CLUSTER ANALYSIS OF THE SPREAD OF COVID-19 IN NIGERIA

Ajibode, I. A. & Adeboye, N. O.
Department of Mathematics/Statistics, Federal Polytechnic, Ilaro, Ogun state, Nigeria
Ilesanmi.ajibode@federalpolyilaro.edu.ng; nureni.adeboye@federalpolyilaro.edu.ng

Abstract

Coronaviruses are viruses that can cause illnesses such as the common cold and severe acute respiratory syndrome. Shortly after declaring the COVID-19 as a pandemic, there was a massive increase in the growth of the virus across the countries including Africa and Nigeria is no exception. However, little is known about pattern of the virus in Nigeria especially the States. Thus, the needs for this research where hierarchical clustering has been employed to classified states according to the impact of COVID-19 cases using number of cases, number on admission, number discharged and number of deaths. The data for the research work was extracted from the record of Nigeria Centre for Disease Control up to 27nd of August, 2020. Clustering analysis was used to classify the states into their respective groups and the result revealed three classes which were classified as extremely affected (cluster 1), highly affected (cluster 2) and moderately affected (cluster 3). The results revealed that Lagos state is in cluster 1, Abuja is in cluster2 while the remaining states are found in cluster 3. The ANOVA tests carried out equally revealed the significance difference in the 3 classifications in relations to all the considered cases. The paper then conclude that there is a need for the government to intensify efforts in sustaining policies that could help flattening the curve of the pandemic

Keywords: ANOVA, COVID-19 Pandemic, Classification, Hierarchical Cluster

Introduction

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered corona virus (WHO, 2020). The victims of COVID-19 virus do experience mild to moderate respiratory illness and recover without requiring special treatment. However, older victims and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness which might eventually lead to death. The transmission of the disease could be checkmated by informing people about the virus on causes and its spreads. The virus spreads through droplets of saliva or discharge from the nose when an infected person coughs or sneezes.

On March 11, 2020, the WHO officially declared the COVID-19 pandemic and since then it has affected the socioeconomic activities of many nations as well as causes financial activities to be crippled. The World Health Organization records on Covid-19 indicated 17,039,160 confirmed cases and 667,084 deaths globally as of 30th July 2020. The outbreak spread from the Chinese city of Wuhan to more than 180 countries and territories, affecting every continent except Antarctica. Several efforts to stop the spread of the virus include lockdown of cities; with no activities on air space of countries.

Africa is not out of the experience of COVID-19 as the continent is also experiencing geometric spread of the virus across countries within weeks. Governments and health authorities have since been on their toes trying to limit widespread of the disease. The index case in Nigeria was announced on 27 February, 2020. The index case was an Italian citizen who works in Nigeria returning from Milan, Italy through the Murtala Muhammed International Airport, Lagos and was transferred to Lagos State biosecurity facilities for isolation and testing. Since the reported case of the virus in Nigeria, the Government through the Ministry of Health has been trying at curtailing the outbreak of the virus in Nigeria since there is no immediate vaccine for the virus.

Since the lockdown was announced, there have been several efforts from Government to provide palliatives so that the citizens will not feel much impact economically, thanks to international donors that partnered with the Federal Government in bringing succour to the less privilege and some state governors also played their parts in making sure that majority of the populace get some food items.

However, despite Government effort to curb the spread of COVID-19, most of the citizen are of the opinion that the virus is not real in Nigeria and that the Government is using the opportunity to get money from the international donors. In addition, there is believe that majority of the patience that tested positive to COVID-19 were having normal malaria because the treatment given to them at the isolation center is nothing but malaria drugs. Darlington and McNeil (2018) illustrate how conspiracy theories about disease can spread despite attempts by governments to correct misinformation. False information circulated widely in the country about the causes of the disease, the reasons for the spread, and the consequence they could have for human health.

Surprisingly, these aforementioned beliefs by majority have given room for citizens to jettison government directives and people still find their ways in and out their place of abode even crossing the state and regional borders. These in and out movements were made possible through the connivance of the corrupt security personal that were expected to uphold law.

The rate of COVID-19 incidence in Nigeria has been associated to failure to checkmate movement of the people and to make citizens abide with the laid down rules and regulations. However, these lackadaisical and carefree attitudes of people encourage the spread of corona virus within the states, across the states and even across regional boundaries.

Susan, Chigozie, Emenike and Zara (2020) in their study examine the impact of Corona Virus on Nigeria Political Economy, the researchers concluded that Nigeria needs to collaborate with the private sector stakeholders to garner resources and procure quality medical, commodity relief packages and create fiscal policy measures that can support rapid deployment of responders and provide socio-economic succor for the most vulnerable population, as it gets the political-economy wheels back, and running after the pandemic. The study underscores the urgency for the government's implementation of appropriate economic stimuli to free the nation's economy from over-dependence on oil.

Olusanya and Ahamuefula (2020) studied the impact of the corona virus (COVID-19) on Nigeria economy. The researchers concluded that Nigeria's macroeconomic fundamentals indicate that Nigeria, just like most developing countries, is at a risk of an impending recession. This may not be unconnected with the government's policy response to the pandemic; by way of imposing movement restrictions in terms of partial or total lockdown, social distancing, providing citizenry with some palliatives, and also trying to contain the spread of the corona virus by quarantining and treating infected individuals; as most economic activities are somewhat halted. Amidst low economic productivity and negatively affected foreign exchange earnings, the country has to commit large funds to provide necessary supplies that are required to combat the spread and attend to already infected persons. They also noted that the existing imbalance between revenue and expenditure will definitely increase, as the former will inevitably continue to dwindle except economic activities resume and return to normalcy. Also, accrued debts prior to and during this pandemic, and consequently, debt servicing will influence the country's economic recovery. Hence, given the inevitable slide into recession, necessary actions need to be taken to facilitate recovery. Provision of infrastructures that will encourage diversification of the economy from oil dependence cannot be overemphasized and should be pursued vigorously, with focus on the major contributing sectors to moderate variability in the real GDP growth rate.

Otitoloju, et al (2020) examined the faces of spatial differences in the morbidity and mortality in sub-Saharan Africa, Europe and USA. The researchers concluded that the wide variation in the outcome of the COVID-19 disease burden in the selected countries are attributable largely to climatic condition (temperature) and differential healthcare approaches to management of the disease.

The need to examine the similarity of the spread of COVID-19 pandemic in Nigeria states call for the use of cluster analysis as there was no extant literature on this. Cluster analysis can also be described as data reduction technique which is unique because its goal is to reduce the number of cases or observations by classifying them into

homogeneous clusters, identifying groups without previously knowing group membership or the number of possible groups.

In this paper, hierarchical clustering is employed. Hierarchical cluster analysis can either be agglomerative or divisive. Agglomerative hierarchical clustering separates each case into its own individual cluster in the first step so that the initial number of clusters equals the total number of cases (Norusis, 2010). At successive steps, similar cases–or clusters–are merged together (as described above) until every case is grouped into one single cluster. Divisive hierarchical clustering works in the reverse manner with every case starting in one large cluster and gradually being separated into groups of clusters until each case is in an individual cluster. This latter technique is not popular because of its heavy computational load. The focus of the present paper is on the method of hierarchical agglomerative cluster analysis and this method is defined by two choices: the measurement of distance between cases and the type of linkage between clusters (Bratchell, 1989).

In Nigeria, the impact of COVID-19 spread has not been demonstrated in a more advance quantitative terms. This in essence might convince policy makers to devote the needed attention and more resources to combating this dreadful virus.

Hence, the need to uncover similarities in quantitative effect of COVID-19 data among the states, for meaningful grouping.

Methodology

The data for this study was obtained mainly from records kept by Nigeria Centre for Disease Control up to 27nd of August, 2020. It comprises of cumulative number of reported cases of Corona virus victims, Number on admission, Number discharged and number of deaths.

The researcher employed Hierarchical cluster analysis which is always used for similarity grouping of items. In addition, One Way Analysis of Variance way used to examine whether there exists significant difference in the groups identified.

Hierarchical Cluster Analysis

The hierarchical technique adopted for this research is agglomerative cluster. In order to build clusters, we need to define the distance between two objects and eventually between clusters. The units of measure of the p variables are quite different; the variables were normalized by forming z-scores of the variables as in subtracting the sample means from the original variables and dividing the deviations by their respective sample standard deviations. The most often used measure of distance (dissimilarity) between the two cases is the Euclidean distance defined by:

$$d_{IJ} = \sqrt{(A_{i1} - N_{j1})^2 + (A_{i2} - N_{j2})^2 + \cdots + (A_{ip} - N_{jp})^2} \quad (1)$$

This algorithm allows for the distance between two cases to be calculated across all variables and reflected in a single distance value. At each step in the procedure, the squared Euclidean distance between all pairs of cases and clusters is calculated and shown in a proximity matrix. At each step, the pair of cases or clusters with the smallest squared Euclidean distance will be joined with one another. This makes hierarchical clustering a lengthy process because after each step, the full proximity matrix must once again be recalculated to take into account the recently joined cluster. The squared Euclidean distance calculation is straightforward when there is only one case per cluster. However, an additional decision must be made as to how best to calculate the squared Euclidean distance when there is more than one case per cluster. This is referred to as the linkage measure and there is a need to know how to best calculate the link between two clusters.

Linkage Measure

The problem that arises when a cluster contains more than one case is that the squared Euclidean distance can only be calculated between a pair of scores at a time and cannot take into account three or more scores simultaneously. In

line with the proximity matrix, the goal is still to calculate the difference in scores between pairs of clusters, however in this case the clusters do not contain one single value per variable. This suggests that one must find the best way to calculate an accurate distance measure between pairs of clusters for each variable when one or both of the clusters contains more than one case. Once again, the goal is to find the two clusters that are nearest to each other in order to merge them together. There exist many different linkage measures that define the distance between pairs of clusters in their own way. Some measures define the distance between two clusters based on the smallest or largest distance that can be found between pairs of cases (single and complete linkage, respectively) in which each case is from a different cluster (Mazzocchi, 2008). In average linkage, the distance between two clusters is defined to be the average of the distances between all pairs of objects, where each pair is made up on one object from each cluster. If cluster P is the set of objects $P_1, P_2, \ldots . P_m$ and cluster S is $S_1, S_2, \ldots . S_m$, the Average Linkage distance between clusters P and S is:

$$D(P, S) = \frac{T_{PS}}{N_P . N_S} \quad (2)$$

Where T_{PS} is the sum of all pairwise distances between cluster P and Cluster S. N_p *and* N_S are the sizes of the clusters P and S respectively. At each stage of hierarchical clustering, the clusters P and S are merged such that, after merger, the average pairwise distance within the newly formed cluster, is minimum.

Average linkage was utilized and it referred to as the Unweighted Pair-Group Method. The method is most suitable in order to overcome the limitations of single and complete linkage methods. This method is supposed to represent a natural compromise between the linkage measures to provide a more accurate evaluation of the distance between clusters. For average linkage, the distances between each case in the first cluster and every case in the second cluster are calculated and then averaged. The method is accurate reflection of the distance between two clusters of cases. Each linkage measure defines the distance between two clusters in a unique way. The next section of the paper revealed the result of hierarchical cluster analysis performed using SPSS version 23.

The comparison of mean for the clusters was obtained using:

$$SS_{Cluster} = \frac{\sum_{i=1}^{n_i} X_i^2}{n_i} - \frac{X^2}{N} \quad (3)$$

$$TSS = \sum_i \sum_j X_{ij}^2 - \frac{X^2}{N} \quad (4)$$

$$ESS = TSS - SS_{Cluster} \quad (5)$$

$$MSS_{Cluster} = \frac{SS_{cluster}}{k-1} \quad (6)$$

$$MESS = \frac{ESS}{N-k} \quad (7)$$

$$F_{cal} = \frac{MSS_{Cluster}}{MESS} \quad (8)$$

When $F_{critical} > F_{cal}$, the null hypothesis of no difference within the clusters is accepted.

n_i is the sample size per cluster

k is the number of clusters

N is the total number of samples in all the clusters

X_{ij} the jth response sampled from the ith cluster

Results and Discussion

The agglomeration schedule is a numerical summary of the cluster solution presented in Table 1.

Table 1: Agglomeration Schedule for COVID-19 Pandemic

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	31	36	.000	0	0	2
2	31	34	.000	1	0	8
3	33	35	.000	0	0	8
4	29	30	.450	0	0	7
5	22	27	.453	0	0	9
6	20	24	.455	0	0	28
7	28	29	.456	0	4	12
8	31	33	.457	2	3	13
9	22	26	.458	5	0	15
10	18	21	.459	0	0	20
11	15	17	.460	0	0	18
12	23	28	.461	0	7	24
13	31	37	.463	8	0	16
14	9	10	.466	0	0	27
15	22	25	.467	9	0	19
16	31	32	.469	13	0	19
17	6	7	.470	0	0	23
18	14	15	.472	0	11	22
19	22	31	.473	15	16	25
20	18	19	.475	10	0	24

21	12	13	.477	0	0	22
22	12	14	.479	21	18	26
23	5	6	.480	0	17	32
24	18	23	.485	20	12	25
25	18	22	.490	24	19	28
26	12	16	.510	22	0	27
27	9	12	.523	14	26	30
28	18	20	.524	25	6	30
29	3	11	.530	0	0	33
30	9	18	.540	27	28	31
31	8	9	.785	0	30	32
32	5	8	1.720	23	31	33
33	3	5	1.817	29	32	34
34	3	4	2.376	33	0	35
35	2	3	2.390	0	34	36
36	1	2	104.521	0	35	0

At the first stage, cases 31 and 36 are combined because they have the smallest distance. The cluster created by their joining next appears in stage 2. In stage 8, the clusters created in stages 1 and 2 are joined. The resulting cluster next appears in stage 8. However, because the agglomeration is long, we use coefficients column for large gaps rather than scan the dendrogram. A good cluster solution sees a sudden jump (gap) in the distance coefficient. The solution before the gap indicates the good solution. The largest gaps in the coefficient's column occur between stages 3 and 4, indicating a 3-cluster solution, and stages 35 and 36, indicating a 3-cluster solution.

Examining the Vertical Icicle in Figure 1, it is noticed that case 1 (Lagos State) and case 2 (FCT) form a cluster and other states in Nigeria (thirty-five states) forms the third cluster. This result is same with the result in the dendrogram in figure 2.

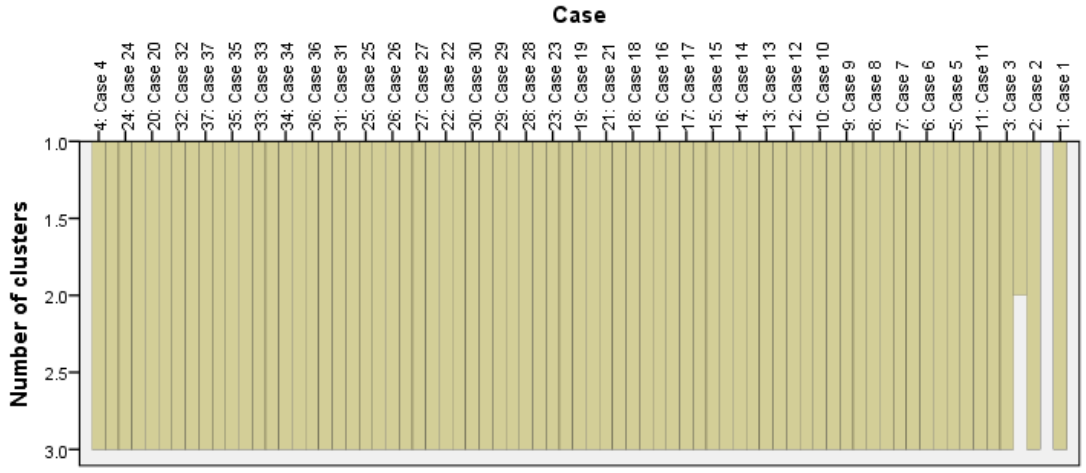

Figure 1: Icicle for COVID 19 Pandemic

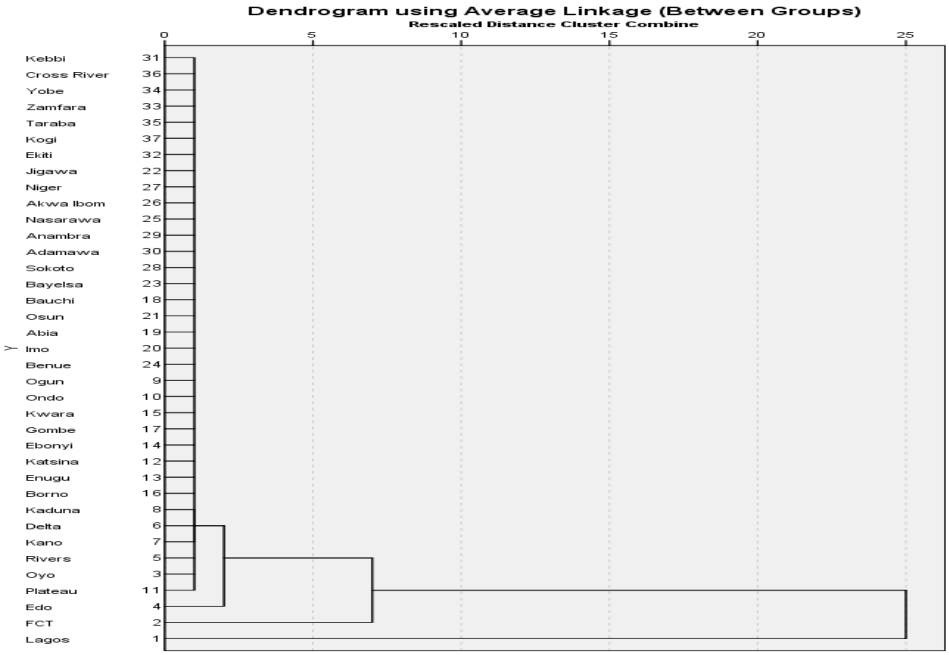

Figure 2: Dendrogram using Average Linkage for COVID 19 Pandemic

The states were classified as extremely affected (cluster 1), Highly affected (cluster 2) and moderately affected (cluster 3). The descriptive statistics from the three cluster solutions with respect to number of reported COVID-19 cases, number on admission, number discharged and number of deaths is as presented in table 3 showing the average, minimum and maximum for each case.

Table 3: Descriptive analysis for COVID-19 data

		N	Mean	Minimum	Maximum
Number of Cases	Extremely Affected	1	18056.00	18056	18056
	Highly Affected	1	5094.00	5094	5094
	Moderately Affected	35	861.91	5	3091
	Total	37	1441.00	5	18056
Number on Admission	Extremely Affected	1	2627.00	2627	2627
	Highly Affected	1	3572.00	3572	3572
	Moderately Affected	35	153.74	0	1136
	Total	37	312.97	0	3572
Number Discharged	Extremely Affected	1	15227.00	15227	15227
	Highly Affected	1	1472.00	1472	1472
	Moderately Affected	35	686.49	3	2269
	Total	37	1100.70	3	15227
Number of Deaths	Extremely Affected	1	202.00	202	202
	Highly Affected	1	50.00	50	50
	Moderately Affected	35	21.77	2	100
	Total	37	27.41	2	202

Furthermore, one-way analysis of variance was utilized to affirm whether there is significant difference in the three classification with respect to number of cases, number on admission, number of discharged and number of deaths. The result (table 4) shows that there is significant difference in the three classifications for all the considered cases, that is, number of infected cases, number on admission, number discharged, and number of deaths.

Table 4: One Way ANOVA table for COVID-19 Grouping

		Sum of Squares	df	Mean Square	F	Sig.
Number of Cases	Between Groups	301139543.257	2	150569771.629	221.847	.000
	Within Groups	23076136.743	34	678709.904		
	Total	324215680.000	36			
Number on Admission	Between Groups	16863376.287	2	8431688.144	130.927	.000
	Within Groups	2189598.686	34	64399.961		
	Total	19052974.973	36			
Number Discharged	Between Groups	205695286.987	2	102847643.493	235.225	.000
	Within Groups	14865846.743	34	437230.787		
	Total	220561133.730	36			
Number of Deaths	Between Groups	32104.747	2	16052.374	42.792	.000
	Within Groups	12754.171	34	375.123		
	Total	44858.919	36			

Conclusions

The findings of this research give credence to the high significant impacts of cluster analysis in COVID-19 pandemic classification of States in Nigeria using number of cases, number on admission, number discharged and number of deaths across the 36 states and federal capital territory.

The results also confirmed the opinions of both the government and majority of the citizenry that Lagos is the epic center followed by the federal capital territory, Abuja. Hence, the need for the government to intensify efforts in sustaining policies that could help flattening the curve of the pandemic.

References

Bratchell, N. (1989). Cluster analysis. *Chemometrics and Intelligent Laboratory Systems, 6*, 105–125.

Darlington, S. & McNeil, D. G. (2018). *Yellow fever circles Brazil's huge cities*. New York Times, https://www.nytimes.com/2018/03/05/health/brazil-yellow-fever.html.

Mazzocchi, M. (2008). *Statistics for Marketing and Consumer Research*. London, UK: Sage Publications Ltd.

Norusis, M. J. (2010). *Chapter 16: Cluster analysis. PASW Statistics 18 Statistical Procedures Companion, 361-391*. Upper Saddle River, NJ: Prentice Hall.

Susan, D.A., Chigozie, E., Emenike, J.O. & Zara, I.B. (2020). The Impact of Corona Virus (COVID-19) on the Nigeria Political Economy: Government Support and Economic Relief Packages. *International Journal of Management Studies and Social Science Research*, *2*(2), 23-35.

Olusanya, E. O. & Ahamuefula, E. O. (2020). *COVID-19 and the Nigeria Economy: Analyses of Impacts and Growth Projections*. Retrieved from https://www.researchgate.net/publication/342439011. Accessed July 27, 2020.

Otitolojua A. A., Okaforb, I. P., Fasonac, Ifeoma P., Bawa-Allaha, K. A., Isanbord, C., Chukwudoziee, O.S., Folarina, O.S., Adubif, T.O., Sogbanmua, T.O. & Ogbeibu, A.E. (2020). *COVID-19 pandemic: examining the faces of spatial differences in the morbidity and mortality in sub-Saharan Africa, Europe and USA.* Retrieved from https://doi.org/10.1101/2020.04.20.20072322

WHO (2020). *Africa report COVID-19* ***case.*** https://www.afro.who.int/health-topics/coronavirus-covid-19. Accessed July 28, 2020.