International Journal of Combinatorial Optimization Problems and Informatics, 14(2), May-Aug 2023, 67-76. ISSN: 2007-1558.

www.editada.org

How to determine whether the covid-19 infection series are stationary and can be predicted or whether they are non-stationary and cannot be predicted?

Mitzunory Enrique Medrano Meneses[1], Zeus Salvador Hernández Veleros[1], Miguel Ángel Torres González[2]*
[1] Universidad Autónoma del Estado de Hidalgo, México.
[2] Universidad Politécnica de Pachuca, México.
E-mails: me295462@uaeh.edu.mx, mtorres@upp.edu.mx
* corresponding author: zeus_hernandez@uaeh.edu.mx

Abstract. Lopez-Gatell, who has managed the public policy to control the SARS-CoV-2 virus, has on several occasions made forecasts on the dynamics of infections and deaths; but if such series are non-stationary, this implies a very serious error. In our opinion, many of these series have a non-stationary data generating process and, therefore, forecasts cannot be made. To determine this, we will use various econometric techniques such as unit root tests and, in addition, we will see if the series responds to a regime shift process. As results we have that the series of weekly contagions by COVID-19 in 8 entities of the country, only one of them is stationary, in addition, when analyzing the trajectory of the contagions through Markov chains to determine the performance of these states to control the contagions, we found that one of them had a very bad performance, 5 with bad performance and 1 with good performance.

Keywords: Unit roots, Stationarity, Markov chains, Covid-19.

Article Info
Received Dec 26, 2022
Accepted May 11, 2023

1 Introduction

A new, unknown virus arrived as an epidemic in the city of Wuhan, China in late 2019 [1]. The clinical picture of the disease it causes is represented by a severe respiratory disorder leading to pneumonia and death in the most severe cases [2]. Subsequently, the World Health Organization announced COVID-19 as a global pandemic on March 11, 2020 [3].

The first case detected in Mexico occurred on February 27, 2020, as an imported case. Until March 30, 2020, the federal government decreed a one-month national health emergency. However, on April 20, 2020, with almost 10,000 accumulated cases and more than 800 deaths, this period was extended to May 30, 2020 [4] and the economic reopening would begin in search of the "New Normality" with the introduction of the idea of the epidemiological risk traffic light, thus transferring the corresponding powers to the state governments to decide how to carry out the economic reopening [5].

One of the main contradictions of the economic reopening on that date is that only 24 hours before, the highest number of new infections had been reported up to that time. Hugo López-Gatell, undersecretary of Health, defended this decision by arguing that the epidemic curve would already be in decline between September and October. However, it was in mid-october that we saw the beginning of the second wave of contagions, according to data from the Ministry of Health, evidencing that the undersecretary of health's estimate was erroneous. However, this was not the first time that the pandemic controllers' estimates were incorrect, initially, José Luis Alomía, director of Epidemiology, on February 27, 2020, stated that it could take up to 40 days to go from imported cases to local transmission, but less than a month has elapsed since the first case was announced at the start of the National Healthy Distance. On April 23, one month after the beginning of the social distancing, Lopez-Gatell estimated that the peak of the epidemic would be between May 8 and 10, but by the 28th of that month, the turning point had not yet been reached [6].

2 Problem description and state of the art

Trying to predict the future values of a non-seasonal time series is a mistake, since there is no pattern or way to determine the next values that our variable will take, in the best case, our best estimate would be the last value that the series had. Lopez-Gatell, in different press conferences, claimed that his estimates were correct, but none of them were fulfilled, making a mistake in this type of estimates should not be taken lightly since people's lives are at stake.

This investigation aims to determine that the time series of weekly COVID-19 infections in the selected entities (figure 1) are not stationary, and their trajectory will be analyzed in the period from January 4, 2020, to August 9, 2021 (83 weeks), through the application of Markov chains.

Figure 1. Selected states of each macro region.

Macro region	State selected
Northwestern Region	Sinaloa
Northeastern Region	Nuevo León
West Region	Jalisco
East Region	Hidalgo
Center north	Guanajuato
Center south	Ciudad de México
Southwest	Chiapas
Southeast	Tabasco

Source: Own elaboration.

We can mention some examples of research in which Markov chains have been applied, where the system is a health area, as well as in planning models, whose objective is to plan the stay of patients in a specialty area, hospitalization costs and decision making.

In 2005, [7] analyzed the incidence, that is, the probability of transition from a state of depression to a state of no depression, detecting the population standards with the highest probability of incidence of depression, finding that they are those persons of the female gender and individuals with no sentimental condition. The research is oriented to decision making by relating the epidemiological data.

Likewise, [8] designed a Markov chain to predict the number of kidney transplant therapy patients in Greece. Setting three states: hemodialysis, peritoneal dialysis, renal transplantation, and death. According to its results, 26% of the population will have an incidence of the disease.

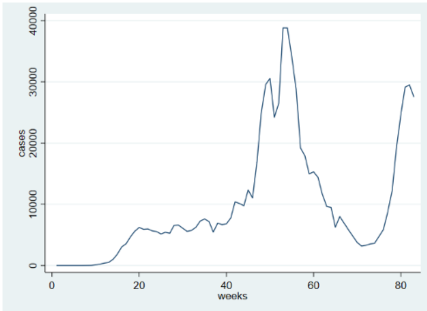
A Mexican case, [9] used Markov chains to model the disease process in patients with AH1N1 influenza in 2009, by determining severity states and average length of stay, to plan the costs and materials needed to adequately treat the disease.

3 Methodology

The Augmented Dickey Fuller test is used to determine whether time series are nonstationary or stationary; that is, if its mean, variance and covariance are independent of time, it will be stationary; if this is not the case, implies that these statistics change depending on the time or space of time considered, with the consequence that the best forecast of a nonstationary process of this nature is the last observed data.

A weakness of the Augmented Dickey Fuller test is that the presence of breaks in the series favors the conclusion that the data generating process is nonstationary. To respond to this situation: on many occasions the test establishes that the process is nonstationary when there are one or more breaks, the Clemente, Montañés and Reyes test was applied with two different types of breakages: additives, which involve sudden changes; or innovative, which are suitable for incremental changes.

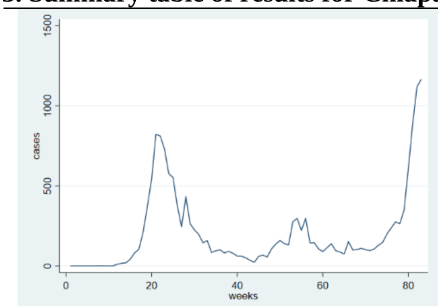
Figure 2. Summary table of results for Mexico City



DFA with intercept and trend	DFA with intercept	DFA	Clemao1
Non stationary.	Non stationary.	Non stationary.	Non stationary.
Clemio1	Clemao2	Clemio1	Zandrews
Non stationary.	Non stationary.	Non stationary.	Non stationary.

Source: Prepared by the authors with data from the Secretary of Health.

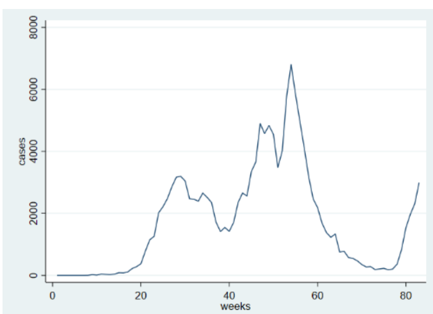
Figure 3. Summary table of results for Chiapas



DFA with intercept and trend	DFA with intercept	DFA	Clemao1
Non stationary.	Non stationary.	Non stationary.	Non stationary.
Clemio1	Clemao2	Clemio1	Zandrews
Non stationary.	Non stationary.	Non stationary.	Non stationary.

Source: Prepared by the authors with data from the Secretary of Health.

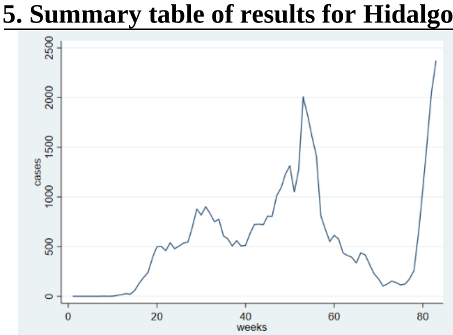
Figure 4. Summary table of results for Guanajuato



DFA with intercept and trend	DFA with intercept	DFA	Clemao1
Non stationary.	Non stationary.	Non stationary.	Non stationary.
Clemio1	Clemao2	Clemio1	Zandrews
Non stationary.	Non stationary.	Non stationary.	Non stationary.

Source: Prepared by the authors with data from the Secretary of Health.

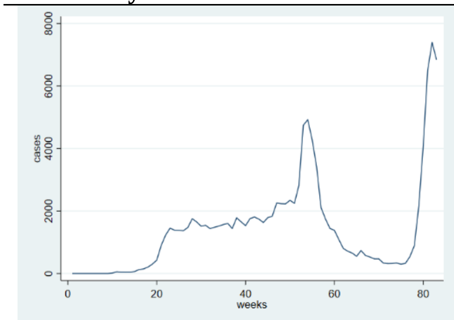
Figure 5. Summary table of results for Hidalgo



DFA with intercept and trend	DFA with intercept	DFA	Clemao1
Non stationary.	Non stationary.	Non stationary.	Non stationary.
Clemio1	Clemao2	Clemio1	Zandrews
Non stationary.	Non stationary.	Non stationary.	Non stationary.

Source: Prepared by the authors with data from the Secretary of Health.

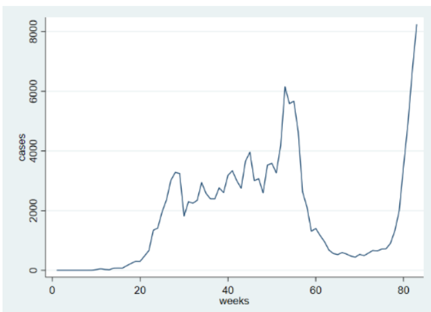
Figure 6. Summary table of results for Jalisco



DFA with intercept and trend	DFA with intercept	DFA	Clemao1
Non stationary.	Non stationary.	Non stationary.	Non stationary.
Clemio1	Clemao2	Clemio1	Zandrews
Non stationary.	Non stationary.	Non stationary.	Non stationary.

Source: Prepared by the authors with data from the Secretary of Health.

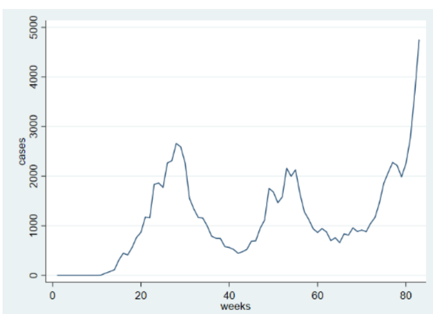
Figure 7. Summary table of results for Nuevo León



DFA with intercept and trend	DFA with intercept	DFA	Clemao1
Non stationary.	Non stationary.	Non stationary.	Non stationary.
Clemio1	Clemao2	Clemio1	Zandrews
Non stationary.	Non stationary.	Non stationary.	Non stationary.

Source: Prepared by the authors with data from the Secretary of Health.

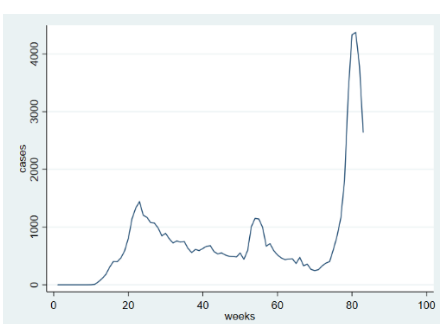
Figure 8. Summary table of results for Tabasco



DFA with intercept and trend	DFA with intercept	DFA	Clemao1
Non stationary.	Non stationary.	Non stationary.	Non stationary.
Clemio1	Clemao2	Clemio1	Zandrews
Non stationary.	Non stationary.	Non stationary.	Non stationary.

Source: Prepared by the authors with data from the Secretary of Health.

Figure 9. Summary table of results for Sinaloa



DFA with intercept and trend	DFA with intercept	DFA	Clemao1
If stationary.	If stationary	Non stationary.	If stationary
Clemio1	Clemao2	Clemio1	Zandrews
If stationary.	Non stationary.	If stationary.	If stationary

Source: Prepared by the authors with data from the Secretary of Health.

After applying these tests, we conclude that seven of the eight series are nonstationary; therefore, the best forecast to have is the last data of each series; no other forecast can be made. Only the Sinaloa series is stationary and from it build a time series model which can be used for forecasting.

4 What can we discover with nonstationary series?

A If the series are non-stationary, we can extract useful information by estimating a Markov regimen switching model, which makes it possible to determine the probabilities of moving into a low-infection regime, identified with number 1, toward a regime of high contagion, identified with number 2; or what is the probability of moving from regime 2 to regime 1; or remain in the same regime.

For example, in the likelihood of being in the low infection regime and remain in the low contagion in Mexico City is 0.37, the probability of moving from a low infection to a high infection regimen is 0.63; that is, there is a high likelihood of an increase in the number of infections. While the probability of moving from a high to a low regime is 0.42, while the probability of staying in a high regime is 0.58 (figure 10).

Figure 10. Summary table of regime change probabilities.

Mexico City		
	Subsequent moment (t+1)	
Starting time (t)	Low contagion rate (1)	High contagion rate (2)
Low contagion rate (1)	0.37	0.63
High contagion rate (2)	0.42	0.58
Chiapas		
	Subsequent moment (t+1)	
Starting time (t)	Low contagion rate (1)	High contagion rate (2)
Low contagion rate (1)	0.94	0.06
High contagion rate (2)	0.12	0.88
Guanajuato		
	Subsequent moment (t+1)	
Starting time (t)	Low contagion rate (1)	High contagion rate (2)
Low contagion rate (1)	0.98	0.02
High contagion rate (2)	0.11	0.89
Hidalgo		
	Subsequent moment (t+1)	
Starting time (t)	Low contagion rate (1)	High contagion rate (2)
Low contagion rate (1)	0.97	0.03
High contagion rate (2)	0.15	0.85
Jalisco		
	Subsequent moment (t+1)	
Starting time (t)	Low contagion rate (1)	High contagion rate (2)

Low contagion rate (1)	0.94	0.06
High contagion rate (2)	0.23	0.77
Nuevo León		
	Subsequent moment (t+1)	
Starting time (t)	Low contagion rate (1)	High contagion rate (2)
Low contagion rate (1)	0.95	0.05
High contagion rate (2)	0.01	0.99
Tabasco		
	Subsequent moment (t+1)	
Starting time (t)	Low contagion rate (1)	High contagion rate (2)
Low contagion rate (1)	0.9	0.1
High contagion rate (2)	0.06	0.94

Source: own elaboration

5 What can we discover with stationary series?

A With a stationary series we can formulate a time series model, that if you can forecast, that's what we'll do with the Sinaloa data series (figure 11), The red line indicates the observed series and the blue line the estimated series.

Figure 11. Regression suggested by Automatic ARIMA Forecasting for Sinaloa.

Dependent Variable: CASOS
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 08/31/21 Time: 16:55
Sample: 1/04/2020 7/31/2021
Included observations: 83
Convergence achieved after 35 iterations
Coefficient covariance computed using outer product of gradients

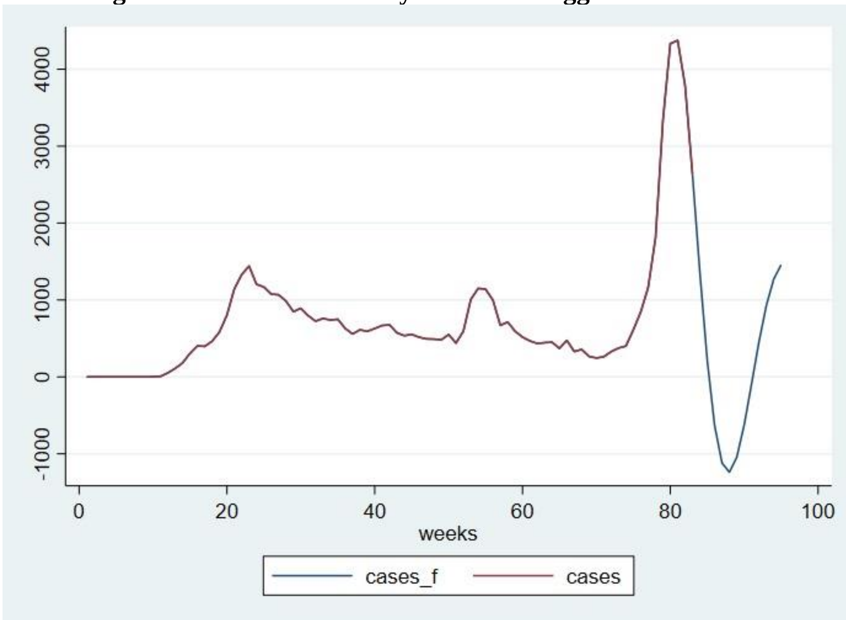
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	678.4153	247.1218	2.745267	0.0075
AR(1)	1.893025	0.126338	14.98385	0.0000
AR(2)	-1.201604	0.245517	-4.894174	0.0000
AR(3)	0.193356	0.157340	1.228903	0.2228
SIGMASQ	29901.45	3568.783	8.378612	0.0000

R-squared	0.958357	Mean dependent var	755.5783
Adjusted R-squared	0.956221	S.D. dependent var	852.5244
S.E. of regression	178.3766	Akaike info criterion	13.32223
Sum squared resid	2481820.	Schwarz criterion	13.46795
Log likelihood	-547.8727	Hannan-Quinn criter.	13.38077
F-statistic	448.7651	Durbin-Watson stat	1.959536
Prob(F-statistic)	0.000000		

Inverted AR Roots	.83-.35i	.83+.35i	.24

Source: own elaboration

.Figure 12. Forecast made by the model suggested in Sinaloa.



Source: Prepared by the authors with data from the Secretary of Health

Figure 13. Average duration of each regimen in Chiapas.

Expected Duration	Estimate	Std. Err.	[95% Conf.	Interval]
State1	**17.42861**	**10.18593**	**5.873555**	**56.38038**
State2	**8.34666**	**5.135672**	**2.866632**	**29.91486**

Source: own elaboration.

Figure 14. Average duration of each regimen in Mexico City.

Expected Duration	Estimate	Std. Err.	[95% Conf.	Interval]
State1	**24.31812**	**18.64024**	**5.866902**	**112.7209**
State2	**22.98932**	**17.80959**	**5.495798**	**108.5516**

Source: own elaboration.

Figure 15. Average duration of each regimen in Guanajuato.

Expected Duration	Estimate	Std. Err.	[95% Conf.	Interval]
State1	**56.44853**	**60.34387**	**7.569602**	**468.9948**
State2	**9.373843**	**8.26511**	**2.209974**	**58.95269**

Source: own elaboration.

Figure 16. Average duration of each regimen in Hidalgo.

Expected Duration	Estimate	Std. Err.	[95% Conf.	Interval]
State1	31.08113	22.97624	7.731914	135.4156
State2	6.658306	5.253874	1.916907	35.91787

Source: own elaboration.

Figure 17. Average duration of each regimen in Jalisco.

Expected Duration	Estimate	Std. Err.	[95% Conf.	Interval]
State1	18.02887	10.17955	6.276584	55.95644
State2	4.286331	2.297546	1.834867	13.93615

Source: own elaboration.

Figure 18. Average duration of each regimen in Nuevo León.

Expected Duration	Estimate	Std. Err.	[95% Conf.	Interval]
State1	19.22558	16.20482	4.190486	105.1132
State2	73.96604	75.91473	10.49556	561.6877

Source: own elaboration.

Figure 19. Average duration of each regimen in Tabasco.

Expected Duration	Estimate	Std. Err.	[95% Conf.	Interval]
State1	9.659659	7.596246	2.551771	49.32524
State2	15.92272	11.93333	4.112888	72.53732

Source: own elaboration.

6 Results, analysis, and discussion

Weekly infections across the states show heterogeneous behavior, this is due to the social and territorial characteristics of each of them, but this does not mean that one is better than the other.

The main findings are that the best regime switching model for all entities was the autoregressive of degree two, that is, past two weeks' infections influence current infections, but only the first delay causes contagions to rise, while those of the second delay cause a decrease in present infections, this could be due to an increase in people's perception of risk, knowing that infections are increasing, they are increasing their sanitary measures.

Thus, of the entities analyzed, Mexico City is the one with the worst pandemic control, the probability of moving to a high level of contagion is of 63% and the probability of these decreases is only 42%, the entity behind Mexico City is Nuevo León, with a probability of remaining in the low-infection state is 80%, while that of state (2) is 92%.

The entity that best controlled the disease is the state of Chiapas, having a 94% probability of remaining in the state with low infections, but in addition to that, the probability of returning to stage (1) due to accelerated infection is 88%.

Outside the aforementioned entities, we have similar results, the probability of moving from one state to another is very low, being greater than 90% in most cases, which would be the cause of the fact that coming out of the peaks with accelerated contagions is very complicated.

Why did the COVID-19 forecasts fail?

The columnist of The Atlantic [10], presents the case studies of the projections made by the Aspen Institute and McKinsey & Company, for the United States, that claimed a "tsunami" of evictions in up to 10% of the population, and, on the other hand, that approximately 25 percent of the female workforce would have to leave their jobs or "drop down the career ladder". However, current publications from Harvard University, the New York Times, among other experts, found that these statements were not close to reality.

For the United Kingdom, [11], made a similar comparison with the Scientific Pandemic Influenza Group projections (SPI-M) among other mathematical models on daily cases and cumulative deaths in that country. Again, all these projections were criticized by the population, among other newspapers, since they are different from what actually happened.

Lastly, [12], based on the arguments of Graham Medley, professor in infectious disease modeling in London, continued the theme of why the covid-19 predictions failed, especially, those related to the number of cases per day and accumulated deaths.

How are these publications related? The main argument of the 3 authors is that modeling experts make their projections based on the current situation without expecting changes in the behavior of individuals, however, decision-makers and public policymakers, when they see such alarming forecasts, they establish policies and measures to prevent such scenarios from occurring, thus changing the original variables of the models. In second place, established that the media only highlighted the fatal consequences of their forecasts without mentioning the assumptions on which they were based, making the population believe that these scenarios would occur without explanation. Finally, the excess of information: at the beginning of the pandemic, charts, graphs and tables were published without any filtering, that is, all this information generated may not be useful, and may not even be true and, despite this, would reach the population in the form of magazine articles, social media posts and word-of-mouth effect, information that, in one way or another, would end up affecting the mathematical models.

7 Conclusions

As we have shown, 7 of the 8 series are nonstationary; therefore, the Ministry of Health and the Undersecretary of Prevention and Health Promotion, headed by Dr. Lopez-Gatell, have made incorrect estimates and forecasts; errors have been made that have implications beyond the time series technique as they relate to human lives, not only with economic or social issues. Hence the need to study the data generating process of the series of interest.

The management of the pandemic across the selected entities shows that most of them performed poorly in controlling infections, only with the exception of Chiapas, which, in addition to having weekly infections of less than 1,000 cases, manages to slow down infections when they tend to rise rapidly.

The impact of public policies on the behavior of contagion should be analyzed, since these changes could change the trajectory of new cases in an unexpected way, for example.

References

1. Zhu Y, Xie J (2020) Association between ambient temperature and COVID-19 infection in 122 cities from China. Sci Total Environ: 138201. https://doi.org/10.1016/j.scitotenv.2020.138201.
2. Wang J, Tang K, Feng K, Lv W (2020a) High temperature and high humidity reduce the transmission of COVID-19. SSRN Electron J3551767. https://doi.org/10.2139/ssrn.3551767.
3. OMS (2020) WHO Director-General's opening remarks at the media briefing on COVID-19.
4. Suárez, V., Suarez Quezada, M., Oros Ruiz, S., & Ronquillo De Jesús, E. (2020). Epidemiología de COVID-19 en México: del 27 de febrero al 30 de abril de 2020. *Revista clínica española*, 220(8), 463–471. https://doi.org/10.1016/j.rce.2020.05.007.
5. Mejía Reyes, P., Hurtado Jaramillo, A., & Rendón Rojas, L. (2020). Efecto de factores socio-económicos y condiciones de salud en el contagio de COVID-19 en los estados de México. Contaduría y Administración, 65(5), 223. https://doi.org/10.22201/fca.24488410e.2020.3127.
6. Los números del COVID-19 en México ¿Cómo han cambiado las estimaciones? (29 de mayo de 2020). *Expansión digital*. *https://politica.expansion.mx/mexico/2020/05/29/los-numeros-del-covid-19-en-mexico-como-han-cambiado-las-estimaciones.*
7. Patten S. B. (2005). Markov models of major depression for linking psychiatric epidemiology to clinical practice. Clinical practice and epidemiology in mental health : CP & EMH, 1(1), 2. https://doi.org/10.1186/1745-0179-1-2.
8. Rodina, A., Bliznakova, K., & Pallikarakis, N. (2010). End stage renal disease patients' projections using Markov Chain Monte Carlo simulation. In *XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010* (pp. 796-799). Springer, Berlin, Heidelberg.
9. Orozco, T. (2013). Modelación con cadenas de Markov del proceso de enfermedad en pacientes infectados con influenza norteamericana A(H1N1). (Tesis de Maestría). Universidad Nacional Autónoma de México, México. Recuperado de https://repositorio.unam.mx/contenidos/87945.
10. Demsas, J. (2022, 11 abril). Why So Many COVID Predictions Were Wrong. The Atlantic. https://www.theatlantic.com/ideas/archive/2022/04/pandemic-failed-economic-forecasting/629498/
11. Yates, K. (2022, 27 enero). Why mathematicians sometimes get Covid projections wrong. The Guardian. https://www.theguardian.com/commentisfree/2022/jan/26/mathematicians-covid-projections-modelling
12. Whipple, T. (2022, 11 febrero). Did the Covid modellers get it wrong? News | The Times. https://www.thetimes.co.uk/article/did-the-covid-modellers-get-it-wrong-fthr3x5zc?utm_medium=Social&utm_source=Facebook&fbclid=IwAR3jNfQM11tH2HdI4NomAXoJtroUcupytJkOAsdgRajMYIxcArqFhFD36rQ#Echobox=1644604657