Global Scientific JOURNALS

GSJ: Volume 9, Issue 2, February 2021, Online: ISSN 2320-9186

www.globalscientificjournal.com

STATISTICAL ANALYSIS OF COVID-19 INCIDENCE IN NIGERIA

[1]Matthew, Chukwuma Michael; [2]Odiachi, Ifeanyi James; [3]Ashinze, Akudo Nkpuruoma; [4]Omokri Peter

[1,3-4]Department of Industrial Mathematics, Delta State Polytechnic, Ogwashi-Uku.

[2]Department of Science Laboratory Technology, Delta State Polytechnic, Ogwashi-Uku

Telephone: +2348034124558; Email: chukwuma_mathematics@adun.com; megawaves4life@yahoo.com

Abstract

Covid-19 is a novel disease that is rampaging all the countries and continents of the world including Nigeria. So many lives have been lost and the entire world has been thrown into confusion mainly because there is no cure and, no vaccines are known for treatment and prevention of the disease respectively. The incidence of COVID-19 is on the rise in Nigeria but no statistical basis has been laid for its growth; hence, this study aimed at statistically analyzing COVID-19 cases in Nigeria. Data were collected from the National Centre for Disease Control from 27th February 2020 to 1st June 2020. The R statistical package (R, 64, 4.0.0) and MINITAB 17 were used to analyze the collected data. The data were represented graphically for the entire country and specifically by states. Statistical models were fitted for the COVID-19 cases and forecast of future occurrences were made. A downward trend was observed from the models and, the state specific death and recovery rates were computed showing Kogi State with the highest proportion of active cases, Zamfara with the highest proportion of recovery followed by Bauchi while Ondo had the highest proportion of death. In conclusion, the research has revealed that there is a general downward trend indicating possible reduction in incidence of COVID-19 in the future.

Keywords: COVID-19, Nigeria, pandemic, epidemic, incidence, Coronavirus and specific rates

1.0 INTRODUCTION

Coronaviruses are a large family of viruses that are known to cause illness ranging from the common cold to more severe diseases such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS) (WHO, 2020). The name *Coronavirus* was coined from the outer fringe, or *Corona* of the embedded envelop protein of the virus(es) (Frieman, 2014). Human coronavirus are mainly the most significant viral causes of upper respiratory tract infections. The coronavirus, which history has that they surfaced in 1956, have also been found to be associated with some lower respiratory tract infections (Kenneth, 2005). Coronavirus disease, otherwise called COVID-19, is an infectious disease caused by a new type of corona virus discovered recently in Wuhan, China (WHO, 2020). There are no known vaccines for the coronavirus and no treatment has been discovered too. The coronavirus is easily transmitted from human to human through droplets of fluid/mucous from the respiratory of an infected person(s). The rate of spread of the virus is so high that it has caused so much panic and the World Health Organization has declared it a pandemic.

Nigeria is a naturally blessed country situated on the Golf of Guinea with a population of 195.9 million persons as at 2018. The country is made of thirty-six states and the Federal Capital Territory. It is comprised of many ethnic groups and diversity of cultures. The Nigerian cultural structure/system is such that communal life is highly encouraged while schools, churches, markets and many public places are overcrowded. These outrightly give rise to the spread of diseases due to uncontrolled body contact.

2.0 METHODOLOGY

The data for this study were collected from the home page of the National Centre for Disease Control of Nigeria and the Wikipedia page on COVID-19 cases in Nigeria on 4th day of June 2020. The data were available on the number of new confirmed cases per day from 27th February 2020 to 1st June 2020 and also distributed on the total confirmed cases, active cases, recovered and deaths per state. The data will be represented graphically in order to visualize the form of relationship between; the current total number of confirmed cases and new cases at each particular time in the period under study, the one step lag variable and the current number of newly confirmed cases and, time period and the number of new cases. The simple linear regression model will be used to estimate the relationship between current total and new cases, the one-step auto-regressive model will be used to estimate the relationship between the one-step lag cases and the current new cases and, the least squares and the moving average methods will be used to model the relationship between time period and current new confirmed cases.

2.1 The Simple Linear Regression Model

The Simple Linear Regression Model is given as

$$y_i = b_0 + b_1 x_i + e_i \qquad (1)$$

where y_i is the i^{th} observation on the dependent variable, b_0 is the intercept parameter, b_1 is the slope parameter that measures the change in the dependent variable due to a unit change in the independent variable, x_i is the i^{th} observation on the independent and e_i is the i^{th} error or residual associated with the prediction of y_i using the model. The least square method or the Maximum Likelihood Method can be used to estimate parameters of the model. The test of significance of the regression parameters can be performed for the goodness-of-fit of the model using the F-ratio generated from the ANOVA table below:

Table 1: ANOVA for Test of Significance of Simple Linear Regression Model

Source of Variation	Degree of Freedom	Sum of Squares	Mean Squares	F-Ratio
Reg	1	$SS_R = \sum_{i=n}^{n} (y_i - \hat{y}_i)^2$	$MS_R = \frac{SS_R}{1}$	$F_{cal} = \frac{MS_R}{MS_e}$
Error	$n-2$	$SS_e = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$	$MS_e = \frac{SS_e}{n-2}$	
Total	$n-2$	$SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2$		

The calculated F – value, F_{cal}, can be compared with a tabulated value, F_{tab}, at a specified level of significance and degree of freedom defined for the numerator and the denominator. A calculated F – value, F_{cal}, less than its tabulated counterpart leads to the acceptance of the null hypothesis, H_0, of insignificance the model while a calculated F – value, F_{cal}, greater than the tabulated value, F_{tab}, leads to the rejection of the null hypothesis, H_0, of insignificance of the model. That is, accept H_0 if

$$F_{cal} = \frac{MS_R}{MS_e} \leq F_{tab} = F_{1,n-2}^{1-\alpha} \qquad (2)$$

where $F_{1,n-2}^{1-\alpha}$ is the $(1-\alpha)100$ percentile of the F – distribution (Neter, Wasserman, & Kutner, 1983).

2.2 Autogressive Time Series Model

Non-inclusion of one or several key variables and systematic coverage errors in the dependent variable time series, which errors are possibly positively correlated over time lead to autocorrelation in the error terms (Neter, Wasserman, & Kutner, 1983; Gujarati & Porter, 2009). Autocorrelation leads to serious setbacks of the Ordinary Least Square procedures because the minimum variance property does not hold even if the estimates are unbiased. In the same way, the non-inclusion of other key independent random variables in the time series model thus regressing the time series data on its one-step lagged variable may lead to a situation where the time series data and its one-step lagged value are autocorrelated (Gujarati & Porter, 2009).

The autoregressive model is given as

$$y_t = \alpha + \beta y_{t-1} + e_t \qquad (3)$$

where y_t is a time ordered dependent variable, α and β are regression parameters, y_{t-1} is a one period lag of the dependent variable and e_t is the residual /error associated with the estimation of y_t (Neter, Wasserman, & Kutner, 1983; Gujarati & Porter, 2009).

y_t and y_{t-1} are stochastic. This violates the assumption of non-stochastic independent variables in classical least squares theory. Now,

$$e_t = \rho e_{t-1} + \mu_t \qquad (4)$$

The h statistic given as

$$h = \underline{\hat{\rho}} \sqrt{\frac{n}{1 - n[var(\rho)]}} \qquad (5)$$

which exists within the interval $0 < n[var(\rho)] < 1$; where $\underline{\hat{\rho}}$ is the estimate of first-order serial correlation $\underline{\rho}$, n is the sample size and $var(\rho)$ is the variance of the coefficient of the lagged variable y_{t-1}; is used to detect first-order serial correlation in Equation 2.3.8 because the traditional Durbin-Watson d statistic may yield a d value that usually tends to 2, the value of d expected in a truly random sequence (Gujarati & Porter, 2009).

Where the sample size is large, under the null hypothesis that $\underline{\rho} = 0$,

$$h \sim N(0,1) \qquad (6)$$

More so, $\underline{\rho}$ may be estimated as

$$\underline{\hat{\rho}} \approx 1 - \frac{d}{2} \qquad (7)$$

(Gujarati & Porter, 2009), where d is the Durbin-Watson d statistic.

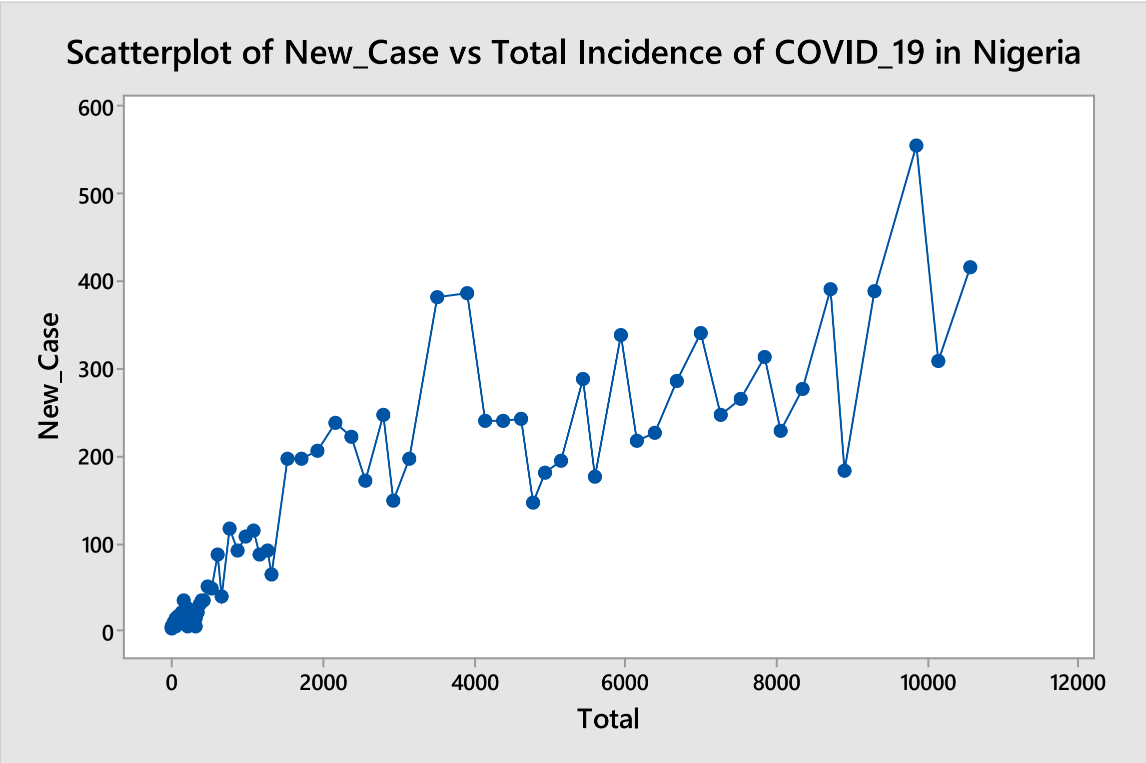

Figure 1: A relationship between New Cases and Current Totals of COVID – 19 in Nigeria.

Figure 1 shows the relationship between current totals and new cases of the virus in which there is a visible direct relationship between the two variables.

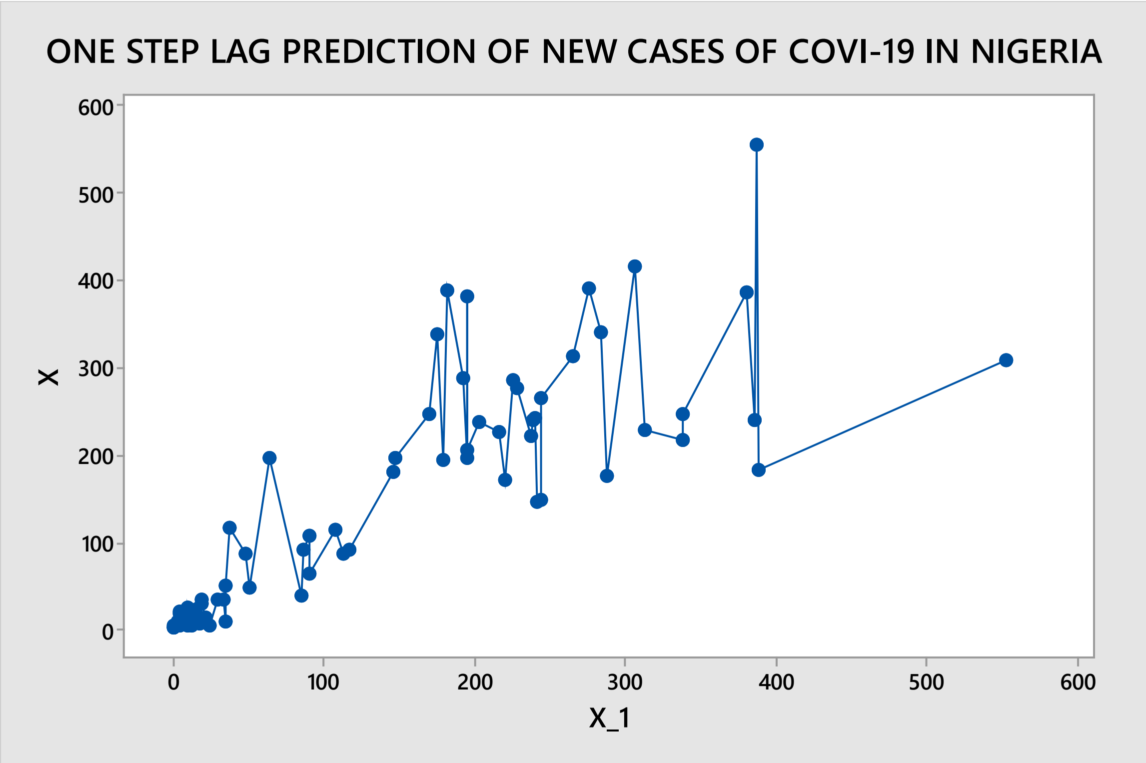

Figure 2: One Step Lag relationship for new cases of COVID-19.

Figure 2 above shows a direct relationship between current cases and its one step lag.

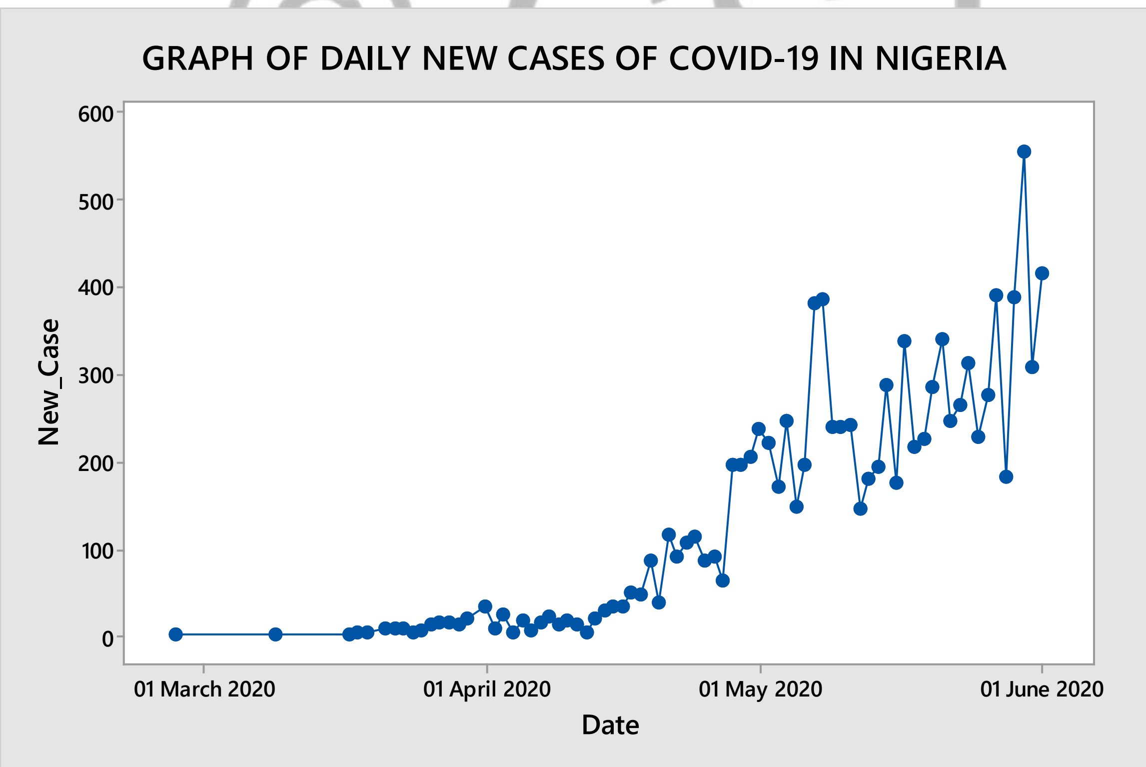

Figure 3: Time Trend of New Cases of COVID – 19 from March to June 2020

The trend shows an upward relationship between New Cases and Time.

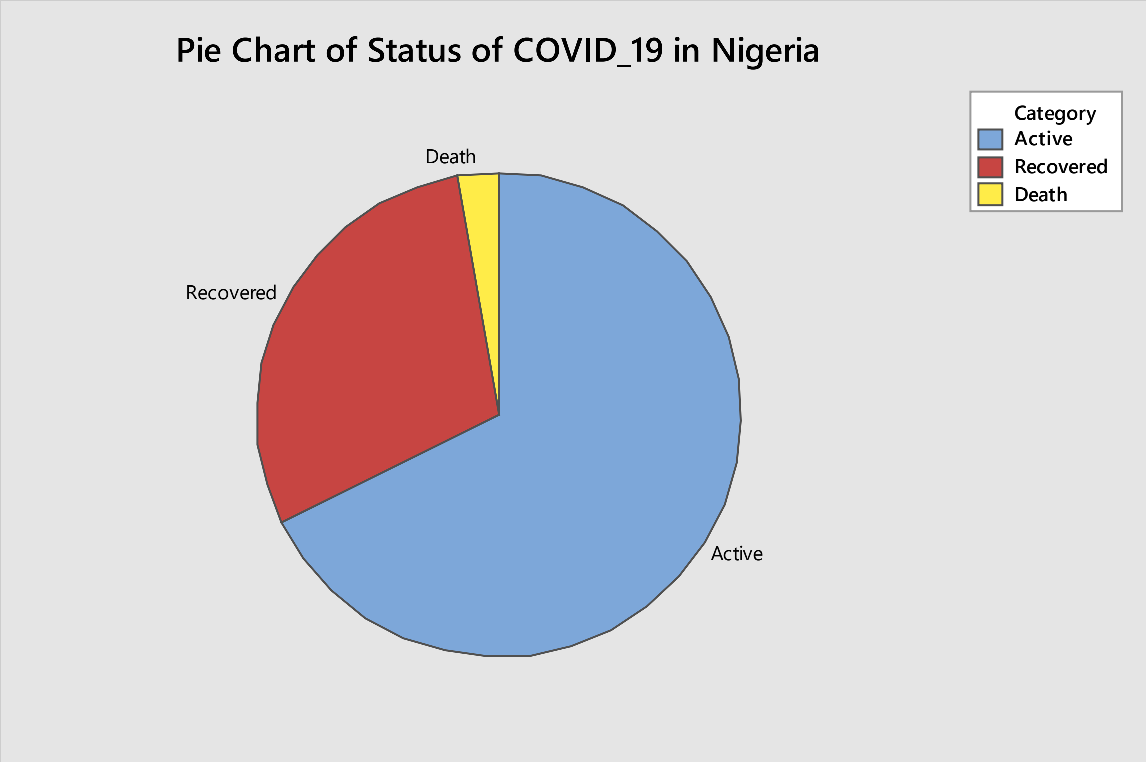

Figure 4: Pie Chart of COVID-19 status in Nigeria

The chart above shows Actives Cases to be the highest followed by Recovered and Death.

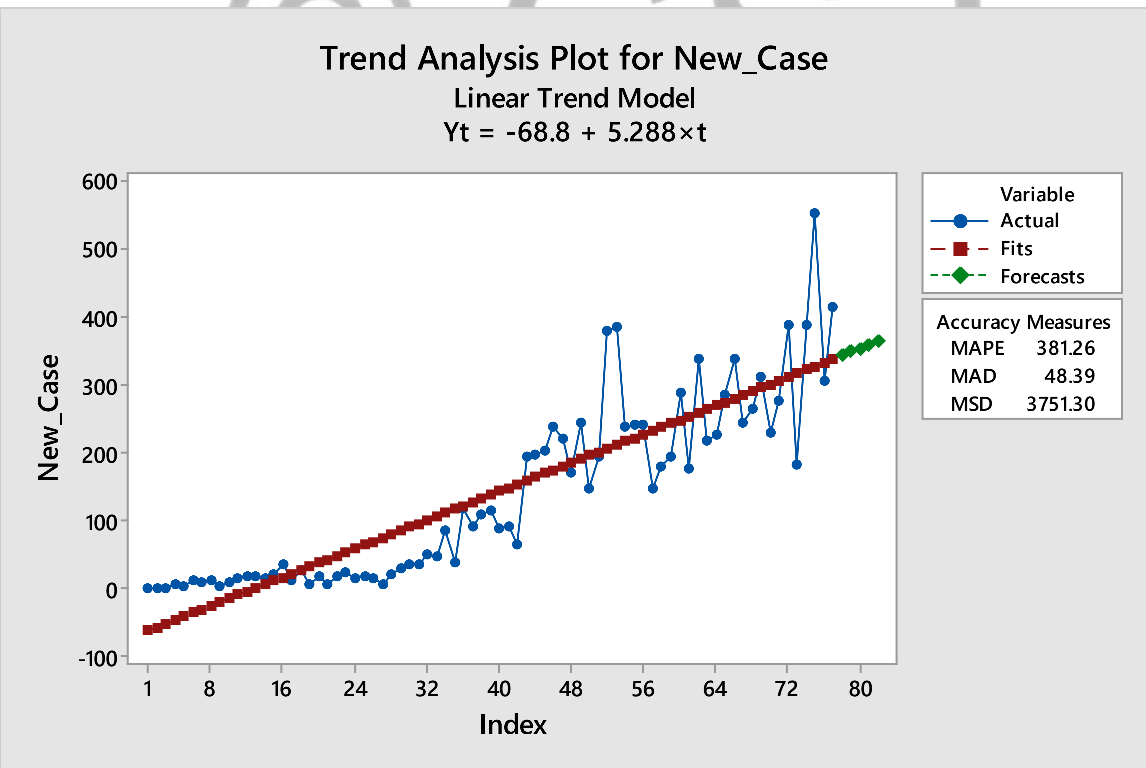

Figure 5: Linear Trend of New Cases of COVID – 19 in Nigeria.

The Linear Trend is

$$y_t = -68.8 + 5.288t \quad (8)$$

7

The trend forecasted new cases for five periods as 344, 349, 354, 359 and 365. All the five forecasts are less than the last observed value showing infection rate was already dropping.

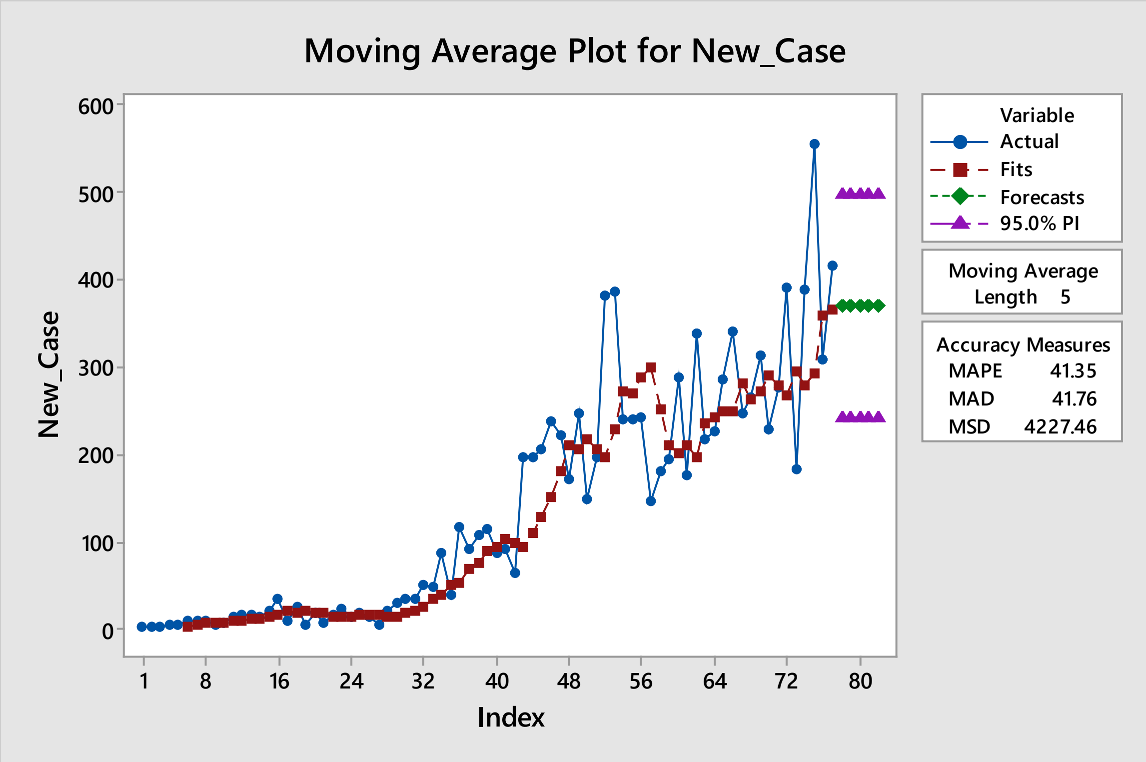

Figure 6: Five Period Moving Average of New Cases of COVID – 19 in Nigeria

The five period moving average forecast is constant at 369 for the next five periods. This is actually less than the last observed New Case showing a drop in future cases.

The Simple Linear Regression Model

The simple linear regression model for the relationship between New Cases and Total reported cases is

$$New\ Case_i = 36.64 + 0.03669 \times Total_i + e_i \qquad (9)$$

An F – value of 248.29 and a P – value of 0.00 show that the regression parameter is significant implying that the model is a good-fit for the data at 0.05 level of significance.

The Autogressive Model:

The autoregressive model was estimated as

$$X_t = 23.6 + 0.8642X_{t-1} + e_t \qquad (10)$$

while the autocorrelation model for the error term was estimated as

$$e_t = -0.29 - 0.354e_{t-1} + \mu_t \qquad (11)$$

The Durbin Statistic was estimated as

$$h = -3.7109$$

At 0.5% level of significance, the null hypothesis of insignificance of the autocorrelation coefficient was rejected showing that autocorrelation exists between the error terms.

Proportional Incidence Analysis

Kogi State recorded a total of 2 cases, 2 active cases, no recovery, and no death giving a 1.00 proportion of active cases in the state. Zamfara State recorded a total of 76 cases, 0 active cases, 1 recovered cases and 5 deaths giving 0.00 proportion of active cases, 0.9342 of recovery and 0.06578 of deaths. Bauchi recorded a total of 240 cases, 11 active cases, 221 recovered cases and 8 deaths giving 0.04583 proportion of active cases, 0.92083 cases recovery and 0.08333 of deaths. Finally, Ondo State recorded a total of 28 cases, 3 active cases, 21 recovered cases and 4 cases yielding 0.10714 proportion of active cases, 0.75 of recovery and 0.1429 proportion of deaths. It then implies that Kogi State has the highest state specific recovery rate followed by Zamfara and Bauchi states while Ondo State has the highest state specific proportion of deaths.

CONCLUSION/RECOMMENDATIONS

It is therefore apt to conclude that the daily new cases of COVID-19 in Nigeria was increasing monotonously within the studied period; the Linear Regression between daily new cases and total existing cases, the Time Series Trend Analysis of daily new Cases, the Autoregressive Time Series Model and the Five Year moving Average Model all forecasted a reduction in the daily New Cases pointing to a future reduction of Daily New Cases of COVID-19 in Nigeria. Analysis of Proportions has shown that Kogi State has the highest state specific recovery rate followed by Zamfara and Bauchi states while Ondo State has the highest state specific proportion of deaths. It is also obvious that the preventive measures in the country are paying off. Hence, it would be recommended that the current preventive measures should be sustained if not improved upon, citizens should continue to observe the preventive measures and the anxiety and fear associated with the coronavirus should be doused since all the predictive models point to a reduction in future occurrence of the virus.

References

Frieman, C. M. (2014). Coronavirus: important emerging human pathogens. *Journal of viroogy, 88*, 5209 - 5212.

Gujarati, N. D., & Porter, C. D. (2009). *BAsic Econometrics.* Singapore: McGraw Hill.

Kenneth, K. J. (2005, November). History and Recent Advances in Coronavirus Discovery. *The Pediatric Infectious Disease Journal*, S223 -S227. doi:10.1097/01.inf.0000188166.17324.60

Neter, J., Wasserman, W., & Kutner, M. H. (1983). *Applied Linear Regression Models.* Illinois: Richard D. Irwin, INC.

WHO. (2020). *COVID-2019: Question and Answers.* Switzerland: World Health Organization. Retrieved June 4, 2020

WHO. (2020). *There is a current outbreak of coronavirus (COVID-19) disease.* Switzerland: World Health Organization. Retrieved June 4, 2020