SciencePG
Science Publishing Group

3^2 Factorial Design and Model for the Spread of COVID-19 in West Africa

Richard Ayamah[1, *], Samuel Kwarteng[1], Awudu Obeng[2], Patrick Kwame Babah[3]

[1]Mathematics/ICT Department, St. Joseph's College of Education, Bechem, Ghana

[3]Mathematics/ICT Department, Wesley College of Education, Kumasi, Ghana

[4]Mathematics/ICT Department, Ada College of Education, Ada, Ghana

Email address:

r.ayamah@yahoo.com (R. Ayamah), samuelkwarteng45@yahoo.com (S. Kwarteng), aobeng17@gmail.com (A. Obeng),
pkbabah@adacoe.edu.gh (P. K. Babah)

[*]Corresponding author

To cite this article:

Richard Ayamah, Samuel Kwarteng, Awudu Obeng, Patrick Kwame Babah. 3^2 Factorial Design and Model for the Spread of COVID-19 in West Africa. *Mathematical Modelling and Applications*. Vol. 6, No. 2, 2021, pp. 17-28. doi: 10.11648/j.mma.20210602.11

Abstract: The coronavirus diseases 2019 (COVID-19) is a worldwide pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-COV-2). This research is therefore aimed at using experimental design to derive a model for the spread of COVID-19 in West Africa. In effect, 3^2 factorial design with mixed factors was used to design the data layout with country as the random factor and status of COVID-19 patient as the fixed factor. Under country, Nigeria, Senegal and Ghana were randomly selected as the three levels from the seventeen West African countries while under status of COVID-19 patient, recovery, death and active cases were the three fixed levels used. The data for the study were collected based on the monthly recorded number of the various COVID-19 cases (i.e. recovery cases, death cases and active cases out of the total confirmed cases) for the period from February to September, 2020. The data were retrieved from the reports on COVID-19 given by the respective country's health authorities and published on their various websites. In effect, a mixed effect model was derived through formulation processes for the prediction of the various COVID-19 cases across West Africa. Also, different residual analyses were conducted for the model adequacy checking and it proved that the model was adequate in the estimation of the various COVID-19 cases in the West African Sub-region. This therefore makes it the first time experimental design and analysis has successfully been used for a study of this nature on the spread of COVID-19 in West Africa.

Keywords: COVID-19, Experimental Design, 3^2 Factorial Design, Mixed Effect Model, Random Factor, Fixed Factor, Residual

1. Introduction

1.1. What Is COVID-19

Coronavirus disease 2019 (COVID-19) is a worldwide pandemic caused by severe acute respiratory syndrome corona virus 2 (SARS-COV-2). A research conducted by Bette Korber and associates at Los Alamos National Laboratory and published in the journal cell shows that a specific change in the SARS-COV-2 genome, previously associated with increased viral transmission and the spread of COVID-19, is more infectious in cell culture. The variant in question, D614G, makes a small but effective change in the virus's 'spike' protein which the virus uses to enter human cells. The virus with D614G change in spike-out completes original strain, but may not increase the severity of the patient's illness. [1].

Although COVID-19 causes only mild illness in most people, it can make some people very ill. More rarely, the disease can be fatal, especially in older people and those with such pre-existing medical conditions as high blood pressure, heart problems, diabetes, etc. This is from WHO report in February 2020, as in reference [2].

The most common symptoms of COVID-19 are fever, tiredness and dry cough. Some patients too may have aches and pains, nasal congestion, runny nose or sore throat. These symptoms are usually mild and begin gradually. Most people (about 80%) recover from these symptoms without needing

special treatment, of which majority are children and young adults. [3].

The best way to prevent and slow down transmission is to be well informed about the COVID-19 virus, the disease it causes and how it spreads. Protect yourself and others from infection y washing your hands or using an alcohol based sanitizer frequently and not touching your face. Also, since the COVID-19 virus spreads primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes, it is important to practice respiratory etiquette (for example, by coughing or sneezing into a flexed elbow, wearing face/nose mask, etc). [4].

1.2. Brief History of COVID-19

1.2.1. In The World

When a new virus is discovered, it is important to understand where it comes from. This is critical to be able to identify and isolate the source and prevent further introductions of the virus into the human population. It also helps to understand the dynamic of the beginning of the outbreak, which can be used to inform the public health response. Understanding the origin of the virus may also aid the development of therapeutics and vaccines.

From WHO report on 26th March 2020, all available evidence suggests that the virus, SARS-COV-2, which causes COVID-19 has a natural animal origin and as such not created from laboratory. [5].

The first human case involving COVID-19 was reported at Wuhan Central Hospital in December 2019 in Wuhan, China. During the early stages of the outbreak, the number of cases doubled at approximately every seven and a half days and the spread moved to other Chinese provinces by mid-January 2020. The infection rate kept on increasing to the extent that by 20th January, China was recording about 140 infections a day, which accrued to 6,174 infected people.

On 30th January, with 7818 confirmed cases across 19 countries, the WHO declared the outbreak a Public Health Emergency of Internal Concern (PHEIC) and then a pandemic on 11th March 2020 as Italy, Iran, South Korea and Japan reported increasing numbers of cases. Later that month, the number of cases outside China quickly surpassed the number of cases inside China. Italy for instance overtook China as the country with the most reported deaths on 19th March and United States of America (USA) took over from Italy as the country with the highest number of confirmed cases in the world. Before this, WHO considered Europe the active centre of the pandemic as at 13th March 2020. According to the Johns Hopkins COVID-19 interactive map, as at 30th September, globally, there were 33706888 confirmed cases, 1008993 deaths and 23432488 recoveries. [6].

1.2.2. In Africa

The COVID-19 pandemic in Africa was confirmed to have spread to the continent on 14th February 2020 and Egypt happened to be the first country to have recorded the first confirmed case. [7]. Chronologically, Egypt was followed by Algeria who had her first case reported on 25th February and then Nigeria who happened to be the third on the continent and first in the Sub-Saharan Africa, with her first case recorded on the 27th February. Apart from these three countries, the first cases in the other African countries were detected in March. However, Lesotho is the last African to have reported her first confirmed case which happened in 13th May 2020. [8].

As at September 30 2020, the most COVID-19 affected African countries were South Africa (confirmed cases: 671, 669 and deaths: 16, 586), Morocco (confirmed cases: 121, 183 and deaths: 2, 152), Egypt (confirmed cases: 103, 079 and deaths: 5, 914), Ethiopia (confirmed cases: 73, 944 and deaths: 1, 177), Nigeria (confirmed cases: 59, 345 and deaths: 1, 113), Algeria (confirmed cases: 51, 213 and deaths: 1, 719), Ghana (confirmed cases: 46, 829 and deaths: 303). The least affected was Seychelles with 41 confirmed cases and no death recorded. Also, the total confirmed cases in Africa was 1, 570, 055 with 35, 554 deaths. [9]

1.2.3. In West Africa

In West Africa where this study is focused, Nigeria happened to be the first country to have reported a COVID-19 case on the 27th February, 2020, followed by Senegal on 29th February. [10]. By March ending, all the seventeen West African countries had reported confirmed COVID-19 cases. As at 30th September 2020, the total number of confirmed COVID-19 cases had reached 187,290 with 2,882 deaths, 166,501 recoveries and 17,980 active cases across the West African sub-region. The five countries with the highest confirmed caseloads are Nigeria (59345), Ghana (46, 829), Côte d'Ivoire (19, 849), Senegal (15, 094) and Guinea (10, 754). However, Ghana and Côte d'Ivoire recorded the highest recovery rates of 98.5% and 97.8% respectively, while Chad and Liberia the highest fatality rates of 7.1% and 6.1% respectively. [11].

1.3. Related Studies

Some researchers across the world have carried out some studies elsewhere which are directly or indirectly related to this study. Anwar Zeb and others for instance used non-standard finite difference (NSFD) and Runge-Kutta fourth order method to derive a mathematical model for COVID-19 which contain isolation class and concluded that their derived model was used to show that the corona virus spreads through contact. It was also used to describe how fast something changes by counting the number of people who are infected and the likelihood of new infections. [12].

Also, in reference [13], Wang and his partners developed epidemiological models to forecast the evolution of the coronavirus and estimate the effectiveness of various intervention measures and their impacts on the economy. In their study, they reviewed a range of mathematical models for the outbreak of COVID-19 which can be employed to estimate the epidemiological trends, including severity, basic reproduction number and herd immunity as well as the potential effects of interventions on COVID-19. They also found that most of the existing epidemiological models of

COVID-19 are typically based on epidemic-dynamic models rather than the statistical models or machine learning.

Ivorra and associates mathematically developed a new θ-SEIHRD model, for the spread of the COVID-19 in China which takes into account the known special characteristics of the COVID-19, as the existence of infectious undetected cases and the different sanitary and infectiousness conditions of hospitalized people. Their model included a novel approach that considers the fraction (θ) of detected cases over the real total infected cases which allows studying of the importance of the ratio and its impact on COVID-19. The model was also able to estimate the needs of beds in hospitals. They further used the reported data on the spread of COVID-19 in about five provinces in China to identify the model parameters which can be of interest for estimating the spread of COVID-19 in other countries. [14].

In reference [15], Li and others established dynamic models of six chambers and established time series models based on different mathematical formulas according to the variation law of an original data on COVID-19 in China. Their results from the time series kinetic model analysis showed that the COVID-19 infection rate and the basic regeneration number of COVID-19 continued to decline while the results from the sensitivity analysis showed that the time it takes for a suspected population to be diagnosed as a confirmed population can have a significant impact on the peak size and duration of the cumulative number of diagnoses. They then explained the results from the model analysis that the emergency intervention measures (such as blocking Wuhan, restricting the flow of people in Hubei province and increasing the support to Wuhan) adopted by the Chinese Government in the early stages of the epidemic had a crucial restraining effect on the original spread of COVID-19.

In this study, the researchers want to answer the following questions:

1. What effects do the countries and patient's status have on the spread of COVID-19 in West Africa?
2. Could there be a model for the prediction of the COVID-19 infection rate across the West African Sub-Region?
3. How would the model for the prediction of the COVID-19 infection rate across the West African Sub-Region be derived?

It is in this direction that this research was conducted by employing, using the available, to derive a model for the spread of COVID-19 in West Africa.

2. Materials and Methods

The purpose of this research was to employ experimental design in modeling the spread of COVID-19 in the West African Sub-Saharan region. In reference [16], Sweeney defined experimental design as a branch of statistics that deals with the design and analysis of experiments.

In particular, factorial design was used for the whole process involved in the model derivation. In his book, Chris Spatz (2005) described factor in experimental design as independent variable, as in reference [17]. Factorial design is

a design in which each complete trial (replication) of an experiment has all possible combinations of the levels of the factors being investigated. Factorial design can be classified into three based on the nature of the factors involved in the experiment. They are factorial design with fixed factors, factorial design with random factors and factorial design with mixed factors. In factorial design with fixed factors, all the factors contain small number of levels and as such all the respective levels are used in the experiment (design). Here, conclusions drawn covers only the levels used in the experimental design. In the case of the factorial design with random factors, the factors involved have levels which have been randomly selected from a large population of possible levels. Here, conclusions drawn cover the entire population levels and not just those which were used in the experimental design. In factorial design with mixed factors, some of the factors are fixed while others are random. In this design, the conclusions drawn about the levels involved in the experiment cover the entire population levels. [18].

In this research, two factors were involved, namely, countries in West Africa as factor A and status of COVID-19 patients as factor B. With factor A, three countries (Nigeria, Senegal and Ghana) were randomly selected from the seventeen West African countries (assumed as population) as the levels while factor B had the status of the COVID-19 patients (Recovery cases, Death cases and Active cases) as the levels.

Since factor A (countries) is a random factor and factor B (status of the COVID-19 patients) is a fixed factor, two-factor factorial design with mixed factors was used for the design, hence the analysis of mixed effect model was employed. More to the point, since there are only two factors with each having three levels, the precise two-factor factorial design used is the 3^2 factorial design.

2.1. Data Collection

The data on the number of confirmed COVID-19 cases recorded under the recovery cases, death cases and active cases were collated from the COVID-19 situation reports and timeline for the months from February to September published on the respective websites associated with each of the three countries involved. In the case of Nigeria, the data was retrieved from Nigeria Centre for Disease Control (NCDC) situation report, as in reference [19]. Ghana's own was retrieved from Ghana Health Service (GHS), as in reference [20]. That of Senegal was retrieved from Wikipedia, as in reference [21]. Each replication (observation) is the cumulated number of COVID-19 status per month in each country for the period from February to September 2020.

2.2. Concept of Two-Factor Factorial Design

Generally, the two-factor factorial design is as given in table 1 below. Passing onto the general case, let y_{ijk} be the observed response when factor A is at the ith level (i=1, 2, …, a) and factor B is at the jth level (j=1, 2, …, b) for the kth replicate (k=1, 2, …, n). Example, y_{124} means the 4th observation (replicate) taken at the first level of factor A, and

the second level of factor B. The order in which the *n* observations per cell are arranged is in correspondence with the order of the recordings. It follows that the $y_{i..}$ is the total of all the observations under the *i*th level of factor A; $y_{.j.}$ is the total of all the observations under the *j*th level of factor B; $y_{ij.}$ is the total of all the observations in the *ij*th cell; and $y_{...}$ is the grand total of all the observations in the whole data set.

Also, $\bar{y}_{i..}$, $\bar{y}_{.j.}$, $\bar{y}_{ij.}$ and $\bar{y}_{...}$ are the corresponding row, column, cell and grand means. Mathematically, the terms are defined as follows:

$$y_{i..} = \sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} => \bar{y}_{i..} = \frac{y_{i..}}{bn}, (i = 1, 2, \ldots a)$$

$$y_{.j.} = \sum_{i=1}^{a}\sum_{k=1}^{n} y_{ijk} => \bar{y}_{.j.} = \frac{y_{.j.}}{an}, (j = 1, 2, \ldots b)$$

$$y_{ij.} = \sum_{k=1}^{n} y_{ijk} => \bar{y}_{ij.} = \frac{y_{ij.}}{n}, (i = 1, 2, \ldots a; j = 1, 2, \ldots b)$$

$$y_{...} = \sum_{ic=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} => \bar{y}_{...} = \frac{y_{...}}{abn}$$

***Table 1.** General layout for the two-factor factorial design.*

FACTOR A	FACTOR B					Row Totals	Row Means
	1	2	...	B			
1	$y_{111}, y_{112}, y_{113}, \ldots, y_{11n}$ $y_{11.}, \bar{y}_{11.}$	$y_{121}, y_{122}, y_{123}, \ldots, y_{12n}$ $y_{12.}, \bar{y}_{12.}$	...	$y_{1b1}, y_{1b2}, y_{1b3}, \ldots, y_{1bn}$ $y_{1b.}, \bar{y}_{1b.}$		$y_{1..}$	$\bar{y}_{1..}$
2	$y_{211}, y_{212}, y_{213}, \ldots, y_{21n}$ $y_{21.}, \bar{y}_{21.}$	$y_{221}, y_{222}, y_{223}, \ldots, y_{22n}$ $y_{22.}, \bar{y}_{22.}$	...	$y_{2b1}, y_{2b2}, y_{1b3}, \ldots, y_{21n}$ $y_{2b.}, \bar{y}_{2b.}$		$y_{2..}$	$\bar{y}_{2..}$
:	...		...	...		...	...
A	$y_{a11}, y_{a12}, y_{a13}, \ldots, y_{a1n}$ $y_{a1.}, \bar{y}_{a1.}$	$y_{a21}, y_{a22}, y_{a23}, \ldots, y_{a2n}$ $y_{a2.}, \bar{y}_{a2.}$	....	$y_{ab1}, y_{ab2}, y_{ab3}, \ldots, y_{abn}$ $y_{ab.}, \bar{y}_{ab.}$		$y_{a..}$	$\bar{y}_{a..}$
Column Totals	$y_{.1.}$	$y_{.2.}$	...	$y_{.b.}$		$y_{...}$	$\bar{y}_{...}$
Column Means	$\bar{y}_{.1.}$	$\bar{y}_{.2.}$	...	$\bar{y}_{.b.}$			

2.3. Model Formulation and Parameters Estimation

Now considering the situation where one of the factors (A) is random and the other (B) is fixed, a mixed effects model arises and it is given by:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + E_{ijk} \quad (1)$$

Where μ is the overall mean effect, α_i is the effect of the ith level of factor A, β_j is the effect of the jth level of factor B, $(\alpha\beta)_{ij}$ is the effect of the interaction between α_i and β_j, and E_{ijk} is the random error component.

Using the least squares method, it can be shown that the normal equations are:

$$\mu: abn\hat{\mu} + bn\sum_{i=1}^{a}\hat{\alpha}_i + an\sum_{j=1}^{b}\hat{\beta}_j + n\sum_{i=1}^{a}\sum_{j=1}^{b}\widehat{\alpha\beta}_{ij} = y_{...} \quad (2)$$

$$\alpha_i: bn\hat{\mu} + bn\hat{\alpha}_i + n\sum_{j=1}^{b}\hat{\beta}_j + n\sum_{j=1}^{b}\widehat{(\alpha\beta)}_{ij} = y_{i..} \quad (3)$$

$$\beta_j: an\hat{\mu} + n\sum_{j=1}^{b}\hat{\alpha}_i + an\hat{\beta}_j + n\sum_{i=1}^{a}\widehat{(\alpha\beta)}_{ij} = y_{.j.} \quad (4)$$

$$(\alpha\beta)_{ij}: n\hat{\mu} + n\hat{\alpha}_i + n\hat{\beta}_j + n(\widehat{\alpha\beta})_{ij} = y_{ij.} \quad (5)$$

In order to obtain a unique solution, the following constraints are imposed.

$$\sum_{i=1}^{a}\hat{\alpha}_j = 0,$$

$$\sum_{j=1}^{b}\widehat{\beta}_j = 0,$$

$$\sum_{i=1}^{a}(\widehat{\alpha\beta})_{ij} = 0,$$

$$\sum_{j=1}^{b}\widehat{(\alpha\beta)}_{ij} = 0$$

Hence, the normal equations, (2) to (5), simplified considerably, give the results

$$\hat{\mu} = \bar{y}_{...} \quad (6)$$

$$\hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{...} \quad (7)$$

$$\hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...} \quad (8)$$

$$\widehat{(\alpha\beta)}_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{...} \quad (9)$$

Now assuming E_{ijk} is negligible, and substituting equations (6), (7), (8) and (9) into (1), the estimated value of y_{ijk} may be found as:

$$\hat{y}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \widehat{(\alpha\beta)}_{ij}$$

$$= \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{...})$$

$$= \bar{y}_{ij.}$$

This means that any observation (y_{ijk}) in a cell is estimated by the cell mean $(\bar{y}_{ij.})$. [18].

2.4. Model Adequacy Checking

Before a conclusion is drawn on the accuracy of a model, the adequacy of the underlying model is checked. The primary diagnostic tool for model adequacy checking is the residual analysis. The residual of each observation for two-factor factorial model is:

$$E_{ijk} = y_{ijk} - \hat{y}_{ijk}.$$

$$\text{But } \hat{y}_{ijk} = \bar{y}_{ij.}$$

$$=> E_{ijk} = y_{ijk} - \bar{y}_{ij.}$$

The residual analysis consists of two main categories. They are:
1. the numerical residual analysis, and
2. the graphical residual analysis.

In the numerical residual analysis, the standardized residual (d_{ijk}) is calculated using the equation:

The standardized residual $\left(d_{ijk}\right) = \frac{E_{ijk}}{\sqrt{MS_E}}$.

Where E_{ijk} is the largest residual among the other residuals and MS_E is the mean sum of the squares of the error component, which is also given by:

$$MS_E = \frac{SS_E}{ab(n-1)}$$

Where SS_E is the sum of the squares of the error component, a is the number of levels under factor A, b is the number of levels under factor B and n is the number of observation (replicates) per cell.

If the value of d_{ijk} falls within ± 3 (ie $-3 \leq d_{ijk} \leq 3$), then the normality assumption of the residuals (errors) is assured, hence the residuals would not reveal anything troublesome. Therefore, the model is adequate.

In the case of the graphical residual analysis, three main graphs were considered. They are the normal probability plot, the residuals plot in time sequence and the plot of residuals versus fitted values $(\hat{y}_{ijk})$. In the normal probability plot, if the underlying error distribution is normal, the plot would portray some kind of linearity. This would be ensured by placing more emphasis on the dots of the central values than on the dots of the extreme values.

Plotting the residuals in order of the time the data were collected is helpful in detecting the correlation between the residuals. This would be portrayed through the uniform spread out of the dots about the mean of the residuals. This would also imply that there is no violation of the independency or constant variance assumptions. Hence, the model is adequate.

When a graph of the residuals against the fitted values is plotted, the model would be adequate and the various assumptions satisfied if the dots reveal no obvious pattern (that is the plot would be structureless).

3. Results

The detail of the raw data on the COVID-19 cases, entailing the number of recovery cases, death cases and active cases for Nigeria, Senegal and Ghana are as presented in table 2 below. The first column is headed by country, under which we have Nigeria, Senegal and Ghana; the second column is headed by month, under which we have the listed months from February to September for each country; and the third column is headed by the status of COVID-19 patient, under which we have recovery cases, death cases and active cases.

Table 2. The raw data showing the monthly COVID-19 recovery, death and active cases in Nigeria, Senegal and Ghana for the period February to September 2020.

COUNTRY	MONTH	STATUS OF COVID-19 PATIENT		
		Recovery Cases	Death Cases	Active Cases
Nigeria	February	0	0	1
	March	9	2	128
	April	310	56	1555
	May	2688	229	6868
	June	6739	303	15358
	July	9819	289	18707
	August	18073	134	11357
	September	8720	99	7378
Senegal	February	0	0	0
	March	40	0	135
	April	294	9	455
	May	1467	33	1212
	June	2630	70	448
	July	2345	93	1001
	August	4227	79	637
	September	1228	25	2379
Ghana	February	0	0	0
	March	52	5	104
	April	160	12	1845
	May	2774	21	5273
	June	10573	79	4458
	July	16754	74	3308
	August	9165	85	904
	September	2528	27	494

From table 2 above, all the three countries involved started with lower recordings under all the COVID-19 cases in February and this continued to increase tremendously to a point and then began to decrease after some months time. For instance, Nigeria's recovery cases started with 0 patients in February increased through the consecutive months to 18,073 in August and then started reducing to 8720 in September ending. Senegal's own was 0 in February, increased though the months to 4227 and then reduced to 1228 in September. Also, Ghana started with 0 patients in February, increased to 16754 in July and then started reducing through August to 2528 in September.

The two-factor factorial design with mixed factors on the raw data above is as shown in table 3 below. The row factor

is country with level headings Nigeria, Senegal and Ghana. The column factor is the status of the COVID-19 patient with level headings Recovery cases, Death cases and Active cases. Each cell contains eight replications (observations) arranged in correspondence with the order of the months in which they are recorded for the period from February to September with the k=1, 2, 3, …, 8 term of y_{ijk} assigned to the months correspondingly. The last two observations in each cell are also the cell total ($y_{ij.}$) and cell mean ($\bar{y}_{ij.}$) respectively. The last two columns of the table contain the row totals ($y_{a..}$) and row means ($\bar{y}_{a..}$) respectively while the last two rows of the table contain the column totals ($y_{.b.}$) and column means ($\bar{y}_{.b.}$) respectively. Moreover, the last cell at the right down corner contains the grand total ($y_{...}$) and the mean ($\bar{y}_{...}$).

The totals and means were calculated using the corresponding equations preceding table 1 above.

Table 3. *Two-factor factorial design with mixed factors for the data on COVID-19 cases in Nigeria, Senegal and Ghana.*

COUNTRY	STATUS OF COVID-19 PATIENT			Row Totals ($y_{i..}$)	Row Means ($\bar{y}_{i..}$)
	Recovery Cases	Death Cases	Active Cases		
Nigeria	0 6739	0 303	1 15358	116822	4867.583
	9 9819	2 289	128 18707		
	310 18073	56 134	1555 11357		
	2688 8720	229 99	6868 7378		
	$y_{11.}$ =50358 $\bar{y}_{11.}$ =6294.75	$y_{12.}$ =1112 $\bar{y}_{12.}$ =139	$y_{13.}$ =65352 $\bar{y}_{13.}$ =8169		
Senegal	0 2630	0 70	0 448	18807	783.625
	40 2345	0 93	135 1001		
	294 4227	9 79	455 637		
	1467 1228	33 25	1212 2379		
	$y_{21.}$ =12231 $\bar{y}_{21.}$ =1528.875	$y_{22.}$ =309 $\bar{y}_{22.}$ =38.625	$y_{23.}$ =6267 $\bar{y}_{23.}$ =783.375		
Ghana	0 10573	0 79	0 4458	62695	2612.292
	52 16754	5 74	104 3308		
	160 9165	12 85	1845 904		
	2774 2528	21 27	5273 494		
	$y_{31.}$ =46006 $\bar{y}_{31.}$ =5750.75	$y_{32.}$ =303 $\bar{y}_{32.}$ =37.875	$y_{33.}$ =16386 $\bar{y}_{33.}$ =2048.25		
Column Totals ($y_{.j.}$)	108595	1724	88005	$y_{...}$ =198324	
Column Means ($\bar{y}_{.j.}$)	4524.792	71.833	3666.875	$\bar{y}_{...}$ =2754.5	

Since each observation in a cell is estimated by the mean of the cell it is well to use the column means ($\bar{y}_{.j.}$) to estimate the various COVID-19 cases across the West African Sub-region. Not all, the ground mean ($\bar{y}_{...}$) is used to estimate the population's average COVID-19 infection rate. Hence, from table 3 above, the number of COVID-19 recovery, death and active cases across the West African Sub-region for the period between February and September 2020 were estimated at about 4524.792, 71.833 and 3666.875 respectively per month. Also, the monthly COVID-19 infection rate across West Africa was estimated at 2754.5.

3.1. Computational Process

The computational processes involved here are the calculation of the respective sum of squares, mean sum of squares and the estimated errors/variances (i.e. σ^2's)

Finding the respective sum of squares, with reference to table 3, we have:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y^2{}_{ijk} - \frac{y^2{}_{...}}{abn} = (0^2 + 9^2 + 310^2 + \cdots + 2379^2) - \left[\frac{198324^2}{(3)(3)(8)}\right]$$

$$= (712404136 + 258668 + 984496180 + \cdots + 8962309) - 546283458$$

$$= 1897636748$$

$$SS_A = \frac{1}{bn}\sum_{i=1}^{a} y^2{}_{i..} - \frac{y^2{}_{...}}{abn} = \left[\frac{1}{3(8)}(116822^2 + 62695^2 + 18807^2)\right] - \frac{198324^2}{(3)(3)(8)}$$

$$= 747156081.6 - 546283458 = 200872623.6$$

$$SS_B = \frac{1}{an}\sum_{j=1}^{b} y^2{}_{.j.} - \frac{y^2{}_{...}}{abn} = \left[\frac{1}{3(8)}(108595^2 + 1724^2 + 88005^2)\right] - \frac{198324^2}{(3)(3)(8)}$$

$$= 814196926.1 - 546283458 = 267913468.1$$

$$SS_{Subtotals} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y^2_{ij.} - \frac{y^2_{...}}{abn} = \left[\frac{1}{8}(50358^2 + 1112^2 + 65352^2 + \cdots + 65352^2)\right] - \frac{198324^2}{(3)(3)(8)}$$

$$=1172770198 - 546283458 = 626486740$$

$$SS_{AB} = SS_{Subtotals} - SS_A - SS_B$$

$$= 626486740 - 200872623.6 - 267913468.1$$

$$= 157700648.3$$

$$SS_E = SS_T - SS_{AB} - SS_A - SS_B = SS_T - SS_{Subtotals}$$

$$= 189763748 - 626486740$$

$$= 1271150008$$

Calculating the corresponding mean sum of squares, we have:

$$MS_A = \frac{SS_A}{a-1} = \frac{200872623.6}{3-1} = 100436311.8$$

$$MS_B = \frac{SS_B}{b-1} = \frac{267913468.1}{3-1} = 133956734.1$$

$$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)} = \frac{157700648.3}{(3-1)(3-)} = 39425162.08$$

$$MS_E = \frac{SS_E}{ab(n-1)} = \frac{1271150008}{(3)(3)(8-1)} = 20176984.25$$

Not all, the variability of the effects about the random factor A (country), interaction between country and COVID-19 status, and that of the error are respectively estimated as:

$$\widehat{\sigma^2}_A = \frac{MS_A - MS_E}{an} = \frac{100436311.8 - 20176984.25}{3(8)} = 3344138.648$$

$$\widehat{\sigma^2}_{AB} = \frac{MS_{AB} - MS_E}{n} = \frac{39425162.08 - 20176984.25}{8} = 2406022.229$$

$$\widehat{\sigma^2}_E = MS_E = 20176984.25$$

These are the variance components of the model after eliminating the mean squares containing the fixed factors. And it is clear that none of these variance components is zero or negative.

3.2. The Prediction Model

Since the three levels of factor A (country) were randomly selected from seventeen countries in West Africa and the three levels of factor B (the status of COVID-19 patient) were fixed, two-factor factorial design with mixed factors was used. Hence, the mixed effect model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + E_{ijk}.$$

Through the model formulation processes and under some constraints, the number of COVID-19 patients under a particular status per country may be estimated by the corresponding cell mean. This is mathematically given by:

$$\hat{y}_{ijk} = \bar{y}_{ij.}$$

On the broader perspective, the number of COVID-19 patients under a particular status in West Africa may be estimated by the column mean. This is:

$$\hat{y}_{ijk} = \bar{y}_{.j.}$$

Precisely, the number of the recovery, death and active COVID-19 cases per month in West Africa could be estimated as 4524.792, 71.833 and 3666.875 respectively. These are approximated to the nearest whole figures as 4525, 72 and 3667 respectively.

Also, the population mean of the number of confirmed COVID-19 cases per month in West Africa is estimated by the sample grand mean. That is:

$$\hat{\mu} = \bar{y}_{...} = 2754.5 \approx 2755 \text{ (to the nearest whole figure)}$$

3.3. Results of the Model Adequacy Checking

Having derived the model for the spread of COVID-19 in West Africa, there is the need to check the adequacy of the model. This is done by first finding the residuals from the

prediction model using the relation:

$$E_{ijk} = y_{ijk} - \hat{y}_{ij.} = y_{ijk} - \bar{y}_{ij.}$$

That is the difference between the actual value and the estimated/predicted value. The detail of the results is as shown in table 4 below. In table 4, the first column is labeled standard order, that is the order in which the observations (in

table 3) are arranged, starting from the first cell down to the last cell of the recovery case column through the death case column to the last cell of the active case column. The second, third and fourth columns of table 4 are headed by Actual value (y_{ijk}), Predicted value ($\bar{y}_{ij.}$) and Residual (E_{ijk}) respectively.

***Table 4.** Residuals from the prediction model.*

Standard Order	Actual value (y_{ijk})	Predicted value ($\bar{y}_{ij.}$)	Residual (E_{ijk})
1	0	6294.75	- 6294.75
2	9	6294.75	- 6285.75
3	310	6294.75	- 5984.75
4	2688	6294.75	- 3606.75
5	6739	6294.75	444.25
6	9819	6294.75	3524.25
7	18073	6294.75	11778.25
8	8720	6294.75	2425.25
9	0	5750.75	- 5750.75
10	52	5750.75	-5698.75
11	160	5750.75	- 5590.75
12	2774	5750.75	- 2976.75
13	10573	5750.75	4822.25
14	16754	5750.75	11003.25
15	9165	5750.75	3414.25
16	2528	5750.75	- 3222.75
17	0	1528.875	- 1528.875
18	40	1528.875	- 1488.875
19	294	1528.875	- 1234.875
20	1467	1528.875	- 61.875
21	2630	1528.875	1101.125
22	2345	1528.875	816.125
23	4227	1528.875	2698.125
24	1228	1528.875	- 300.875
25	0	139	- 139.00
26	2	139	-137.00
27	56	139	-83.00
28	229	139	90.00
29	303	139	164.00
30	289	139	150.00
31	134	139	-5.00
32	99	139	-40.00
33	0	37.875	-37.875
34	5	37.875	-32.875
35	12	37.875	-25.875
36	21	37.875	-16.875
37	79	37.875	41.125
38	74	37.875	36.125
39	85	37.875	47.125
40	27	37.875	- 10.875
41	0	38.625	- 38.625
42	0	38.625	- 38.625
43	9	38.625	- 29.625
44	33	38.625	- 5.625
45	70	38.625	31.375
46	93	38.625	54.375
47	79	38.625	40.375
48	25	38.625	- 13.625
49	1	8169	-8168.00
50	128	8169	-8041.00
51	1555	8169	-6614.00
52	6868	8169	-1301.00
53	15358	8169	7189.00
54	18707	8169	10538.00
55	11357	8169	3188.00
56	7378	8169	-791.00

Standard Order	Actual value (y_{ijk})	Predicted value ($\bar{y}_{ij.}$)	Residual (E_{ijk})
57	0	2048.25	-2048.25
58	104	2048.25	- 1944.25
59	1845	2048.25	- 203.25
60	5273	2048.25	3224.75
61	4458	2048.25	2409.75
62	3308	2048.25	1259.75
63	904	2048.25	- 1144.25
64	494	2048.25	- 1554.25
65	0	783.375	- 783.375
66	135	783.375	- 648.375
67	455	783.375	- 328.375
68	1212	783.375	428.625
69	448	783.375	- 335.375
70	1001	783.375	217.625
71	637	783.375	-146.375
72	2379	783.375	1595.625

From table 4 above, it is clear that the two extreme residuals (11778.25 and 11003.25) were recorded under recovery cases of COVID-19 in August and July in Nigeria and Ghana respectively. The third extreme residual (10538.00) was recorded under the active cases of COVID-19 in July and in Nigeria.

Now using the largest residual ($E_{117} = 11778.25$), the standardized residuals (d_{ijk}) is given by:

Standardized residuals ($d_{117}) = \frac{E_{117}}{\sqrt{MS_E}} = \frac{11778.25}{\sqrt{20176984.25}} \approx 2.622$

The standardized residuals (d_{117}) ≈ 2.622 falls within the interval ± 3, indicating that the normality assumption of the residuals is assured. Also, the errors are independent of the variability of the country's effects, COVID-19 status's effects, and the effects of the interaction between country and COVID-19 status.

The normal probability plot presents normal probability (in percentage) on the vertical axis and residuals on the horizontal axis. The detail of the plot is as shown in the excel output in figure 1 below:
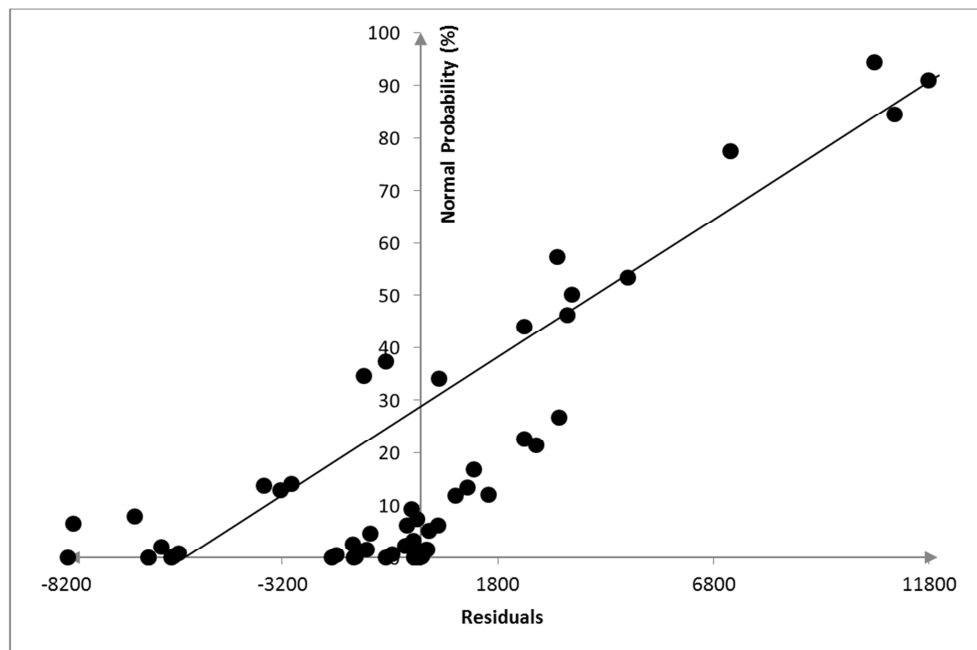

Figure 1. Normal probability plot of the residuals (E_{ijk}) of the predicted number of COVID-19 cases.

In figure 1 above, the normal probability generally exhibits some kind of linearity from left to right in ascending order. Some plots (dots) too are scrambled around the foot of the vertical axis. However, the effect of these scrambled residuals may not be significant on the variance than that of those who portray the linearity. Hence, the independency of the errors (residuals) was not violated in a significant manner.

Plotting residuals against time in months, the excel output

is as shown in figure 2 below. The months are coded in a way that February is assigned 1, March is assigned 2 up to September which is assigned 8. This plot of residuals in time order of data collection on the number of COVID-19 cases from February to September 2020 shows that the plots are uniformly distributed about the horizontal axis. This implies that the independence and constant variance assumption on the errors had not been violated.
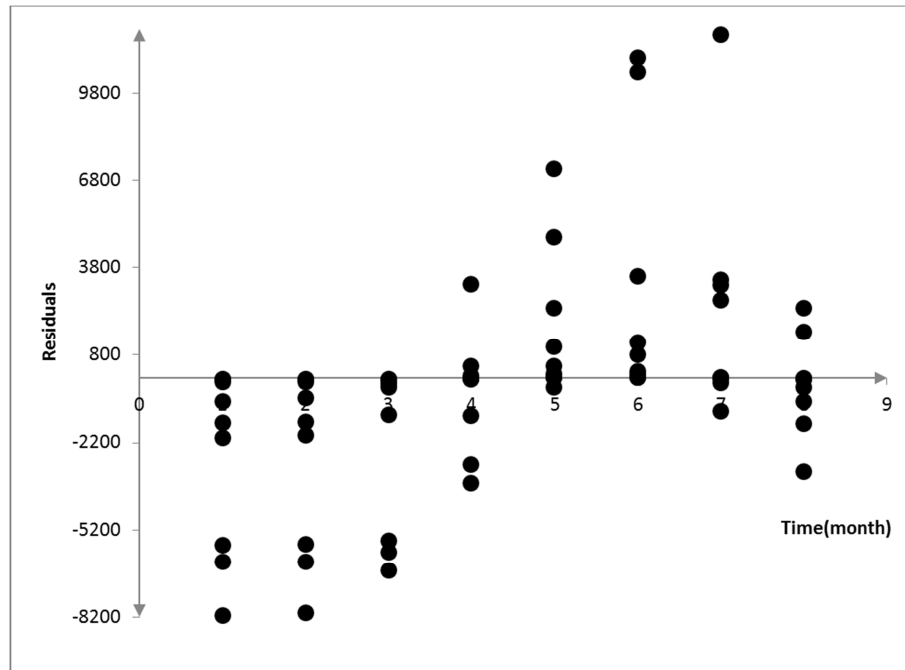
Figure 2. *Graph of residuals of predicted number of COVID-19 cases against time in months.*

Another way of graphical residual analysis for the model adequacy checking is plotting the residuals against the predicted values ($\hat{y}_{ijk}$). This is as shown by the excel output in figure 3 below. The predicted values ($\hat{y}_{ijk}$) are the cell means ($\bar{y}_{ij.}$) in table 3. In the graph, even though the plots exhibit series of vertical linearity about the horizontal axis, they do not together follow a unique pattern, hence they are structureless. This is a clear indication that the normality and independency assumptions of the errors (residuals) were accurate.
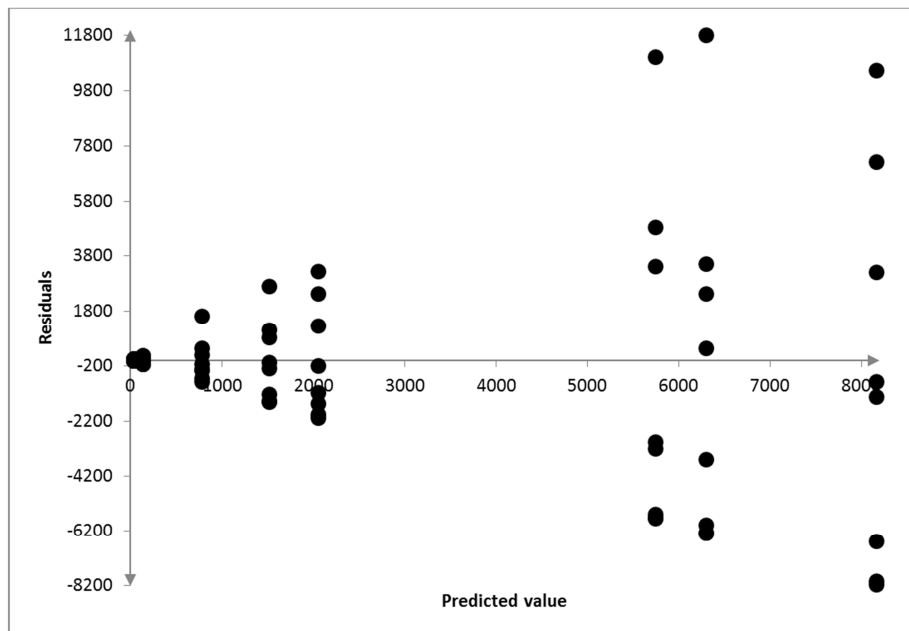
Figure 3. *Graph of residuals against predicted values of the number of COVID-19 cases.*

4. Discussion

This research was purposed at using experimental design to derive a model for the spread of COVID-19 across the West African Sub-region. In effect, two-factor factorial design with mixed factors was used, resulting in the design shown in table 3.

From table 3, it is clear that the average monthly COVID-19 recovery, death and active cases per country in West Africa were about 4525, 72 and 3667 respectively for the period between February and September, 2020. Also, the

average monthly number of confirmed COVID-19 cases was 2755 per country in West Africa right from February to September 2020. These values were estimated by using the various terms in the mixed effect model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + E_{ijk}.$$

Since none of the estimated variance components ($\widehat{\sigma^2}_A = 3344138.648$, $\widehat{\sigma^2}_{AB} = 2406022.229$ and $\widehat{\sigma^2}_E = 20176984.25$) was zero or negative, it implies that the effect of the country, the effect of COVID-19 patient status, the effect of interaction between country and COVID-19 patient status, and the effect of the error incurred during the process all helped in explaining the spread of COVID-19 in West Africa, just as Ivorra and associates' were able to use the reported data on the spread of COVID-19 in about five provinces in China to identify the model parameters which could be of interest in estimating the spread of COVID-19 in other countries, as in reference [14].

Also, since the standardized error (d_{117}) = 2.622 falls within the interval ± 3, the normality and the independency of the residuals are assured. Hence, the mixed effect model is adequate for the prediction of the various COVID-19 cases per country across the West African Sub-region.

From the normal probability plot in figure 1, the linearity exhibited by the plots is an indication that the residuals (errors) recorded are normal and independent. Meaning that the mixed effect model is appropriate for the prediction of the various COVID-19 cases. Also, in figure 2, the plots were uniformly spread out about the horizontal axis, confirming the independence and normal assumptions of the errors (residuals). Hence, the adequacy of the mixed effect model. Moreover, the plot of residuals verses predicted values in figure 3 being structureless proves the normality and independency assumptions of the residuals, signifying the adequacy of the mixed effect model.

From table 3, it is clear that Nigeria led in all the COVID-19 cases among the three randomly selected West African countries followed by Ghana and Senegal. With this, it was revealed that Nigeria happens to have the highest number of contact times followed by Ghana and Senegal. Through this, further investigation revealed that most of the West African countries with higher contact times recorded higher COVID-19 infection rate, and this is in line Zebs and his associates' findings through their derived model that coronavirus spreads much through higher number of contact times, as in reference [12].

Last but not the least, since the processes involved in the derivation and formulation of the mixed effect model were basically statistical, the derived model is classified as a statistical model and this would help in the increase of the few number of epidemiological models for COVID-19 based on statistical process, as revealed by Wang and co in reference [13].

5. Conclusion

3^2 factorial design with mixed factors was successfully used for the data on the number of the various COVID-19 cases. The two factors involved serve as the exponent 2 while the three levels of each factor serve as the base 3, hence the 3^2 factorial design. Also, the levels of factor A (country) were randomly selected while those of factor B (status of COVID-19 patient) were fixed, hence the mixed effect aspect of the design.

From table 3, the 3^2 factorial design used to construct the table helped in the adequate estimation of the average number of COVID-19 recovery cases, death cases and active cases per month in West Africa. These are the corresponding column means $\bar{y}_{.1.} = 4524.792$, $\bar{y}_{.2.} = 71.833$ and $\bar{y}_{.3.} = 3666.875$, and these are approximated to the nearest whole figures as $\bar{y}_{.1.} = 4525$, $\bar{y}_{.2.} = 72$ and $\bar{y}_{.3.} = 3667$ respectively since there is no part human being. Also, the average number of COVID-19 confirmed cases per month in West Africa was adequately estimated by using the grand mean ($\bar{y}_{...}$) = 2754.5, which is 2755 to nearest number of human beings.

All the residual analysis conducted for the model adequacy checking proved that the residuals (errors) incurred in the study are normal and independent. Therefore, there was enough evidence to neglect the residual (error) effect, E_{ijk}, in the mixed effect model. Hence, the model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

is adequate for the prediction of the various COVID-19 cases in West Africa. Also, the various constraints imposed during the model formulation were in order. This means that the various estimations made were correct and adequate.

The uniqueness of this research is portrayed in the fact that it is the first time experimental design and analysis is being used in the study of COVID-19 issue since its discovery in December 2019 in China.

The researchers would also wish to recommend that future researches should be directed towards the use of factorial design to study the effect of any possible vaccine on the spread of COVID-19 in a locality.

Acknowledgements

References

[1]   Bette Korber, et al, (2020). Newer Variant of COVID-19 Causing Virus Dominates Global Infections. Science Daily. (www.sciencedaily.com).

[2]   WHO (2020, February). COVID-19: Situation Reports in 2020. WHO. (www.who.int/situation-reports).

[3]   WHO (2020, March 8). Knowing the Risk for COVID-19; Media Statement. WHO. (www.who.int/news>detail>08-03-2020).

[4]   WHO. (2020). Coronavirus. WHO. (www.who.int/health-topics).

[5]   WHO. (2020, March 26). Origin of SARS-COV-2. Google.com. (apps.who.int>bitstreams>retrieve).

[6]   Sydney Jennings (2020, September 30). COVID-19 Update: US and Global Cases, Deaths and Recoveries as of September 30, 2020. Patient Care.

[7]   Wikipedia. (2020). Coronavirus Disease 2019. Wikipedia. (en.m.wikipedia.org).

[8]   Shabir Ahmad Lone and Aijaz Ahmad. (2020, June 15). COVID-19 Pandemic- an African Persspective. 9 (1). pp 1300-1308. Emerging Microbes and Infections. doi: 10. 1080/22221751. 2020. 177513. (www.tandfoline.com).

[9]   WHO. (2020, September 30). Coronavirus-Africa: WHO COVID-19 Africa Update. Africanews. (www.africanews.com).

[10]   WHO. (2020, March 2). COVID-19 Pandemic in Senegal. Wikipedia. (www.en.m.wikipedia.org).

[11]   UNFPA (2020, September 30). COVID-19 Situation Report Number 8 for West and Central Africa. UNFPA. (www.unfpa.org.>resources>covid-19).

[12]   Anwar Zeb, Ebraheem Alzahrani, Vedat Suat Erturk and Gul Zaman. (2020, June 29). Mathematical Model for Coronavirus Disease 2019 (COVID-19) Containing Isolation Class. Biomed Research International. Vol. 2020. pp 1-7. (doi.org/10.1155/2020/3452402).

[13]   Ning Wang, Yuting Fu, Hu Zhang and Huipeng Shi. (2020, May 22). An Evaluation of Mathematical Models for the Outbreak of COVID-19. Precision Clinical Medicine. 3 (2). pp 85-93. (Google Scholar).

[14]   B. Ivorra, M. R. Ferrandez, M. Vela-Perez and A. M. Ramos. (2020, April 30). Mathematical Modeling of the Spread of the Coronavirus Disease 2019 Taking into Account the Undetected Infections. Elsevier Public Health Emergency Collection. 2020(88). pp 1-13. (www.nih.gov/coronavirus).

[15]   Yichi Li, Bowen Wang, Ruiyang Peng, Chen Zhou, Yonglong Zhan, Zhuoxun Liu, Xia Jiang and Bin Zhao. (2020). Mathematical Modeling and Epidemic Prediction of COVID-19 and Its Significance to Epidemic Prevention and Control Measures. Ann Infect Dis Epidemiol. 5 (1): 1052. (Remedy Publications).

[16]   Dennis J. Sweeney. (2020, October 20). Statistics. UK. Science Britannica. (www.britannica.com/science).

[17]   Christ Spatz. (2005). Basic Statistics: Tales of Distribution. (8th edition). USA. Wadsworth Thomson Learning Inc. pp 251-281.

[18]   Douglas C. Montgomery (2001). Design and Analysis of Experiments. (5th edition). New York, USA. John Wiley and Sons Inc. pp 170.

[19]   Nigeria Centre for Disease Control. (2020. October 15). COVID-19 Situation Report. NCDC. (covid-19.ncdc.gov.ng).

[20]   Ghana Health Service. (2020, October 27). COVID-19: Ghana's Outbreak Management Updates. Ghana Health Service. (www.ghanahealthservices.org).

[21]   Wikipedia. (2020, October 26). COVID-19 Pandemic in Senegal. Wikipedia. (www.en.m.wikipedia.org).