

GSJ: Volume 8, Issue 8, August 2020, Online: ISSN 2320-9186 www.globalscientificjournal.com

Modeling Fatality Rate of COVID – 19 in Nigeria Using Multiple Linear Regression Analysis.

1*Adejumo, T. J., ¹Akomolafe, A. A., ²Owolabi, A. T., ²Okegbade, A. I., ²Oladapo, O. J., ²Idowu, J. I. and ³Gbolagade S. D.

1,2,3 Department of Statistics, Ladoke Akintola University of Technology Ogbomoso Oyo State

¹Department of Statistics, Federal University of Technology Akure Ondo State

tjadejumo@lautech.edu.ng, akomolafeayotade@gmail.com, atowolabi@lautech.edu.ng, aiokegbade@lautech.edu.ng, ojoladapo@lautech.edu.ng, jiidowu@lautech.edu.ng and sundaydanielgbolagade@gmail.com

Keywords

COVID – 19, Confirmed cases, Critical cases, fatality rate and Multiple linear regression

Abstract

Coronavirus pandemic known, to be caused by severe acute respiratory syndrome CoV- 2 (SARS-CoV-2) remains the global pandemic threatening hundreds of countries this day. It was said to first breakout in Wuhan, Hubei province of China late December, 2019. The first confirmed case in Nigeria was announced on 27th February, 2020 when an Italian citizen in Lagos was tested positive for the virus. Hence, sequel to the escalation of confirmed cases on daily basis and fatality rate, this study modeled the effect of confirmed cases and critical cases on the fatality rate using multiple linear regression analysis. Monitored are the data released by the Nigeria Centre for Disease Control (NCDC) for complete fifty-two (52) days spanning from May 1st – June 21st, 2020. Assumptions of Ordinary Least Squares were examined and it was observed that multicollinearity does not exist among the explanatory variables (Confirmed Cases and Critical Cases), meanwhile the Durbin-Watson test statistic indicates that there is no autocorrelation. Kolmogorov-Smirnov and Shapiro-Wilk statistics confirmed that the regression residual follows a normal distribution. From the estimated model, it was observed that confirmed cases have lesser effect on the fatality rate compared with cumulative critical cases. Moreover, other factors that contribute to the fatality rate of COVID–19 are so enormous. Test of the overall model indicates that there is significant difference in the regression coefficients of the model. All other patients with other health challenges should be attended to, in order to curb fatality rate in the country through COVID -19.

1. Introduction

It is no longer news that Coronavirus, also known as COVID-19, is currently a major pandemic threatening many countries across the globe including Nigeria. The history had it that the outbreak began late December 2019, Wuhan City in China. Since then the number of confirmed cases has been escalating and increasing on a daily basis most especially in African countries where Nigeria is not left out. Here, in Nigeria, sequel to the monitoring of confirmed cases of the pandemic by the Nigeria centre for Diseases Control (NCDC), it has been spotted that Lagos State is now the epicenter of the disease being a state with the highest number of confirmed cases, followed by Federal Capital Territory (FCT) Abuja in this order as shown in figure 1. Out of thirty-six (36) states including FCT, only one state which is Cross Rivers State is yet to record any case as at the time of carrying out this research. The sporadical increase in the number of confirmed cases here in Nigeria was as a result of inability to put necessary control measures, prevention, low testing capacity and state of our health institution among others. Though, several measures have been put in place before, such as lockdown of states either partially or completely, and this has been helpful in curtailing the spread of the pandemic. However, now that the easing of lockdown has began the number of confirmed cases started increasing, also, the number of recovered cases is increasing. This was due to the risen up of government and some voluntary organizations toward the health institution and testing capacity. Hence, this study put into consideration the implication of cumulative confirmed cases based on testing capacity and cumulative critical cases on fatality rate (Death rate) using multiple linear regression analysis.

Many well-known scholarly researchers have written and published papers on this on-going pandemic. Their views and contributions are reviewed; Kayode *et al* (2020) considered the daily cumulative confirmed cases of COVID-19 in Nigeria. In their analysis, Nine (9) curve estimation statistical models in quartic, cubic, quadratic, and simple were used, whereby the best of the thirty-six (36) models considered were selected and used for forecasting. However, with autocorrelation or order (1), the analysis identified quartic linear regression model as the best model. In addition, they identified Ordinary Least Squares (OLS), Cochrane Orcutt, Hildreth-Lu, Praise-whisten, and Least Absolute Deviation (LAD) estimators as estimators to estimate the model's parameters. They later recommended LAD estimator as the best for daily cumulative forecast values of the pandemic with 99% CL for May and June, 2020 because of its precision.

At the early stage of the outbreak of COVID – 19 in Philipine, Michael (2020) carried out a simple linear regression analysis on daily confirmed cases with fatality rate based on public available data. From the regression model, its slope was used as the fatality rate whereby its confidence interval was as well estimated from the standard error. He claimed that in the Country, at the first few weeks of the pandemic, 4.35% with 0.95Cl of 4.2% to 4.55% was recorded. It was found out that in mainland China and Italy the estimate is lesser. The outcome of the interstate migrant movement workers on the spike of coronavirus at their state of residence in India was considered by (Maji *et al*, 2020) where they made use of Modified SEIR model to examine the implication of inter-state migrant workers' movement during the restriction of movement in India. In their research, a series of levels of COVID – 19 transmissions during travel, daily arrival, and total number of migrant workers were considered. The Numbers of confirmed and active cases of different states in the country were as well considered. They finally suggested that stringent screening and isolation of migrant workers after reaching their destination must of paramount as this will cub the spreading of the pandemic.

A viable model was proposed in (2020) by Mohamed *et al*, tagged Fractional-order Susceptible – Exposed- Infected – Quarantined – Recovered- Death - *Insusceptible* (F-SEIQRDP) to estimate the transmissibility of the pandemic. The proposed model took into account the subdiffusion process of both confirmed and recovered cases. In the new model, the entire populations were subdivided into seven populations, viz; S(t), E(t), I(t), Q(t), R(t), D(t) and P(t). Taken into consideration is the history of the state because each state follows the fractional – order behaviour. The parameters of the proposed model were estimated by a nonlinear least squares minimization with the aids of MATLAB. However, Root Mean Square Error (RMSE) was used as a criterion to identify the best model for forecasting. The results of SEIQRDP were compared with F-SEIQRDP and found out that the RMSE of F-SEIQRDP is lesser.

Another SEIR based model was proposed by Sheldon *et al* (2020) which actually fits for pandemic modeling. They made use of short history data to estimate the trend of the changing disease dynamics for the dead, recovered and infected in such away that the adaptive real-time changes of disease mitigation, business and societal population can be naturally accommodated. Furthermore, a generalized SEIR was proposed by Liangrony *et al* (2020) to analyze the ongoing Coronavirus pandemic. The data used in their analysis was collected from the National Health Commission of China from January, 20th to 9th of February, 2020.

The main key epidemic parameter were estimated and as well made forecast on the inflection point. In their analysis, they predicted the possible ending period for the pandemic in five (5) different regions of China. According to their study, they proposed that in Beijing and Shanghai in China, the epidemic will come to an end within two weeks. In the same vein, they discovered that by inference, the outbreak of the pandemic will be experienced in Mainland Hubei and Wuhan at the beginning of April.

To study the mode of transmission of COVID – 19 globally, Nita *et al* (2020) worked on the SEIR model and proposed a generalized SEIR model under different control strategies. In their research, all means of human-to-human transmission were looked into and the number of reproductions is formulated to analyze the perfect transmission dynamics of the pandemic. Also, applied to the model was the optimal control theory to prevent the influence of various intervention strategies including; Isolation of infected persons, Voluntary quarantine, improvement of individual immunity, and hospitalization. In addition, through simulation study, the numerically graphical significant effect of the control strategies on the proposed model was analyzed. Epidemiological data was collected by Bilin *et al* (2020) to analyze the trend of COVID -19 in China. The data were collected through the Chinese and official global website. In their study, a total number of confirmed cases were recorded on a daily basis in some major cities in China such as Wuhan, Hubei province, among others. They found out that the number of daily cases increases except in Hubei province.

Atangana and IgretAraz (2020) worked on Mathematical model of COVID – 19 spread in Turkey and South Africa: theory, methods, and applications approach. Some basic statistical methods which include Arithmetic Mean (AM), Geometric Mean (GM), Standard deviation (Sd), Skewness, Variance, Pearson, and Spearman rank correlation were applied to the available data with the two countries. Hence, sequel to the results of Spearman correlation for Turkey, daily infections could decrease, meanwhile, daily death and recovered patients could also increase. However, in the case of South Africa, the number of daily deaths and infectious people could decrease under the current situation. Likewise, in their study, a new mathematical model was suggested and extended to the scope of nonlocal operators for each case. To compare the significant difference between the cases of South Africa and Turkey is Atangana Seda numerical method.

To forecast the epidemic trend of Poland under different scenarios of lockdown and its removal is the extended Susceptible – Infected Removal (eSIR) model by Magdalena and Andrzej (2020) used is the time series of data collected on SARS – Cov- 2 infection commencing from March 4 to may 22nd, 2020. Using eSIR algorithm, basic reproductive number was estimated including a number of infections at different epidemic trend scenarios. In their further claims, they concluded that under different precautions removal, in the next one year, a million or more cases of infectious cases may be recorded.

2. Methodology

NCDC has been monitoring and reporting the cumulative and new number of confirmed cases, number of critical cases, and recorded death since the first occurrence of COVID-19 in Nigeria. These data were collected on a daily basis from Nairametric sites over a period of one month- twenty one days beginning from 1st of May, 2020 (first day of occurrence) till 21th June, 2020. The multiple linear regression was employed to analyze the contribution of the cumulative number of confirmed cases and the cumulative number of critical cases to the cumulative number of deaths of the coronavirus pandemic in Nigeria.

The statistical multiple linear regression is an extension of simple linear regression (relationship between one dependent variable and only one independent variable), which involves the relationship between one dependent variable (number of coronavirus deaths) and two or more independent variables, in this study, only two independent variables were used, the number of coronavirus confirmed cases and the number of coronavirus critical cases. The dependent variable (number of deaths) in this study is denoted as Y, and the independent variables, the number of confirmed cases and the number of critical cases are denoted as X_1 and, X_2 respectively.

However, the general multiple linear regression models with response Y and terms X_{1,...}X_p has the form

$$E(Y|X=x) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \tag{1}$$

The residual sum of squares as a function β is written in matrix form as:

$$RSS(\beta) = (Y - X\beta)^{1}(Y - X\beta)$$
⁽²⁾

The least squares estimate is given as:

$$\hat{\beta} = (X^1 X)^{-1} X^1 Y$$

The fitted or predicted value is defined as:

$$\hat{Y} = X\hat{\beta}$$

and the estimated residual is defined as:

$$\hat{e} = Y - X\hat{e}$$

Where,

$$\mathbf{X} = \begin{cases} 1 & x_{11}x_{21} \\ 1 & x_{12}x_{22} \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & x_{np}x_{np} \end{cases}, \ \mathbf{Y} = \begin{cases} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_p \end{cases}, \ \mathbf{X}^{1}\mathbf{X} = \begin{pmatrix} n & \sum_{i=1}^{p} x_{1i} \sum_{i=1}^{p} x_{2i} \\ \sum_{i=1}^{p} x_{1i} \sum_{i=1}^{p} x_{2i} \\ \sum_{i=1}^{p} x_{2i} \sum_{i=1}^{p} x_{2i} \\ \sum_{i=1}^{p} x_{2i} \sum_{i=1}^{p} x_{2i} \\ \sum_{i=1}^{p} x_{2i} y_{i} \\ \sum_{i=1}^{p} x_{2i} y_{i} \end{cases} \text{ and } \mathbf{X}^{1}\mathbf{Y} = \begin{cases} \sum_{i=1}^{p} y_{i} \\ \sum_{i=1}^{p} x_{1i} y_{i} \\ \sum_{i=1}^{p} x_{2i} \\ \sum_{i=1}^{p} x_{2i$$

In this study,

y = Number of Total Death (TDE) x_1 = Number of Confirmed Cases (COC) x_2 = Number of Critical Cases (CRC)

2.1 Multicollinearity Test:

Multicollinearity refers to the extent to which independent variables are correlated; it is usually exist when one independent variable is correlated with another independent variable or when one independent variable is correlated with a linear combination of two or more independent variables.

There are two popular ways to measure multicollinearity, one is by computing a coefficient of multiple determination for each independent variable and the other one is by computing a variance inflation factor for each independent variable. In this research, variance inflation factor method of measure multicollinearity was employed.

It is computed by the equation given below:

$$VIF_J = \frac{1}{(1-R_J^2)}$$

where VIF_J is the variance inflation factor for variable j, and R_j^2 is the coefficient of multiple determination for variable j.

and R_i^2 is calculated by:

$$R_j^2 = \frac{\beta' x' x_1 - n \bar{x}_1^2}{x' x_1 - n \bar{x}_1^2} \tag{7}$$

Multicollinearity exists if VIF > 10.

2.2 Test for Significant Overall Regression:

The null hypothesis for this test may be generally stated as Ho: "All k independent variables considered together do not explain a significant amount of the variation in y". Equivalently, the null hypothesis can still be stated as; Ho:"There is no significant overall regression using all k-independent variables in the model" or as Ho: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$. Table 1 displays the computation of the Analysis of Variance (ANOVA).

Table 1: Analysis of Variance (ANOVA) table

Source	d.f	SS	MS	F
Regression	Κ	SST-SSE	MSR	MSR/MSE
Error	n-k-1	SSE	MSE	
Total	n-1	SST		

(3)

(4)

(5)

The calculated F-statistic

 $F = \frac{Regression \ MS}{Residual \ MS} = \frac{(SST - SSE)/k}{SSE/(n-k-1)}$ (8)

where SST = $\sum_{i=1}^{n} (y_i - \bar{y})^2$ and SSE = $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ are total and error sums of squares, respectively.

3.0 Analysis and Results

The spread of novel Corona Virus Disease (COVID-19) in Nigeria has now touched a new milestone as the latest statistics provided by the Nigeria Centre for Disease Control reveals as shown in figure 1 as at 29th June, 2020. However, as at the time that the analysis is being carried out, Nigeria has 20,242 confirmed cases, 518 total deaths as at the 21st of June, 2020. **See table 2.** In addition, figures; 2, 3a and 3b reveal that the high rate or spike of the pandemic in the country is alarming and requires urgent attention to curb the spreading.



Figure 1: Nigeria Map showing the total number of confirmed cases across the country.



Figure 2: Trend of Confirmed, Death and Critical cases in Nigeria



Figure 3a: Pyramid Frequency Chart of Confirmed Cases and Total Death till 21st of June 2020



Figure 3b: Pyramid Frequency Chart of Confirmed Cases and Total Death till 21st of June 2020

Table 2: Descriptive Statistics of COVID-19

	Confirmed Cases	Total Death	Critical Cases
Ν	52	52	52
Mean	9468.4231	265.75	5.9615
Mode	2170.00 ^ª	221	7
Minimum	2170	69	1
Maximum	20242	518	9

Source: computer output

A multiple linear regression model was employed in estimating the fatality rate of COVID – 19 in Nigeria. The cumulative number of COVID-19 confirmed cases and the number of those in critical state were treated as independent variables, while the cumulative COVID-19 number of death cases was treated as the dependent variable. The slope of the regression model (regression line) is used as the case fatality rate. The results in the table 3 below suggest the model that best fits the data.

	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics		
(Constant)	B 22.044	Std. Error	Beta	t 6 471	Sig.	Tolerance	VIF	
Confirmed Cases	0.023	0	0.961	89.208	0	0.538	1.86	
Critical Cases	3.716	0.737	0.054	5.044	0	0.538	1.86	

Table 3: table of test of multicollinearity and parameters estimation

Source: Computer output

The model is given below:

Fatal cases = 22.044 + 0.023Confirmed Cases + 3.716Critical Cases

The model suggests that to every unit increment in confirmed cases of coronavirus, there will be 0.023 increment in the number of deaths, this regression coefficient is significant to the fitted model as its p-value (0.000) less than alpha value. At the same time, to every unit increment in the number of critical cases, there will be 3.716 increment in the number of deaths, the regression coefficient is also significant to the fitted model as its p-value (0.000) also less than alpha value and that 22.044 increment in fatality rate are due to other underline health challenges that are not put into consideration in the study for analysis while confirmed cases and critical cases are held constants. It can also be observed that multicollinearity does not exist among the explanatory variables (Confirmed Cases and Critical Cases) in the data set as the Variance Inflation Factors (VIFs) are less than 10.

3.2 Overall Significant of the Model:

The overall significance test was also carried out to check if there is significant difference in the regression coefficients of the model as contained in table 4. The results from the table show that the p-value (0.000) is less than the alpha value. Hence, there is sufficient evidence that the regression coefficients differ to one another. Thus, the regression model best fit for the data.

100.

Table 4: Table of test of overall significant of the model							
Model Regression	SS	Df	MS	F	Sig.		
	811531.9	2	405766	7991.965	.000 ^b		
Residual	2487.815	49	50.772				
Total	814019.8	51					

Source: Computer output

3.3 Coefficient of Determination and Autocorrelation:

The coefficient of determination, R-square is 0.997, showing that the number of deaths is 99.7% explained by the number of confirmed cases and the number of critical cases. The Durbin-Watson value also indicates that there is no autocorrelation in the data set since the value does not exceed 2. (See table 5).

			Adjusted	R	Std. Error of	Durbin-			
	R	R Square	Square		the Estimate	Watson			
	0.998	0.997	0.997		7.12543	1.008			

Table 5: Table of Coefficient of Determination and Autocorrelation

Source: Computer output

3.4 Normality Test of the Regression Residuals:

Normality test of the regression residuals was carried out to check if the residuals are normally distributed at the preselected alpha level (α) 0.05 as it can be seen in table 6. Since the p-value of both Kolmogorov-Smirnov and Shapiro-Wilk test statistics are greater than the alpha value, hence the regression residual follows a normal distribution. To establish this fact is figure 4.



Figure 4: Normality Plot of Regression Standardized Residual

4.0 Conclusion

A multiple linear regression analysis was employed to model the fatality rate of COVID - 19 in Nigeria. From the estimated model, it was observed that cumulative cases based on testing capacity have lesser effect on the fatality rate compared with critical cases that have higher effect. Additionally, other factors that contribute to the fatality rate of COVID - 19 are so enormous; this simply means that the patient identified with the virus may die of the disease if such patient has underlined health challenges. Therefore, attention should not be given to the patient with COVID - 19 alone, all other patients with other health challenges should be attended to and be given proper treatment as the case may be to curb fatality rate in the country. Hence, it becomes imperative for the government of the country to wake up to their responsibilities, especially to the health institution of the country so that the rate at which people are dying in the country through poor health care will be drastically reduced.

Funding: Not applicable

Availability of data: ncdc website; www.ncdc.gov.ng

Conflict of interest: The authors hereby declare that there is no conflict of interest.

Acknowledgement

The authors appreciate all the peer reviewers for their contributions.

References

- [1] A. P. M. Michael. (2020, April). Preliminary Estimate of COVID-19 Case Fatality Rate in the Philippines using Linear Regression Analysis. : https://www.researchgate.net/publication/340505831.DOI: 10.2139/ssrn.3569248 .
- [2] A. Atangana and S. Igret Araz. (2020, may). Mathematical model of COVID-19 spread in Turkey and South Africa: Theory, methods and applications. *https://www.researchgate.net/publication/341360578. DOI: 10.1101/2020.05.08.20095588*.
- [3] A. Maji, M. B. Sushma, T. Choudhari. (2020). Implication of Inter-State Movement of Migrant Workers during COVID19 Lockdown using Modified SEIR Model. 1-16.
- [4] A.B. Mohamed, C. Abderrazak, L. K Taous-Meriem. (2020). Fractional-order SEIQRDP model for simulating the dynamics of COVID-19 Epidemic. *https://www.researchgate.net/341175202*.
- B. Chen, H.Zhong, Y. Ni, L. Liu and X. Su. (2020). Epidemiological Trends of Coronavirus Disease 2019 in China. Journal of Frontiers in Medicine. www.frontiersin.org. doi: 10.3389/fmed.2020.00259, VII (259).
- [6] H. S. Nita, H. S. Ankush and N. J. Ekta. (2020). Control Strategies to Curtail Transmission of COVID-19. doi: https://doi.org/10.1101/2020.04.04.20053173.
- [7] k.Ayinde; A.F., Lukman; R.I., Rauf; O. O., Alabi; C. E., Okon; O.E. Ayinde. (2020). Modeling Nigerian Covid-19 cases: A comparative analysis of models. *Chaos, Solitons and Fractals. .doi:org/10.101/j.chaos.2020.109911.*, 138-153.
- [8] O. Magdalena and B. Andrzej. (2020, May 29). Forecasting COVID-19 pandemic in Poland according to government r egulations and people behavior. *Frontiers. doi: https://doi.org/10.1101/2020.05.26.20112458*.
- [9] P. Liangrong, W. Yang, D. Zhang, C. Zhuge, and L. Hong. (2020, February). Epidemic analysis of COVID-19 in China by dynamical modeling. *https://www.researchgate.net/publication/339323999*. DOI: 10.1101/2020.02.16.20023465.
- [10] X. sheldon, D. Tan and L. Chen. (2020). Real-Time Differential Epidemic Analysis and Prediction for COVID19.Pandemic.https://www.researchgate.net/publication/340662745_RealTime_Differential_Epidemic_Analysis_and_Prediction_for_COVID-19_Pandemic.