JP Journal of Biostatistics
© 2021 Pushpa Publishing House, Prayagraj, India
http://www.pphmj.com
http://dx.doi.org/10.17654/JB018020149
Volume 18, Number 2, 2021, Pages 149-163 **P-ISSN: 0973-5143**

A REVIEW ON THE COVID-19 EFFECTS OF THE MOBILITY TRENDS

Özlem Deniz Başar and Seda Bağdatlı Kalkan

Department of Statistics
Istanbul Commerce University
İstanbul, Turkey

Abstract

Due to the COVID-19 outbreak which showed its deadly results all around the world in the first quarter of 2020, all the countries have taken different measures. It is planned with lockdown, which has been a primary measure, to minimize human contact and decrease transmission rates. Various lockdown measures have been taken and implemented in different regions of Turkey, too. This research studies the effect of the change in densities in different locations on COVID-19 health data, based on the mobility trends in a certain time period and COVID-19 health data obtained regarding the related time period in Turkey. Thus, it is shown that decreasing the mobility trend has a positive effect on statistics concerning human life.

1. Introduction

COVID-19 virus, which was first seen in China towards the end of 2019, has spread around the world in a very short time and today it shows its deadly results on humans. All the countries have taken various measures in order to be able to keep under control the spread process of the virus and these measures are still strictly implemented today.

Received: February 4, 2021; Accepted: March 1, 2021
Keywords and phrases: canonical correlation, COVID-19, Google Mobility Report, mobility trends.

While the spread process of the virus continues at a very rapid pace in the world, the first case in Turkey was officially reported in March 2020 [1]. Following this report, various measures have been taken to reduce the spread of the virus and these measures are still implemented from time to time with updates/improvements.

There are several factors which affect the spread of the virus. The existing evidence shows that the most obvious proved reason for the spread of the virus is through respiratory droplets between people who are in close contact with one another. The first studies of the World Health Organization on COVID-19 indicate that the transmission of respiratory droplets between people is most likely to occur in certain environments, especially in crowded and poorly ventilated indoor areas such as restaurants, schools, sport areas, offices, shopping centers, groceries, etc. where infected people spend long hours with others [2]. In the light of this information, countries firstly restricted mass events in which their population move collectively and took decisions to prevent people from going out from their houses in different periods.

This research studies the effect of mobility changes in different locations of people who live in Turkey on the number of persons diagnosed with COVID-19 and/or the death toll caused by this virus. For this reason, Google Mobility Report data indicate that the rate of presence in different environments increase or decrease based on a reference value. The health data obtained from the Republic of Turkey Ministry of Health were used in this research. The data obtained was studied with the canonical correlation analysis; and, following the completion of the study, the effect of the change in densities in different locations on COVID-19 data was evaluated.

2. Literature

In the last 6 months, since the number of observations has increased considerably, the number of COVID-19 related publications in different scientific fields has also increased. While, in general, the studies in the

medical field help to find out the effects of the virus on humans, and the studies in social sciences provide an opportunity to introduce different approaches on the effects of COVID-19.

Saha et al. [3] studied community mobility in different areas of India, based upon Google Mobility Report. They studied the impact of COVID-19 lockdown and unlock periods on mobility trends, by employing time-series trends, spatial analyses, and different mapping techniques in their research. Vokó and Pitter [4] studied the effect of social distance measures on the COVID-19 epidemic. For this reason, they calculated the Social Distance Index for 28 European countries based on Google Mobility Report, and also conducted the Poisson regression analysis in their research. Russo et al. [5] carried out their studies in Lombardy, Italy, and observed the growth of the infected cases in the Lombardy region based on the restrictions implemented in Italy, by using a medical approach for case detection. To this end, they benefited from Google Mobility Report and simulated this process with the mixed-integer optimization model. Day [6] used the data obtained for the United Kingdom from Google Mobility Report. In this research, Day studied the economic effects of curfews on Greenspace, by benefiting from Latent Class Regression methods. Upon the modeling of the mortality data and the mobility trends obtained from Google Mobility Report data with different regression methods, in order to evaluate the social distancing policies implemented in Brazil, Valenti et al. [7] concluded that the implemented social distancing policy contributes to the reduction in mortality rates. Chernozhukov et al. [8] assessed the mobility trends for places of work and residence by using Google Mobility Report data, for the purpose of identifying the impact of implemented mask mandates in the United States. Counterfactual Structural Equations Model was used to assess the relationship between variables and it was concluded that the impact of implemented mask mandates is more effective than that of workplace closures and stay-at-home policies. DucHuynh [9] found out that, among 58 countries, the ones with a higher “Uncertainty Avoidance Index” value have a lower mobility trend data according to Google Mobility Report. DucHuynh

also included some economic variables in the study as a control group and used the factor analysis and regression analysis for modeling.

3. Data

The data used in this study were obtained from the Google Mobility Report published on November 8, 2020, and the General Coronavirus table published on the same date by the Republic of Turkey Ministry of Health.

Google Mobility Report is published on a daily basis [10]. The report used in this study was prepared by comparing the mobile mobility in locations determined in countries for a 5 week-period from January 3 to February 6 with the reference values calculated as the median of the related day of a week. For this reason, mobility trends for retail and recreation, grocery and pharmacy, parks and transit stations, workplaces, and residences were calculated and reported on a daily basis by Google based on signal data from users who have opted-in to "location history" of their devices.

Table 1. Variables used in the research

Dependent variable set (the COVID-19 set)	Independent variable set (mobility trends)
Total number of tests	Retail and recreation
Total number of patients	Grocery and pharmacy
Total number of deaths	Parks
Pneumonia rate in patients	Transit stations
Number of critically ill patients	Work places
Total number of recoveries	Residential
Number of cases reported today	
Number of tests conducted today	
Number of deaths reported today	
Number of recoveries today	

An increase or decrease of mobility in these locations is considered as a factor affecting the transmission rate of the virus. To find out if this is the case, the variables given in the general coronavirus tables updated daily by the Republic of Turkey Ministry of Health and shown in Table 1 were used in this study. The mobility trends cover the period from June 26, 2020 to November 24, 2020. Since it was understood from scientific findings that a person who is exposed to an infected person shows symptoms in 4 days on

average, the health data was prepared based on the period from July 1, 2020 to November 29, 2020. It was considered that the relationship between health data could be examined and evaluated based on mobility trends.

4. The Canonical Correlation Analysis

The Canonical Correlation Analysis is a multivariate statistical analysis method, and the objective of which is to analyze the linear relationship between two sets of variables. It could be considered as a generalized version of the classic multivariate linear regression analysis. It is stated in the literature that it is not necessary to define dependent and independent variable sets. However, if it is possible in practice to make a distinction between the dependent and independent variable sets, the main objective of the analysis is to determine to what extent the independent variable set defines the dependent variable set. If the number of dependent variables is 1, the Canonical Correlation becomes a multivariate regression, and if the number of dependent and independent variables is 1, then it becomes a simple correlation. The variance analysis and multivariate analysis of variance are also special conditions of the canonical correlation. The canonical correlation analysis is similar to factor analysis, for it determines the sets as a sole variable, too [11]. In the canonical correlation analysis, it is aimed to find the maximum correlations between the linear functions of variable sets. Thus, firstly, linear components of variables of each set, one with maximum correlation and one with unit variance are found. Then the independent second linear component pair with maximum correlation and unit variance is obtained from this pair. The pairs with the maximum correlation are chosen in a specific process after being tested [12]. The basis for the canonical correlation analysis is the correlation matrix. By standardizing variances of variables as 1, the covariance to be calculated between variables would be equal to correlation. Thus, if variables are standardized, the result would not change depending on the choice of correlation or covariance and their matrices [13].

U and V, respectively, represent linear variables of X and Y, so the canonical variables. If p corresponds to a variable set of either X or Y, one with less variables, then canonical variable pair p times U and V are obtained as follows:

$$\begin{aligned}
U_1 &= \alpha_1' X = \alpha_{11} X_1 + \alpha_{12} X_2 + \cdots + \alpha_{1m} X_m \\
U_2 &= \alpha_2' X = \alpha_{21} X_1 + \alpha_{22} X_2 + \cdots + \alpha_{2m} X_m \\
&\vdots \\
U_p &= \alpha_p' X = \alpha_{p1} X_1 + \alpha_{p2} X_2 + \cdots + \alpha_{pm} X_m \\
V_1 &= \beta_1' Y = \beta_{11} Y_1 + \beta_{12} Y_2 + \cdots + \beta_{1n} Y_n \\
V_2 &= \beta_2' Y = \beta_{21} Y_1 + \beta_{22} Y_2 + \cdots + \beta_{2n} Y_n \\
&\vdots \\
V_p &= \beta_p' Y = \beta_{p1} Y_1 + \beta_{p2} Y_2 + \cdots + \beta_{pn} Y_n.
\end{aligned}$$

The Pearson correlation coefficient between the specified canonical variables is called canonical correlation [14]. Canonical correlations are found due to eigenvalues. After the calculation of canonical correlations, they should be tested for significance. In this method, the objective is to test how many of the relationships between canonical correlation pairs found could be considered as significant [15]. Many tests have been developed to test the significance of canonical correlation coefficients. The one most used is the Bartlett test. Roy's Largest Characteristic Roots Method and a graphical approach developed by Heck could be given as an example for other tests used in the literature. When the significance of canonical correlation coefficients is tested, if the number of observations is high, weak canonical correlation coefficients could be considered statistically significant. If the canonical correlation coefficients are strong, it does not necessarily mean that the correlation coefficient between two variable sets will be high. This is because, in the canonical correlation, the objective is not to maximize the amount defined by one set for the other variable set but the correlations between linear combinations in variable sets [16]. The step following the significance test is to calculate the eigenvectors corresponding to significant eigenvalues.

While evaluating the results of the canonical correlation analysis, the standardized coefficients, canonical loadings, and cross canonical loadings are used [17]. The standardized coefficients indicate the effect of the original variables of the related variable set on the occurrence of the canonical variable in a variable set [18]. The canonical loadings show the simple linear correlation coefficient between the original variable and its canonical variable. It could be stated as $\mathrm{Cor}(U_1, X)$ and $\mathrm{Cor}(V_1, Y)$, and allows to determine the original variables which have the biggest effect on the canonical variables, and consequently, on the canonical correlation coefficient [19]. The cross canonical loadings are the correlation coefficients between the independent variables and dependent canonical variables, and between the dependent variables and independent canonical variables; they could be stated as $\mathrm{Cor}(V_1, X)$ and $\mathrm{Cor}(U_1, Y)$. If these coefficients are high, it indicates the importance of the effect of the related variables on defining the canonical variable [14]. The strength of the relationship is determined based on the absolute value of these values. The proportion of variance explained and the redundancy index have an important effect on defining the results of the canonical correlation. The proportion of variance explained shows the part of the related dependent or independent variables, explained by the canonical variables in the dependent or independent variable set. The average of squared canonical loadings of each canonical variable in the related set gives the proportion of variance explained [17]. The redundancy indexes found by using the redundancy analysis show the part of the cross set, explained by the canonical variables. It indicates to which extent a variable set defines the variance of the other set [11].

5. Findings

COVID-19 set, which is the dependent variable set in this research, includes 10 variables, and the mobility trends, which stand for the independent variable set, including 6 variables. Thus, 6 canonical correlation coefficients will be calculated. Firstly, statistically significant canonical correlation coefficients should be determined in the analysis. The canonical correlation coefficients and significance tests are given in Table 2.

When Table 2 is examined, it is seen that the first four canonical correlation coefficients are statistically significant. However, only the function with the highest eigenvalue will be taken into consideration in the evaluation. There is a 96% strong linear relationship between the COVID-19 set, which is the dependent variable set, and the mobility trends, which stand for the independent variable set. The mobility trends, the independent variable set, define the dependent COVID-19 set at the rate of 92% (0.9612^2) (according to the first explained variance).

Table 2. The canonical correlation coefficients and significance tests

	Correlation	**Eigenvalue**	**Wilks statistics**	**F**	**Num. D.F**	**Denom. D.F**	**Sig.**
1	,961	12,016	,008	14,336	60,000	570,901	,000
2	,848	2,555	,106	7,122	45,000	490,686	,000
3	,679	,856	,375	3,875	32,000	407,255	,000
4	,507	,345	,696	2,042	21,000	319,282	,005
5	,228	,055	,937	,620	12,000	224,000	,824
6	,108	,012	,988	,269	5,000	113,000	,929

In the canonical correlation analysis, each variable set should be examined separately for determining the importance level of the variables in the dependent and independent variable sets. To this end, Table 3 and Table 4 were prepared.

Table 3. The standardized canonical coefficients of the variables in COVID-19 set

Variables	**1**	**2**	**3**	**4**
Total number of tests	**,083**	-,800	14,077	7,114
Total number of patients	**1,140**	5,645	1,490	30,198
Total number of deaths	**3,005**	1,357	-25,287	-13,321
Pneumonia rate in patients	**,205**	-,427	-1,091	-2,483
Number of critically ill patients	**-1,335**	2,745	3,218	,311
Total number of recoveries	**-2,373**	-8,950	10,099	-20,990
Number of cases reported today	**-,050**	,060	,296	,504
Number of tests conducted today	**,393**	,670	-3,745	-3,852
Number of deaths reported today	**,151**	-,080	-,800	-3,601
Number of recoveries today	**,079**	-,563	-,504	,962

When Table 3 is examined, the equation regarding the first canonical variable of the dependent variable set is as follows:

$V_1 = 0.083 *$ total number of tests + 1.140 $*$ total number of patients + 3.005 $*$ total number of deaths + 0.205 $*$ pneumonia rate in patients – 1.335 $*$ number of critically ill patients – 2.373 $*$ total number of recoveries – 0.050 $*$ number of cases reported today + 0.393 $*$ number of tests conducted today + 0.151 $*$ number of deaths reported today + 0.079 $*$ number of recoveries today.

It is seen that the first variable having the biggest effect on the first canonical variable is “total number of deaths”, the second variable is “total number of recoveries” and the third variable is “number of critically ill patients”.

Table 4. The standardized canonical coefficients of the variables in mobility trends set

Variables	1	2	3	4
Retail and recreation	**-,469**	,424	1,830	-,552
Grocery and pharmacy	**-,065**	,135	-1,078	1,191
Parks	**-,084**	,095	-2,341	-1,166
Transit stations	**,726**	-1,483	,195	-,343
Work places	**,327**	,193	-,525	-1,259
Residential	**,940**	-,061	-,820	-1,609

When Table 4 is examined, the equation regarding the first canonical variable of the independent variable set is as follows:

$U_1 = -0.469 *$ retail and recreation – 0.065 $*$ grocery and pharmacy – 0.084 $*$ parks + 0.726 $*$ transit stations + 0.327 $*$ work places + 0.940 $*$ residential.

It is seen that the first variable having the biggest effect on the first canonical variable is “residential”, the second variable is “transit stations” and the third variable is “retail and recreation”.

The canonical loadings should be calculated for the evaluation of the

canonical variables. The canonical loadings of the dependent variable set are given in Table 5.

Table 5. The canonical loadings of the dependent variable set

Variables	1	2	3	4
Total number of tests	**,939**	,297	,115	,063
Total number of patients	**,906**	,386	,079	,096
Total number of deaths	**,919**	,339	,054	,128
Pneumonia rate in patients	**-,905**	-,335	-,113	-,070
Number of critically ill patients	**,740**	,625	,074	,147
Total number of recoveries	**,914**	,358	,089	,092
Number of cases reported today	**,586**	,684	-,031	,190
Number of tests conducted today	**,838**	,428	,156	-,221
Number of deaths reported today	**,775**	,567	,073	,038
Number of recoveries today	**,674**	,604	-,042	,251

When Table 5 is examined, it is seen that the variable which determines COVID-19 dependent set the most, is “total number of tests” and the following is “total number of deaths”. However, when the coefficients are examined, it is seen that “total number of recoveries”, “total number of patients”, and “pneumonia rate in patients” variables have a considerably high determination rate, too.

The canonical loadings of the independent variable set are given in Table 6.

Table 6. The canonical loadings of the independent variable set

Variables	1	2	3	4
Retail and recreation	**-,520**	-,543	,359	,039
Grocery and pharmacy	**,222**	-,487	,156	,440
Parks	**-,861**	-,396	-,251	-,139
Transit stations	**-,179**	-,939	,123	-,052
Work places	**,285**	-,163	,174	-,633
Residential	**,781**	,441	,012	,298

When Table 6 is examined, it is seen that the variable which determines mobility trends independent set the most is “parks” and the following is “residential”.

After this phase, the cross loadings have been created and evaluated.

It is seen in Table 7 that the dependent original variable having the strongest linear relationship with the first independent canonical variable is "total number of tests".

Table 7. The cross loadings between the independent canonical variable and the dependent original variable

Variables	1	2	3	4
Total number of tests	**,902**	,251	,078	,032
Total number of patients	**,870**	,327	,054	,049
Total number of deaths	**,883**	,288	,037	,065
Pneumonia rate in patients	**-,869**	-,284	-,077	-,036
Number of critically ill patients	**,711**	,529	,050	,075
Total number of recoveries	**,878**	,303	,061	,046
Number of cases reported today	**,563**	,580	-,021	,096
Number of tests conducted today	**,805**	,362	,106	-,112
Number of deaths reported today	**,745**	,481	,049	,019
Number of recoveries today	**,648**	,512	-,028	,127

Table 8. The cross loadings between the dependent canonical variable and the independent original variable

Variables	1	2	3	4
Retail and recreation	-,499	-,460	,244	,020
Grocery and pharmacy	,213	-,412	,106	,223
Parks	-,827	-,336	-,171	-,070
Transit stations	-,172	-,796	,083	-,026
Work places	,274	-,138	,118	-,321
Residential	,751	,374	,008	,151

It is seen in Table 8 that the independent original variable having the strongest linear relationship with the first dependent canonical variable is "parks".

Table 9. Proportion of variance explained and the redundancy indices

Canonical variables	1. Proportion of variance explained	1. Redundancy index	2.Proportion of variance explained	2. Redundancy index
1	**,297**	**,274**	**,685**	**,632**
2	,298	,214	,232	,167
3	,044	,020	,008	,004
4	,118	,030	,021	,005

For the proportion of variance explained for the independent set by the canonical variable obtained from the movement trends variable, the independent set is 29.7%. The proportion of variance explained for the dependent set by the canonical variable obtained from the dependent COVID-19 set is 68.5%.

The part explained for the movement trends, the independent set, by the dependent canonical variable is calculated as 27.4% and the part explained for the dependent COVID-19 set, by the independent canonical variable, is calculated as 63.2%.

6. Conclusion

During the COVID-19 pandemic which the whole world is still in a fight with, scientists mainly focus on social distancing, the use of masks, and hygiene guidelines. However, many researches indicate that the transmission rates are considerably high in environments where a significant number of people are present together, no matter how much strict attention they pay to the mentioned measures. Given the above information, the canonical correlation analysis has been conducted to find out if there is a statistical relationship between COVID-19 data published daily and the change in densities in different locations, which means the mobility trends. Upon the analysis conducted, 6 canonical variable sets have been obtained and four of them have been considered statistically significant. Only the function with the highest eigenvalue has been evaluated; however, the results of all these four functions are shown in the tables. Upon the analysis conducted, it is seen that the variable which determines COVID-19 set the most is total number of tests and the following is total number of deaths. It is seen that the variable which determines mobility trends set the most is parks and the following is residential. When cross loadings are examined, it is seen that the dependent original variable having the strongest linear relationship with the first independent canonical variable is total number of tests. It is also seen that the independent original variable having the strongest linear relationship with the first dependent canonical variable is parks. When the redundancy index is examined, the part explained for the COVID-19 set by the mobility

trends canonical variable is calculated as 63.2%. In conclusion, as the presence in retail and recreation, parks, and transit stations decreases, the pneumonia rate among the patients also decreases. As mentioned above, these areas are where people are mostly exposed to one another and the exposure period is the longest. Thus, as the presence in these areas decreases, the pneumonia rate in patients, which is the most important effect of COVID-19 on humans, will also decrease; and this will be a natural result. As the presence in groceries and pharmacies, workplaces, and residences increases, the pneumonia rate in patients decreases; however, total number of tests, total number of patients, total number of deaths, number of critically ill patients, total number of recoveries, number of patients reported today, number of tests conducted today, number of tests reported today, number of deaths reported today, and number of recoveries reported today continue to increase.

References

[1] The Republic of Turkey Ministry of Health, 2020, Genel Koronavirüs Tablosu, available online at: https://covid19.saglik.gov.tr/TR-66935/genel-koronavirus-tablosu.html.

[2] World Health Organization, 2020, Coronavirus disease (COVID-19): how is it transmitted? available online at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19-how-is-it-transmitted.

[3] J. Saha, B. Barman and P. Chouhan, Lockdown for COVID-19 and its impact on community mobility in India: an analysis of the COVID-19 community mobility reports, 2020, Child. Youth Serv. Rev. 116 (2020), 105160. doi:10.1016/j.childyouth.2020.105160.

[4] Z. Vokó and J. G. Pitter, The effect of social distance measures on COVID-19 epidemics in Europe: an interrupted time series analysis, GeroScience 42 (2020), 1075-1082.

[5] L. Russo et al., Tracing day-zero and forecasting the COVID19 outbreak in Lombardy, Italy: a compartmental modelling and numerical optimization approach, PLoS ONE 15(10) (2020).

[6] B. H. Day, The value of greenspace under pandemic lockdown, Environmental and Resource Economics 76 (2020), 1161-1185.

[7] V. E. Valenti, et al., Social distancing measures could have reduced estimated deaths related to COVID-19 in Brazil, Revista Brasileira de Crescimento e Desenvolvimento Humano 30(2) (2020), 164-169.

[8] V. Chernozhukov, H. Kasahara and P. Schrimpf, Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S., Journal of Econometrics 220(1) (2020), 23-62.

[9] T. L. DucHuynh, Does culture matter social distancing under the COVID-19 pandemic? Safety Science 130 (2020).

[10] Google, Google, 2020.
Available online at: https://www.google.com/covid19/mobility/.

[11] Y. Karagöz, SPSS ve AMOS 23 Uygulamalı İstatistiksel Analizler, Ankara: Nobel Yayınevi, 2016 (in Turkish).

[12] D. Altaş and S. Giray, Türkiye'nin Eğitim, Sağlık ve İktisadi Göstergeleri Arasındaki Etkileşimin Kanonik Korelasyon Analizi ile İncelenmesi, Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 2005, 45-66 (in Turkish).

[13] R. Khattree and D. D. Naik, Multivariate Data Reduction and Discrimination with SAS Software, Wiley, 2000.

[14] Ö. Yorulmaz, Dayanıklı İstatistiksel Yöntemler ve R Uygulamaları. İstanbul: Beta, 2016 (in Turkish).

[15] S. Yavuz and T. Karabulut, The analysis of the relationship between successive courses via canonical correlation analysis method: a sample from Department of Business Administration, Kafkas University Journal of the Institute of Social Sciences 18 (2016), 459-476.

[16] E. Özgür Güler, H. Güler and S. Şanlı, Sosyal Gelişme Endeksine Dayalı Temel İnsani İhtiyaçlar ve Fırsatlar Altendeksleri Arasındaki İlişki, Ç.Ü. Sosyal Bilimler Enstitüsü Dergisi 2 (2019), 55-70 (in Turkish).

[17] N. Yüceday Erdinç and G. Sayılgan, BİST İmalat Sektöründeki Küçük Boyuttaki ve Orta Ölçekteki İşletmelerin Sermaye Yapisina Etki Eden Faktörlerin Kanonik Korelasyon Analizi İle Karşılaştırmalı Olarak İncelenmesi, Journal of Business Research-Turk 12 (2020), 826-847 (in Turkish).

[18] A. Altın Yavuz and B. Ergül, Sağlam Kanonik Korelasyon Analizi İle İnsani Gelişmişlik ve Dijital Rekabetçilik Endeksleri Arasindaki İlişkinin Değerlendirilmesi, Sosyal ve Beşeri Bilimler Dergisi 12 (2020), 1-18 (in Turkish).

[19] Ç. Arıcıgil and M. Can, Banka Şubelerinin Performanslarını Etkileyen Faktörlerin Kanonik Korelasyon Analizi İle İncelenmesi, Dumlupınar Üniversitesi Sosyal Bilimler Dergisi EYİ 2013 Özel Sayısı, 2013, pp. 285-296 (in Turkish).