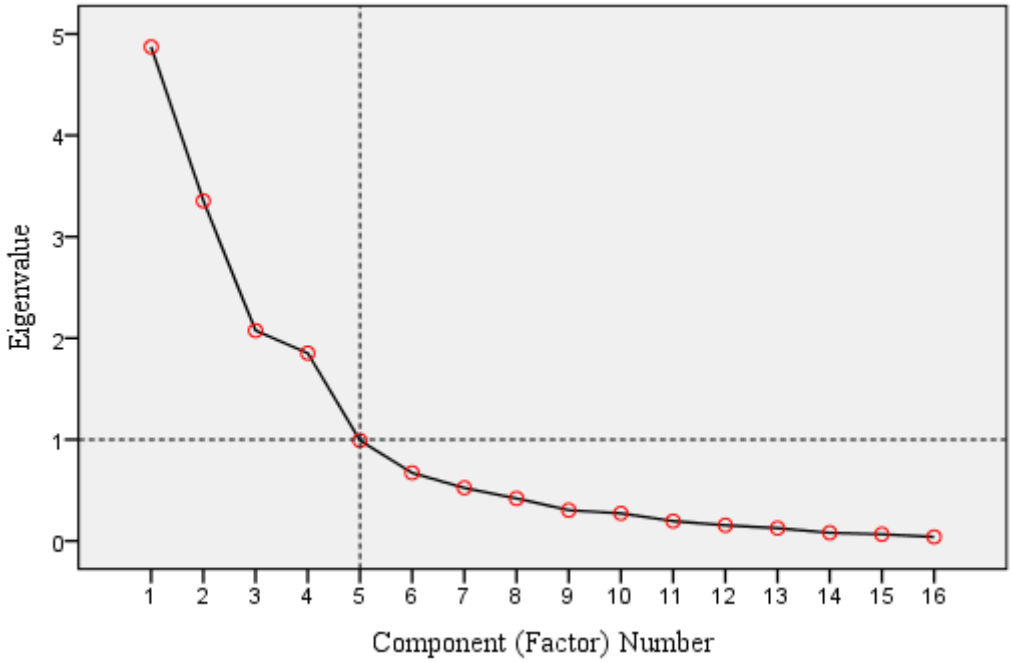


Figure 1. Scree plot for eigenvalue and factors

When Figure 1 is examined, it can be stated that after a significant decrease has occurred up to a certain point, this decrease reduces considerably and the variability between the points is minimized and thus, the component values are at the same point. The point where this variability is minimized is called the elbow, considering the number of components on the left side of the elbow, it was determined that the data set can be explained with 4 factors (F1, F2, F3, F4) instead of 16 original variables. In addition, the number of factors can be determined by considering the factors with eigenvalues of one and greater than one ($\lambda \geq 1$) in the diagram.

After Varimax rotation, the eigenvalues and explained variance of the factors with eigenvalues greater than 1 are given in Table 4.

Table 4. Descriptive statistics for dimensions after Varimax rotation

Factor	Eigenvalue	Variance Explained (%)	Cumulative Variance (%)
F1	3.33	20.82	20.82
F2	3.27	20.44	41.26
F3	3.07	19.17	60.43
F4	2.49	15.54	75.97

When Table 4 is examined; 4 factors with eigenvalues greater than 1 were taken into account. Of these, the 1st factor with 3.33 eigenvalue accounted for 20.82% of the total variance, the 2nd factor with 3.27 eigenvalue accounted for 20.44% of the total variance, the 3rd factor with 3.07 eigenvalue and 4th factor with 2.49 eigenvalues accounted for 19.17% and 15.54% of the total variance, respectively Thus, 75.97% of the total variance was explained by 4 factors. Since the cumulative explained variance is more than 67%, the original data set can be explained by using 4 factors instead of 16 original variables. Although there are considerable differences between the explained variance of the factors before the rotation, these values approached after the rotation.

Regardless of whether a variable has a higher loading under which factor, regardless of whether it is positive or negative, the variable is more correlated with that factor. It is also known that loadings higher than 0.50 are also very important (Alpar, 2011: 261-307). Factor loadings showing the relationship between the original variables and their factors are given in Table 5.

Table 5. Factor loadings with eigenvalues greater than 1 after rotation

	Factors			
	F1	F2	F3	F4
Brazil	**0.811**	-0.374	0.105	0.173
USA	**0.799**	-0.016	-0.002	0.442
Italy	**-0.771**	0.408	0.315	0.229
India	**0.699**	-0.031	*0.643*	0.064
Germany	**-0.546**	*0.529*	0.375	0.360
Iran	**0.451**	-0.036	0.286	0.078
UK	-0.138	**0.925**	0.121	-0.188
Canada	-0.107	**0.868**	-0.238	-0.215
Turkey	-0.174	**0.800**	-0.043	0.122
Singapore	0.011	0.348	**-0.783**	-0.151
France	0.238	0.422	**0.767**	-0.143
S. Korea	0.121	-0.214	**0.727**	0.003
Spain	-0.067	0.171	**0.687**	0.084
China	0.050	-0.016	-0.022	**0.898**
Austria	0.134	-0.217	0.073	**0.881**
Japan	*0.541*	0.096	0.202	**0.584**

According to Table 5, a high correlation was found between Brazil, USA, Italy, India, Germany and Iran with the first factor. Among these variables (countries), a negative correlation was observed between F1 and Italy as well as F1 and Germany, while a positive correlation was observed between F1 and the other 4 countries. The second factor is defined by the UK, Canada and Turkey, and it was found a positive corralations between F2 and these three countries. The third factor is composed of Singapore, France, South Korea and Spain. Among these variables, a negative correlation was found between Singapore and F3, and a positive corrleation between F3 and the other three countries. The fourth factor has been identified with China, Australia and Japan to be positively correlated. Also; it can be observed that the loadinds of Japan's in F1, Germany's in

F2 and India's in F3 are quite good. Thus, the data set could be easily interpreted with a certain loss by using 4 factors instead of 16 original variables.

Coronavirus, which has turned into a global problem, continues its effect strongly despite all global measures. In order to determine the global effect of this virus that causes COVID-19 disease and to struggle elimimate it, research should be done in many different scientific fields about this virus and its effects. Because this virus causes serious problems not only in the health sector but in all sectors, especially in the economy. For this reason, countries are struggling with the virus both nationally and globally. However, due to both economic and other reasons, there are differences between countries' methods of combating the virus. Thus, the spread of the virus also varies from country to country.

In this study, the daily case numbers of 16 selected countries were analyzed by factor analysis, and it was determined which countries are associated with each other and which country is under which class (factor).

According to the factor analysis results, the original data set consisting of sixteen variables in total was reduced to four factors with a loss of approximately 24% in the total variance. After applying Varimax rotation, countries with similar characteristics in terms of daily case numbers were included under the same factor and these factors were listed as F1, F2, F3 and F4.

Conlussion

As a result, it can be stated that the factors affecting the virus are close to each other in countries under the same factor or that countries' methods of combating the virus are close to each other. It is emphasised that the policies implemented by the countries in the fourth factor are more successful and therefore other countries can achieve more successful results by using similar methods of struggle.

References

Akbıyık, A. & Avşar, Ö. S. (2020). Epidemiology and control of coronavirus ınfection disease (COVID-19). *İzmir Kâtip Çelebi Üniversitesi Sağlık Bilimleri Fakültesi Dergisi, 5*(2), 109-116.

Akgül, A. & Çevik, O. (2003). *İstatistiksel analiz teknikleri spss'te işletme yönetimi uygulamaları*. *2nd ed.,* Emek Ofset.

Alpar, R. (2011). *Uygulamalı çok değişkenli istatistiksel yöntemler*. *3nd ed.,* Detay Yayıncılık.

Atlaş, D. (2002). Hayat ve hayat dışı sigorta şirketlerinin çok değişkenli istatistiksel yöntemlerle karşılaştırılması. *Öneri Dergisi*, *5*(18), 147-154. https://doi.org/10.14783/maruoneri.683290

Büyüköztürk, Ş. (2020). *Sosyal bilimleri için veri analizi el kitabı istatistik araştırma deseni spss uygulamaları ve yorum*. *27nd ed.,* Pegem A Yayınları.

Child, D. (1975). *The essentials of factor analysis.* Billing & Sons Ltd.

Comrey, A. L. & Lee, H. B. (2009). *A first course in factor analysis. 2nd ed.,* Psychology Press.

Cureton, E. E & D'Agostino, R. B. (2009). *Factor analysis: an applied approach.* Psychology Press.

Durak, İ., Bayat, M. & Arslan, H. M. (2017). Customer preferences in bank selection: an empirical analysis on participation banks. *The Journal of Social Science, 1*(2), 60-70. https://doi.org/10.30520/tjsosci.363391

ECDC. (2020). European Centre for Disease Prevention and Control. COVID-19 2019. https://www.ecdc.europa.eu/en/novel-coronavirus-china (Date of Access: 20.09.2020)

Github. (2020). https://github.com/owid/covid-19-data/blob/master/public/data/owid-covid-data.xlsx (Date of Access: 29.09.2020)

Gorbalenya, A. E., Baker, S. C., Baric, R. S, et al. (2020). The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol, 5*, 536–544. https://doi.org/10.1038/s41564-020-0695-z

Hanlıoğlu, E. B. (2011). *Kontrol edilebilir risk* [Master Thesis]. Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü.

Karagöz, Y., Kıngır, S., Mesci, M. & Akbaş, Z. (2010). Zamanın etkin kullanımını sağlayan faktörlerin belirlenmesine yönelik bir araştırma. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 23,* 97-08.

Mulaik, S. A. (2010). *Foundations of factor analysis. 2nd ed.,* CRC Press.

Özdamar, K. (2010). *Paket programları ile istatistiksel veri analizi-2 (çok değişkenli analizler). 7nd ed.,* Kaan Kitapevi.

Özen, H. (2013). Okul müdürlerine yönelik motivasyonel dil ölçeği: türk kültürüne uyarlama, dil geçerliği ve faktör yapısının incelenmesi. *Journal of Educational Sciences Research, 3*(1), 87-03.

Patır, S. (2009). Faktör analizi ile öğretim üyesi değerleme çalışması. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi, 23*(4), 69-86.

T.R. Ministry of Health. (2020). COVID-19 (SARS-CoV2 enfeksiyonu) rehberi (bilim kurulu çalışması). TC. Sağlık Bakanlığı. https://dosyamerkez.saglik.gov.tr/Eklenti/37175, covid-19rehberipdf.pdf?0 (Date of Access: 25.08.2020)

Tatlıdil, H. (1992). *Uygulamalı çok değişkenli istatistiksel analiz.* Akademi Matbaası.

WHO. (2019). Coronavirus disease (COVID-19) outbreak. https://www.who.int/emergencies /diseases/novel-coronavirus-2019 (Date of Access: 02.09.2020)

WHO. (2020). Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19). https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf (Date of Access: 02.09.2020)