Covid-19 Deaths in Mexico: Exploratory Data Analysis

Gerardo Reyes Guzmán
Universidad DelaSalle Bajío, Campus Salamanca Guanajuato, México

Marco Antonio Escobar Acevedo
Universidad DelaSalle Bajío, Campus Salamanca Guanajuato, México

Perla Esperanza Rostro Hernández
Universidad DelaSalle Bajío, Campus Salamanca Guanajuato, México

Abstract: The purpose of this paper is to conduct an Exploratory Data Analysis (EDA) on Covid-19 deaths and confirmed cases in Mexico, from February 27th to December 31st, 2020. In spite of the fact that the information published by the authorities concerning the pandemic contains numerous flaws and setbacks, it represents hitherto the only official source in the country. By means of EDA we found that it is possible to forecast the number of deaths for the next 31 days; that fatality rate changed as the number of confirmed cases increased; that Tuesday and Wednesday were the days of the week in which most deaths were computed, whereas Sundays and Mondays were the days with the least deaths; and that deaths took place mostly in those federal states with a large share on Mexico´s GDP and/or those located near the border with USA. Furthermore, we also found that deaths per 100,000 people were correlated with health infrastructure, remittances and homicides.

Keywords: Mexico, Covid-19, Exploratory Data Analysis, times series ARIMA, deaths per 100,000 people,

Introduction

Covid-19 deaths are of significant importance for a vast range of studies. The Economist reported one million deaths at the beginning of October 2020. With that number, Covid-19 deaths surpassed 620,000 yearly deaths of malaria, 794,000 of suicide, and 954,000 of HIV/AIDS. The same source pointed out that the official numbers of infected people and deaths were highly underestimated. The number of infected people could have been around 10% of the world population, and the number of deaths, 30% more than officially computed (Briefing, 2020, 26 September). In this regard and already at the beginning of 2021, the Mexican Office for Statistics (INEGI) reported 120,503 deaths vs 64,414 published by the health authorities on August 31st 2021, that means 87% more (Staff, 2021). According to the New York Times, Latin America was the hardest hit continent by the COVID-19 death toll; Mexico, Brazil, and Peru were by that time the countries with the highest deaths per capita (Ahmed, 2020).

Further research interest focuses on death profiles. For instance and with data collected until May 27th, Hector Hérnandez Bringas found that for each woman between the ages of 40 and 49, three men died. Also, the geographical location of deaths has been used as a criteria to relax or restrict the population's mobility utilizing a color-coded system. According to the government, red means people should stay at home; orange: if possible, stay home; yellow: activities are allowed keeping sanitary protocols and green: back to normality with hygiene protocols (Gobierno de México, 2020). Thus, Bringas pointed out that 55.7% of the deaths took place in Mexico City and the State of Mexico together, entities with the largest GDP (Hernández Bringas, 2020). On April 30th 2020, Mexico City, Baja California North, the State of Mexico and Sinaloa, made 51% of a total of 1859 deaths; on 23rd June 2020, Mexico City, the State of Mexico, Baja California North and Veracruz made 52% of a total of 23,377 deaths; this same set of entities concentrated 47% of a total of 39,462 deaths on July 21st and 42% of 46,688 deaths on July 31st 2020. The share of the top four federal states remained on August 10th 2020 except that Puebla took the fourth place instead of Baja California North.

An Exploratory Data Analysis (EDA) (Dearing, 1979; Jeffers, 1994) is a key procedure to understand a phenomenon in science. For instance, Hartwig and Dearing state that researchers should work deeply with data to examine the main variables and possible relationships among them before they delve into any hypothesis-testing at all. The authors resort to data distribution such as skewness, outliers, gaps, and multiple peaks. This paper works with basic statistical indicators and graphs to shed some light on the Covid-19 deaths in Mexico from 27th of February to 31st December, 2020. It is divided in methodology, results and discussion and conclusions. In the first one, we explain the model used to predict deaths for the next 31 days; in the second one we present the results of ARIMA model, the EDA on deaths and correlation; we also address the flaws and setbacks of the information published by the authorities and finally we draw some conclusions.

Methodology

Covid-19 Deaths: Forecasts. At the beginning of the pandemic, Lopez Gatell made some predictions about the number of people expected to die of Covid-19. He started with 6,000; then he said 8,000; and he kept increasing his forecasts naming 30,000 in a worst case scenario (Ann, 2021). Efforts to predict the behavior of Covid-19 deaths have animated researchers to prove several models around the world (Arias Velásquez & Mejía Lara, 2020; Pham, 2020; Schüttler, Schlickeiser, Schlickeiser, & Kröger, 2020; Yunus et al., 2020). In Mexico, there have been some efforts in this direction (Mena et al., 2020; Torrealba-Rodriguez, Conde-Gutiérrez, & Hernández-Javier, 2020). Here we resort to times series models to make predictions only for the next 31 days. The ARIMA model e.g. describes one or more variables over time. It has forecasted exchange rates (Tseng, Tzeng, Yu, & Yuan, 2001), wind speed (Kavasseri & Seetharaman, 2009), and electricity price (Contreras, Espínola, Nogales, & Conejo, 2003). Moreover, there have been proposals worldwide to predict the pandemic's behavior (Benvenuto, Giovanetti, Vassallo, Angeletti, & Ciccozzi, 2020; Singh et al., 2020). Box and Jenkins (1994) developed ARIMA, a statistical model for the time series, where each observation value is a function of previous values. ARIMA consist of three components: 1) autoregressive (AR), 2) integrand (I) and 3) moving average (MA). Sometimes the model requires seasonal components turning into a SARIMA model. The ARIMA model predicts time series values based on historical behavior without counting the linked dependent variable's underlying factors. The p, d, q, values must be assigned appropriately to model the behavior of the time series and then select a reduced set to try to adjust the series. The ARIMA model is composed of 3 values (p, d, q), p represents the value of the autoregressive component (AR), d corresponds to the order of the integrand component (I), and q is the order value of the moving average (MA). ARIMA models are expressed as:

$$Y_t = \varphi_1 Y_{t-1} + \cdots \varphi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \quad (2)$$

Where:
$\boldsymbol{\varphi}$ is the autoregressive coefficient
$\boldsymbol{\theta}$ moving average coefficient
$\boldsymbol{\varepsilon}$ error
$\boldsymbol{Y_{t-1}}$ normalized series value
AIC (Akaike´s Information Criterion) measures the goodness of fit
BIC (Bayesian Information Criterion) tends to choose smaller models

We use the ARIMA implementation available in the forecasting package of the R programing language. R is a programming language for statistical computing. (Hyndman & Khandakar, 2008). We tried the model in five different intervals along the period in question and we found three episodes in which ARIMA model correctly predicted deaths within the next 31 days.

Results and Discussion

a) Forecasts

Interval 1. We fed the model with numbers published by the health authorities from February th 27th until September 14th which makes a total of 201 observations to predict accumulated deaths from **September 15th to October 15th**. The results are shown in table 1. As can see, ARIMA model has one integrated component equal to 2, or a differencing order of 2 and a moving average component (ma1) of order one equal to -0.9264; a seasonal autoregressive component (sar1) of order one equal to 0.8682 and two seasonal moving average components (sma1 and sma2), each of -0.7183 and 0.1078 with a seven day-week frequency. This optimal model has an AIC of 2708.55, AICc of 2708.86, and a BIC of 2725.02, speaking for the best model.

As Fig 1 displays, ARIMA predicted 84,496 deaths vs. 85,285 officially reported deaths. That makes an error of -0.93%, which is quite good.

Interval 2. We fed the same model with updated numbers until October 14th gathering a total of 231 observations to predict once more accumulated deaths from **October 15th to November 24th** 2020. We obtained the results shown in table 2. The ARIMA model has one integrated component equal to 2, or a differencing order of 2 and a moving average component (ma1) of order one equal to -0.92501; a seasonal autoregressive component (sar1) of order one equal to 0.80849 and two seasonal moving average components (sma1 and sma2), each of -0.7589 and 0.2937 with a seven day-week frequency. This optimal model has an AIC of 3126.43, AICc of 3126.7 and a BIC of 3143.2.

As we can see from Fig 2, ARIMA predicted 104,757 deaths vs. 102,739 officially reported deaths on November 24th, making an error of -1.95%, which is also quite good.

Interval 3. We fed the same model with updated numbers until November 24th, making a total of 271 observations to predict once more accumulated deaths from **November 24th to January 4th**. The results are displayed in table 3. The ARIMA model has one autoregressive component of 0.9577; one integrated component equal to 2, or a differencing order of 2; three moving average components, each of -1.77, 0.615 and 0.1678 and two seasonal moving average components of 0.166 and 0.160 with a frequency of 7 days. This model has an AIC of 3676.24; AICc of 3676.6 and BIC of 3701.43.

As we can see from Fig 3, ARIMA predicted 123,976 deaths vs. 127,757 officially reported deaths on January 4th. That makes an error of -2.96%, which falls within the tolerance of 5%.

b) EDA on Covid-19 deaths

From the first death occurring on February 27th until December 31st, 2020, a total of 125,807 people died of Covid-19 in Mexico. As we can see from Fig. 4 and table 4, daily deaths were reported mostly on Tuesday (19%), followed by Wednesday (18%), Thursday (16%), Friday (16%), Saturday (13%), and Sunday (7%). The highest number of deaths reported on Monday took place on October 5th with 2789 (an outlier depicted in Fig. 4); Tuesday reported its largest number on 29th of December with 990; Wednesday, 1092 on the 3rd of June; Thursday, 910 on the 31st of December; Friday, 794 on the 7th of August; Saturday, 784 on the 1st of August and Sunday, 1044 on 21st of June (an outlier depicted in Fig 4). The highest mean was detected on Tuesday with 537.1 followed by Wednesday with 515. Sundays perform in general with the lowest number of deaths (mean of 204.8) and show an uptrend slope. 53% of the deaths were registered between Tuesday and Thursday and the rest between Friday and Monday. Almost all days of the week, except Saturdays, soared in terms of deaths by the end of the year; showing Tuesday, Wednesday and Thursday the sharpest spikes.

Fig. 5 represents confirmed cases and deaths (left) and fatality rate (right) per month along 2020. According to Ximenes Fivie Laurie, the fatality rate worldwide is 2.2% (Ann, 2021). This indicator results from dividing the total deaths by the total of confirmed cases. The fatality rate changed as the numerator and the denominator took off at different pace. For instance, in March the number of cases was 41 times the number of deaths, but since the number of cases was not augmenting at the same pace as the number of deaths, the fatality rate increased constantly reaching its highest rate in June with 13%, or just 7 times the number of cases in comparison with the number of deaths. From that month on, the number of cases soared more rapidly with respect to deaths making the fatality rate descend to 6% in December 2020, when the number of cases were 13 times the number of deaths. In January, this indicator goes up again to 7%, and the number of cases were 8 times the number of deaths. Thus, the slower the number of cases increases with respect to the number of deaths, the higher the fatality rate and the sooner the number of deaths increases with respect to the number of cases the lower the fatality rate.

Fig. 6 represents the number of days needed to gather 10,000 deaths. Thus, it took 75 days to reach the first 10,000 deaths; 18 to get 20,000; 15: 30,000; 17: 40,000; 16: 50,000; 16: 60,000; 20: 70,000; 24: 80,000; 23: 90,000; 22: 100,000; 17: 110,000, 17: 120,000; 14: 130,000; 10: 140,000; 8: 150,000 and so on. The continuous line represents the average daily deaths for every interval; so to a smaller number of days corresponds a higher number of daily deaths in average. We can see a smoothing effect between the 60,000 and 90,000 deaths, but from the 100,000 deaths on, the pandemic soared reaching a maximum level as the number of deaths turned to 150,000 in which 10,000 people died in a period of only 8 days and 1254 daily deaths in average. Thus, we can see two waves of the pandemic in terms of average daily deaths: 665 from June 20th to July 4th (15 days) and the second 1254 along 8 days from January 17th to January 25th, 2021.

The total number of deaths until the 31st of December were concentrated in 7 out of 32 federal entities, making it 51% of the total (see Fig. 7). These were the State of Mexico, Mexico City, Veracruz, Jalisco, Puebla, Baja California and Guanajuato. The least affected entities were Nayarit, Campeche, Colima and Baja California Sur.

As we can see from Fig. 8, the federal entities with the highest number of deaths per 100,000 people on December 31st 2020 were Mexico City: 181.91; Baja California: 146.36; Sonora: 141.30; Sinaloa: 139,74 and Tabasco: 133.19. On the other hand, the federal states with the least deaths per 100,000 people were Michoacán: 58.06; Oaxaca: 51.45 and Chiapas: 21.66; country wide the average was 99.83. With that number and according to the John Hopkins University of Medicine (The John Hopkins University School of Medicine, 2020), Mexico took second place in Latin America after Peru, which had 117.46 on December 31st. We can also compare this numbers by taking a look at other countries in this regard by that time, e.g. UK: 183; Italy: 165; USA: 158; Spain: 152; France: 135; Sweden: 129; Swiss: 116; Austria: 97; Holland: 93; Germany: 86 and Israel: 68 (Schieritz, 2021: 2)

The entities that reported the largest growth per 100,000 people from June the 23rd to December the 31st 2020, were San Luis Potosí: 26.8 times; Coahuila: 25; Zacatecas: 22; Nuevo León: 20.4 and Guanajuato: 19.1 times. The Federal States with the least increments per 100,000 people were Morelos: 2.3; Mexico City: 3; Chiapas: 3 und Baja California North: 3.1.

c) Correlation with socioeconomic variables

According to an article published by the German newspaper Die Zeit, the Robert Koch Institute found that Covid-19 was imported to Germany by wealthy people travelling around the world and latter transmitted to poor people (Mayr, 2020). If this was also the case in Mexico, Covid-19 most be correlated with socioeconomic variables at different moments. Mexico has different levels of development and the best living conditions are located in states whose share in the country´s GDP is the highest. We correlated several socioeconomic indicators with Covid-19 deaths per 100,000 people on December 31st and on April 30th, 2020. The socioeconomic indicators chosen for this purpose were: Extreme Poverty; Labor Poverty; Labor Income per Head; Remittances and Homicides. We converted original numbers in logarithms to standardize the data.

$$Y_{Log(CCH\_31Dec)} = \beta_0 + \beta_{Log(Ing\_Lab\_Pca)}$$

According to table 5, on December 31st for every unit of income labor per capita, the number of deaths per 100,000 people increased on average by 0.7662.

$$Y_{Log(CCH\_30ABR)} = \beta_0 + \beta_{Log(Health)} + \beta_{Log(Rem\_20)} + \beta_{Log(Hom\_2019)}$$

According to table 6, on April 30th 2020 for every unit health infrastructure increased, the number of deaths per 100,000 augmented on average by 1.8647; for every unit remittances went up, the number of deaths per 100,000 people diminished of average by -1.0317 and finally, for every unit the homicides increased, the number of deaths did it in average by 0.4911.

$$Y_{Log(Inc\_Jun\_Dec)} = \beta_0 + \beta_{Log(Health)} + \beta_{Log(Rem\_20)}$$

Concerning the increment of deaths per 100,000 people from June to December 2020, we found that for every unit the health infrastructure increased, the number of deaths per 100,000 people diminished by 2.5291 and for every unit the remittances augmented, the number of deaths per 100,000 grew by 0.3336. This findings deserve further research.

Caveats and flaws on the Covid-19 official numbers. The information published by the Mexican health authorities (*Secretaría de Salud*) have been questioned by experts. According to Laura Ximenez, Hugo Lopez Gatell, responsible to conduct the government policy to tackle the pandemic, has been misinterpreting the real menace of Covid-19 for the population (Ann, 2021). From the beginning the model chosen to estimate the contagiousness and fatality of the illness was "*Centinela*". This model calculates the number of people infected without conducting a PCR-test. The lack of resources impeded the government to conduct massive tests to spot infected people and made them stay home. Only 475 clinics out of a total 26,000 operating medical centers administered by the Health Ministry made PCR-tests properly. Arturo Erderly, a mathematician working at UNAM, published an article outstanding the most important inconsistencies of official reports, concerning the calculations of fatality rates and contagious rates. For instance, authorities stated that contagious rate resulting from the ratio of estimated cases divided by confirmed cases, was 8.3 times. Erderly pointed out that this number was the result of taking 26,519 estimated cases, reported on March the 28th, divided by 3,181 confirmed cases reported on April the 8th, which is not consistent. Once the flaw was corrected, the number became 31.3 times (Erderly, 2020). That made a big difference and gave the public a more objective idea of the pandemic´s dimension. Furthermore, the number of victims reported every day gathered people who died at least 20 days before. Rojas Gonzalez explains that neither the number of deaths nor the number of contagions were accurate. The fatality rate was not 10.8% like the authority said around August 2020, but 2%, if we consider a more realistic number of people carrying the virus, either with symptoms or without symptoms. For instance, China reported by that time a fatality rate ranging from 0.5% to 1.5%, and the USA, 2.6% (Gonzalez, 2020; The Economist, 21 March 2020).

It is believed that testing the highest number of people is the only way to efficiently reduce contagiousness and deaths (Ann, 2021). It is known that there are two diagnostic tests: one using mucus or saliva and the second one with a blood sample (Duncan, 2020). For instance, In Germany a study of the Technische Universität in Berlin, showed that by implementing massive quick tests the pandemic´s contagious velocity slows down rapidly. The German government was planning to conduct massive tests that could spot highly contagious people and isolate them so that the contagious chain could be substantially cut. Quick tests should be available in schools, factories, businesses, drugstores, etc. Michael Mina, a US epidemiologist stated that quick tests could have a similar effect as a vaccine, if 50% de population would do the test twice a week. However, the disadvantage is that if a quick test gives positive, the patient has to back it up with a PCR-test, which is more reliable but neither rapid nor cheap. The Robert Koch Institute points out that an antigen-Test requires a high level of virus content in order to turn positive, which means that a negative result given by a quick test is not very reliable after all (Menne, 2021). In this regard, the Mexican authorities saw no reason to test people because knowing or not knowing who was infected made no difference since by that time there was no other cure than the natural inmune system. That means the people were supposed to be resilient and thus, restrictions of any kind were obsolete. In other words, they resorted to the *immune herd* solution. According to this principle, 70% of the population in Mexico needed to be infected or approximately 89.6 million people. This strategy turned to be macabre if we take a fatality rate ranging from 2.2% to 3.9% average, because that would mean that authorities were preparing to accept around 1.97 or 3.5 million deaths. Another mistaken policy was to declare that using a mask was not necessary and that asymptomatic people were not in the position to transmit the virus. For Favie-Laurie that was the evidence of the government wanting to have the highest number of people with Covid-19 so that the herd immunity effect could work.

The health authorities decided to set the number of beds and ventilators available to attend Covid-19 patients as an indicator to prevent crowds in hospitals. However and in order to keep this indicator low, many people were not admitted in those clinics so they had to go back home because their symptoms were not considered serious. Therefore, it was believed that many of those died at home and were not included in the official statistics. Ann explains that a person with Covid-19 can infect three in her surroundings. She states that 40% of all infected persons are asymptomatic; 40% don´t present symptoms at the beginning but afterwards and in the meantime, they can transmit the disease to several other people. From this 80% asymptomatic people, 40% are transmitting the disease permanently until they heal whereas the other 40% are contagious along 7 days. The illness remains in humans 10-20 days before it disappears, but among people with cancer it can last from 30 to 45 days (Ann, 2021: 432).

Conclusion

The way the authorities managed the pandemic during 2020 in Mexico was not correct. They were not prepare to tackle a phenomena of this magnitude. The shortcomings were numerous e.g., the lack of capacity to conduct PCR-test in proper numbers; to estimate the number of infected people by the Centinela model; to report the number of deaths with setbacks and imprecisions and to enforce the herd immunity by ignoring the basic protocols like wearing a mask and conducting massive quick-tests.

So far, the only official numbers to estimate the dimension of the pandemic are those published by the authorities. If they were true, several conclusions could be drawn: a) it is possible to predict the deaths within the next 31 day using times series ARIMA model; b) deaths are concentrated in two days of the week: Tuesday and Wednesday; c) the fatality rate changes as the confirmed cases move at a different velocity in relation to deaths; d) there were two waves by which the average daily deaths reached a maximum of 665 in summer and 1254 in winter; e) Mexico ranged in the second place in Latin America by the end of the year with 99.83 deaths per 100,000 people and f) deaths per 100,000 people are correlated with income labor per capita, health infrastructure and homicides. This EDA is the first step to conduct further analysis in regards to Covid-19. Hopefully, the results shown here will help to formulate more accurate hypotheses concerning the pandemic.

References

Ahmed, A. (2020, September 23). *In the Epicenter of Mexico´s Epicenter, Feeling like a "Trapped Animal"*. The New York Times, págs. https://www.nytimes.com/2020/09/23/world/americas/mexico-coronavirus-covid-iztapalapa.html?action=click&module=Well&pgtype=Homepage&section=World%20News.

Ann, X.-F. L. (2021). *Un daño irreparable. La criminal gestión de la Pandemia en México.* CDMX: Planeta.

Arias Velasquez, R. M. (2020). Forecast and evaluation of COVID-19 spreading in the USA wieh reduced-space Gassian progess regression. *Chaos, Solitions an Fractals, 136, 109924*, https://doi.org/10.1016/j.chaos.2020.109924.

Benvenuto, D. G. (2020). Application of the ARIMA model on the Covid-2019 epidemic data set. *Data in Brief, 29, 105340*, https://doi.org/10.1016/j.dib.2020.105340.

Box, G. E. (1994). Time Series Analysis: forecasting and Control. *Journal of Times Series Analysis (Vol. Third)*, https://doi.org/10.1111/j.1467-9892.2009.00643.x.

Briefing. (2020, September 26). The covid-19 pandemic is worse than official figures show. *The Economist*, https://www.economist.com/briefing/2020/09/26/the-covid-19-pandemic-is-worse-than-official-figures-show.

Contreras, J. E. (2003). ARIMA models to predict next day electricity prices. *IEEE Transactions on Power Systems*, https://doi.org/10.1109/TPWRS.2002.804943.

Dearing, F. H. (1979). *Exploratory Data Analysis.* California: SAGE publications.

Duncan, A. (2020, October 7). 3 Types of COVID-19 Tests, and Which One Is the Most Accurate. *Explore Health*, págs. https://www.health.com/condition/infectious-diseases/coronavirus/covid-19-test-types.

Erderly, A. (2020, April 19). Destapa Doctor de la UNAM Fraude Estadístico de López Gatell. *Demócrata*, págs. https://democratacoahuila.com/2020/04/19/destapa-doctor-de-la-unam-fraude-estadistico-de-lopez-gatell/.

Gobierno de Mèxico. (2020, August, 25). *Semaforo Covid-19.* Obtenido de El semaforo de riesgo epidemiològico: https://coronavirus.gob.mx/semaforo/

Gonzalez, R. R. (2020, August, 20). *Modelos Matematicos y el Covid-19.* Obtenido de 3.ER ENCUENTRO DE INVESTIGADORES LASALLISTAS: https://youtu.be/T0WtJKRyKpM

Hernández Bringas, H. H. (2020, June). Mortalidad por Covid-19 en México. Notas preliminares para un perfil sociodemográfico. *Notas de Coyuntura del CRIM, No. 36*, 1-7.

Hyndman, R. J. (2008). Automatic Times Series Forecasting: The forecast package for R. *Journal of Statistical Software*, Retrieved from https://www.jstatsoft.org/article/view/v027i03.

Jeffers, J. (1994). The Importance of Exploratory Data Analysis Before the Use of Sofisticated Procedures. *Biometrics 50, September*, 881-883. https://doi.org/10.2307/2532803

Kavasseri, R. G. (2009). Day-ahead wind speed forecasting unisg f-ARIMA , models. *Renewable Energy, 34(5)*, 1388-1393 https://doi.org/10.1016/j.renene.2008.09.006.

Mayr, A. (2020, November 26). Politik. Die Angreibaren. *Die Zeit*, pág. 2.

Mena, R. V.-H.-B.-S. (2020). Using posterior predictive distribution to analyze epidemic models: Covid-19 in Mexico. *Physical Biology, 17(6),065001*, https://doi.org/10.1088/1478-3975/abb115.

Menne, S. R. (2021, März). Ein Land tested den Alltag. *Die Zeit, Wissen*, pág. 31.

Pham, H. (2020). On estimating the number of deaths related to Covid-19.

Mathematics, 8(5), 655 https://doi.org/10.3390/MATH8050655.

Schieritz, M. (2021, März). Wie Deutschland dasteht. *Die Zeit, Politk*, pág. 2.

Schuttler J., S. R. (2020). Covid_19 predictions Using a Gauss Model, Based on data from 2 April. *Physics, 2(2)*, 197-212. https://doi.org/10.3390/physics2020013.

Singh, R. K. (2020). The Prediction of Covid-19 Pandemic for top-15 Affected Countries using advanced ARIMA model. *JMIR Public Health and Surveillance, 6(2), e19115*, https://doi.org/10.2196/19115.

Staff, F. (2021, January 27). *Inegi reporta exceso de mortalidad: 184,917 decesos de enero a agosto de 2020*. Forbes Mexico. https://www.forbes.com.mx/noticias-inegi-exceso-mortalidad-184917-decesos-enero-agosto-2020/.

The Economist. (2020, March, 21). Pandemic. The Lockdown and the long haul. https://www.economist.com/briefing/2020/03/19/in-europe-and-around-the-world-governments-are-getting-tougher.

The John Hopkins University School of Medicine. (2020, December, 31). *Coronavirus Resource Center*. Retrieved from:Cases and Mortality by Country: https://coronavirus.jhu.edu/data/mortality

Torrealba-Rodriguez, O., Conde-Gutiérrez, R. A., & Hernández-Javier, A. L. (2020). Modeling and prediction of COVID-19 in Mexico applying mathematical and computational models. Chaos, Solitons & Fractals, 138, 109946. https://doi.org/10.1016/j.chaos.2020.109946

Tseng, F.-M., Tzeng, G.-H., Yu, H.-C., & Yuan, B. J. C. (2001). Fuzzy ARIMA model for forecasting the foreign exchange market. *Fuzzy Sets and Systems*, *118*(1), 9–19. https://doi.org/10.1016/s0165-0114(98)00286-3

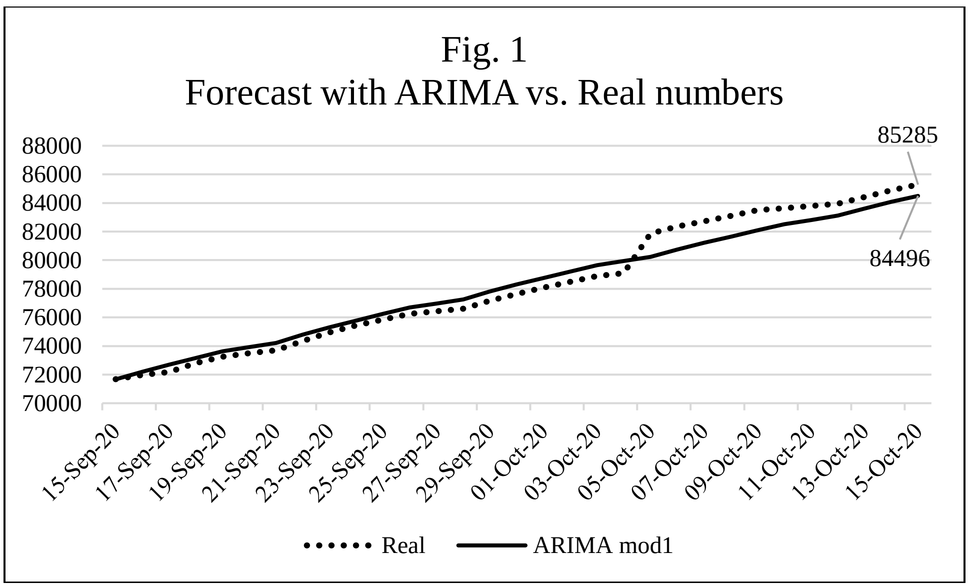
Yunus, M. S. (2020). Prediction of Cumulative Death Cases in the United States due to Covid-19 using mathematical models. *Journal of Enviormental Microbiology, 8 (1)*, 31- 41.

Figures and Tables

Table 1. Best ARIMA model

Best fit Model: **ARIMA(0,2,1)(1,0,2)[7]**				
	ma1	**sar1**	**sma1**	**sma2**
Coefficients:	-0.9264	0.8682	-0.7183	0.1078
s .e.	0.0225	0.0734	0.1061	0.0748
log likelihood=-1349.28				
AIC=2708.55	AICc=2708.86	BIC=2725.02		

Source: own calculations with Data of Health Ministry



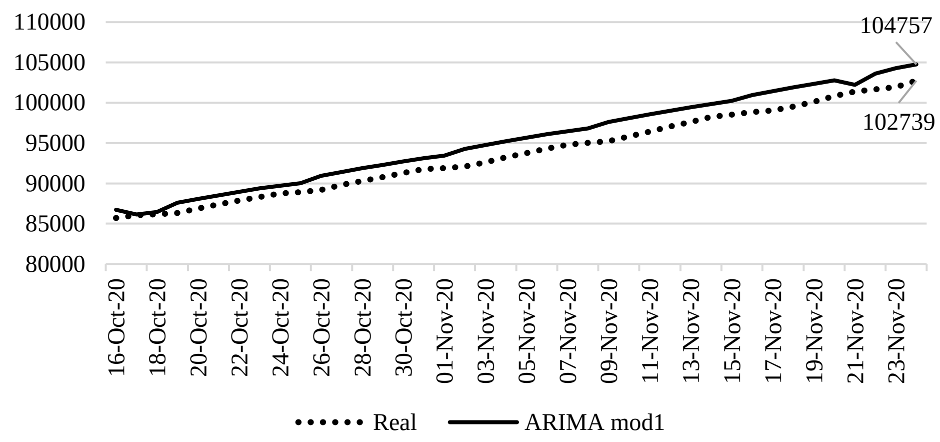
Source: own calculations with Data of Health Ministry

Table 2. Best ARIMA model

Best fit Model: **ARIMA(0,2,1)(1,0,2)[7]**				
	ma1	**sar1**	**sma1**	**sma2**
Coefficients:	-0.92501	0.80849	-0.7589	0.2937
s .e.	0.0260	0.0957	0.1121	0.1101
log likelihood=-1558.21				
AIC=3126.43	AICc=3126.7	BIC=3143.2		

Source: own calculation with Data of Health Ministry

Fig. 2
Forecast with ARIMA vs. Real numbers



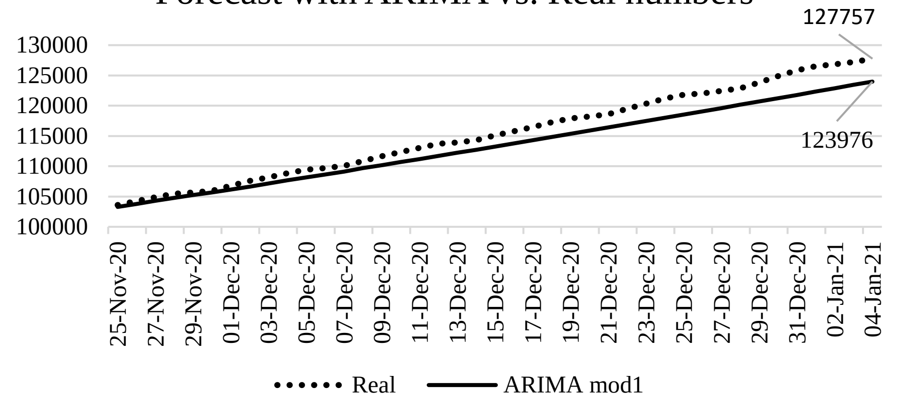
Source: own calculations with Data of Health Ministry

Table 3. Best ARIMA model

Best fit Model: **ARIMA(1,2,3)(0,0,2)[7]**						
	ar1	**ma1**	**ma2**	**ma3**	**sma1**	**sma2**
Coefficients:	0.9577	-1.77	0.615	0.1678	0.166	0.160
s .e.	0.0261	0.063	0.1212	0.0614	0.063	0.055
log likelihood=-1831.12						
AIC=3676.24	AICc=3676.6	BIC=3701.43				

Source: own calculations with Data of Health Ministry

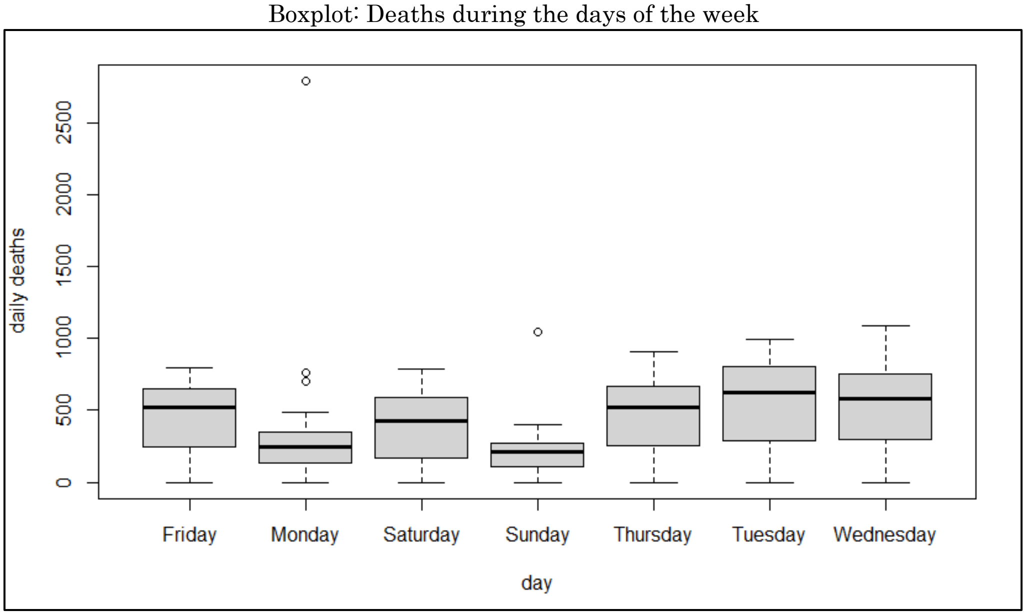
Fig. 3
Forecast with ARIMA vs. Real numbers



Source: own calculations with Data of Health Ministry

Fig. 4

Boxplot: Deaths during the days of the week



Source: own calculations with Data of Health Ministry

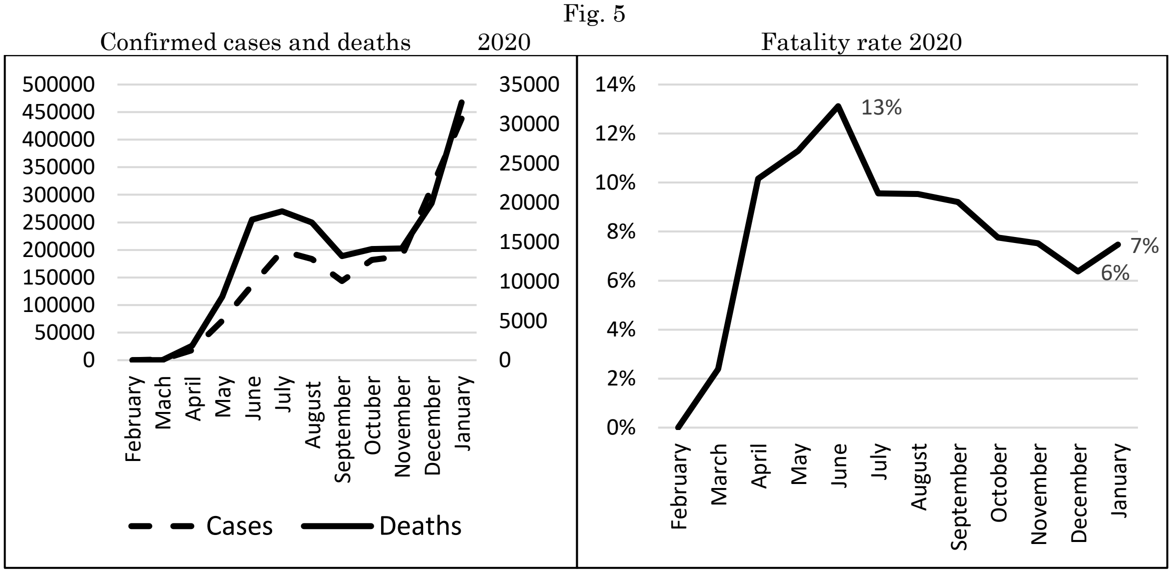
Table 4
Basic statistical indicators

	1st Quartile	Median	Mean	3th Quartile	Maximum	Total
Monday	145.5	243	305.2	347.2	2789 (5th Oct)	13435 (11%)

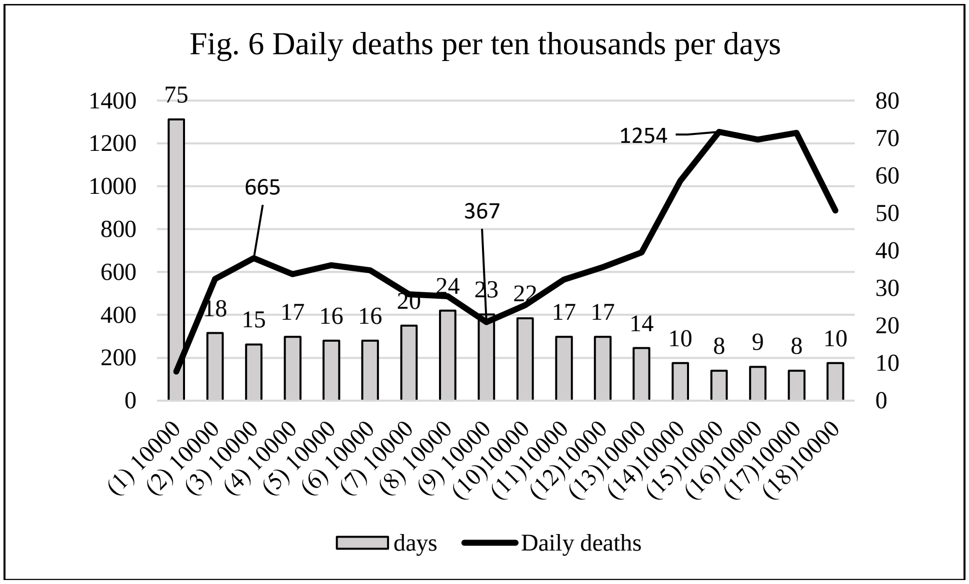
Tuesday	309.5	623	537.1	804	990 (29th Dec)	23632 (19%)
Wednesday	298.5	577	515	748.2	1092 (3rd Jun)	22678 (18%)
Thursday	257	518	456.8	667	910 (31st Dec)	20557 (16%)
Friday	267.2	519	446.4	648.8	794 (7th Aug)	19640 (16%)
Saturday	178	422.5	383	590	784 (1st Aug)	16853 (13%)
Sunday	111	214.5	204.8	274.5	1044 (21st Jun)	9012 (7%)
31st December, 2020					total	125807 (100%)

Source: own calculations with Data of Health Ministry

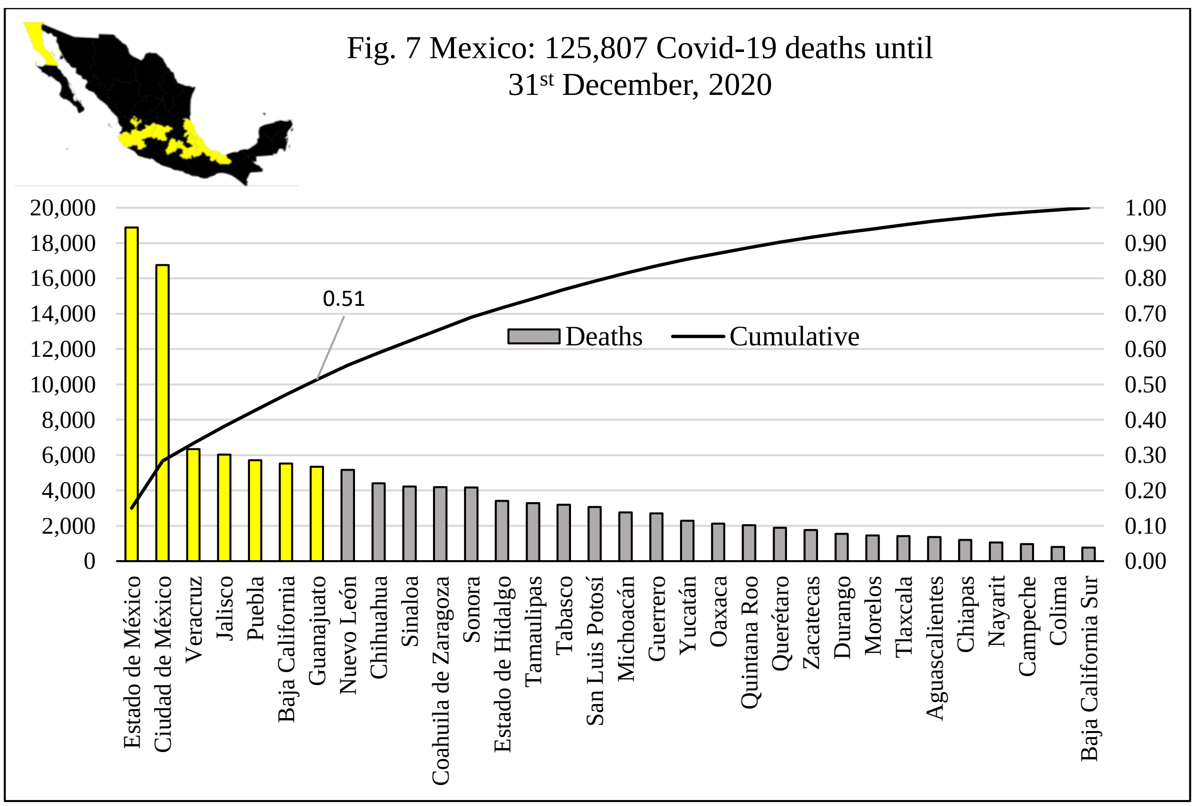
Fig. 5



Source: own calculations with Data of Health Ministry

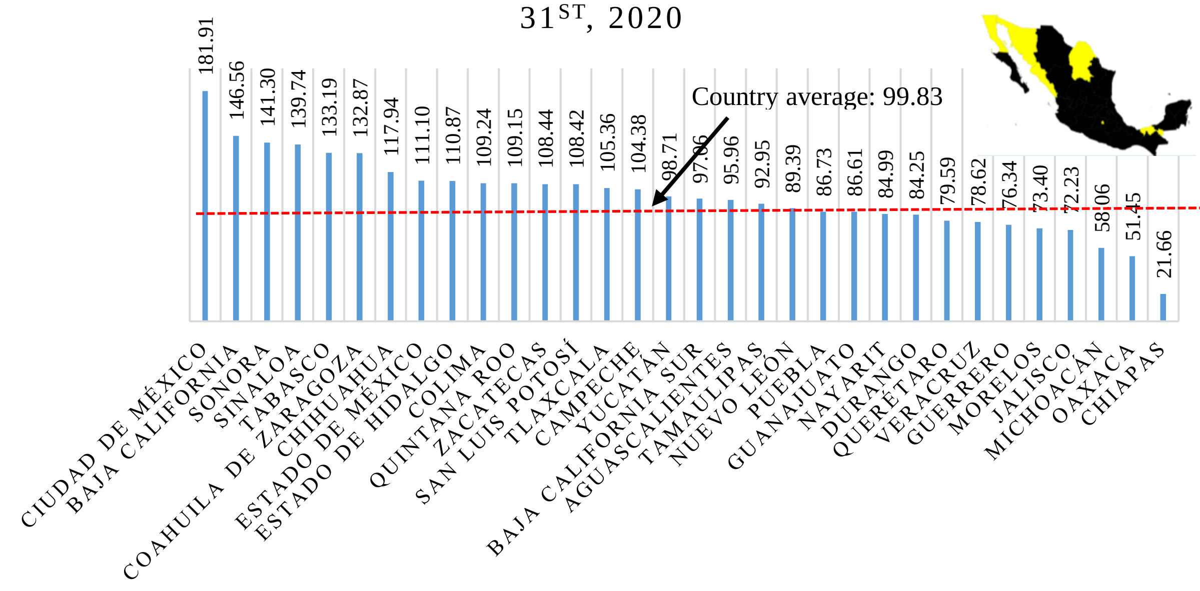


Source: own calculations with Data of Health Ministry



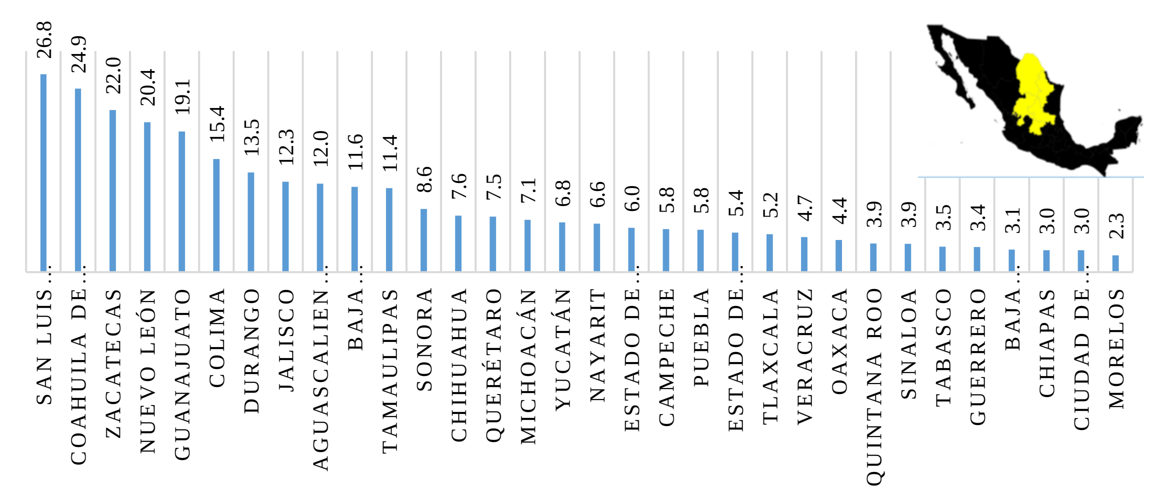
Source: own calculations with Data of Health Ministry

FIG. 8 DEATHS PER 100,000 PEOPLE ON DECEMBER 31ST, 2020



Source: own calculations with Data of Health Ministry

FIG: 9 INCREMENT IN DEATHS PER 100,000 PEOPLE FROM 23RD JUNE TO 31ST DECEMBER, 2020



Source: own calculations with Data of Health Ministry

Table 5
Regression´s results

31st December, 2020	Coeff	Std.Error	t value	Pr(>|t|)	Significance
Intercept	-1.4404	1.4180	-1.016	0.03148	
log(Ing_lab_Pca)	0.7662	0.1814	4.223	0.000206	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 0.3046 on 30 degrees of freedom Multiple R-squared: 0.3729, Adjusted R-squared: 0.352 F-statistic: 17.84 on 1 and 30 DF p-value: 0.000206					

Source: own calculations with Data of Health Ministry

Table 6
Regression´s results

30th April, 2020	Coeff	Std.Error	t value	Pr(>|t|)	Significance
(Intercept)	-1.3040	1.7421	-0.747	0.4612	
log(Health)	1.8647	0.7755	2.405	0.0230	*
log(Rem_20)	-1.0317	0.2259	-4.567	9.06e-05	***
log(Hom_2019)	0.4911	0.1816	2.2704	0.0115	*
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 0.7789, Adjusted R-squared: 0.3848 F-statistic: 7.463 on 3 and 28 DF, p-value: 0.000804					

Source: own calculations with Data of Health Ministry

Table 7
Regression´s results

Increment from June to December, 2020	Coeff	Std.Error	t value	Pr(>|t|)	Significance
(Intercept)	6.4731	1.1746	5.511	6.14e-06	***
log(Health)	-2.5291	0.5160	-4.901	3.34e-05	***
log(Rem_20)	0.3336	0.1153	2.894	0.00715	**
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 0.5264, Adjusted R-squared: 0.4172 F-statistic: 12.1 on 4 and 27 DF, p-value: 0.0001513					

Source: own calculations with Data of Health Ministry