Philippine e-Journal for Applied Research and Development *10(2020), 29-34* ISSN 2449-3694 (Online)
http://pejard.slu.edu.ph/vol.10/2020.12.27.pdf

The Distribution of COVID-19 Cases in the Philippines and the Benford's Law

Aldwin T. Miranda[1,*]

[1]*Southern Philippines Agri-Business and Marine and Aquatic School of Technology, Malita, Davao Occidental*
* *Contact information (aldew2003@gmail.com)*

Received, 6 November 2020; Accepted, 20 December 2020; Published, 27 December 2020

Abstract

As part of the Philippine government's efforts to contain the COVID-19 pandemic, the Department of Health has regularly been reporting the number of positive cases, deaths, and recoveries from all over the country. However, there were doubts and allegations of concealment on the reported COVID-19 data. Hence, a test of fraud was conducted by examining the cumulative distribution of the country's COVID-19 data if this conformed to the principle of Benford's law. Kolmogorov-Smirnov test (KS) was employed to analyse the difference between the distributions. The results show no significant difference between the COVID-19 data's first digit distribution and the distribution set by the Benford's law, thus suggesting no substantial evidence for data manipulation in the reported COVID-19 data of the Philippines.

Keywords: Fraud detection in COVID-19 data, Benford's law applications

Introduction

COVID-19 is a highly infectious respiratory disease caused by a novel coronavirus that was first observed in Wuhan, China in late December 2019. This disease has easily spread around the globe, tallying a number of positive cases in almost every region. Subsequently, the World Health Organization (WHO) declared this a pandemic on March 11, 2020, as it expressed deep concern on the virus' alarming levels of spread and severity as well as the levels of inaction noted [1]. At the time of declaration, there were over 118,000 cases of coronavirus illness in over 110 countries and territories around the world, with a sustained risk of further global spread [2].

In the Philippines, the first COVID-19 case was recorded on January 30, 2020 and the first COVID-19 related death, also the first recorded death outside China, on February 1, 2020 [3]. The coronavirus has spread throughout the country, prompting the government to strictly implement enhanced community quarantines and impose public health measures like the wearing of face masks, proper sanitation, and social distancing. As of April 30, 2020, a total of 8,488 COVID-19 cases has been recorded, with 568 total deaths and 1043 total recoveries, and distributed throughout 55 out of 81 provinces in the country [4]. The Department of Health (DOH) has since been updating the public on the COVID-19 status, as it provided a tracker for total positive cases, deaths, and recoveries in its website (www.doh.gov.ph). However, in spite of the effort to be transparent, there were still doubts and allegations that the government is concealing COVID-19 data [5]. To help clear up this issue, a test of fraud needs to be conducted on the DOH reported data. Towards this, this study proposes to apply Benford's law in order to determine whether there is substantial evidence indicating that the country's COVID-19 data is fraudulent. Specifically, it will calculate the first digit frequency distribution of the total COVID-19 cases, the total number of deaths, and the total number of recoveries in the Philippines, and then verify whether the observed frequency distributions conform to the Benford's expected distribution.

Benford's law is the observation that in many collections of numbers, be they mathematical tables, real-life data, or combinations thereof, the leading significant digits are not uniformly distributed, as might be expected, but are heavily skewed toward the smaller digits [6]. Digits in statistical data produced by natural or social processes are often distributed in this manner [7]. Benford's law is generally used in detecting fraud and error in data sets of naturally occurring numbers. The test is based on the supposition that first, second, third, and other digits in real data follow the Benford distribution while the digits in fabricated data do not [7].

Goodman [8] mentioned the requirements for a data set to be compatible with Benford's law, namely: (a) sufficient sample size, (b) large span of number values, (c) positively skewed distribution of numbers, and (d) not human-assigned numbers. Though it seems difficult to decide on an adequate sample size for Benford's law [9], it has been shown that this holds true for data sets containing as few as 50 to 100 numbers [10]. Some authors even illustrated the law using fewer samples (n < 50) [8] and with COVID-19 data [11]. Moreover, according to Koch & Okamura [12], the spread of COVID-19 demonstrates exponential growth and changes of magnitude, which correspond to the above requirements (c) and (b), respectively. Likewise, the reported COVID-19 data are clearly not human-assigned numbers, unless these were manipulated at some degree. Hence, several studies have applied the Benford's law on COVID-19 data from various countries for more or less the same purpose of establishing possible evidences of data manipulation [12, 13, 14].

Materials and Method

This study employed secondary analysis on the reported Philippine COVID-19 data, which cover total number of cases, total number of deaths, and total number of recoveries. Gathered from the Department of Health (DOH) website - www.doh.gov.ph/covid19tracker, the data came from 83 provinces/cities with at least one case, 49 provinces/cities with at least one death, and 67 provinces/cities with at least one recovery [15]. The data used were those reported as of April 30, 2020, exactly three months after the first recorded confirmed COVID-19 case in the country. This was also when the first Enhanced Community Quarantine (ECQ) was lifted in several provinces/cities.

Selection of the First Digit

Since the data considered in this study were the reported COVID-19 cases, these must be nonzero and non-negative integers. The first digit number is obtained by finding the significand S(x) from the raw data. This is done by moving the decimal point immediately to the right of the first nonzero digit [16]. For example S(329) = 3.29, S(37) = 3.7, both with first digit number 3.

Observed Frequency Distribution

The number of times a particular first digit number occurs in the data set is called the absolute frequency (F_a) and the consolidated F_a's of all the first digit numbers is the total absolute frequency (TF_a). The observed distribution is the relative frequency (F_r) distribution of the first digit numbers of COVID-19 data – total cases, total deaths, and total recoveries. This is done with the following formula:

$$F_r = \frac{F_a}{TF_a},$$

where:

F_r = relative frequency;
F_a = absolute frequency in each first digit number;
TF_a = total absolute frequency.

Expected Frequency Distribution

In this study, the expected distribution is the Benford's law on first digit probability. The Benford's distribution is derived with the following formula [17]:

$$P(d) = \frac{\ln\left(1+\frac{1}{d}\right)}{\ln(10)},$$

where:

P(d) = probability of the occurrence of the digit d;
d = the first digit d = 1,2,3, ... , 9.

Hence, the expected frequency distribution is shown in the following table:

Table 1. The Expected Frequency Distribution of the First Digit Numbers [17]

First Digit (d)	Probability of occurrence in the naturally-generated data sets
1	0.30
2	0.18
3	0.13
4	0.10
5	0.08
6	0.07
7	0.06
8	0.05
9	0.05
TOTAL	1.00

Statistical Analysis

To analyse whether there is a significant difference between the observed frequency distributions and the expected frequency distribution, Kolmogorov-Smirnov (KS) goodness of fit test was employed at 5% level of significance. This test assesses the degree of agreement between an observed distribution and a completely specified theoretical distribution. KS is an exact test, which is sensitive to all characteristics of a distribution including location, dispersion, and shape [18].

Results and Discussion

The absolute frequencies (F_a) and the corresponding relative frequencies (F_r) of the first digit numbers of the three sets of reported COVID-19 data – total cases, total deaths, and total recoveries – are presented in Table 2. For the total cases, the relative frequency (F_r) of the first digit number 1 is higher compared to other leading digits with F_r values of 0.31, 0.18, 0.16, 0.10, 0.07, 0.05, 0.08, 0.01, and 0.04 for the leading digits 1, 2, 3, 4, 5, 6, 7, 8, and 9, respectively. As observed, the F_r values are decreasing as the first digit number is increasing up to the leading digit 4. First digit numbers 5, 6, 7, 8, and 9, in the total cases, obtain a minimal F_r values ranges from 0.01 – 0.08. This observation is also true to the total deaths data whose calculated F_r values are 0.33, 0.24, 0.16, 0.06, 0.06, 0.02, 0.02, 0.04, and 0.06 for the leading digits 1, 2, 3, 4, 5, 6, 7, 8, and 9, respectively. Similarly, for the total recovery, the obtained F_r values are 0.33, 0.19, 0.18, 0.09, 0.07, 0.03, 0.01, 0.03, and 0.06 for the leading digits 1, 2, 3, 4, 5, 6, 7, 8, and 9, respectively.

The corresponding distribution of the F_r values for the leading digits in each set of reported COVID-19 data is now the observed relative frequency distribution, which was subsequently compared to the expected frequency distribution according to Benford's law. The behavior of these distributions is presented in Figure 1.

Table 2. The Observed First Digit Distribution of COVID-19 Data in the Philippines.

First Digit Number	Total Cases		Total Deaths		Total Recoveries	
	Absolute Frequency (F_a)	Relative Frequency (F_r)	Absolute Frequency (F_a)	Relative Frequency (F_r)	Absolute Frequency (F_a)	Relative Frequency (F_r)
1	26	0.31	16	0.33	22	0.33
2	15	0.18	12	0.24	13	0.19
3	13	0.16	8	0.16	12	0.18
4	8	0.10	3	0.06	6	0.09
5	6	0.07	3	0.06	5	0.07
6	4	0.05	1	0.02	2	0.03
7	7	0.08	1	0.02	1	0.01
8	1	0.01	2	0.04	2	0.03
9	3	0.04	3	0.06	4	0.06
TOTAL	83	1.00	49	1.00	67	1.00

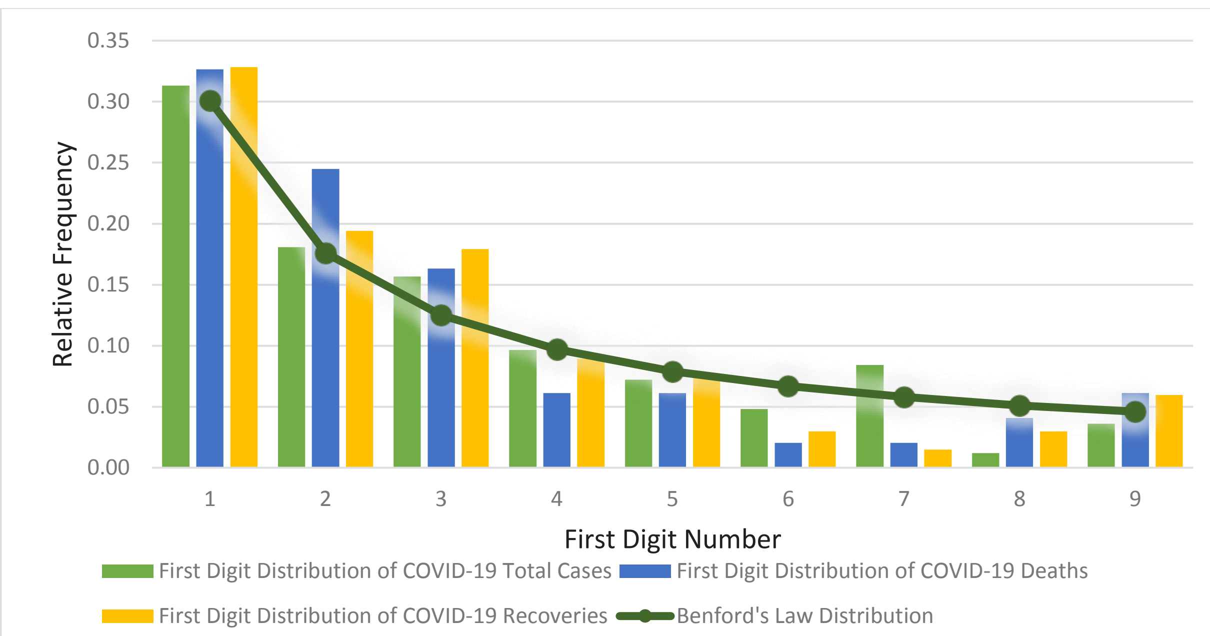

Figure 1. The Observed Relative Frequency (F_r) Distribution of the First Digit Numbers of COVID-19 Total Cases, Total Deaths, and Total Recoveries with the Benford's Law (Expected Distribution).

In Figure 1, the three sets of the observed relative frequency (F_r) distributions for COVID-19 total cases, total deaths, and total recoveries are represented by the bar graphs, while the expected frequency distribution, the Benford's law, is characterized by the line graph. Through visual examination, the observed frequency distributions seem to agree with the expected frequency distribution. To statistically check the similarity of these distributions, Kolmogorov-Smirnov test (KS) was applied at 0.05 level of significance (see Table 3 for results).

Table 3. Test Difference between the Observed Frequency Distribution of COVID-19 Data and the Benford's Law.

COVID-19 Data	p-value	Interpretation
Total Cases	0.979	Not Significant
Total Deaths	0.699	Not Significant
Total Recoveries	0.699	Not Significant

In the three sets of COVID-19 data considered in this study, the statistical p-values are all greater than 0.05 significant level, leading to the conclusion that there is no significant difference between the relative frequency (F) distributions of the first digit number of the reported COVID-19 total cases, total deaths, and total recoveries and the Benford's law distribution. Hence, this suggests no substantial evidence that the reported COVID-19 data in the Philippines is fraudulent. This should help dispel doubts cast upon the country's COVID-19 data as provided by the DOH in its website. However, as the data sets considered were limited only to those reported as of April 30, 2020, it is recommended that this test be conducted on a regular basis until the pandemic is over.

Acknowledgments

The author extends his gratitude to Southern Philippines Agri-Business and Marine and Aquatic School of Technology for its support in completing this study.

References

1. WHO Director-General's Opening Remarks at the Media Briefing on COVID-19. 2020 Mar 11. Retrieved from https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020#.
2. Ducharme, J. World health organization declares COVID-19 a 'pandemic.' Here's what that means. Times; 2020 Mar 11. Available from: https://time.com/5791661/who-coronavirus-pandemic-declaration/.
3. Department of Health. DOH confirms first 2019-ncov case in the country; assures public of intensified containment measures; 2020 Jan 30. Available from: https://doh.gov.ph/doh-press-release/doh-confirms-first-2019-nCoV-case-in-the-country.
4. Santos, E. PH reports 276 new COVID-19 cases, total now 8,488. CNN Philippines; 2020 Apr 30. Available from: https://www.cnnphilippines.com/news/2020/4/30/philippines-covid-19-cases-deaths-recoveries.html.
5. Gotinga, JC. Hospital told not to report coronavirus deaths? DOH probing allegation. Rappler Philippines; 2020 Apr 12. Available from: https://www.rappler.com/nation/doh-probes-hospital-told-not-report-coronavirus-deaths.
6. Berger, A & Hill, T.P. 2011. Benford's law strikes back: No simple explanation in sight for mathematical gem. The Mathematical Intelligencer, 33(1): 85–91; 2011.
7. Diekmann, A. Not the first digit! Using Benford›s Law to detect fraudulent scientific data. Journal of Applied Statistics, 34(3): 321-329; 2007. DOI: https://doi.org/10.1080/02664760601004940.
8. Goodman, W. The promises and pitfalls of Benford's law. Significance, 13: 38-41; 2016. https://doi.org/10.1111/j.1740-9713.2016.00919.x.
9. Kruger, P.S. and Yadavalli, V.S.S.. The power of one: Benford's law. The South African Journal of Industrial Engineering, 28(2): 1-13; 2017. Available from: DOI http://dx.doi.org/10.7166/28-2-1753.

10. Collins, J.C. Using excel and Benford's law to detect fraud. Journal of Accountancy. 2017 April 1, Available from: https://www.journalofaccountancy.com/issues/2017/apr/excel-and-benfords-law-to-detect-fraud.html#:~:text=The%20larger%20the%20better.,for%20this%20type%20of%20analysis.
11. Moreno-Montoya José. Benford´s Law with small sample sizes: A new exact test useful in health sciences during epidemics. Revista de la Universidad Industrial de Santander. Salud. 52(2): 161-163; 2020. DOI http://dx.doi.org/10.18273/revsal.v52n2-2020010.
12. Koch, C., & Okamura, K. Benford's Law and COVID-19 reporting. Economics Letters, 196, 109573; 2020. DOI https://doi.org/10.1016/j.econlet.2020.109573.
13. Zhang, J. Testing case number of coronavirus disease 2019 in China with Newcomb-Benford law; 2020. Available from: https://arxiv.org/abs/2002.05695.
14. Balashov, V.S., Yan, Y., & Zhu, X. Are Less Developed Countries More Likely to Manipulate Data During Pandemics? Evidence from Newcomb-Benford Law; 2020. Available from: https://arxiv.org/pdf/2007.14841.pdf.
15. Department of Health. COVID-19 Tracker Philippines; 2020 April 30. Available from: www.doh.gov.ph/covid19tracker.
16. Berger, A & Hill, T.P. What is… Benford's Law? Notices of the AMS, 64(2):132-134; 2017. DOI: http://dx.doi.org/10.1090/noti1477.
17. Glen, S. "Benford's Law (The First Digit Law): Simple Definition, Examples" From StatisticsHowTo.com**:** Elementary Statistics for the rest of us! Available from: https://www.statisticshowto.com/benfords-law/.
18. InfluentialPoints.com. Kolmogorov-Smirnov Test One & Two-sample, and Related Tests. Available from: https://influentialpoints.com/Training/kolmogorov-smirnov_test-principles-properties-assumptions.htm.