Research Article

PEARSON'S CORRELATION COEFFICIENT-ASSISTED ESTIMATION OF THE CASE-FATALITY RATE OF COVID-19 FOR 16 COUNTRIES

[1, *]Todo, Charles Oghenekparobor, [2]Igabari, John N. and [3]Ojarikre, Harrietta I.

[1]Department of Statistics, Delta State Polytechnic, Otefe-Oghara, Delta State, Nigeria
[2, 3]Department of Mathematics, Delta State University, Abraka, Delta State, Nigeria

Abstract

Increased cases of corona virus infections have been reported in many countries of the world. This work considers the relationship between the number of deaths from COVID-19 and the numbers of infection cases. Using the Pearson's correlation coefficient, we found a very strong positive relationship between the numbers of deaths and detected cases of COVID-19. The value of r we got as at two different dates were 0.9217 and 0.9188. We further got an estimate of the case-fatality rate of COVID-19 by ascertaining the percentage of detected cases that die, giving us a case-fatality rate of 0.0248, which can be used to forecast deaths resulting from COVID-19.

Keywords: Correlation coefficient, COVID-19, Case-fatality rate, Estimate, forecast.

INTRODUCTION

Many researchers have undertaken studies to ascertain the relationships that exist between COVID-19 and a number of known phenomena. A number of other researchers have also focused on providing avenues for more empirical and dependable understanding of COVID-19. According to Wu *et al.* (2020), epidemiological studies have explored the relationship between COVID-19 and meteorological factors. Onder *et al.* (2020) compared COVID-19 cases and mortality rates between different countries. In a work, Russell et al. (2020) stated that fatality risk of COVID-19 may also have been influenced by differences in health care between countries. Based on results of their work in early 2020, Jung *et al.* (2020) were able to argue that COVID-19 which was only then described as an epidemic had a substantial potential for causing a pandemic even before the WHO declared it to be pandemic. Covid *et al.* (2020) observed COVID-19 utilization data from select locations and used them to develop a statistical model forecasting deaths and hospital utilization against capacity by states for the US for 4 months. Rubino *et al.* (2020) have said that research studies on a global scale will be necessary to answer many key questions on COVID-19. The current study is an effort to contribute to this global necessity. Sedgwick (2012) once used the Pearson's correlation coefficient in a medical-related study.

METHODOLOGY

We used line charts and Pearson's correlation coefficient in our investigation. The tables below present the number of detected cases of COVID-19 in 16 countries along with the numbers of those treated, recovered and dead.

Table 1. Number of detected cases of COVID-19 along with numbers of those treated, recovered and dead

	As at July 1st, 2021	Detected (x)	Treated	Recovered	Dead (y)
1	USA	34,174,774	16,916,185	16,306,656	609,529
2	UK	5,519,602	129,167	344	128,823
3	INDIA	31,216,337	29,845,810	29,427,330	418,480
4	BRAZIL	19,419,437	17,323,316	16,779,136	544,180
5	RUSSIA	6,030,240	5,151,098	5,000,393	150,705
6	FRANCE	5,890,062	495,322	383,768	111,554
7	ITALY	4,293,083	4,209,786	4,081,902	127,884
8	CANADA	1,423,889	1,407,520	1,381,016	26,504
9	SPAIN	4,189,136	231,524	150,376	81,148
10	GERMANY	3,748,613	3,457,848	3,366,432	91,416
11	MEXICO	2,678,297	2,234,191	1,997,381	236,810
12	CHINA	119,837	91,325	86,689	4,636
13	JAPAN	847,614	782,370	767,275	15,095
14	SOUTH KOREA	182,265	149,753	147,693	2,060
15	AUSTRALIA	32,129	26,401	25,486	915
16	NIGERIA	169,884	166,112	163,984	2,128

Numbers of detected cases of COVID-19, recovered and dead were compiled from www.tradingeconomics.com

*****Corresponding Author:** *Todo, Charles Oghenekparobor;* Department of Statistics, Delta State Polytechnic, Otefe-Oghara, Delta State, Nigeria.

Table 2. Number of detected cases of COVID-19 along with numbers of those treated, recovered and dead as at August 21ˢᵗ, 2021

	As at August 21ˢᵗ,2021	Detected (x)	Treated	Recovered	Dead (y)
1	USA	38,074,886	16,937,472	16,306,656	630,816
2	UK	6,666,399	132,347	344	132,003
3	INDIA	32,737,939	29,865,540	29427330	438,210
4	BRAZIL	20,645,537	17,355,781	16,779,136	576,645
5	RUSSIA	6,901,152	5,182,822	5,000,393	182,429
6	FRANCE	6,673,336	497,649	383,768	113,881
7	ITALY	4,502,396	4,210,816	4,081,902	128,914
8	CANADA	1,486,436	1,407,906	1,381,016	26,890
9	SPAIN	4,815,205	234,066	150,376	83,690
10	GERMANY	3,835,375	3,458,353	3,366,432	91,921
11	MEXICO	3,249,878	2,251,847	1,997,381	254,466
12	CHINA	122,995	91,325	86,689	4,636
13	JAPAN	1,454,364	783,292	767,275	16,017
14	SOUTH KOREA	250,051	149,977	147,693	2,284
15	AUSTRALIA	51,256	26,485	25,486	999
16	NIGERIA	191,345	166,438	163,984	2,454

Numbers of detected cases of COVID-19, recovered and dead were compiled from www.tradingeconomics.com

Table 3. Number of persons treated, recovered, dead and being treated presented as a percentage of number of detected cases as at July 1st, 2021

Treated	Recovered	Dead	Being Treated
as a % of Detected	as a % of Detected	as a % of Detected	as a % of Detected
49.5	47.72	1.78	50.5
2.34	0.01	2.33	97.66
95.61	94.27	1.34	4.39
89.21	86.4	2.80	10.79
85.42	82.92	2.5	14.58
8.41	6.52	1.89	91.59
98.06	95.08	2.98	1.94
98.85	96.99	1.86	1.15
5.53	3.59	1.94	94.47
92.24	89.8	2.44	7.76
83.42	74.58	8.84	16.58
76.21	72.34	3.87	23.79
92.3	90.52	1.78	7.7
82.16	81.03	1.13	17.84
82.17	79.32	2.85	17.83
97.78	96.53	1.25	2.22

Table 4. As at August 21st, 2021

Treated	Recovered	Dead	Being Treated
as a % of Detected	as a % of Detected	as a % of Detected	as a % of Detected
44.48	42.83	1.66	55.52
1.99	0.01	1.98	98.01
91.23	89.89	1.34	8.77
84.07	81.27	2.79	15.93
75.1	72.46	2.64	24.9
7.46	5.75	1.71	92.54
93.52	90.66	2.86	6.48
94.72	92.91	1.81	5.28
4.86	3.12	1.74	95.14
90.17	87.77	2.4	9.83
69.29	61.46	7.83	30.71
74.25	70.48	3.77	25.75
53.86	52.76	1.1	46.14
59.98	59.07	0.91	40.02
51.67	49.72	1.95	48.33
86.98	85.7	1.28	13.02

Graphs of Numbers of Persons Treated, Recovered, Dead and Being Treated Presented as a Percentage of Detected
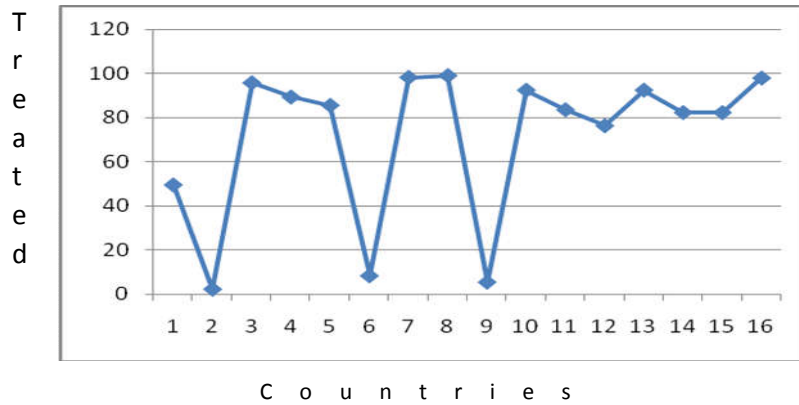
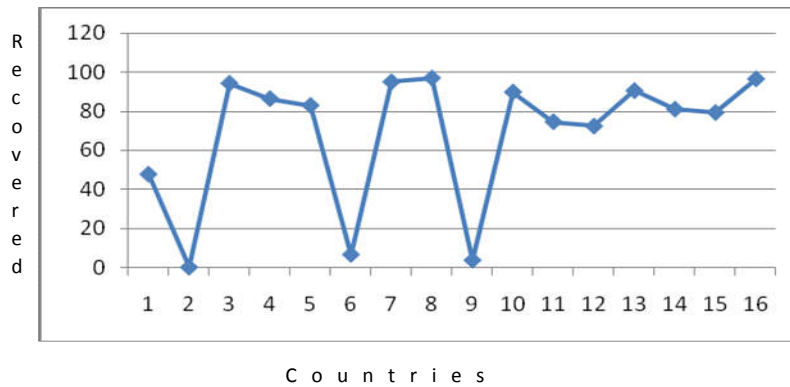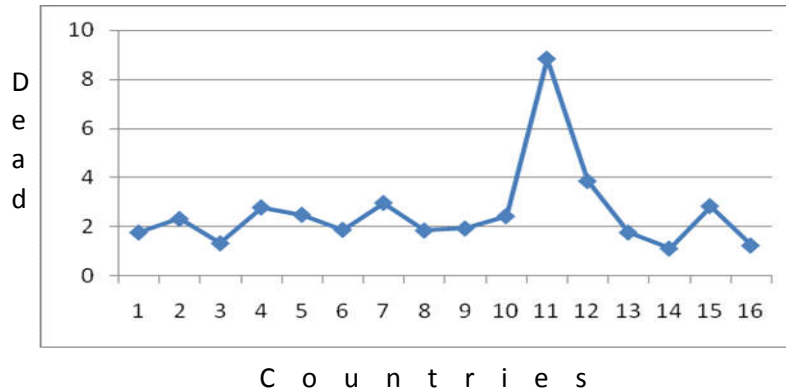### Chart 1. Treated as at 1ˢᵗ July, 2021

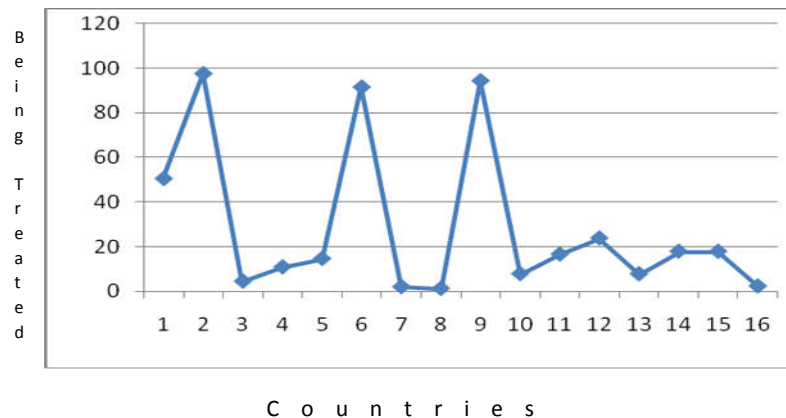### Chart 2. Recovered as at 1st July, 2021

There is no defined range for Charts 1 and 2.

Chart 3. Dead as at 1ˢᵗ July, 2021

There is a clear range (only slightly deviated from by Mexico)

Chart 4. Being Treated as at 1ˢᵗ July, 2021

There is no defined range in Chart 4.

Charts 1, 2, 3, and 4 above show that only the number of deaths taken as a percentage of the number of detected cases of COVID-19 has a defined range (see Chart 3). Out of the 16 countries, only the figures for Mexico went out of the range. The range is 1.10 to 3.87. But for Mexico it is between 7.33 and 8.84. The numbers of the treated, the recovered, and those still being treated, as a percentage of the number of detected show no particular pattern or defined range (as shown in the charts).

The above shows that there is a relationship between the number of COVID-19 deaths and the number of detected cases of COVID-19. The charts for figures as at 21st August, 2021 also show same pattern.

Pearson's Correlation Coefficient: Pearson Product Moment Correlation (PPMC)

Our above observations have shown that there is a relationship between the number of people detected to have COVID-19 and the number of people that die of the disease. Correlation is a statistical measure of how closely two variables are related. There could either be a positive (as one goes up, the other goes up) or a negative (as one goes up, the other goes down) correlation. The degree of correlation can be weak or strong (Emerson, 2015). Pearson's correlation is one of the most common measures of linear dependence (Ly et al., 2018). Correlation coefficient formulas are used to find how strong a relationship is between data. The formulas return a value between -1 and 1, where:

- 1 indicates a strong positive relationship.
- -1 indicates a strong negative relationship.
- A result of zero indicates no relationship at all.

One of the most commonly used formulas is Pearson's correlation coefficient formula. Below is the Pearson's correlation coefficient formula:

$$r = \frac{n(\sum_{i=1}^{n} xy) - (\sum_{i=1}^{n} x)(\sum_{i=1}^{n} y)}{\sqrt{\left[n\sum_{i=1}^{n} x^2 - (\sum_{i=1}^{n} x)^2 \right]\left[n\sum_{i=1}^{n} y^2 - (\sum_{i=1}^{n} y)^2 \right]}} \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \textbf{(*)}$$

We re-present the number of detected (cases) and the number of deaths.

Table 5

		1st July, 2021		21st August, 2021	
		Cases (x)	Deaths (y)	Cases (x)	Deaths (y)
1	US	34,174,774	609,529	38,074,886	630,816
2	UK	5,519,602	128,823	6,666,399	132,003
3	CHINA	119,837	4,636	122,995	4,636
4	JAPAN	847,614	15095	1,454,364	16,017
5	GERMANY	3,748,613	91,416	3,835,375	91,921
6	FRANCE	5,890,062	111,554	6,673,336	113,881
7	INDIA	31,216,337	418,480	32,737,939	438,210
8	ITALY	4,293,083	127,884	4,502,396	128,914
9	BRAZIL	19,419,437	544,180	20,645,537	576,645
10	CANADA	1,423,889	26,504	1,486,436	26,890
11	SOUTH KOREA	182,265	2,060	250,051	2,284
12	RUSSIA	6,030,240	150,705	6,901,152	182,429
13	SPAIN	4,189,136	81,148	4,815,205	83,690
14	AUSTRALIA	32,129	915	51,256	999
15	MEXICO	2,678,297	236,810	3,249,878	254,466
16	NIGERIA	169,884	2,128	191,345	2,454

FINDING AND DISCUSSION

Applying equation (*) to the figures as at July 1st, 2021 yielded a Pearson's correlation coefficient, r, of **0.9217**. Applying it to the figures as at August 21st, 2021 yielded r = **0.9188**. Both coefficients are very close to 1. This shows very strong positive relationship between the number of cases of corona virus detected and the number of deaths among those detected.

Having established that there is a very strong relationship between deaths and number of cases, we sought to find this relationship.

Table 6. Number Deaths as a Percentage of Number of Cases Detected

		July 1st, 2021	August 21st, 2021
1	USA	1.78	1.66
2	UK	2.33	1.98
3	INDIA	1.34	1.34
4	BRAZIL	2.80	2.79
5	RUSSIA	2.50	2.64
6	FRANCE	1.89	1.71
7	ITALY	2.98	2.86
8	CANADA	1.86	1.81
9	SPAIN	1.94	1.74
10	GERMANY	2.44	2.40
11	MEXICO	8.84	7.83
12	CHINA	3.87	3.77
13	JAPAN	1.78	1.10
14	SOUTH KOREA	1.13	0.91
15	AUSTRALIA	2.85	1.95
16	NIGERIA	1.25	1.28
AVERAGE		2.60	2.36

$(2.60 + 2.36)/2 = 2.48$

DISCUSSION

Apart from the figures for Mexico, the percentage of infected people dying out of detected cases fell within a clearly defined range. The average was 2.48%. This then gives a case-fatality rate of 0.0248 for COVID-19 (among 16 countries studied). This Pearson's correlation coefficient-assisted case-fatality rate can be used to estimate the number of people that would die out of the number of people detected to have the virus at any particular time.

Conclusion

The study has shown that the number of COVID-19 infected persons that die as a result of the disease and the number of detected cases are dependent. We have got a Pearson's correlation coefficient-assisted case-fatality rate of COVID-19 for the 16 countries of 0.0248.

References

Covid, I. and Murray, C. J. 2020. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. *MedRxiv*.

Emerson, R. W. 2015. Causation and Pearson's correlation coefficient. *Journal of visual impairment & blindness*, *109*(3), 242-244.

Jung, S. M., Akhmetzhanov, A. R., Hayashi, K., Linton, N. M., Yang, Y., Yuan, B. and Nishiura, H. 2020. Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: inference using exported cases. *Journal of clinical medicine*, *9*(2), 523.

Ly, A., Marsman, M. and Wagenmakers, E. J. 2018. Analytic posteriors for Pearson's correlation coefficient. *Statistica Neerlandica*, *72*(1), 4-13.

Onder, G., Rezza, G. and Brusaferro, S. 2020. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *Jama*, *323*(18), 1775-1776.

Rubino, S., Kelvin, N., Bermejo-Martin, J. F. and Kelvin, D. 2020. As COVID-19 cases, deaths and fatality rates surge in Italy, underlying causes require investigation. *The Journal of Infection in Developing Countries*, *14*(03), 265-267.

Russell, T. W., Hellewell, J., Jarvis, C. I., Van Zandvoort, K., Abbott, S., Ratnayake, R. and CMMID COVID-19 working group. 2020. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. *Eurosurveillance*, *25*(12), 2000256.

Sedgwick, P. 2012. Pearson's correlation coefficient. *Bmj*, *345*.

Wu, Y., Jing, W., Liu, J., Ma, Q., Yuan, J., Wang, Y. and Liu, M. 2020. Effects of temperature and humidity on the daily new cases and new deaths of COVID-19 in 166 countries. *Science of the Total Environment*, *729*, 139051.

****