OPEN ACCESS
Freely Available Online

INTERNATIONAL JOURNAL OF CORONAVIRUSES
ISSN NO: 2692-1537

Review

Doi: 10.14302/issn.2692-1537.ijcv-20-3376

How Valid are the Reported Cases of People Infected with Covid-19 in the World?

Raul Isea[1,*]

[1]Fundacion IDEA, Hoyo de la Puerta, Baruta, Venezuela

Abstract

The goal of this paper is to analyze the registered cases of people who have been infected with Covid-19 registered from throughout the world, using a digital forensic analysis technique that is based on Benford's Law. Twenty-three countries were randomly chosen for this analysis: China, India, Germany, Brazil, Venezuela, Netherlands, Italy, Colombia, Russia, Norway, South Africa, Portugal, Singapore, United Kingdom, Chile, Ecuador, Egypt, Denmark, Ireland, France, Belgium, Australia and Croatia.. We calculate on the p-values based on Pearson χ^2 and Mantissa Arc Test according to the results obtained with the first digit. If any country fails these two tests, a third proof will be carried out based on the Freedman-Watson test. The results indicated that results from Italy, Portugal, Netherlands, United Kingdom, Denmark, Belgium and Chile are suspicions of data manipulation because the numbers fail the Benford's Law according to the results obtained until April 30, 2020. However, it is necessary to carry out further studies in these countries in order to ensure that they countries manipulate or altered the information.

Corresponding author: Raul Isea, Fundación IDEA, Hoyo de la Puerta, Baruta, Venezuela. Email: raul.isea@gmail.com

Keywords: Coronavirus, 2019-nCoV,Covid-19, Benford, Lay, Forensic, data

Received: May 11, 2020 **Accepted:** May 26, 2020 **Published:** May 28, 2020

Editor: Sasho Stoleski, Institute of Occupational Health of R. Macedonia, WHO CC and Ga2len CC, Macedonia

Introduction

In December 2019, the first cases of a new coronavirus (2019-nCoV) responsible for atypical pneumonia began to be registered in Wuhan (China). As of April 30, there are more than three million people infected individuals and there have been almost 230,000 deaths in 180 countries throughout the world. For that reason, On March 11, the disease was declared a pandemic by the World Health Organization.

There is currently no vaccine against this disease, and social distancing measures have been the main recommendation of the World Health Organization to prevent the spread of this disease. Recently, a study (written in Spanish) based on differential equations that simulate the transmission dynamics of the disease was presented from the reported cases of infection in four different countries, according to data recorded at Johns Hopkins University [1]. This paper concludes/indicates that the success of the model will depend on the quality of the data.

For this reason, it is necessary to validate the data obtained from the infected cases of Covid-19, and thus, we can indicate that the data have not been altered or manipulated or even poorly transcribed for unknown reasons. Remember that the Benford's Law has been used in various scenarios to detect, for example, fraud in campaign finances [2], Governmental Economics data [3], in account data [4], fraud in scientific data [5], among others [6,7].

In the scientific literature, we only found one paper published in a repository (arXiv) where the author studied the first contagion outbreaks occurred in China until February 13, 2020 using Benford's Law [8]. This manuscript concluded that until this date, there was no evidence of alteration or manipulation of the cases registered in China.

For this reason, we carry out a more complete study to determine if it is possible to validate the data of people infected by covid-19 using Benford's Law based on Pearson χ2 and the Mantissa Arc Test, and eventually, the Freedman-Watson test to verify that the data has not been manipulated.

Computational Methodology

The data of infected cases were obtained in the database John Hopkins University (available at coronavirus.jhu.edu), from December 31, 2019 to April 30, 2020. The next step was to determine the frequency of appearance of the first digit according to Benford's Law. In order to do that, we employed an algorithm in R employed the library: Benford.analysis according to the following equation:

$$Prob(d_i) = \log_{10}(1 + \frac{1}{d_i})$$

where i corresponds to the values that go from 1 to 9 [see details in 9]. With this distribution, we calculate the Pearson value X^2, which means the goodness of fit statistics according to this equation:

$$\chi^2 = \sum_{k=1}^{9} \frac{[P(k) - b(k)]^2}{b(k)}$$

where *P(k)* and *b(k)* are the proportions obtained from the data and the Benford's Law, respectively. The p-value is simply the probability obtained according to random values as explained in [9], where the p-value should be greater than 0,05 which implied that the numbers have not been altered or manipulated. In addition, the Pearson value χ^2 should tend to zero.

In the Mantissa Arc Test, itwas necessary to calculate a center of mass of the set of values obtained from the mantissa values when considering that the data is distributed in a unit circle, where the center of the circle is given by:

$$x - coordinate = \frac{\sum_{i=1}^{N} \cos(2\pi \cdot (\log_{10}\log(x_i) mod 1))}{N}$$

$$y - coordinate == \frac{\sum_{i=1}^{N} \sin(2\pi \cdot (\log_{10}\log(x_i) mod 1))}{N}$$

where $x_1, x_2, \ldots, x_N$ are the data values.

The next step is to determine the length of the mean values L^2, which is given as

$$L^2 = (x - coordinate)^2 + (y - coordinate)^2$$

And finally, the p-value is simply.

$$p - value = 1 - e^{L^2 \cdot N}$$

Finally, to verify if any country really fails Benford's Law, we will verify with a third test called the Freedman-Watson [10], which is based on the following equation:

$$\frac{N}{9\cdot 10^{i-1}}\left[\sum_{i=10^{k-1}}^{10^{k}-2}\left[\sum_{j=1}^{i}(f_i^0-f_i^e)\right]^2-\frac{N}{9\cdot 10^{k-1}}\left[\sum_{i=10^{k-1}}^{10^{k}-2}\sum_{j=1}^{i}(f_i^0-f_i^e)\right]^2\right]$$

but this equation is complicated to explain and see details in [10].

And remember that the p-value should be greater than 0,05 that indicates that the data has not been altered or manipulated.

Finally, the calculations were carried out for twenty-three countries: from 29 December, 2019 until April 30, 2020: China, India, Germany, Brazil, Venezuela, Netherlands, Italy, Colombia, Russia, Norway, South Africa, Portugal, Singapore, United Kingdom, Chile, Ecuador, Egypt, Denmark, Ireland, France, Belgium, Australia and Croatia, and the results are explained in the next section.

Results

In Table 1, we summarize the results that have been obtained with the two tests according to the data obtained up to April 30, 2020. The results were grouped random into three blocks, where the number of degree of freedom in the Pearson χ^2 and Mantissa Arc Test were 8 and 2, respectively. In addition, we indicate the number of data points by each country (the results were verified *with other module of R called BenfordTest).*

The countries that pass the two tests which means that the p-value greater than 0,05, are China, Germany, Brazil, Venezuela, Norway, South Africa, Singapore, Ecuador, Egypt, Ireland, France and Australia. This means that the information these countries is valid. In fact, China, Singapore and Australia perfectly are agreed with the Benford's Law. On the other hand, Colombia, India, Russia and Croatia pass at least one of the two tests as shown in Table 1, so these countries no manipulate the data.

However, Italy, Portugal, Netherlands, United Kingdom, Denmark, Belgium and Chile do not pass either of the two tests (their values have been highlighted and in red color in the Table 1). For these countries, we calculate the p-value according to the Freedman-Watson test (employed the Benford.analysis library), and the results obtained were: 10^{-3}, 10^{-16}, 10^{-4}, 10^{-16}, 10^{-10}, 10^{-16}, 10^{-4}, correspondent to Italy, Portugal, Netherlands, United Kingdom, Denmark, Belgium and

Table 1. Results obtained according to Benford's law (see text for more details).

	China	Italy	Brazil	Colombia	Venezuela	India	Russia
X^2	3,450	33,383	6,785	16,974	8,557	12,560	22,709
S. size	109	71	58	52	34	62	54
p-value (X^2)	0,903	10^{-5}	0,560	0,030	0,381	0,128	0,004
p-value (Mantissa)	0,522	10^{-6}	0,354	0,061	0,868	0,002	0,118

	Germany	Norway	S. Africa	Portugal	Singapore	Netherlands	UK	Chile
X^2	12,425	7,952	6,619	16,623	4,373	22,725	55,074	26,363
S. size	75	63	54	60	91	64	70	58
p-value (X^2)	0,133	0,438	0,578	**0,034**	0,822	**0,003**	10^{-6}	10^{-4}
p-value (Man)	0,386	0,331	0,372	**0,004**	0,935	10^{-8}	10^{-6}	**0,001**

	Ecuador	Egypt	Denmark	Ireland	France	Belgium	Australia	Croatia
X^2	9,408	10,194	25,535	9,174	14,025	24,605	5,011	7,868
S. size	55	54	64	59	72	62	77	62
p-value (X^2)	0,309	0,252	**0,001**	0,328	0,081	**0,002**	0,756	0,447
p-value (Man)	0,557	0,142	10^{-4}	0,167	0,139	**0,003**	0,445	0,001

Chile, respectively. Therefore, three tests different indicated that these countries may have somewhat or altered the data, because it is not possible to verify their accuracy with these three different tests.

However, it is necessary to wait until the end of the pandemic to be able to analyze all the data and to ensure that these countries have been able to manipulate the data, or perhaps there are failures due to the omission of registered cases.

Conclusions

The results obtained from the analysis based on Benford's Law of infected cases with Covid-19 obtained that China, Germany, Brazil, Venezuela, Norway, South Africa, Singapore, Ecuador, Egypt, Ireland, France, Australia, Colombia, India, Russia, Croatia don't manipulate the information register in the Jonhs Hopking dataset. However, Italy, Portugal, Netherlands, United Kingdom, Denmark, Belgium and Chile do not pass three tests carried out in the paper, and therefore, it is necessary to carry out further studies in these countries in order to ensure that they countries manipulate or altered the information.

In fact, we consider that we must wait until the end of the pandemic until all cases have been registered in all countries, and thus we must ensure the lack of credibility of the data provided in a given country in the world.

Acknowledgment

I'd like to acknowledgment to Karl E. Longreen for your comments in this manuscript.

References

1. Isea, R (2020). La dinámica de transmisión del Covid-19 desde una perspectiva matemática. Revista del Observador del Conocimiento. 5(1): 15-23.
2. Cho W., and Gaines. B. (2007). Breaking the (Benford) Law: statistical fraud detection in campaign finance. *The American Statistician*, 61 (3):218-223.
3. Rauch, B., Gottsche, M., Engel, S. (2011). Fact and Fiction in EU-Governmental Economics data. *German Economics Review*, 12(3): 243-255
4. Durtschi, C., Hillison, W., Pacini, W. (2004). The effective use of Benford's Lawto assist in detection fraud in accounting data. *J. Foresic Accounting*, 5: 17-34.
5. Diekman, A. (2007). Not the first digit! Using Benford's Law to detect fraudulent scientific data. *J. Appl Stat*, 34(3): 321-329.
6. Forman, A.K. (2010) The Newcomb-Benford Law in its relation to some common distributions. *PlOS ONE*, 5:e10541
7. Pietronero, L., Tossati,V., Vespignant, A. (2001). Explaining the uneven distribution of number in nature: the Benford and Zipl. *Physica A*, 293(1-2): 297-304.
8. Zhang, J. (2020). Testimg case number of coronavirus disease 2019 in China with Newcomb-Benford Law. Respository arXiv ID: 2002.05695.
9. Nigrini, J.N. Benford's Law.Applications for Forensic Accounting, Auditing, and Fraud Detection. John Wiley & Sons, Inc.2012. New Jersey.
10. Freedman, L.S. (1981) Watson's Un2 Statistic for a Discrete Distribution. Biometrika. 68: 708–711.