Research article

Statistical characterization of vaccinated cases and deaths due to COVID-19: methodology and case study in South America

Carlos Martin-Barreiro[1,*], **Xavier Cabezas**[1,2], **Víctor Leiva**[3], **Pedro Ramos-De Santis**[1,*], **John A. Ramirez-Figueroa**[1,2] **and Erwin J. Delgado**[1]

[1] Facultad de Ciencias Naturales y Matemáticas, Escuela Superior Politécnica del Litoral ESPOL, Guayaquil 090902, Ecuador

[2] Centro de Estudios e Investigaciones Estadísticas, Escuela Superior Politécnica del Litoral ESPOL, Guayaquil 090902, Ecuador

[3] Escuela de Ingeniería Industrial, Pontificia Universidad Católica de Valparaíso, Valparaíso 2362807, Chile

* **Correspondence:** Email: cmmartin@espol.edu.ec; pramos@espol.edu.ec.

Abstract: Many studies have been performed in different regions of the world as a result of the COVID-19 pandemic. In this work, we perform a statistical study related to the number of vaccinated cases and the number of deaths due to COVID-19 in ten South American countries. Our objective is to group countries according to the aforementioned variables. Once the groups of countries are built, they are characterized based on common properties of countries in the same group and differences between countries that are in different groups. Countries are grouped using principal component analysis and K-means analysis. These methods are combined in a single procedure that we propose for the classification of the countries. Regarding both variables, the countries were classified into three groups. Political decisions, availability of resources, bargaining power with suppliers and health infrastructure among others are some of the factors that can affect both the vaccination process and the timely care of infected people to avoid death. In general, the countries acted in a timely manner in relation to the vaccination of their citizens with the exception of two countries. Regarding the number of deaths, all countries reached peaks at some point in the study period.

1. Introduction

The pandemic caused by COVID-19 affected the world in a significant way, not only in terms of people's health but also from an economic/financial perspective [1–3]. Aspects related to marketing and the social responsibility of organizations were also studied [4, 5]. Much research has been carried out in which behavioral, environmental, psychological, and social issues have been discussed [6–9]. Hospitals have had to change the management of their inventories and models have been proposed in the literature to avoid shortages and supply medicines to patients on time [10].

From the perspective of the dynamics of the phenomenon generated by SARS-Cov2, compartmental models are a strategy widely used to analyze the evolution of an epidemic [11]. Individuals in the population under study are divided into compartments according to their characteristics. This type of models is utilized to predict the spread of the epidemic under different scenarios as well as the introduction of large-scale vaccination [12]. The governments and most important pharmaceutical companies in the world started to work quickly to find a vaccine from the beginning of the pandemic which was available at the end of 2020. However, in South America the vaccination process began in the first months of 2021. Many researchers have worked on issues related to COVID-19 vaccination. Among others, in [13, 14] it was discussed the impact of vaccination in containing the COVID-19 epidemic. In [15], the spread of infections in Italy was analyzed in the midst of vaccination and appearance of new variants.

In [16], through multivariate regression, the relationship between the different macro-economic factors of fully vaccinated health and care personnel [17] between February and June 2021 was investigated. In [18], it were identified post-vaccination risk factors for COVID-19 infection using univariate and multivariate logistic regression of data collected in the United Kingdom between March 2020 and July 2021. In [19], it was used logistic regression to analyze individuals at risk who are reluctant to be vaccinated against COVID-19 utilizing data collected in the last quarter of 2021 in Germany. In [20], it was studied the most important statistical characteristics of populations in two regions relative to total COVID-19 immunization using the maximum likelihood estimation of the parameters of a probability model. In [21], a cluster analysis was carried out using the K-means algorithm on data of the proportion of daily residents in the home, the trips made daily and the dose of vaccines per capita in the 50 US states. In that work, a multivariate regression analysis (fixed effects model) was also performed from panel data on temporally segmented observations.

In [22], it was presented a longitudinal study using multivariate logistic regression on hesitancy to get vaccinated, social norms and acceptance of the vaccine in the US, a country with a high degree of access to inoculation against COVID-19. In [23], it was applied a canonical correlation analysis to data obtained from a cohort of individuals which includes measures of physical and mental wellness of children and their parents as well as demographic and socioeconomic data. In [24], a cluster analysis was conducted to identify patterns of behavior in vaccine data in Brazilian states. In [25], probabilistic vaccine projections about the spread of SARS-CoV-2 infections were established. In [26], the authors identified the key issues associated with vaccinations in the presence of misinformation in rural areas of developing countries. In [27], the barriers to vaccination faced by socially vulnerable groups were analyzed using univariate and multivariate multilevel logistic regression in the Ile-de-France region and in Marseille with data collected between November and December 2021.

In [28], the reported benefits of inoculation on the COVID-19 mortality rate were evaluated by stepwise linear regression isolating the independent effects of treatment and associated comorbidities separating out bias and uncovering beneficial factors. In [29], a multidimensional approach using logistic regression and linear regression was used to identify relationships between the demographic aspects of participants and their knowledge, attitudes and practices. In [30], a multivariate model was utilized to study the association between the variation of vital parameters with lunar cycles in patients with COVID-19 hospitalized in Oklahoma, US, between February 2020 and August 2021. In [31, 32], mathematical models were proposed to optimize the vaccination process. In [33–37], a multivariate analysis in research related to SARS-COV2 was utilized. In [38], the use of K-means was applied to complement a component analysis that was carried out to classify countries according to the number of infected people. Further literature covering related topics can be found in Table 1.

Table 1. Summary of additional literature on the topic for the indicated year and author(s).

Year	Author(s)	Summary points
2021	Coccia, [39]	The study's findings are important for developing effective strategies to prevent future pandemics. The study's recommendations can help governments and other organizations to better prepare for and respond to future pandemics.
2021	Chan, et al., [40]	The Delta variant of COVID-19 became predominant in July 2021, which rises coincided with surges in new cases. The number of deaths varied based on regional vaccination status. In countries with high vaccination rates, the number of deaths was less than in countries with low vaccination rates.
2022	Magazzino, et al., [41]	An algorithm based on neural networks is used to evaluate the incidence of vaccination on the death rate in the COVID-19 pandemic. The results obtained suggest that there is a breaking point in the growth of the number of deaths and it corresponds to the beginning of the vaccination of the population.
2022	Coccia, [42]	This study analyzes what is the adequate percentage of the population to be vaccinated to keep the COVID-19 epidemic under control. It is concluded that on average the percentage should be around 80%. Furthermore, if vaccination is carried out in the early stages of the epidemic, this percentage can be much lower.
2022	Benati, et al., [43]	A global analysis of the relationship between public policies implemented and timely vaccination is performed in the first months of 2021. This analysis will help identify factors and implement better public health strategies.
2022	Coccia, [44]	A comparative analysis of the development of the COVID-19 pandemic in Italy is conducted between two time periods. The first one, at the beginning of the pandemic (April-September 2020), characterized by strong control measures, and the second one (April-September 2021) characterized by massive vaccination campaigns.
2022	Fiori, et al., [45]	This study suggests that the combination of vaccination and natural infection helped to achieve conditional herd immunity in South America and that the containment of regional variants may be due to several factors.

Continued on next page

Year	Authors	Summary points
2022	Coccia, [46]	This investigation advises that by prioritizing good governance, new technology, and health infrastructure investments, countries can better prepare for and respond to future pandemics. This will help reduce future pandemics' impact.
2022	Musa, et al., [47]	This research used a SEIR-based model to assess the COVID-19 pandemic in South America. Twelve countries' mortality rates and transmission dynamics were analyzed. The findings highlighted the importance of increasing vaccination rates and implementing social distancing measures.
2022	Coccia, [48]	This analysis found that vaccination campaigns alone are not enough to reduce the impact of COVID-19. Factors like new variants and socioeconomic conditions influence its spread. A comprehensive approach involving good governance, health investments, and new technology is needed for effective crisis management.
2022	Coccia, [49]	Vaccination was essential during pandemics, but only up to 70% of people were willing to be vaccinated without coercion. Exceeding this percentage can have negative socioeconomic and democratic impacts. Effective communication strategies and trust-building are recommended to increase vaccination rates.
2022	Oyewola, et al., [50]	This article indicates that vaccines were essential for stopping the pandemic and saving lives. COVID-19 vaccine acceptability can be predicted using machine learning. Machine learning can be used to optimize the daily immunization of citizens across the globe.
2023	Lucero-Prisno, et al., [51]	The incidence of vaccination on the achievement of herd immunity in South America is studied over time. Currently available vaccines do not guarantee that herd immunity is maintained in the population.
2023	Coccia, [52]	The article shows that countries with a higher utilization of assisted ventilation devices had lower death rates compared to those with limited usage. Consequently, the author suggests that the availability of an adequate number of ventilators will play a crucial role in mitigating the impact of future respiratory epidemics.
2023	Jen, et al., [53]	In this article, three hypotheses are presented and validated. The first one is that the number of vaccinated people is related to the gross domestic product of each country. Second, vaccines can reduce the death rate, and third, dashboards present more helpful information than classic statistical graphs.
2023	Torales, et al., [54]	It analyzes both the physical and psychological sequelae of the COVID-19 pandemic, also known as long-COVID syndrome. The study found average levels of fatigue and significant levels of anxiety among those who participated in the survey.
2023	Coccia, [55]	This work has two primary objectives. The first one is to determine the conditions and factors that trigger a pandemic, and the second one is to show the advantages and disadvantages of the models used to monitor the evolution of a pandemic.

Continued on next page

Year	Authors	Summary points
2023	Torres, et al., [56]	The dynamics of the third and fourth waves of variants of the Omicron strain of CVID-19 in ARG are studied, based on phylogenetic and phylodynamic sequencing analyses, concluding that the different propagation dynamics would be due to various factors such as decreased immunity and increased population reinforced with vaccination or immunity to previous strains.
2023	Zhao, et al., [57]	This article investigates the relationship between the acceptance of the COVID-19 vaccine by people with chronic diseases and the factors that correlate with their disagreement with vaccination.
2023	Zambrana, et al., [58]	The study's findings are important for understanding the evolution of SARS-CoV-2 and its impact on the COVID-19 pandemic. The study highlights the importance of monitoring virus mutations and implementing effective containment measures to control the spread of the virus.

The objective in this work is to analyze data on vaccinated people and deaths due to COVID-19 during the years 2021 and 2022 in ten South American countries: Argentina (ARG), Bolivia (BOL), Brazil (BRA), Chile (CHI), Colombia (COL), Ecuador (ECU), Peru (PER), Paraguay (PRY), Uruguay (URY) and Venezuela (VEN). As a result of the analysis, the countries were classified into groups and these groups were characterized. The statistical methods employed are component analysis and the K-means method. All computational experiments were conducted using the R software [59].

To show the content in a proper way, the remainder of this document has been distributed as follows. In Section 2, we introduce the materials and methods utilized in this work. The results of our study are presented in Section 3. In particular, the principal component analysis (PCA) can be seen in Subsection 3.1; in Subsection 3.2, a sparse technique known as disjoint principal component analysis (DPCA) is used, which facilitates the grouping and characterization of the countries; and in Subsection 3.3, we present the K-means analysis, which is employed to complement the component analysis. In Section 4, we summarize all the results obtained from this statistical study. Possible future work and some motivations are also discussed in this final section.

2. Materials and methods

To carry out the statistical analysis in this research, we have downloaded the data from the website `https://ourworldindata.org/coronavirus` (accessed on 13 June 2023). The period of this study is two years, from January 2021 to December 2022. The data is divided into two parts: number of vaccinated cases with full doses and number of deaths due to COVID-19. Table 2 shows the vaccinated data matrix. In Table 3, we can see the death data matrix. For each matrix, in the columns we have the countries and in the rows we have the months. In each matrix entry we have the number of inhabitants per million. We have included plots corresponding to Tables 2 and 3 in logarithmic scale in Figure 1.
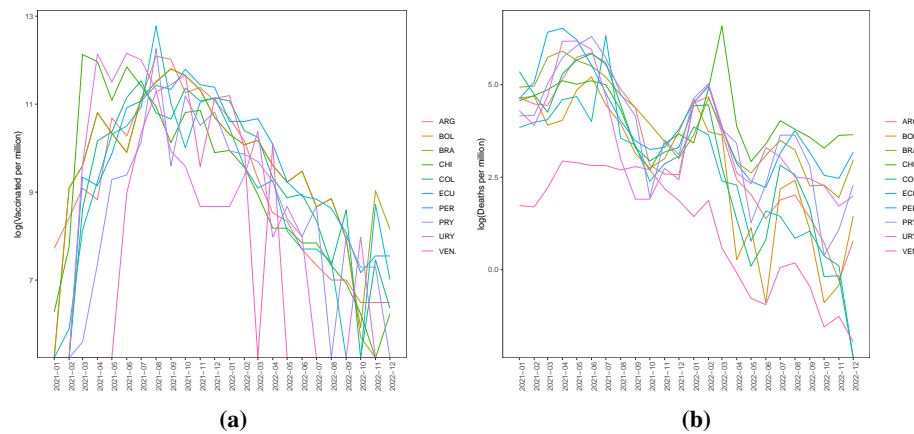
Figure 1. Plots of the number of COVID-19 (a) vaccinated cases and (b) deaths by country in log scale.

Table 2. Data matrix of the number of vaccinated COVID-19 cases per million inhabitants for the indicated month and country.

Month	ARG	BOL	BRA	CHI	COL	ECU	PER	PRY	URY	VEN
2021-01	2253.6	0.0	0.0	531.3	0.0	0.0	0.0	0.0	0.0	0.0
2021-02	4514.7	8900.0	8917.4	2345.2	0.0	363.5	0.0	0.0	0.0	0.0
2021-03	8818.9	14800.0	14723.0	185388.8	5515.1	3410.4	11400.0	268.6	15108.5	0.0
2021-04	6825.4	49200.0	49277.8	157142.9	25909.6	10261.3	9400.0	1579.6	187062.6	0.0
2021-05	43506.9	30800.0	30839.3	65306.1	32003.1	30111.8	19400.0	10782.4	98998.5	0.0
2021-06	29004.6	20000.0	19971.2	139795.9	69597.1	36408.6	55300.0	11970.2	190058.5	7910.2
2021-07	67018.2	69100.0	67251.9	90816.3	101792.9	56111.1	64300.0	25401.0	163742.7	30959.1
2021-08	178202.6	98400.0	100227.6	56632.7	48968.6	353333.3	91700.0	211060.0	81871.3	79151.9
2021-09	165897.6	133400.0	133435.5	25000.0	42606.5	61111.1	83800.0	14749.3	20467.8	93286.2
2021-10	77785.1	114700.0	114671.9	49489.8	86177.0	22222.2	132400.0	72271.4	14619.9	116254.4
2021-11	87453.3	82600.0	82625.1	52040.8	64391.7	62777.8	93100.0	36873.2	5848.0	14487.6
2021-12	64381.5	44300.0	44261.8	19898.0	69018.7	68888.9	87900.0	50147.5	5848.0	68197.9
2022-01	43506.9	30100.0	30142.6	20918.4	64584.5	34444.4	40300.0	20649.0	5848.0	72791.5
2022-02	25928.4	23800.0	23733.2	14285.7	32774.2	14444.4	40500.0	19174.0	11695.9	21908.1
2022-03	10986.6	26300.0	26334.1	7653.1	28147.3	8888.9	43000.0	16224.2	32163.7	0.0
2022-04	5053.8	14900.0	14908.7	3571.4	11181.8	10555.6	24100.0	10324.5	2924.0	24028.3
2022-05	4174.9	10200.0	10124.9	3571.4	7133.2	3333.3	10200.0	4424.8	5848.0	0.0
2022-06	2197.3	13000.0	13097.4	2551.0	7711.6	2222.2	7400.0	2949.9	2924.0	0.0
2022-07	1538.1	5800.0	5805.6	2551.0	4241.4	2222.2	6900.0	5899.7	0.0	0.0
2022-08	1098.7	7000.0	6966.7	1530.6	1542.3	1666.7	5600.0	0.0	0.0	0.0
2022-09	1098.7	2800.0	2740.2	1020.4	5398.1	0.0	3200.0	2949.9	0.0	0.0
2022-10	659.2	300.0	371.6	510.2	0.0	0.0	1300.0	1474.9	2924.0	0.0
2022-11	659.2	0.0	8360.0	0.0	1735.1	6111.1	1900.0	1474.9	0.0	0.0
2022-12	659.2	0.0	3483.3	510.2	578.4	1111.1	1900.0	0.0	0.0	0.0

Table 3. Data matrix of the number of COVID-19 deaths per million inhabitants for the indicated month and country.

Month	ARG	BOL	BRA	CHI	COL	ECU	PER	PRY	URY	VEN
2021-01	104.633	103.999	139.043	96.464	211.664	46.757	101.482	63.931	73.409	5.663
2021-02	88.304	108.795	143.198	110.903	113.657	53.959	159.924	64.912	49.515	5.453
2021-03	84.904	49.755	310.989	129.337	70.484	57.556	614.214	151.180	108.480	9.117
2021-04	174.116	56.792	369.379	165.612	198.535	99.111	683.847	309.145	480.117	18.869
2021-05	312.635	128.723	285.983	150.357	290.226	107.833	501.674	425.369	485.380	18.021
2021-06	350.275	183.552	246.993	165.561	342.587	54.889	251.043	547.050	385.088	16.714
2021-07	253.834	86.252	182.541	148.112	267.746	559.500	118.120	307.670	108.480	16.678
2021-08	132.301	51.637	112.192	75.969	85.811	34.889	56.094	115.929	19.883	14.806
2021-09	78.488	23.159	79.648	27.092	27.299	29.167	32.423	63.569	6.725	16.219
2021-10	16.941	15.548	51.321	14.745	18.932	10.889	25.815	7.080	6.725	14.912
2021-11	13.272	20.131	32.061	30.051	24.041	17.556	27.313	33.333	15.497	8.940
2021-12	13.008	41.653	20.199	39.235	27.260	21.389	44.464	22.419	11.404	6.502
2022-01	90.683	101.391	38.340	30.918	84.018	47.333	92.335	102.802	90.351	4.205
2022-02	107.207	41.489	103.795	134.286	86.119	38.333	142.085	153.097	147.953	6.502
2022-03	41.024	38.298	45.664	725.867	15.963	11.000	45.521	44.985	50.000	1.802
2022-04	11.492	1.309	18.425	48.316	3.894	9.833	17.680	30.383	13.743	0.919
2022-05	7.625	3.110	13.655	18.622	1.099	2.167	10.896	3.540	10.234	0.459
2022-06	3.977	0.409	22.015	29.949	2.236	4.889	9.222	10.177	27.193	0.389
2022-07	6.570	8.838	32.966	56.020	16.869	4.222	23.319	37.906	21.053	1.060
2022-08	7.515	11.211	25.317	44.592	12.994	2.333	42.849	38.053	12.281	1.201
2022-09	4.087	2.946	9.619	35.867	5.302	2.833	23.877	16.667	11.696	0.636
2022-10	2.065	0.409	9.851	26.684	0.829	1.444	12.834	1.475	9.649	0.212
2022-11	0.747	0.655	7.004	37.755	0.848	1.111	11.806	2.950	5.556	0.283
2022-12	2.175	4.255	19.451	38.520	0.000	0.000	24.023	9.882	7.310	0.141

2.1. Measures of variables

The two data matrices shown in the previous section are of order 24×10 (24 months and 10 countries) whose entries are non-negative real numbers. For the dimensional reduction that was carried out to classify the countries into groups we have used the PCA and the K-means methods. With both methods, we have utilized the countries as variables and the months as entities. In the case of PCA, the components define the groups of countries. In the case of K-means, the clusters define the groups of countries. In the next section, we present the procedure that was employed both with the vaccinated matrix and with the death matrix.

2.2. Models and data analysis procedure

Next, we propose a procedure that allows us to classify the countries of South America into groups. Our procedure is based on the PCA and K-means methods. We can see this procedure in Algorithm 1, which is used for analyzing COVID-19 data of the number of vaccinated cases and the number of deaths. We want to highlight step 9 in this 10-step algorithm. In step 9, a comparison of the country grouping obtained by the PCA method and by the K-means method is performed.

A detailed background of the PCA and the component rotation methods can be seen in [60]. In particular, the calculation of rotated components with the VARIMAX method may be found in [61]. In [62] we can see how to calculate disjoint components. DPCA is a recent sparse technique that facilitates the interpretation of the components. During the last years some researchers have proposed algorithms to calculate disjoint components. In [38, 63], we observe a statistical study where DPCA was used. For full details of the DPCA method, see [62]. The DPCA method has even extended for three-way matrices as we can see in [64, 65]. A background of the K-means method is presented in [66]. In [38, 67] we find applications of the K-means method to study COVID-19 data.

Algorithm 1 Proposed procedure for the statistical study of the number of vaccinated cases and number of deaths due to COVID-19 in South America.

1: Collect the data in a matrix X of order $p \times q$ where p is the number of months and q is the number of countries.
2: Pre-process X applying centering and scaling to the data.
3: Apply a PCA and determine the number c of components to use for data analysis ($c < q$).
4: Compute c components from X and fit the model. Go to Step 7 if the components are interpretable.
5: Calculate c rotated components (for example, with the VARIMAX method [61]). Go to Step 7 if the rotated components are interpretable.
6: Obtain c disjoint components [62] and fit the model.
7: Build the country groups and interpret the latent dimensions.
8: Apply a K-means analysis to the matrix X using c clusters.
9: Carry out a comparative analysis between the groups of countries that were generated using components and using clusters.
10: Generate tables, plots and conclude.

---

3. Results and discussion

3.1. Principal component analysis

PCA was used to classify the selected countries. What are the countries that have common characteristics? What are the countries with different characteristics? These research questions are the ones we want to answer for vaccinated people and for deaths due to COVID-19. Next, we discuss the results obtained from the PCA. Figure 2a shows the cumulative proportional variance plot for vaccinated COVID-19 cases. Figure 2b displays the same type of plot but for COVID-19 deaths. Regarding the number of components to be computed, we utilized three components to obtain an explained variability of around 90%. In [38], we can see a similar justification for selecting the number of components. The three components capture 89.91% of the variability of the data in the case of the COVID-19 vaccinated cases and 86.16% for COVID-19 deaths.
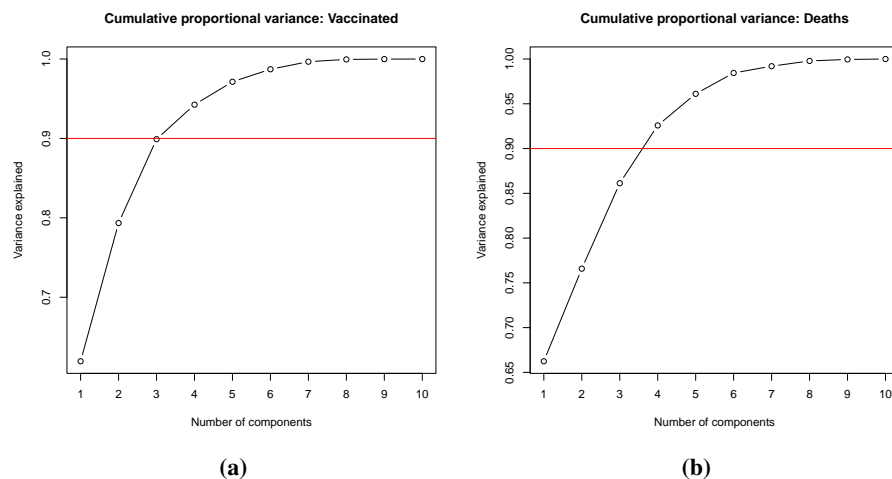
Figure 2. Cumulative variance plots of the number of COVID-19 (a) vaccinated cases and (b) deaths.

Table 4 reports the three principal components calculated (PC1, PC2, and PC3) with the matrix of vaccinated people and Table 5 for COVID-19 deaths. The corresponding columns show the loadings, that is, the correlations between countries and components. Note that there are countries that have similar loadings (in absolute value) for the different components which makes the interpretation difficult (for example with BOL, COL and PER for vaccinated cases, as well as with BRA, PER and URY for deaths). To improve the interpretation, a VARIMAX rotation [61] was performed (the fit of the model remains) but a clear interpretation was not obtained either vaccinated cases or deaths. For this reason, we decide to use a sparse technique known as DPCA to obtain a better interpretation.

Table 4. Loading matrices for data of the number of COVID-19 vaccinated cases with three components for the indicated country and method.

Country	PCA			DPCA		
	PC1	PC2	PC3	DC1	DC2	DC3
ARG	0.376	-0.116	0.095	0.406	0.000	0.000
BOL	0.376	-0.017	-0.203	0.429	0.000	0.000
BRA	0.376	-0.026	-0.193	0.428	0.000	0.000
CHI	0.118	0.650	0.056	0.000	0.000	0.707
COL	0.320	0.212	-0.303	0.364	0.000	0.000
ECU	0.299	-0.089	0.632	0.000	0.707	0.000
PER	0.370	-0.053	-0.242	0.423	0.000	0.000
PRY	0.314	-0.155	0.526	0.000	0.707	0.000
URY	0.132	0.657	0.154	0.000	0.000	0.707
VEN	0.340	-0.226	-0.240	0.396	0.000	0.000

Table 5. Loading matrices for data of the number of COVID-19 deaths with three components for the indicated country and method.

Country	PCA			DPCA		
	PC1	PC2	PC3	DC1	DC2	DC3
ARG	0.372	-0.164	0.079	-0.458	0.000	0.000
BOL	0.324	-0.105	0.150	-0.407	0.000	0.000
BRA	0.353	0.229	-0.174	0.000	-0.588	0.000
CHI	0.101	0.456	0.852	0.000	0.000	1.000
COL	0.364	-0.196	0.066	-0.450	0.000	0.000
ECU	0.209	-0.599	0.278	-0.293	0.000	0.000
PER	0.298	0.449	-0.290	0.000	-0.583	0.000
PRY	0.372	-0.039	0.010	-0.441	0.000	0.000
URY	0.347	0.275	-0.139	0.000	-0.560	0.000
VEN	0.312	-0.161	-0.175	-0.376	0.000	0.000

3.2. Disjoint principal component analysis

As we can see in Algorithm 1 that we propose, the disjoint components are used when the classical components and the rotated components do not allow a clear classification of the countries. Did the PCA method allow us to group countries? The response is no. This the reason we employ DPCA. Now, we report the results obtained after applying the DPCA method.

In the last three columns of Table 4, we can see the computed disjoint components (DC1, DC2, and DC3) for vaccinated COVID-19 cases. The variability captured by the three disjoint components is 87.25% with the matrix of vaccinated cases, that is, 2.66% of explained variability was lost which is not significant but we gain in interpretability. The DPCA loading matrix allows us to group the ten countries as follows: (Group 1) ARG, BOL, BRA, COL, PER and VEN; (Group 2) ECU and PRY; and (Group 3) CHI and URY.

In the case of deaths, in the last three columns of Table 5 we can see the calculated disjoint components. The three disjoint components explain 83.27% of the variability of the data, that is, 2.89% of explained variability has been lost which is not significant and we gain in interpretability. Thus, due to the DPCA method we can group the countries in this way: (Group 1) ARG, BOL, COL, ECU, PRY and VEN; (Group 2) BRA, PER and URY; and (Group 3) CHI.

Figure 3a shows the space of the variables (countries) in the case of number of vaccinated cases due to COVID-19 considering two dimensions. The space of the countries with two dimensions regarding the number of COVID-19 deaths can be seen in Figure 3b. In both plots, the countries are distributed in a plane. However, the countries have been grouped considering the results obtained with the DPCA to facilitate interpretation.

Figure 4a presents the number of vaccinated COVID-19 cases for each month of the analysis period while Figure 4b displays the number of COVID-19 deaths. Note that again the countries have been grouped according to the DPCA results. In this way, we can characterize the three groups of countries for COVID-19 vaccinated cases and deaths.
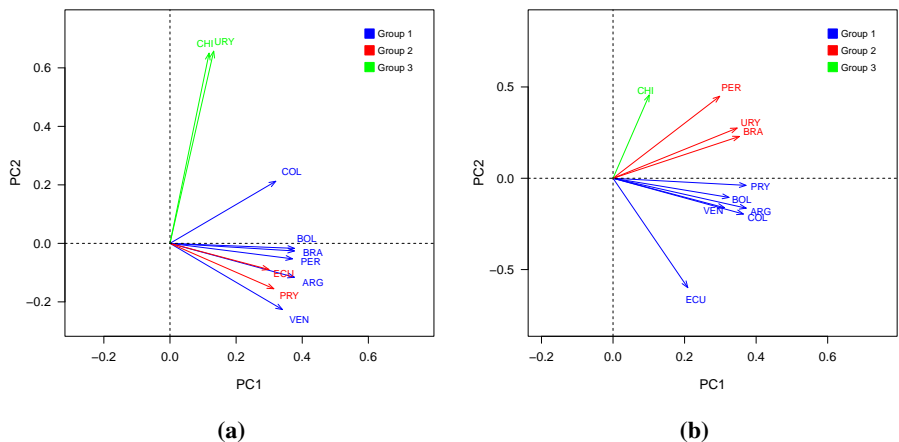
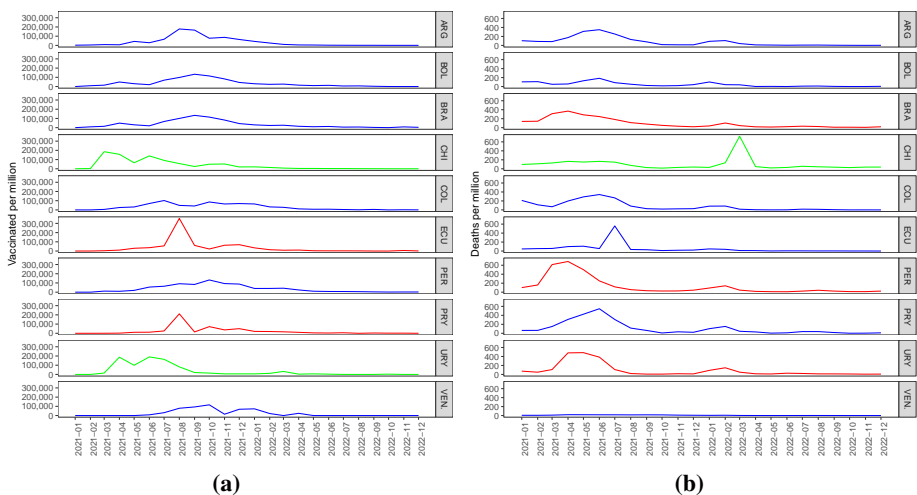**Figure 3.** Space of countries for data of the number of COVID-19 (a) vaccinated cases and (b) deaths.

Figure 4. Plots of the number of COVID-19 (a) vaccinated cases and (b) deaths for the indicated country and month.

3.3. K-means analysis

Was the country grouping obtained with the components correct? This is the research question that we want to answer with the K-means method. Next, we carry out a K-means analysis [38, 66, 67] to compare with the results obtained using components. The computations were made with three clusters for COVID-19 vaccinated cases and deaths to make the corresponding contrast. Nevertheless, note that in Figure 5a the silhouette plot suggests three clusters for vaccinated people but in Figure 5b the silhouette plot suggests only two clusters for COVID-19 deaths.

We now proceed with the analysis of the clusters. On one hand, Figure 6a shows the K-means plot applied to the vaccinated matrix. On the other hand, we can see in Figure 6b the plot of the K-means applied to the matrix of deaths.
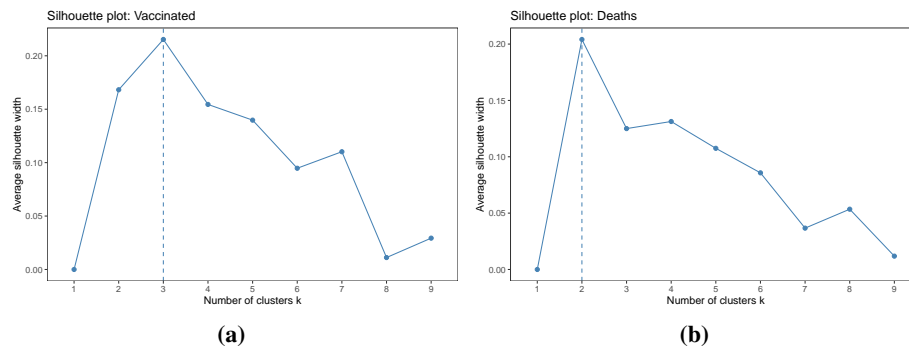
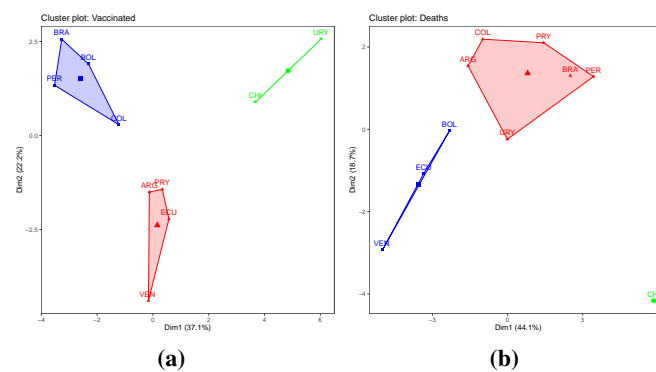**Figure 5.** Silhouette plots for (a) vaccinated cases and (b) deaths.

Figure 6. Cluster plots K-means method with three clusters (a) vaccinated cases and (b) deaths.

The groups obtained with the data set of COVID-19 vaccinated cases are: (Group 1) BOL, BRA, COL and PER; (Group 2) ARG, ECU, PRY and VEN; and (Group 3) CHI and URY. When comparing this grouping with the one obtained using components we notice a single difference: ARG and VEN leave Group 1 and move to Group 2.

Regarding the data set of COVID-19 deaths, the countries are grouped as follows: (Group 1) BOL, ECU and VEN; (Group 2) ARG, BRA, COL, PER, PRY and URY; and (Group 3) CHI. Note that BOL, ECU and VEN stayed in Group 1 but the countries ARG, COL and PRY moved from Group 1 to Group 2.

It is important to have two grouping methods to be able to compare the results obtained by both of them. This is what we considered when designing Algorithm 1. The comparative clustering study between the component analysis and the K-means method gave very similar results but there were a couple of differences that were noted above. To conclude, we chose the grouping that was obtained with the components for two reasons: (i) the components have the advantage of representing latent variables that can be characterized and that maximize the variability of the data (in this way the greatest possible amount of information is captured); and (ii) with the K-means method the number of clusters to be built is known in advance, the variables are grouped solely by their similarity (distances) and the quality of the clusters depends on the initialization of the algorithm. In such a case, the use of the K-means method was important because it complemented and helped to confirm the grouping that was

obtained using components. In addition, in the case of COVID-19 deaths the number of recommended clusters was two.

3.4. Discussion and summary of results

There are various reasons why countries hastened or delayed the start of the vaccination process as budgets, availability of vaccines, public/private health infrastructure and decision-making capacity, among others. The number of deaths is mainly affected by the vaccination process but the transmission rate of the virus due to people's carelessness is also an important factor with our data capturing all of the above. Algorithm 1 that we propose has the ability to group countries according to how the number of vaccinated cases and the number of deaths change over time. Here, we have a discussion of the groups formed both in COVID-19 vaccinated cases and deaths. For vaccinated people the following is observed:

Group 1: ARG, BOL, BRA, COL, PER and VEN. These countries had a prolonged and sustained vaccination process from May 2021 to April 2022, approximately one year. A significant peak was observed between September 2021 and October 2021. We have characterized this group as "countries that had a moderate start in vaccinating their citizens".

Group 2: ECU and PRY. These countries started the vaccination process late. A first segment is observed from July 2021 to September 2021 with a very important peak in August 2021. We can see a second segment from September 2021 to April 2022 with a peak that is not very pronounced in December 2021. We have characterized this group as "countries that took the longest to start vaccinating their citizens".

Group 3: CHI and URY. These countries quickly started the vaccination process which begins in February 2021 and ends approximately in March 2022. Two peaks of a similar size are observed, the first of them between March and April 2021 and the second one in June 2021. We have characterized this group as "countries that react promptly in vaccinating their citizens".

In the ten countries, the vaccination process was mild in the second semester of 2022. Regarding the number of deaths due to COVID-19 we have to conclude the following:

Group 1: ARG, BOL, COL, ECU, PRY and VEN. In this group, VEN is a special case. The component analysis placed VEN in this group but we believe that this country did not report the correct count of deaths due to COVID-19. Leaving VEN aside, the other countries have a pronounced first peak of deaths between June 2021 and July 2021 with a slight second peak between January 2022 and February 2022. We have characterized this group as "countries with a significant number of COVID-19 deaths in the middle of the year 2021".

Group 2: BRA, PER and URY. These countries presented a considerable peak of deaths between April 2021 and May 2021. There is a second milder peak between February 2022 and March 2022. We have characterized this group as "countries with a considerable number of COVID-19 deaths in the fourth month of the year 2021".

Group 3: CHI. In this group there is only CHI. This country presented an extremely considerable peak of deaths in March 2022. We have characterized this group as "countries with a considerable number of COVID-19 deaths in the third month of the year 2022".

The ten South American countries have in common the fact that approximately, in the second semester of 2022, the number of deaths due to COVID-19 decreased considerably. Other important conclusions are the following:

(i) CHI was the country that best managed the pandemic since it was the first country to start the vaccination process very early in February 2021 and only had a peak of deaths in March 2022. Leaving VEN aside, CHI was the country with the smallest number of COVID-19 deaths per million inhabitants.

(ii) ARG, PER and VEN are the countries that took the longest to start the COVID-19 vaccination process.

(iii) URY had an intensive COVID-19 vaccination period in the months of April, May and June 2021. However, in those same months URY had the largest number of deaths due to COVID-19. URY started the vaccination process quickly. Nonetheless, this did not help to control the significant number of deaths in the first half of 2021.

(iv) PER was the country with the largest number of deaths per million inhabitants. This occurred in the months of March, April, and May 2021. In second position is PRY with a very high peak in June 2021.

(v) ECU had a peak of vaccinated cases between July and August 2021, and in those same months it also had the largest number of deaths per million inhabitants.

This comparative study allowed us to obtain the results that we summarize in Table 6 for COVID-19 vaccinated people and the results that we summarize in Table 7 for deaths due to COVID-19.

Table 6. Grouping of the ten South American countries in the case of vaccinated people with the corresponding characterization of each group.

Group	Countries	Characterization
Group 1	ARG, BOL, BRA, COL, PER, VEN	Countries that had a moderate start in vaccinating their citizens
Group 2	ECU, PRY	Countries that took the longest to start vaccinating their citizens
Group 3	CHI, URY	Countries that react promptly in vaccinating their citizens

Table 7. Grouping of the ten South American countries in the case of deaths due to COVID-19 with the corresponding characterization of each group.

Group	Countries	Characterization
Group 1	ARG, BOL, COL, ECU, PRY, VEN	Countries with a significant number of deaths in the middle of the year 2021
Group 2	BRA, PER, URY	Countries with a considerable number of deaths in the fourth month of the year 2021
Group 3	CHI	Countries with a considerable number of deaths in the third month of the year 2022

4. Conclusions

In the present study, we have grouped ten South American countries into three groups. These groups have also been constructed and characterized using two criteria. The first criterion is the number of COVID-19 vaccinated people per million inhabitants and the second one is the number of COVID-19 deaths per million inhabitants. The formed groups permitted us to determine those countries that had a similar behavior as well as those countries that had a different behavior. To carry out the statistical analysis of data related to COVID-19 vaccinated cases and deaths we have utilized two methods: (i) principal component analysis and (ii) K-means analysis. As mentioned, all calculations were performed with the R software. These two methods were combined in a procedure that we have proposed and that was summarized in Algorithm 1. We believe that the procedure we have designed can be used by other researchers to classify entities (countries and companies among others) in their studies.

Regarding the vaccination process, the governments of Chile and Uruguay were the first ones to vaccinate their citizens. Obviously, this showed that they were the most responsible countries in the region. The governments of Ecuador and Paraguay are the ones that took the longest to start vaccination. Regarding the number of COVID-19 deaths, Brazil, Peru and Uruguay had a large number of deaths in April 2021.

In the cases of Ecuador and Paraguay, these countries had a considerable peak of deaths in mid-2021. It should be noted that Chile, despite being the first country in South America to start vaccination showed a significant number of deaths in March 2022. In addition, Venezuela was the country that took the longest to start vaccination. However, this country did not present any peak regarding the number of deaths. Despite the restrictions imposed on citizens by the governments of the ten South American countries analyzed, the indiscipline of the people was a factor that raised the number of deaths, even in countries that started the vaccination process early.

Our comparative study where countries were grouped using one or more variables is very important because it allowed us to evaluate the behavior of governments in the face of certain events. In particular, the COVID-19 pandemic has been the most important health event for the last 3-4 years. For this reason, regarding future work we recommend applying the methodology shown in this article to countries that are in other regions of the planet. For example, a study of the vaccination process and deaths due to COVID-19 in Central and North American countries can be carried out as well as in countries from Europe, Asia or Africa. We could also consider incorporating another statistical methods that complement the use of components and the K-means method. Another important future work is to build an R package that includes the two methods used in this work so that diverse practitioners may employ it and make their analyses easier to them when conducting similar studies.

Use of AI tools declaration

The authors declare they have not used artificial intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors thank to the Editors and Reviewers for their valuable comments that helped to improve the quality of this article.

Conflict of interest

The authors declare there are no conflicts of interest.

References

1. K. Chahuán-Jiménez, R. Rubilar, H. De La Fuente-Mella, V. Leiva, Breakpoint analysis for the COVID-19 pandemic and its effect on the stock markets, *Entropy*, **23** (2021), 100. https://doi.org/10.3390/e23010100

2. Y. Liu, C. Mao, V. Leiva, S. Liu, W. A. Silva Neto, Asymmetric autoregressive models: Statistical aspects and a financial application under COVID-19 pandemic, *J. Appl. Stat.*, **49** (2022), 1323–1347. https://doi.org/10.1080/02664763.2021.1913103

3. E. Mahdi, V. Leiva, S. Mara'Beh, C. Martin-Barreiro, A new approach to predicting cryptocurrency returns based on the gold prices with support vector machines during the COVID-19 pandemic using sensor-related data, *Sensors*, **21** (2021), 6319. https://doi.org/10.3390/s21186319

4. H. He, L. Harris, The impact of COVID-19 pandemic on corporate social responsibility and marketing philosophy, *J. Bus. Res.*, **116** (2020), 176–182. https://doi.org/10.1016/j.jbusres.2020.05.030

5. J. C. Hoekstra, P. S. Leeflang, Marketing in the era of COVID-19, *Ital. J. Mark.*, **2020** (2020), 249–260. https://doi.org/10.1007/s43039-020-00016-3

6. C. Stamu-O'Brien, S. Carniciu, E. Halvorsen, M. Jafferany, Psychological aspects of COVID-19, *J. Cosmet. Dermatol.*, **19** (2020), 2169–2173. https://doi.org/10.1111/jocd.13601

7. T. Rume, S. D. U. Islam, Environmental effects of COVID-19 pandemic and potential strategies of sustainability, *Heliyon*, **6** (2020), e04965. https://doi.org/10.1016/j.heliyon.2020.e04965

8. L. Bera, M. Souchon, A. Ladsous, V. Colin, J. Lopez-Castroman, Emotional and behavioral impact of the COVID-19 epidemic in adolescents, *Curr. Psychiatry Rep.*, **24** (2022), 37–46. https://doi.org/10.1007/s11920-022-01313-8

9. E. Kiran, Prominent issues about the social impacts of COVID 19, *Gaziantep Uni. J. Soc. Sci.*, **19** (2020), 752–766. https://doi.org/10.21547/jss.787779

10. F. Rojas, V. Leiva, M. Huerta, C. Martin-Barreiro, Lot-size models with uncertain demand considering its skewness/kurtosis and stochastic programming applied to hospital pharmacy with sensor-related COVID-19 data, *Sensors*, **2** (2021), 5198. https://doi.org/10.3390/s21155198

11. M. Rangasamy, C. Chesneau, C. Martin-Barreiro, V. Leiva, On a novel dynamics of SEIR epidemic models with a potential application to COVID-19, *Symmetry*, **14** (2022), 1436. https://doi.org/10.3390/sym14071436

12. J. Yego, R. Korom, E. Eriksson, S. Njavika, O. Sane, P. Kanorio, et al., A comparison of strategies to improve uptake of COVID-19 vaccine among high-risk adults in Nairobi, Kenya in 2022, *Vaccines*, **11** (2023), 209. https://doi.org/10.3390/vaccines11020209

13. A. B. Hogan, P. Winnskill, O. J. Watson, P. G. T. Walker, C. Whittaker, M. Baguelin, et al., Within-country age-based prioritisation, global allocation, and public health impact of a vaccine against SARS-CoV-2: a mathematical modelling analysis, *Vaccines*, **39** (2021), 2995–3006. https://doi.org/10.1016/j.vaccine.2021.04.002

14. O. J. Watson, G. Barnsley, J. Tool, A. B. Hogan, P. Winskill, A. C. Ghani, Global impact of the first year of COVID-19 vaccination: a mathematical modelling study, *Lancet Infect. Dis.*, **22** (2022), 1293–1302. https://doi.org/10.1016/S1473-3099(22)00320-6

15. N. Parolini, L. Dede', G. Ardenghi, A. Quarteroni, Modelling the COVID-19 epidemic and the vaccination campaign in Italy by the SUIHTER model, *Infect. Dis. Model.*, **7** (2022), 45–63. https://doi.org/10.1016/j.idm.2022.03.002

16. M. S. Nabaggala, T. S. Nair, M. Gacic-Dobo, A. Siyam, K. Diallo, M. Boniol, The global inequity in COVID-19 vaccination coverage among health and care workers, *Int. J. Equity Health*, **21** (2022), 147. https://doi.org/10.1186/s12939-022-01750-0

17. P. Galanis, I. Vraka, A. Katsiroumpa, O. Siskou, O. Konstantakopoulou, T. Katsoulas, et al.. COVID-19 vaccine uptake among healthcare workers: a systematic review and meta-analysis, *Vaccines*, **10** (2022), 1637. https://doi.org/10.3390/vaccines10101637

18. M. Antonelli, R. S. Penfold, J. Merino, C. H. Sudre, E. Molteni, S. Berry, et al., Risk factors and disease profile of post-vaccination SARS-CoV-2 infection in UK users of the COVID symptom study app: a prospective, community-based, nested, case-control study, *Lancet Infect. Dis.*, **22** (2022), 43–55. https://doi.org/10.1016/S1473-3099(21)00460-6

19. P. R. Wratil, K. Kotter, M. L. Bischof, S. Hollerbach, E. Apak, A. L. Kalteis, et al., Vaccine-hesitant individuals accumulate additional COVID-19 risk due to divergent perception and behaviors related to SARS-CoV-2 testing: a population-based, cross-sectional study, *Infection*, 2022. https://doi.org/10.1007/s15010-022-01947-z

20. A. Ahmadini, M. Elgarhy, A. W. Shawki, H. Baaqeel, O. Bazighifan, Statistical analysis of the people fully caccinated against COVID-19 in two different regions, *Appl. Bionics Biomech.*, **2022** (2022), 7104960. https://doi.org/10.1155/2022/7104960

21. Y. Guo, B. Li, T. Duan, N. Yao, H. Wang, Y. Yang, et al., A panel regression analysis for the COVID-19 epidemic in the United States, *PloS One*, 2022. https://doi.org/10.1371/journal.pone.0273344

22. C. Latkin, L. Dayton, J. Miller, G. Yi, A. Balaban, B. Boodram, et al., A longitudinal study of vaccine hesitancy attitudes and social influence as predictors of COVID-19 vaccine uptake in the US, *Hum. Vacc. Immunother.*, **18** (2022), 2043102. https://doi.org/10.1080/21645515.2022.2043102

23. S. W. Yip, A. Jordan, R. J. Kohler, A. Holmes, D. Bzdok, Multivariate, transgenerational associations of the COVID-19 pandemic across minoritized and marginalized communities, *JAMA Psychiat.*, **79** (2022), 350–358. https://doi.org/10.1001/jamapsychiatry.2021.4331

24. M. da P. Harb, L. Silva, T. Ayass, N. Vijaykumar, M. Silva, C. R. Francês, Dendrograms for clustering in multivariate analysis: applications for COVID-19 vaccination infodemic data in Brazil, *Computation*, **10** (2022), 166. https://doi.org/10.3390/computation10090166

25. R. Borchering, L. Mullany, E. Howerton, M. Chinazzi, C. P. Smith, M. Qin, et al., Impact of SARS-CoV-2 vaccination of children ages 5–11 years on COVID-19 disease burden and resilience to new variants in the United States. November 2021–March 2022: A multi-model study, *Lancet Regional Health-Americas*, **17** (2023), 100398. https://doi.org/10.1016/j.lana.2022.100398

26. K. I. Kasozi, A. Laudisoit, L. O. Osuwat, G. El-Saber Batiha, N. E. Al Omairi, E. Aigbogun, et al., A descriptive-multivariate analysis of community knowledge, confidence, and trust in COVID-19 clinical trials among healthcare workers in Uganda, *Vaccines*, **9** (2021), 253. https://doi.org/10.3390/vaccines9030253

27. T. Roederer, B. Mollo, C. Vincent, G. Leduc, J. Sayyad-Hilario, M. Mosnier, et al., Estimating COVID-19 vaccine uptake and its drivers among migrants, homeless and precariously housed people in France, *Commun. Med.*, **3** (2022), 30. https://doi.org/10.1038/s43856-023-00257-1

28. G. H. Murata, A. E. Murata, D. J. Perkins, H. M. Campbell, J. T. Mao, B. Wagner, et al., Effect of vaccination on the case fatality rate for COVID-19 infections 2020–2021: multivariate modelling of data from the US Department of Veterans Affairs, *BMJ Open*, **12** (2022), e064135. http://dx.doi.org/10.1136/bmjopen-2022-064135

29. J. Cheng, S. Loong, C. Min-Ho, K. Jing Ng, M. Min Qi Ng, R. Choon Hoe Chee, et al., Knowledge, attitudes, and practices of COVID-19 vaccination among adults in Singapore: A cross-sectional study, *Am. J. Trop. Med. Hyg.*, **107** (2023), 540–550.

30. S. Koya, S. Ponnam, S. Salenius, S. Pamidighantam, A Markov chain Monte Carlo multivariate analysis of the association of vital parameter variation with the lunar cycle in patients hospitalized with COVID-19, *Cureus*, **15** (2023), e34290. http://dx.doi.org/10.7759/cureus.34290

31. X. Cabezas, S. García, C. Martin-Barreiro, E. Delgado, V. Leiva, A two-stage location problem with order solved using a Lagrangian algorithm and stochastic programming for a potential use in COVID-19 vaccination based on sensor-related data, *Sensors*, **21** (2021), 5352. https://doi.org/10.3390/s21165352

32. E. J. Delgado, X. Cabezas, C. Martin-Barreiro, V. Leiva, F. Rojas, An equity-based optimization model to solve the location problem for healthcare centers applied to hospital beds and COVID-19 vaccination, *Mathematics*, **10** (2022), 1825. https://doi.org/10.3390/math10111825

33. M. R. Mahmoudi, M. H. Heydari, S. N. Qasem, A. Mosavi, S. S. Band, Principal component analysis to study the relations between the spread rates of covid-19 in high risks countries, *Alex. Eng. J.*, **60** (2021), 457–464. https://doi.org/10.1016/j.aej.2020.09.013

34. I. Noy, N. Doan, B. Ferrarini, D. Park, Measuring the economic risk of COVID-19, *Global Policy*, **11** (2020), 413–423. https://doi.org/10.1111/1758-5899.12851

35. W. Ye, W. Lu, Y. Tang, G. Chen, X. Li, C. Ji, et al., Identification of covid-19 clinical phenotypes by principal component analysis-based cluster analysis, *Front. Med.*, **7** (2020), 782. https://doi.org/10.3389/fmed.2020.570614

36. A. Ramadan, A. Kamel, A. Taha, A. El-Shabrawy, N. A. Abdel-Fatah, A multivariate data analysis approach for investigating daily statistics of countries affected with COVID-19 pandemic, *Heliyon*, **6** (2020), e05575. https://doi.org/10.1016/j.heliyon.2020.e05575

37. B. E. Zinsou, D. Letourneur, J. Siko, R. M. de Souza, F. Adjagba, P. Pineau, Main modulators of COVID-19 epidemic in sub-Saharan Africa, *Heliyon*, **9** (2022), e12727. https://doi.org/10.1016/j.heliyon.2022.e12727

38. C. Martin-Barreiro, J. A. Ramirez-Figueroa, X. Cabezas, V. Leiva, M. P. Galindo-Villardón, Disjoint and functional principal component analysis for infected cases and deaths due to COVID-19 in South American countries with sensor-related data, *Sensors*, **21** (2021), 4094. https://doi.org/10.3390/s21124094

39. M. Coccia, Pandemic prevention: lessons from COVID-19, *Encyclopedia*, **1** (2021), 433–444. https://doi.org/10.3390/encyclopedia1020036

40. W. S. Chan, M. Lam, J. H. Y. Law, T. L. Chan, E. S. K. Ma, B. S. F. Tang, Geographical prevalence of SARS-CoV-2 variants, August 2020 to July 2021, *Sci. Rep.*, **12** (2021), 4704. https://doi.org/10.1038/s41598-022-08684-1

41. C. Magazzino, M. Mele, M. Coccia, A machine learning algorithm to analyze the effects of vaccination on COVID-19 mortality, *Epidemiol. Infect.*, **150** (2022), e168. https://doi.org/10.1017/S0950268822001418

42. M. Coccia, Optimal levels of vaccination to reduce COVID-19 infected individuals and deaths: a global analysis, *Environ. Res.*, **204** (2022), 112314. https://doi.org/10.1016/j.envres.2021.112314

43. I. Benati, M. Coccia, Global analysis of timely COVID-19 vaccinations: Improving governance to reinforce response policies for pandemic crises, *Int. J. Health Gov.*, **27** (2022), 240–253. https://doi.org/10.1108/IJHG-07-2021-0072

44. M. Coccia, COVID-19 pandemic over 2020 (with lockdowns) and 2021 (with vaccinations): similar effects for seasonality and environmental factors, *Environ. Res.*, **208** (2022), 112711. https://doi.org/10.1016/j.envres.2022.112711

45. M. Fiori, G. Bello, N. Wschebor, F. Lecumberry, A. Ferragut, E. Mordecki, Decoupling between SARS-CoV-2 transmissibility and population mobility associated with increasing immunity from vaccination and infection in South America, *Sci. Rep.*, **12** (2022), 6874. https://doi.org/10.1038/s41598-022-10896-4

46. M. Coccia, Effects of strict containment policies on COVID-19 pandemic crisis: lessons to cope with next pandemic impacts, *Environ. Sci. Pollut. Res.*, **30** (2023), 2020–2028. ttps://doi.org/10.1007/s11356-022-22024-w

47. S. S. Musa, A. Tariq, L. Yuan, W. Haozhen, D. He, Infection fatality rate and infection attack rate of COVID-19 in South American countries, *Infect. Dis. Poverty*, **11** (2022), 40. https://doi.org/10.1186/s40249-022-00961-5

48. M. Coccia, COVID-19 vaccination is not a sufficient public policy to face crisis management of next pandemic threats, *Public Organiz. Rev.*, 2022. https://doi.org/10.1007/s11115-022-00661-6 .

49. M. Coccia, Improving preparedness for next pandemics: max level of COVID-19 vaccinations without social impositions to design effective health policy and avoid flawed democracies, *Environ. Res.*, **213** (2022), 113566. https://doi.org/10.1016/j.envres.2022.113566

50. D. O. Oyewola, E. G. Dada, S. Misra, Machine learning for optimizing daily COVID-19 vaccine dissemination to combat the pandemic, *Health Technol.*, **12** (2022), 1277–1293. https://doi.org/10.1007/s12553-022-00712-4

51. D. E. Lucero-Prisno, D. O. Shomuyiwa, G. R. Vicente, M. J. González Méndez, S. Qaderi, J. C. Lopez, et al., Achieving herd immunity in South America, *Glob. Health Res. Policy*, **8** (2023), 2. https://doi.org/10.1186/s41256-023-00286-2

52. M. Coccia, High potential of technology to face new respiratory viruses: mechanical ventilation devices for effective healthcare to next pandemic emergencies, *Technol. Soc.*, **73** (2023), 102233. https://doi.org/10.1016/j.techsoc.2023.102233

53. T. H. Jen, J. W. Wu, T. W. Chien, W. Chou, Using dashboards to verify coronavirus (COVID-19) vaccinations can reduce fatality rates in countries/regions: Development and usability study, *Medicine*, **102** (2023), e33274. https://doi.org/10.1097/MD.0000000000033274

54. J. Torales, I. González-Urbieta, I. Barrios, M. Waisman-Campos, A. Terrazas-Landivar, L. Viola, et al., Pandemic fatigue in South America: A multi-center report from Argentina, Bolivia, Paraguay, Peru, and Uruguay, *Brain Sci.*, **13** (2023), 444. https://www.mdpi.com/2076-3425/13/3/444

55. M. Coccia, Sources, diffusion and prediction in COVID-19 pandemic: lessons learned to face next health emergency, *AIMS Public Health*, **10** (2023), 145–168. https://doi.org/10.3934/publichealth.2023012

56. C. Torres, M. Nabaes Jodar, D. Acuña, R. M. Zambrana Montaño, A. C. Alberto Culasso, A. Fernando Amadio, et al., Omicron waves in Argentina: dynamics of SARS-CoV-2 lineages BA.1, BA.2 and the emerging BA.2.12.1 and BA.4/BA.5, *Viruses*, **15** (2023), 312. https://doi.org/10.3390/v15020312

57. Y. Zhao, J. Du, Z. Li, Z. Xu, Y. Wu, W. Duan, et al., It is time to improve the acceptance of COVID-19 vaccines among people with chronic diseases: A systematic review and meta-analysis, *J. Med. Virol.*, **95** (2023), e28509. https://doi.org/10.1002/jmv.28509

58. R. Zambrana-Montaño, A. C. A. Culasso, F. Fernández, N. Marquez, H. Debat, M. Salmerón, et al., Evolution of SARS-CoV-2 during the first year of the COVID-19 pandemic in Northwestern Argentina, *Virus Res.*, **323** (2023), 198936. https://doi.org/10.1016/j.virusres.2022.198936

59. R Core Team, *R: A Language and Environment for Statistical Computing*, Vienna: R Foundation for Statistical Computing, 2021.

60. P. Giordani, Principal component analysis, In *Encyclopedia of Social Network Analysis and Mining*, New York: Springer, 2018.

61. I. T. Jolliffe, *Principal Component Analysis*, New York: Springer, 2002.

62. J. A. Ramirez-Figueroa, C. Martin-Barreiro, A. B. Nieto-Librero, V. Leiva, M. P. Galindo-Villardón, A new principal component analysis by particle swarm optimization with an environmental application for data science, *Stoch. Env. Res. Risk Assess.*, **35** (2021), 1969–1984. https://doi.org/10.1007/s00477-020-01961-3

63. P. Sharma, A. K. Singh, V. Leiva, C. Martin-Barreiro, X. Cabezas, Modern multivariate statistical methods for evaluating the impact of WhatsApp on academic performance: methodology and case study in India, *Appl. Sci.*, **12** (2022), 6141. https://doi.org/10.3390/app12126141

64. C. Martin-Barreiro, J. A. Ramirez-Figueroa, A. B. Nieto-Librero, V. Leiva, A. Martin-Casado, M. P. Galindo-Villardón, A new algorithm for computing disjoint orthogonal components in the three-way tucker model, *Mathematics*, **9** (2021), 203. https://doi.org/10.3390/math9030203

65. C. Martin-Barreiro, J. A. Ramirez-Figueroa, X. Cabezas, V. Leiva, A. Martin-Casado, M. P. Galindo-Villardón, A new algorithm for computing disjoint orthogonal components in the parallel factor analysis model with simulations and applications to real-world data, *Mathematics*, **9** (2021), 2058. https://doi.org/10.3390/math9030203

66. J. Wu, Cluster analysis and K-means clustering: an introduction, In: *Advances in K-Means Clustering: A Data Mining Thinking*, Berlin: Springer, 2012

67. D. Abdullah, S. Susilo, A. S. Ahmar, R. Rusli, R. Hidayat, The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data, *Qual. Quant.*, **56** (2022), 1283–1291. https://doi.org/10.1007/s11135-021-01176-w