Análisis y estimación gráfica del comportamiento del COVID-19 en Colombia, Santa Marta y Cartagena enfocado a la letalidad

Analysis and Graphic Estimation of the Behavior of Covid-19 in Colombia, Santa Marta, and Cartagena Focused on Lethality

Jorge Gómez-Rojas; Jeraldyn Redondo-Pérez; María Tovar-Hernández

Introduction

COVID-19 is an infectious disease caused by type 2 coronavirus, discovered in late 2019, responsible for severe acute respiratory syndrome, better known as SARS-CoV-2, which affects the respiratory system with frequent symptoms such as nasal congestion, choking sensation, and intense cough [1]. In addition, it can spread rapidly in the population because it can enter through the nostrils, eyes, or mouth to infect a person, in response to this protective measures have been taken as social distancing, hand washing, and use of masks that cover the nose and mouth. This highly contagious disease due to its zoonotic capacity originated in the Wuhan market, the People's Republic of China, which reported on December 12, 2019, that there were 27 cases of patients with a febrile respiratory condition [3]. However, it was not notified to the WHO until December 31, 2019, so that the next day the Wuhan market is closed and finally on January 9, 2020, is published the existence of Coronavirus by the People's Republic of China. Subsequently, the virus spread in
such a way that on January 13 the first case was reported in Thailand, i.e. it had already migrated to other places in the region. Finally, on January 30, WHO declared the Coronavirus as a global health threat [2].

Approximately one month later, exactly on February 26th, the first case was published in South America, specifically in Brazil, so that all countries in the region including Colombia began to prepare to face the arrival of possible cases of this virus to the nation [3]. Even though Latin American countries in comparison with most European and Asian countries have higher levels of poverty and health centers were not well prepared to face a pandemic, they managed to take advantage of the time it took to arrive, taking preventive measures, such as mandatory quarantine of travelers coming from other countries, border closures and airport closures [4]. The virus arrived in Colombia on March 6, 2020, when Europe was experiencing the most critical moments of the Pandemic. By that date, the average number of people infecting others in the country was 2.28, relatively lower than in other countries in the region, and on March 23 a national quarantine was decreed [5]. Months earlier the Pandemic had reached Europe, leaving among the most affected countries Italy, which is the European state with the highest number of deaths so far, December 2020, an estimated 55,000 [6]. European countries were hit hard by the Pandemic due to several factors, including the fact that the majority of the population is elderly, it has been estimated that by 2050 36% of the population will be older than 54 years [7], and this virus has a high capacity to lethally affect the elderly and/or people with comorbidities [8]. Among the most affected countries was Spain, which implemented the Daily Mortality Monitoring System (MoMo), which is responsible for estimating mortality for the following days based on death data from January 1, 2008, to one year before the date of analysis, this was very useful because it was able to show the "excess deaths" [9]. Most of the cases from Latin America were imported and originated from the northern cities of Italy [1].

The appearance of COVID-19 in the world has not only brought a new lifestyle but also many technological elements have been created for the control and mitigation of this. Among these, there have been mobile applications focused on the pandemic, which have different objectives, from the daily display of the numbers of infections and death to the tracking of contacts or exposures of those infected. Among these apps we have "TraceTogether", which is responsible for tracking the exposures of an infected person with people around him, was created by the government of Singapore [10], another is one developed in Spain by a research group of the University of Santiago de Compostela (USC) which is called "Modestya", this app can predict the behavior of Coronavirus in 3 days, according to data taken 15 days before [11]. Likewise, a mobile application was created in Colombia that aims to track the health status of people in search of possible asymptomatic cases in the population or mild symptoms, this in the company of the Polymerase Chain Reaction (PCR) tests and thus stop the spread of the virus. This mobile application provides users with updated information on the progress of diagnosed cases and deaths in the country, department, and municipalities, as well as offers educational aids to ensure timely care for the well-being of the individual and the collective benefit [12]. In addition, Colombia is also carrying out an analysis called "COVID-19 coronavirus transmission model" which aims to make the adjustment to a SIR model with a probabilistic component and to help eradicate the virus, estimates possible transmission scenarios in the country [13]. On the other hand, three students from different recognized universities in the country created a "Polynomial regression application for the characterization of the curve of COVID-19, using Machine Learning techniques", which allows the treatment of data by polynomial regression leading to obtaining a prediction of the future behavior of the data [14]. Likewise, a "COVID 19 APP" was developed with MATLAB software, which allows an analysis of cases in the country of interest, as well as curve fitting for the coming days [15]. Likewise, collaborators from different countries created an interactive APP that allows us to manage very easily because it offers
us the world map with the different figures of total cases, recovered and deceased instantly and compares through graphs the figures of cases, deaths, and other specific countries [16]. The INS of Colombia also has a web page, very dynamic, which covers a lot of information through different types of graphics, from which we can obtain figures of cases and deaths, classified by regions, gender of people, and others [17]. These Apps have been developed because the information systems and digital media allow us to address the data on time, as the situation merits it, thanks to which prevention and control measures can be taken in time [18].

The excessive number of deaths from Coronavirus since its inception is evidence of its high degree of lethality. However, comparing this with previous pandemics in history, it does not manage to be the most deadly, because to date the COVID-19 has claimed 1.5 million lives [6], and although it is a considerably large figure, the Black Death left approximately 200 dead [19]. In Colombia, when the virus had just arrived, an analysis was conducted to discern how it would proceed, having observed its behavior since its beginnings in Europe and the factors that had made its population more susceptible to have higher numbers of cases and deaths, among them were differentiated: the geographical and climatic conditions, the percentage of the population at risk, among others [20].

In this article, we will show the operation of an App that allows observing the prediction of the behavior of the curve having as a base tool the use of MATLAB and the commands polyfit and polyval for the adjustment and extrapolation of the curve using the polynomial regression, according to what has been investigated in Google Scholar some investigations use this method, but with a different technique. As well as the implementation of the use of mathematical-computational models to examine how the behavior by COVID-19 develops, in what time it will develop and what levels of infection it can reach, is vital for the control of infection, the good use of resources and health services and the reopening of normal daily life [21]. For the development of these mobile applications, prediction techniques must be taken into account, which aims to know future values using the information observed up to the present time, among them, we find the simple exponential smoothing, the winters method, and the hold method [22]. One of the mathematical bases for the realization of this project was the Gaussian function or normal distribution, which models the figure of a bell and requires data such as standard deviation and mean [23], this was applied in conjunction with the polynomial regression that was performed in a way that modeled the curve, because this, being the number of deaths and daily cases, at some point should fall by the behavior of the pandemic.

**Materials and Methods**

**A. Data collection, first stage.**

To obtain the necessary data to model the covid-19 curve in Colombia, Santa Marta and Cartagena were exported the reports on the official page of the Ministry of Health of Colombia [24]. The most relevant data are the cases and deaths accumulated each day, which are treated interactively to give an organized and compact form.

For the analysis of these, they were plotted cumulatively and individually (bar graph).

**Mathematical modeling, second stage.**

To estimate the behavior of deaths from Covid-19 in Colombia, two mathematical models were used, the
normal distribution and polynomial regression, which yield statistical values such as the mean and standard deviation that facilitate the treatment of the data. The normal distribution or Gaussian bell is a statistical model capable of approximating the value of a random variable to an ideal situation, in which the successive standard deviations concerning the mean establish reference values to estimate the percentage of observations of the data [25].

To work with this model we use (1) and (2).

\[
\int_{n-1}^{n} \frac{1}{\sqrt{(2\pi)\sigma}} e^{-\frac{1}{2}(x-\mu)^2} (1)
\]

Where:
- \(\mu\) is the average
- \(\sigma\) is the standard deviation
- \(x\) is the variable

\[
\int_{-\infty}^{+\infty} \left(\frac{1}{2\pi\sigma^2}\right) e^{-\frac{1}{2}(x-\mu)^2} \right) (2)
\]

Where:
- \(\mu\) is the average
- \(\sigma\) is the standard deviation
- \(x\) is the variable

Polynomial regression is the prediction of a quantitative response variable from a quantitative predictor variable, where the relationship is modeled as a polynomial function of order \(n\), polynomial models can be fitted by linear least-squares regression because, although they generate nonlinear models, their formulation is still a linear equation with predictors. [26] The objective of this is to determine values for the parameters of the equation which make the curve fit the data points better [27]. Therefore, starting from the linear model (3).

\[
: \beta_0 + \beta_1 x + \epsilon_i (3)
\]

A polynomial model of degree \(n\) is obtained from (4).

\[
Y_i = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \ldots + \beta_d x^d + \epsilon_i (4)
\]

B. Strategy for visualization and extrapolation, third stage.

In MATLAB, there are several tools for curve fitting and one of them is the "fit" command which allows you to fit a curve in various ways such as linear, cubic, and many more. To work with the normal distribution we used the "Gauss1" mode which is the Gaussian bell of order 1 that fits the curve and interpolates the missing data to shape it; at most it yields three variables that correspond to the components of the equation
that describes this behavior.

**General model Gauss1:**

\[
\text{func}(x) = a_1 \exp \left(-\left(\frac{x-b_1}{c_1}\right)^2\right)
\]  \hspace{1cm} (5)

where:

- \(a_1\) is the expression accompanying (1)
- \(b_1\) corresponds to the average
- \(c_1\) the standard deviation

As you can see (1) and (5) are equivalent.

The "polyfit" and "polyval" commands are also used, which are related to polynomial regression, which consists of adding curvature to the model by introducing new predictors that are obtained by raising the original ones to different powers.

The polyfit function returns the coefficients for a polynomial \(p(x)\) of degree \(n\) that is the best solution (in the least-squares sense) for the \(y\) data. This function also returns \(\mu\) which is a two-element vector with centering and scaling values, \(\mu (1)\) corresponds to the mean, and \(\mu (2)\) is the standard deviation [28]. For this process, the software uses (6).

\[
p(x) = p_1 x^n + p_2 x^{n-1} + \ldots + p_n x + p_{n+1}
\]  \hspace{1cm} (6)

Polyval allows you to give values to the previous result.

After fitting the curve to a polynomial of degree 7 which was the best fit to the data since the degree varies according to the dispersion of the data. This process was done with the cases and deaths. As for the approximation of the curve, as of December 31, the two commands were handled separately, both were adjusted to the normal distribution formula with certain variants, and only the estimate of daily deaths by COVID-19 was made.

An important tool in the extrapolation process is Matlab's anonymous function, which allows mathematical expressions to take on future values or be treated as a variable that depends on the values entered by the user.

By having the data reported up to the end of November, the number of days remaining until the end of the year was calculated and this figure was added to the number of data reported. With this, the data found were plotted. For example, in Colombia from the first case which was reported on March 6, 2020, until November 24 of the same year there are 249 reported days, from this last date until December 31 there are 37 days, therefore, for the estimation 286 evaluated and reported days will be taken into account.

The only variation of the process between the fit command and the polyfit is that for the second process (2) is used since it works more generally with polynomially analyzed data.
Results and Analysis

From the data tables with cases and deaths, we obtained the growth curve and the bar graph, which describe the behavior of this in Colombia, Santa Marta, and Cartagena.
The curve is then fitted concerning the two mathematical models mentioned above.
In this process we observe that in the two cities analyzed there is a similar behavior until November, being Cartagena the one that presents a range of high data. In general, in Colombia, the data is more uniform.

Concerning curve fitting, the Gaussian is a little more accurate in what you want to model, but the polynomial fits a little more to the real data because it can present more than one peak, while the Gaussian is faithful to its principle, which is based on an exponential rise followed by a peak and an exponential fall. Particularly, when trying to fit the curve polynomially, if very scattered data is presented the curve tries to get as close as possible to the real thing but it always moves away from little from the approximations, this is mainly because the equation that describes the polynomial is being adapted to the form of the gaussian bell and although the standard deviation and the mean are similar to the results of the first method, the expression that multiplies the complex exponential varies and modifies the amplitude of the graph, but even so, the polynomial curve adjustment provides a slightly slower decent speed, which is a little closer to the behavior of coronavirus mortality.

When the curve is fitted, it is extrapolated and shaped up to 31 December 2020.
Análisis y estimación gráfica del comportamiento del COVID-19 en Colombia, Santa Marta y Cartagena enfocado a la letalidad

**Figure 25.** Gaussian extrapolation for daily covid-19 deaths in Colombia. Normal distribution modeling by method 1 with data from the entire country.

**Figure 26.** Polynomial extrapolation for daily covid-19 deaths in Cartagena. Normal distribution modeling by method 2 with data from the entire country.

**Figure 27.** Gaussian extrapolation for daily covid-19 deaths in Santa Marta. Normal distribution modeling by method 1 with the city-data.
Figure 28. Polynomial extrapolation for daily covid-19 deaths in Santa Marta. Normal distribution modeling by method 2 with the city-data.

Figure 29. Gaussian extrapolation for daily covid-19 deaths in Cartagena. Normal distribution modeling by method 1 with the city-data.

Figure 30. Polynomial extrapolation for daily covid-19 deaths in Cartagena. Normal distribution modeling by method 2 with the city-data.
In the extrapolation, the graphs made by the Gaussian method present a sharper shape this is presented due to the equation that prioritizes the peak to predict how the graph goes down, but the polynomial extrapolation corresponds a little more to the real model of the coronavirus. Although, the polynomial analysis handles a very small scale that does not fit the real data.

**Conclusions**

With the results presented in this article we note that, the data could be successfully materialized and show the behavior of COVID-19 in Santa Marta and Cartagena along with the mortality associated with this virus, thanks to this we can see that being the two cities tourist and cultural districts, in addition to having different population density presented its first date of infection relatively close to that of the country, specifically in Bogotá D.C., but Cartagena also reported before the contagion had a higher peak and closer to the original date of diagnosis [24]. These differences are due to the flow of tourists in each city and the rapid spread of the virus from Bogota to these urban centers.

Emphasizing the extrapolation, it can be inferred that it is an ideal case because the graphs represent the approximate behavior of the curve if the mandatory quarantine were still in force and many economic sectors of society had not been reactivated. The graph shows that in December, deaths decrease notably. For this reason, the extrapolation is rather an estimate of deaths in Colombia until the end of 2020, because the curve is adapting to the shape of the mathematical model of normal distribution and represents the behavior of deaths with the guidelines and measures necessary to weaken the impact of COVID-19 on the health of Colombians; currently, the real curve has new peaks.

Similarly between Colombia and Spain, which are countries with a similar population (46.94 million for Spain and 49.65 million for Colombia); their behavior differs in many ways, Spain for its part has a rather irregular curve of deaths as it suffers unexpected fluctuations and even a constant period of low cases, however, then presents the highest peak. These differences are due to the first interaction of each country with the virus, the preparation they had to face it and the speed of response to the collapse of the health system, taking into account that in Spain the first case of COVID-19 was recognized in late January [29] and in Colombia, it was reported in early March. It is worth noting that even at their highest peaks there is a difference of almost 500 deaths (approximately 1000 for Spain and 400 for Colombia).

At an international level, the behavior of COVID-19 mortality in Argentina presents a curve very similar to the Gaussian bell, since they began with increasing mortality and upon reaching the peak began to decay exponentially, which shows that with appropriate response and prevention measures the curve can be shaped to the ideal situation, knowing that the first dates of infection between Argentina and Colombia are similar [30].

To conclude, this research shows that deaths from COVID-19 cannot be avoided because it is a utopian situation, but with strict measures by the control agencies, quality care, good management of medical resources, and essentially the compliance and responsibility of the prevention rules by the community, it is possible to reach a fairly low lethality.
References


