

Statistical Modeling of COVID-19 Mortality Trends in Iran

Ehsan Ghasemi¹ , Ali Khorshidi^{2,3} , Mehdi Omid⁴ , Kourosh Sayehmiri⁵ 

¹ School of Health, biostatistics department, Ilam University of Medical Sciences, Ilam, Iran

² Health and Environment Research Center, Ilam university of medical sciences, Ilam, Iran

³ School of Health, epidemiology department, Ilam University of Medical Sciences, Ilam, Iran

⁴ Department of mathematics, Ilam University, Ilam, Iran

⁵ Psychosocial Injuries Research Center, Ilam University of Medical Sciences, Ilam, Iran

Article Info

Article type:

Original article

Article History:

Received: Feb. 07, 2024

Revised: Mar. 13, 2024

Accepted: Jun. 12, 2024

Published Online: July. 21, 2024

✉ Correspondence to:

Kourosh Sayehmiri
Psychosocial Injuries Research
Center, Ilam University of
Medical Sciences, Ilam, Iran

Email:

kourosh86@gmail.com

ABSTRACT

Introduction: The proliferation of the COVID-19 virus has become a significant global public health concern. Iran was among the countries severely affected by this virus, facing considerable challenges in managing and treating COVID-19 infections. To implement more efficient care and prevention strategies, it is critical to understand the progression of the illness and the mortality rate of those affected. This study focuses on analyzing the mortality trends of COVID-19 patients in Iran. The primary aim is to apply statistical models to characterize and predict fatalities caused by COVID-19 in Iran.

Material & Methods: Data on COVID-19-related deaths in Iran were analyzed, encompassing the daily number of new cases and the cumulative number of cases reported between February 19, 2020, and May 15, 2022. The data were divided into six periods to develop more accurate models. Ten time series and regression models were fitted to the data, with the best model for each variable in each period identified using the coefficient of determination (R^2) index. The significance of the models was assessed using the F-test.

Results: Throughout the study period, the ARIMA (4,1,4) model and the cubic regression model were the time series models that best fit the mortality data. The cubic model provided the best fit during the first, second, third, fourth, and fifth periods, while the quadratic model was the best fit during the sixth period. For the cumulative death data, the cubic model was the most accurate.

Conclusion: The study's findings demonstrate that time series and regression statistical models can effectively model and forecast COVID-19 mortality data on both a daily and cumulative basis.

Keywords: COVID-19, Forecasting, best fit, regression models, Time Series Studies

➤ How to cite this paper

Ghasemi E, Khorshidi A, Omid M, Sayehmiri K. Statistical Modeling of COVID-19 Mortality Trends in Iran. J Bas Res Med Sci. 2024; 11(3):73-83.

Introduction

The outbreak of the COVID-19 pandemic has presented a significant global public health challenge, with Iran emerging as one of the severely impacted nations by the virus. Understanding the trajectory of deaths among COVID-19 patients through statistical modeling is crucial for guiding public health responses and healthcare strategies in Iran (1,2). As the number of COVID-19 cases continues to rise in Iran, there is a pressing need to assess and analyze the factors influencing the mortality rate of affected individuals. Utilizing advanced statistical modeling tools allows us to glean valuable insights into the patterns and dynamics of COVID-19-related deaths in Iran. Statistical modeling of COVID-19 patient mortality trends represents a critical area of research aimed at comprehending and forecasting the pandemic's progression. Predicting the trajectory of this disease is essential for adopting more effective preventive and management measures. The role of statistical modeling in comprehending and predicting trends and patterns associated with various phenomena, including the spread and impact of infectious diseases, is well established. Statistical models have been instrumental in analyzing diverse aspects of the COVID-19 pandemic, including trends in mortality rates, and in providing invaluable insights for public health responses (3-5).

Various studies have utilized statistical models to analyze data and predict the number of deaths related to COVID-19 (6). One study employed a mixed linear model to analyze data from multiple countries, including Peru, Brazil, Chile, Bolivia, Colombia, and Argentina (6). Another focused on statistical modeling of daily confirmed COVID-19 cases and deaths in Europe and the United States, emphasizing the importance of understanding disease trends to mitigate its impact (4,7,8). Additionally, the World Health Organization has been estimating excess mortality due to the COVID-19 pandemic, highlighting challenges in accurately reporting COVID-19 deaths and the importance of tracking

excess mortality to fully understand the pandemic's impact (9).

These studies underscore the critical role of statistical modeling in elucidating COVID-19 patient death trends and informing public health responses. Commonly used statistical models for analyzing COVID-19 death trends include linear mixed effect models (6), time series models (10,11), and models adjusted for population size (10,12). These models help assess the influence of various factors on death rates, evaluate intervention effectiveness such as social distancing measures, and forecast COVID-19-related deaths (13).

Numerous studies have effectively utilized statistical models to predict COVID-19 cases globally (12,14). For example, Fanelli and Piazza (2020) employed statistical and mathematical models to analyze and forecast the COVID-19 epidemic in China, Italy, and France (15). Their findings underscored the importance of prompt intervention to mitigate the virus's spread. Similarly, to predict COVID-19 cases in India, Khajanchi, Sarkar, and Mondal (2020) developed a modified SEIRD (Susceptible-Exposed-Infectious-Recovery) model (11). Their model provided accurate projections for different regions of the country, considering the impact of social distancing measures.

Yang et al. (2020) proposed a modified SEIR model to analyze the trajectory of the COVID-19 epidemic in China under various public health interventions (16). While their study focused on China, the methodologies and insights are relevant to the Iranian context as well. Yadav et al. (2020) utilized statistical methods to forecast the spread of COVID-19 in India, highlighting the critical importance of early detection and swift preventive measures (17).

In his article, Abdul Hosseini Morteza estimated the number of individuals expected to be affected by the sixth wave of the disease, considering the current circumstances (18). He highlighted the use of eigenvalue selection in optimizing the grouping

process of the SSA algorithm, which significantly improved the forecasted time series, as evidenced by the error index (18). The study demonstrated that the forecasting accuracy was satisfactory and indicated the viability of using the SSA method for predictions, comparing favorably against other techniques such as Neural Network Autocorrelation (NNAR), TBATS, Fractional ARIMA (ARFIMA), and Integrated Autoregressive Moving Average (ARIMA) (5).

In their research conducted by Zare Zahra and Vasegh Nastaran, it was asserted that employing a single model with uniform parameters across all regions is impractical due to the presence of multiple peaks in Iran's COVID-19 data, whereas the standard SIR (Susceptible-Infectious-Recovered) model assumes a single peak. To address this discrepancy, Iran's data was segmented into five distinct time periods, and specific parameters were estimated for each period accordingly (19).

Materials and methods

In this cross-sectional study, all individuals infected with COVID-19 across Iran from February 19, 2020, to May 15, 2022, were included. The study examined both the daily count of new death cases and the cumulative count of deaths within this population. Ethical approval for this research was obtained under registration code IR.MEDILAM.REC.1399.319.

Research population

Data for the study were extracted from the websites <https://behdasht.gov.ir> and <https://www.worldometers.info/coronavirus/>. These sources provided comprehensive data necessary for the analysis.

Statistical analysis

The variables in this study were characterized using tables, graphs, and descriptive statistics. Subsequently, various regression models including Linear, Quadratic, Cubic, Curve (S-), Growth, Compound, Exponential, Logarithmic, Inverse, Power, and Time Series were fitted to the data. The significance of the fitted models was evaluated using the F test. The model demonstrating the best fit to the data was selected based on the coefficient of determination index, and its formula was determined using the calculated parameters. Statistical analysis was performed using SPSS 25.

Results

The average daily death rate was 132 persons, totaling 141,224 persons who lost their lives (Table 1). The trend of COVID-19 deaths in Iran is illustrated in Figure 1, highlighting 6 distinct peaks, prompting the division of data into 6 time periods (Table 2).

Table 1. Descriptive Statistics of Daily COVID-19 Deaths in Iran (From the Beginning of the Pandemic to May 15, 2022).

Variable	Count	Mean	Median	Mode	Standard Deviation	Minimum	Maximum	Total
Death	816	173	132	63	142	1	709	141224

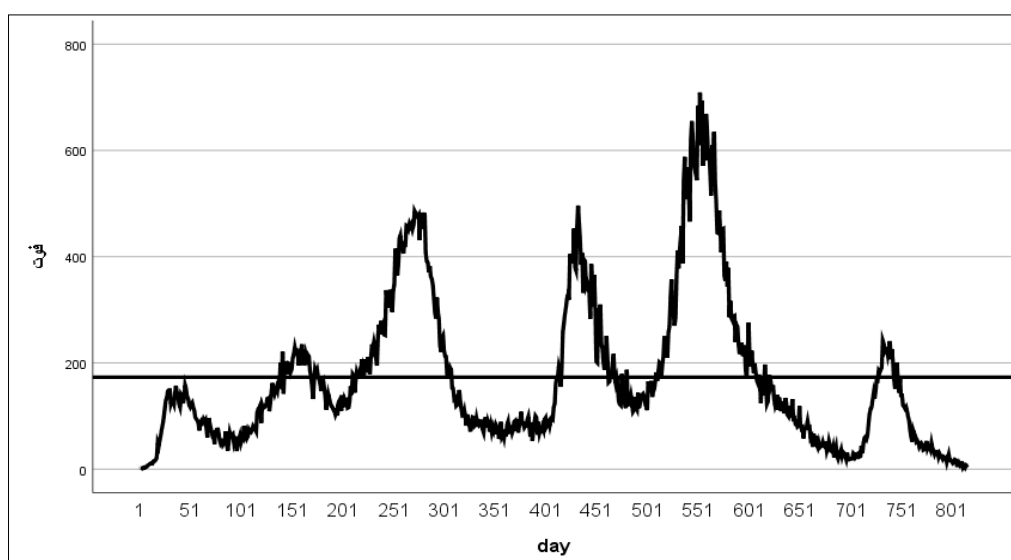


Figure 1. Death Trends in People with COVID-19 in Iran from the Beginning of COVID-19 to May 15, 2022

In each period, a consistent upward trend was observed until reaching the maximum point (peak) on the graph, followed by a decline to a minimum level. It's important to note that each period differed in terms of its peak and trough values, as well as its

duration. By dividing the entire study period into 6 distinct periods, it became apparent that analyzing and fitting models to data within each period could yield more accurate models compared to fitting a single model to the entire study period.

Table 2. Determined Intervals of 6 Time Periods of Daily Death Data Due to COVID-19 in Iran.

Period number	Time Period based on day number	Time Period based on date
1	From the 1 st day to the 89 th day	From February. 19, 2020 to May. 17, 2020.
2	From the 90 th day to the 200 th day	From May. 18, 2020 to September. 5, 2020.
3	From the 201 st day to the 364 th day	From September. 6, 2020 to February. 16, 2021.
4	From the 365 th day to the 489 th day	From February. 17, 2021 to June. 21, 2021.
5	From the 490 th day to the 699 th day	From June. 22, 2021 to January. 17, 2022.
6	From the 700 th day to the 817 th day	From January. 18, 2022 to May. 15, 2022.

Regression model fitting

Ten regression models were applied to analyze the variable data, including the cumulative total number of deaths and daily death counts throughout the entire research period. All regression models, except the simple linear model, demonstrated statistically significant fits when tested against the daily death data ($P < 0.05$).

However, the cubic regression model was identified as the best fit among all models due to its highest coefficient of determination ($R^2 = 0.244$) value (Table 3). The quadratic model exhibited the second highest coefficient of determination, followed by the cubic model.

The graph depicting the ten regression models fitted to the daily death data (Figure 2), as well as the graph showcasing the two regression models with the

highest coefficient of determination fitted to the same data (Figure 3), corroborate the findings presented in Table 3.

Table 3. Regression Models Fitted to Daily Death Data in Iran from the Beginning of COVID-19 to May 15, 2022.

Equation	Model Summary					Parameter Estimates			
	R^2	F	df1	df2	Sig.	Constant	β_1	β_2	β_3
Linear	0	0.023	1	814	0.878	174.396	-0.003		
Compound	0.011	9.348	1	814	0.002	142.808	1		
Growth	0.011	9.348	1	814	0.002	4.962	0		
Exponential	0.011	9.348	1	814	0.002	142.808	0		
Inverse	0.017	13.708	1	814	0	176.824	-451.778		
Logarithmic	0.027	22.703	1	814	0	34.71	24.201		
Power	0.039	33.234	1	814	0	37.731	0.201		
Curve S	0.11	101.049	1	814	0	4.847	-8.059		
Quadratic	0.235	124.663	2	813	0	18.766	1.133	-0.001	
Cubic	0.244	87.575	3	812	0	56.723	0.582	0	-0.00000137

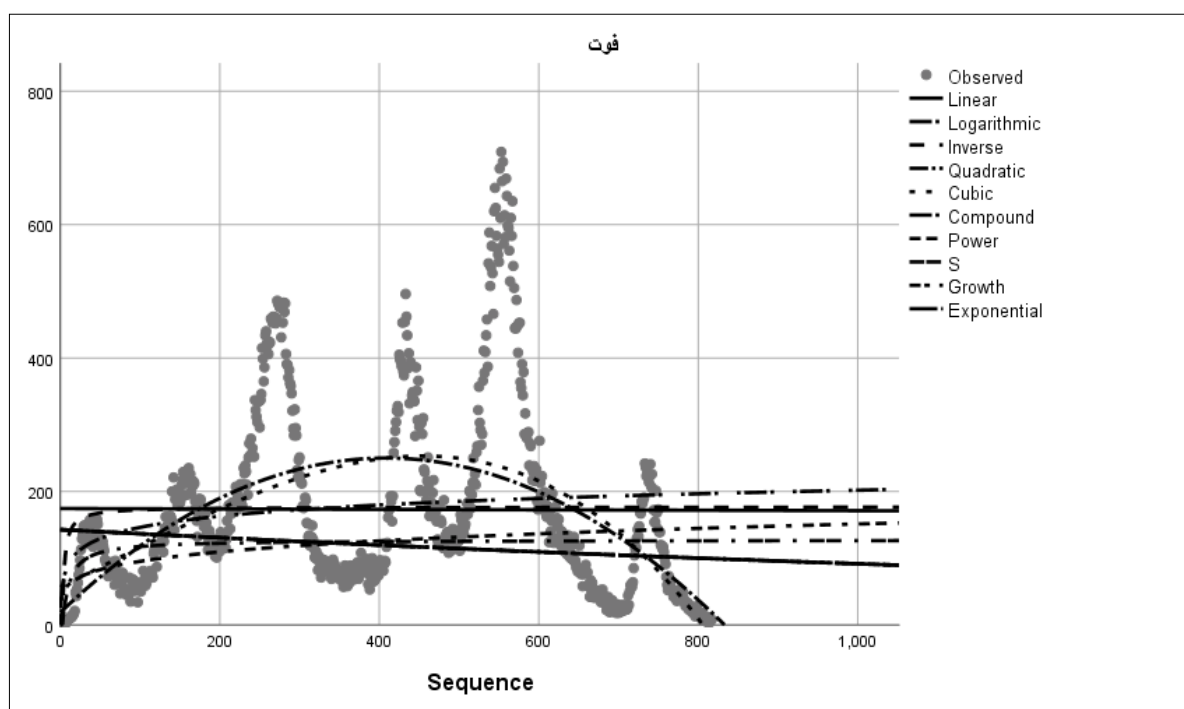


Figure 2. Daily Death Counts and Ten Regression Models Fitted to Data in Iran from the Beginning of COVID-19 to May 15, 2022

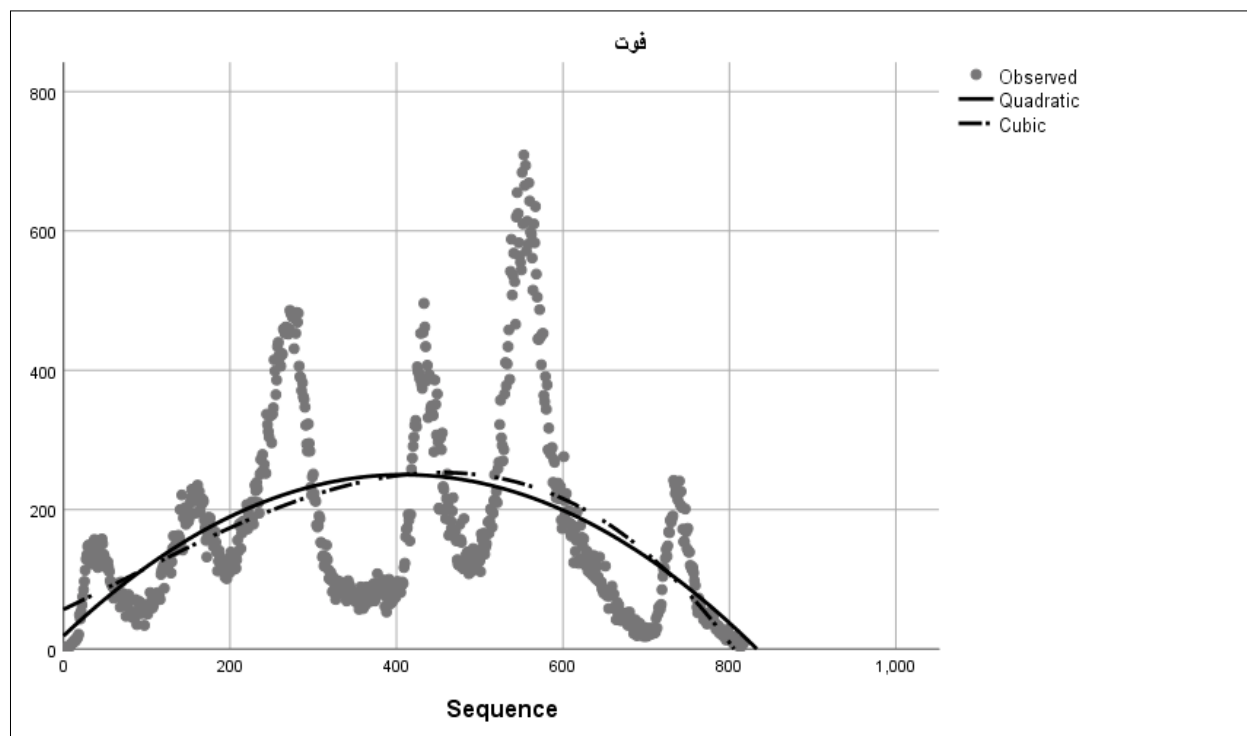


Figure 3. Daily Death Counts in Iran from the Beginning of COVID-19 to May 15, 2022

The formula for the best model fitted to the data on the daily death variable, based on the findings from Table 3, is as follows:

$$death = 56.723 + 0.582t - 0.00000137t^3 \quad (1)$$

Considering that the 3rd degree model, which was fitted among the ten models, was the best model fitted to the data, it does not have a high coefficient of determination. Therefore, in order to achieve fits with a higher coefficient of determination, ten regression models were fitted to the data in 6 time periods.

Considering that the 3rd-degree model, which was selected as the best fit among the ten models, did not achieve a high coefficient of determination, ten regression models were subsequently fitted to the data across 6 distinct time periods to seek higher fits.

The results from fitting ten regression models to the daily death data within these specified time periods indicated that the coefficient of determination of the best-fitted model for each period surpassed that of the models fitted to the data across the entire study period (Table 4).

Table 4. Comparison of Regression Models with the Highest Coefficient of Determination among the Ten Models Fitted to the Variable Data.

Time period	Equation	Model Summary			Parameter Estimates			
		R^2	F	Sig.	Constant	β_1	β_2	β_3
1	Cubic	0.799	111.37	0.000	-64.355	10.072	-0.162	0.001
	Quadratic	0.762	136.29	0.000	-36.274	6.609	-0.068	
2	Cubic	0.81	230.03	0.000	-554.04	7.018	0	-0.0000935
	Quadratic	0.785	197.67	0.000	-795.2	12.47	-0.04	
3	Quadratic	0.66	156.59	0.000	-3113.6	25.559	-0.047	
	Cubic	0.636	140.5	0.000	-1853	12.059	0	-0.0000543

4	Cubic	0.548	73.869	0.000	۸۱۸۳/۴-	29.057	0	-0.0000507
	Quadratic	0.537	70.76	0.000	۱۱۹۳۰-	55.966	-0.064	
5	Quadratic	0.531	117.29	0.000	-8088	30.055	-0.027	
	Cubic	0.527	110.64	0.000	-4783.6	13.715	0	-0.0000145
6	Quadratic	0.501	57.68	0.000	-20208	54.547	-0.037	
	Cubic	0.493	55.902	0.000	-13027	26.454	0	-0.0000159

Based on the results of this test, the formula of the best-fitted model for the daily death data in the first period is as follows:

$$death_1 = -64.355 + 10.072t - 0.162t^2 + 0.001t^3 \quad (2)$$

The formula for the best-fitted model in the second period of daily death data is as follows:

$$death_2 = -554.039 + 7.018t - 0.0000935t^3 \quad (3)$$

The formula for the best-fitted model in the third period of daily death data is as follows:

$$death_3 = -3113.632 + 25.559t - 0.047t^2 \quad (4)$$

The formula of the best model to fit these data in the fourth period is as follows:

$$death_4 = -8183.37 + 29.057t - 0.0000507t^3 \quad (5)$$

The formula of the best model for fitting these data in the fifth period is as follows:

$$death_5 = -8087.95 + 30.055t - 0.027t^2 \quad (6)$$

The formula of the best model to fit these data in the sixth period is as follows:

$$death_6 = -20207.65 + 54.547t - 0.037t^2 \quad (7)$$

Ten regression models were fitted to the cumulative death data to conduct further analysis and refine the models (Table 5). The table reveals that the cubic and quadratic models demonstrate the strongest fit. Both models exhibit notably high coefficients of determination, suggesting they are the most suitable for describing the data accurately.

Table 1. Descriptive Statistics of Daily COVID-19 Deaths in Iran (From the Beginning of the Pandemic to May 15, 2022).

Equation	Model Summary					Parameter Estimates			
	R^2	F	df1	df2	Sig.	Constant	β_1	β_2	β_3
Linear	0.983	47764	1	815	0	-12539.394	203.744		
Compound	0.635	1420.5	1	815	0	4127.069	1.006		
Growth	0.635	1420.5	1	815	0	8.325	0.006		
Exponential	0.635	1420.5	1	815	0	4127.069	0.006		
Inverse	0.051	44.046	1	815	0	73017.173	-249622.1		
Logarithmic	0.71	1993.5	1	815	0	-166516.91	41553.883		
Power	0.967	23748	1	815	0	3.429	1.638		
Curve S	0.361	460.86	1	815	0	10.788	-22.382		
Quadratic	0.984	24344	2	814	0	-10555.61	189.211	0.018	
Cubic	0.992	32743	3	813	0	1165.541	17.786	0.541	0.000427

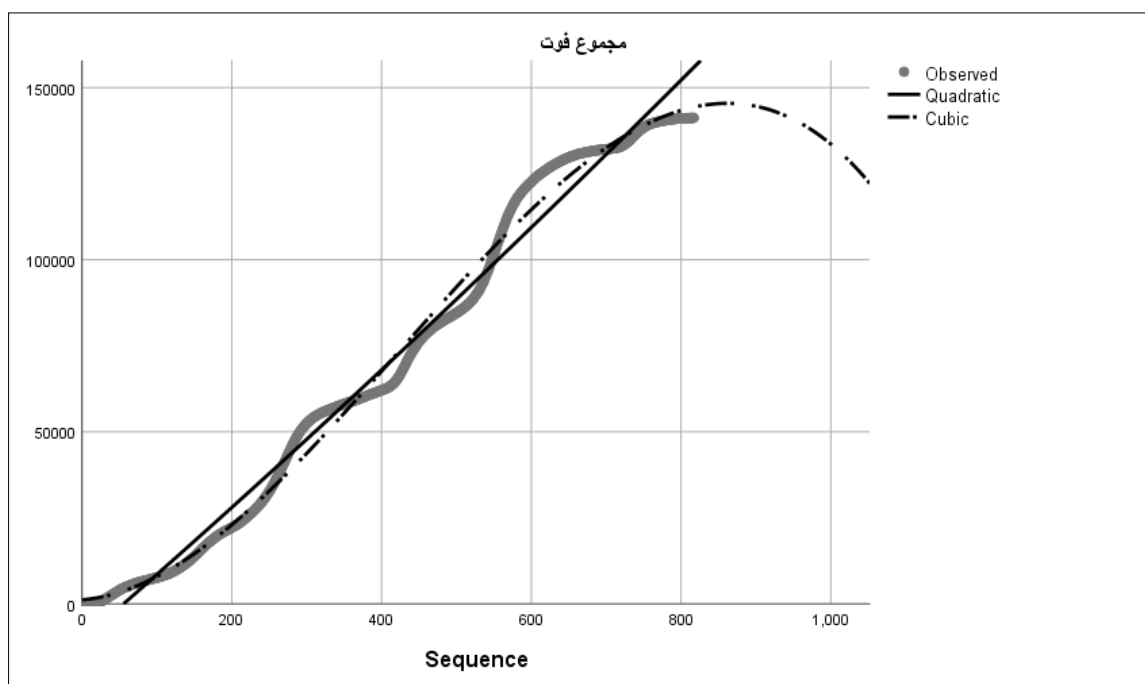


Figure 5. Cumulative Number of Deaths in Iran from the Beginning of COVID-19 to May 15, 2022

The formula for the best fitted model is as follows, and it is based on the outcomes of fitting the regression models to the variable data of the total number of deaths:

$$D_c = 1165.541 + 17.786t + 0.541t^2 + 0.000427t^3 \quad (8)$$

Time Series Fitting Model to Data

To achieve a better fit of the daily death data over the entire study period, a time series model was applied. Based on the results from the time series model fitting test (Tables 6 and 7), the ARIMA (4,1,4) model was found to significantly fit these data.

Table 6. Comparison of Time Series Models Fitted to Daily Death Data in Iran from the Beginning of COVID-19 to May 15, 2022.

Model Statistics					
Model	Model Fit Statistics		Ljung-Box Q (18)		
	Stationary R-squared	R-squared	Statistics	DF	Sig.
Death-Model_1	0.208	0.972	110.478	11	.000

Table 7. Coefficients of the Time Series Model Fitted to Daily Death Data in Iran.

Model Description			
			Model Type
Model ID	Death	Model_1	ARIMA (4,1,4)

ARIMA (4,1,4)					Estimate	SE	T	Sig.
Death-Model_1	Death	Square Root	AR	Lag 2				
				Lag 3				
				Lag 4				
			Difference		1			
			MA	Lag 1	.508	.025	20.072	.000
				Lag 2	.288	.042	6.860	.000
				Lag 3	.725	.043	16.866	.000
				Lag 4	-.766	.038	-20.124	.000

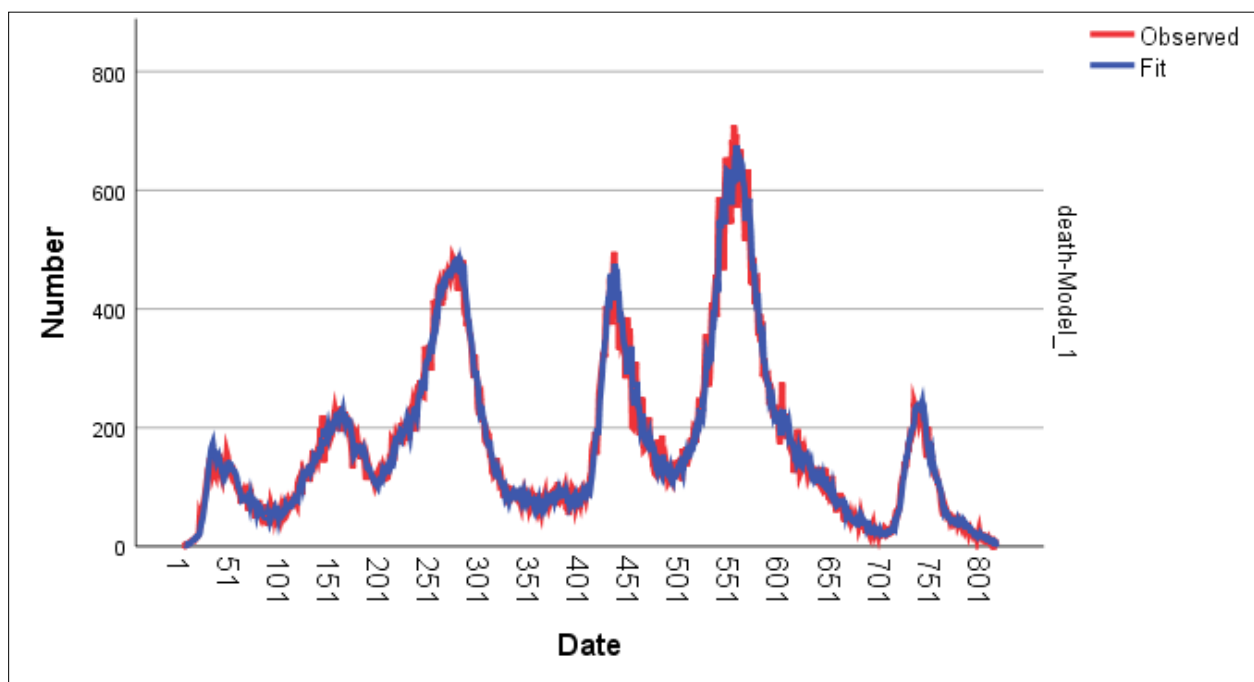


Figure 6. Number of Daily Deaths and Time Series Model Fitted Data in Iran (February 19, 2020 to May 15, 2022)

The Figure 6 displays the fitted time series models, demonstrating a high coefficient of determination ($R^2 = 0.972$), that surpasses those of all regression models fitted in previous sections. This indicates superior data fit and underscores the efficacy of the time series model in capturing the data trends more accurately than regression models.

Discussion

The graphs depicting daily COVID-19 deaths in Iran show six distinct peaks, suggesting that the data can be divided into six distinct time periods. Zare Zahra and Nastaran Vasegh highlighted that due to these multiple peaks in Iran's data, a single-parameter model like the SIR model is insufficient. Therefore,

they proposed dividing Iran's data into five time periods and determining specific parameters for each period (14).

In our study, we found that among the regression models tested, the third-degree polynomial regression model provided the best fit for the overall death data across the entire study period. Additionally, the ARIMA model (4,1,4) was identified as the best fit among the time series models. Across the six identified time periods, the cubic regression model performed best in the first, second, and fourth periods, while the quadratic regression model was optimal for the third, fifth, and sixth periods. Moreover, for cumulative death data,

both the cubic and quadratic regression models demonstrated the best fit characteristics.

These findings are significant as they offer valuable insights into the optimal models for predicting and interpreting the pattern of COVID-19 deaths in Iran during the study period. The identification of the third-degree polynomial regression model and the ARIMA model (4,1,4) as the best fits for the death data across specific time periods enhances our understanding of temporal variations and trends in COVID-19 mortality.

The utilization of the ARIMA model aligns with previous research demonstrating its efficacy in forecasting COVID-19 cases and deaths through time series analysis (12). These results contribute to ongoing efforts to employ advanced statistical and mathematical models for precise prediction and comprehension of the pandemic's impact on mortality.

The identification of different best-fitting models for specific time periods, such as the third-degree polynomial model for the first five periods and the quadratic model for the sixth period, underscores the dynamic nature of the pandemic and the importance of adapting modeling approaches to capture these variations. This highlights the need for ongoing monitoring and analysis to inform timely and context-specific public health responses (12, 15).

In conclusion, the study's findings regarding the best-fitting regression and time series models for COVID-19 death data in Iran offer valuable insights for researchers, policymakers, and public health practitioners. These results contribute to the growing body of evidence on the use of advanced modeling techniques to understand and predict COVID-19 mortality, with potential implications for informing targeted interventions and resource allocation.

Acknowledgements

This research originated from an MSc thesis conducted at Ilam University of Medical Sciences, Ilam, Iran. The authors express their gratitude for the

collaborative efforts of all participants involved in this research.

Financial support

This research is part of a dissertation by a MSc student at Ilam University of Medical Sciences and was supported by Ilam University of Medical Sciences.

Conflict of interest

There is no conflict of interest to declare.

Authors' contributions

KS and EG conceived the study idea, performed the statistical analysis, and wrote the manuscript. AKH contributed to the study design. MO assisted with the statistical analysis. All authors read and approved the final manuscript.

References

- Ministry of Health and Medical Education of Iran. (2021). COVID-19 updates and guidelines. Retrieved from <https://www.behdasht.gov.ir/>.
- World Health Organization. (2021). Coronavirus disease (COVID-19) pandemic. Retrieved from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
- Farooq M, Ijaz M, Atif M, Abushal T, El-Morshedy M. Statistical analysis of Covid-19 mortality rate via probability distributions. *PLoS One*. 2022 Oct 27;17(10):e0274133. doi: 10.1371/journal.pone.0274133. PMID: 36301849; PMCID: PMC9612498.
- Zhang Z, "Statistical Modeling of Daily Confirmed COVID-19 Cases and Deaths in Europe and United States" (2021). Master's Theses (2009 -). 664. https://epublications.marquette.edu/theses_open/664.
- Khajanchi, Subhas and Kankan Sarkar. "Forecasting the daily and cumulative number of cases for the COVID-19 pandemic in India.2020 Chaos 30 (7): doi: 10.1063/5.0016240.
- da Silva RA dSFL, Leite JMRS, Tiraboshi FA, Valente TM, de Paiva Roda VM, Duarte Sanchez JJ. Statistical Modeling of Deaths from COVID-19 Influenced by Social Isolation in Latin American Countries. *Am J Trop Med Hyg*. 2022 Mar 14;106(5):1486–90. doi: 10.4269/ajtmh.21-0217. Epub ahead of print. PMID: 35292589; PMCID: PMC9128698.
- Source: <https://covid19.healthdata.org>.
- Khan N, Arshad A, Azam M, Al-Marshadi AH, Aslam M. Modeling and forecasting the total number of cases and deaths due to pandemic. *J Med Virol*. 2022 Apr;94(4):1592-1605. doi: 10.1002/jmv.27506. Epub 2021 Dec 18. PMID: 34877691; PMCID: PMC9015266.
- Msemburi, W., Karlinsky, A., Knutson, V. et al. The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature* 613, 130–137 (2023). <https://doi.org/10.1038/s41586-022-05522-2>.
- Chyon FA, Suman MNH, Fahim MRI, Ahmmed MS. Time series analysis and predicting COVID-19 affected patients by ARIMA model using machine learning. *J Virol Methods*. 2022 Mar;301:114433. doi: 10.1016/j.jviromet.2021.114433.
- Somyanonthanakul, R., Warin, K., Amasiri, W. et al. Forecasting COVID-19 cases using time series modeling and association rule mining. *BMC Med Res Methodol* 22, 281 (2022). <https://doi.org/10.1186/s12874-022-01755-x>.
- El-Sherpieny EA, Almetwally EM, Muse AH, Hussam E. Data analysis for COVID-19 deaths using a novel statistical model: Simulation and fuzzy application. *PLoS One*. 2023 Apr 10;18(4):e0283618. doi: 10.1371/journal.pone.0283618.
- Nuutinen M HI, Niemi AJ, Rissanen A, Ikivuo M, Leskelä RL. Statistical model for factors correlating with COVID-19 deaths. *Int J Disaster Risk Reduct*. 2022 Nov;82:103333. doi: 10.1016/j.ijdr.2022.103333. Epub 2022 Sep 29.
- Nopour R, Erfannia L, Mehrabi N, Mashoufi M, Mahdavi A, Shanbehzadeh M. Comparison of Two Statistical Models for Predicting Mortality in COVID-19 Patients in Iran. *Shiraz E-Med J*. 2022;23(6):e119172.
- Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos Solitons Fract*. 2020 May;134:109761. DOI: 10.1016/j.chaos.2020.109761.
- Chen, S., Yang, J., Yang, W., Wang, C., & Bärnighausen, T.W. COVID-19 control in China during mass population movements at New Year. *Lancet*, 2020 Mar 7;395(10226):764-766. doi: 10.1016/S0140-6736(20)30421.
- Kotwal A, Yadav AK, Yadav J, Kotwal J, Khune S. Predictive models of COVID-19 in India: A rapid review. *Med J Armed Forces India*. 2020;76(4):377-386. doi:10.1016/j.mjafi.2020.06.001.
- Abdolhosseini, M. (2023). Forecasting of COVID-19 sixth peak in Iran based on singular spectrum analysis. *Journal of decisions and operations research*, 8(1), 123-132. <https://doi.org/10.22105/dmor.2022.312966.1517>
- Zare Z, Vasegh N. Modeling and analysis of the spread of the COVID-19 pandemic using the classical SIR model. *JoC* 2021; 14 (S1) :89-96. DOI: 10.52547/joc.14.5.89 URL: <http://joc.kntu.ac.ir/article-1-821-en.html>.