



History:

Received: September. 05, 2021
Reviewed: September 9, 2021:
Accepted: October. 15, 2021
Published: November . 04, 2021
Collection year: 2021
Status: Published

Identifiers and Pagination:

Year: 2021
Volume: 13
First Page: 55
Last Page: 60
Publisher ID:10.21065.19204159.13.55
doi:http://dx.doi.org/10.21065/19204159.13.55

Correspondence:

Anab Department of Pharmaceutics, Faculty of Pharmacy, Dow college of Pharmacy, DUHS, Karachi, , Pakistan. E.: anabfatima@gmail.com
Tel: 923022896103

Citation:

Anab Fatima, Sheikh Abdul Khaliq , Ishrat younas A statistical analysis of covid-19 data of Pakistan by applying benfords law .J App Pharm, 2021 vol 13, p 55-60

Funding: The authors received no direct funding for this research.

Competing Interests: The authors declare no competing interests

Additional information is available at the end of the article.

Mini-Research Article

A STATISTICAL ANALYSIS OF COVID-19 DATA OF PAKISTAN BY APPLYING BENFORD'S LAW

Anab Fatima¹, Sheikh Abdul Khaliq², Ishrat Younas³

1. Department of Pharmaceutics, Faculty of Pharmacy, Dow college of Pharmacy, D UHS. Karachi. Pakistan.
2. Department of Pharmaceutics, Hamdard University Madinat al-Hikmah Muhammad bin Qasim Avenue,76400, Karachi. Pakistan.
3. Hamdard University Madinat al-Hikmah Muhammad bin Qasim Avenue,76400, Karachi. Pakistan

ABSTRACT

We applied New-comb Benford's Law to validate the reliability of Covid -19 figures in Pakistan. Official data were taken from March 2020 till November 2020 and the first digit test is applied for the national aggregate dataset of total cases. There are shreds of evidence that the dataset of Pakistan does not conform to the New-comb Benford's Law theoretical expectations. The results are robust to the goodness of fit by applying the chi-square test. Although it is appreciating to concern over evidence-based policymaking, we found that the Pakistani epidemiological surveillance system fails to provide trustful data as per New-comb Benford's Law assumption.

Key words: Benford's Law, COVID -19, Pakistan, epidemiology, SARS

Authors' Contributions: 1 Conception, study design, data collection & processing, critical review. 2 & 4 Conception & study design, data collection & processing. 3-4 Data analysis and/or interpretation. 3 Drafting of manuscript. 4 Drafting of manuscript, critical review.

Acknowledgement: The authors acknowledge all those personnel who help and support this study in general.

INTRODUCTION

In December 2019 the city of Wuhan (China) suffered from corona virus epidemic and gradually it becomes a global pandemic of serious concern in January 2020 (World Health Organization, 2020a, 2020b). World Health Organization (WHO) declared this outbreak as public health emergency of international concern (PHEIC) and in March it was stated as global pandemic. Since from that time there were 38964805 cases reported and 1095978 death were observed worldwide till October 17, 2020 due to novel corona virus(<http://covid.gov.pk/>). The situation created a socioeconomic anarchy across whole world as various political, religious, sports and cultural activities postponed (The New York Times, 2020). Such anarchy become more worst due to panic purchase resulted in shortage of goods(Jade Scipioni, 2019). It causes national or local closure of academic institution worldwide (UNESCO 2020).

Our study focuses on applying Benford's law also called the Newcomb–Benford Law, on investigating the trend and the authenticity of COVID-19 data. It was firstly observed by Newcomb and popularized by Benford (Newcomb S.1881, Benford F.1938). The law usually applied to find the authenticity of data in the area of financial matter, investigating accounting fraud and political issues(Beber B, Scacco A :2012, Brady HE:2005, Cho W, Gaines B:2007, Deckert J,etal:2011, Durtschi C,etal:2004, Mebane Jr WR:2006, Winter C, Schneider M, Yannikos Y:2011).

Mainly it based on the principle that first digit of decimal number is likely to be equal to 1 and its possibility to be equal to subsequent numbers i.e. 2 ~ 9 decrease progressively. It is strongly evident that Benford's distribution followed by common disease numbers(Sambridge M,etal:2010). On this ground basis studies attempted to analyze COVID-19 Benfordness to check how well the data fits Benford's law. Here we mentioned data of published studies related to COVID-19 as number of cases may change over time. The data collected by Sambridge and Jackson till 9th April 2020 and found that data from USA, Japan, Indonesia and other European countries follows benford distribution but with certain variances. The study also mentioned that total death data of Czech Republic showed high peak at "9" that is contrary to law indicating that it does not fulfilled Benford's law. However it has been confirmed that the time line for evaluation of data set was very critical. Further that if cases become in thousands, first digit eventually become constant due to flattening of daily confirmed cases when virus growth is reduced. Lee, Han reflected similar thoughts in the paper submitted on 28 April 2020. In a study demonstrated by Koch and Okamura on 28 April 2020, the number of confirmed COVID cases at USA, Italy and China match the law with high Benfordness with no any data manipulation. A study having data up to 15 March 2020 by Idrovo and Manrique-Hernández didn't found data manipulation in number of cases of China. At the same time Raul checked the confirmed COVID cases up to 30 April 2020 among data of 23 countries and found discrepancy in the data of Italy, Portugal, Netherlands, the United Kingdom, Denmark, Belgium and Chile. Hence it is proved from these studies that Benford's law can apply to analyze COVID but with contradictory conclusion concerning possible data manipulation in any country. In this study we apply Benford's law to view COVID-19 cases data in Pakistan.

METHODOLOGY

I. Data collection:

Data were collected from March 2020 till 11th November 2020 from <http://covid.gov.pk/> and it mentioned that there were 348,184 total confirmed cases in Pakistan to date with 320,065 (98%) total recovery cases/ discharged and 7,021 (2%) total deaths. Also find out daily confirmed cases and daily deaths. Data of country is divided into regions(Provinces) and summed the country data per day.

II. Law of Benford and d factor:

In our study we apply Benford's law which is based on the fact that mantissa of 53 logarithms of exponentially growing numbers tends to be uniformly distributed. A decimal number x can be written in the logarithmic form as:

$$\log_{10} x = \log_{10}(10^a \cdot b) = a + \log_{10} b, (1 \leq b < 10)$$

Where

Where a = integer part

$\log_{10} b$ = mantissa of the logarithm of x .

The first digit of x is obvious by the log-mantissa $\log_{10} b$. First digit with numbers equal to

$d = 1, \dots, 9$, respective log-mantissa will be at the domain of $[\log_{10} d, \log_{10} (d + 1))$.

Therefore the probability of first digit to be equal to d is shown by equation:

$$P(d) = \log_{10}(d + 1) - \log_{10} d = \log_{10} \left(1 + \frac{1}{d}\right), \text{ for } (d = 1, \dots, 9)$$

In this we mention probability distribution for numbers ranges from 1-9 as the first number as per Benford's law.(Table 1)

Table 1. First digit distribution according to the Benford's Law

d	1	2	3	4	5	6	7	8	9
P(d)	30.1%	17.6%	12.5%	9.7%	7.9%	6.7%	5.8%	5.1%	4.6%

Here we quantified the deviation from Benford's curve by d^* factor i.e. it is measurement of Euclidian distance between country numbers and Benford's distribution by normalizing it by maximum possible distance. The value of d^* is 0.0 if data exactly matches the Benford's distribution. According to study proposed by Goodman, 2016 a value of d^* higher than 0.25 is evident for data manipulation.

RESULT AND DISCUSSION

Result is addressed in graphical form after application of Benford's Law. Figure 1 and 2 illustrate the result and data comparison of Pakistan to the Benford's Law distribution (line in Black).

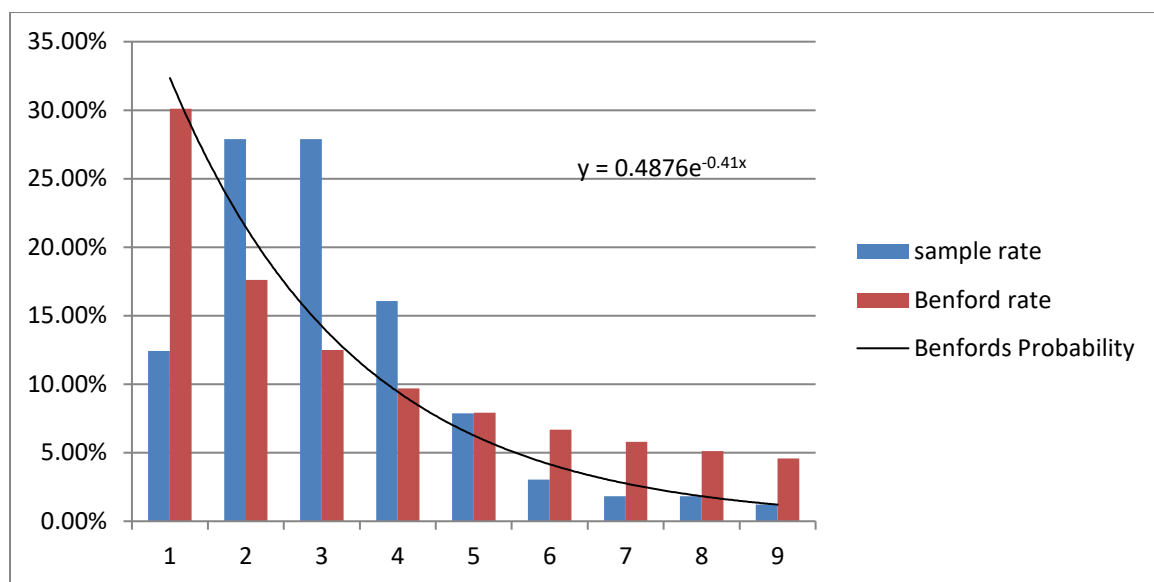


Figure 1: Total confirmed cases of Country. The black curve refers to Benford's Law Probability

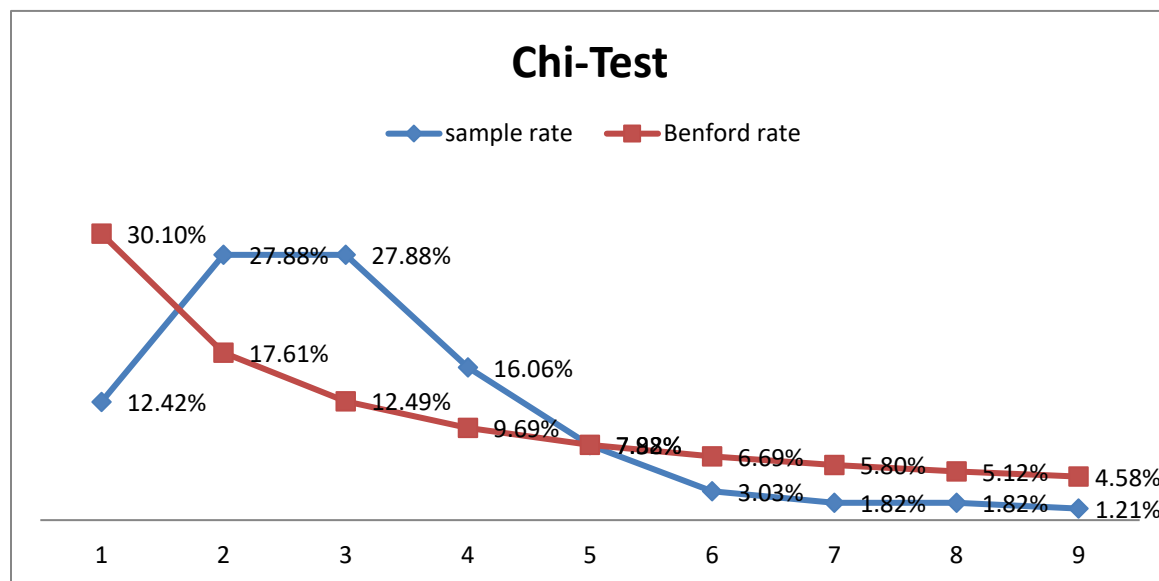


Figure 2: Chi-square distribution

The figure includes daily confirmed cases. Pakistani data showed high Benfordness in the data (Figure 1 and 2) with a d^* of 0.48 which cannot be explained. The data is partially in agreement with Benford with an odd peak at 2'3 and 4 (Figure 1 & 2).

CONCLUSION

The application of the Benford's Law for assessment of Covid-19 data was confirmed. It is one of the valid method to find pitfalls and variations in the data sets of countries and to suggest and identify possible chances of data manipulations. Our analysis suggested a possibility of manipulation on Pakistan's Covid-19 numbers of total deaths till November 11th 2020, but it is not a proof of manipulation and illegality but we found a constant departure of Pakistani Covid 19 reports from Benford's law. In our findings we performed different epidemiological test and data source and it is being evident that local epidemiological surveillance system did not trustful data on SARS-Covid 19 epidemic in Pakistan. For effective evidence based policy making much reliable and valid data is needed which is in relation to transparency of institutional workouts. In order to evaluate interventions of Government affecting the outcomes they designed, an open data is a pre-requisite. Although it is evident that there are several technical issues in quantifying Covid-19 but effective strategies and policies rely on data quality and accessibility. In conclusion it is promising that all stake holders and policy makers of Health and medicine follow more sophisticated guidelines and practices in data collection for more careful and comprehensive analysis of these critical data.

REFERENCES

1. Beber B, Scacco A. What the numbers say: A digit-based test for election fraud. *Political analysis*. 2012;20(2):211-34.
2. Brady HE. Comments on Benford's Law and the Venezuelan Election. MS dated January. 2005;19:2005. 26
3. Benford F. The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society*. 1938;78(4):551-72.
4. Cho W, Gaines B. Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance. *The American Statistician*. 2007;61:218-23.
5. Deckert J, Myagkov M, Ordeshook PC. Benford's Law and the Detection of Election Fraud. *Political Analysis*. 2011;19(3):245-68.
6. Durtschi C, Hillison WA, Pacini C, editors. *The effective use of Benford's Law to assist in detecting fraud in accounting data* 2004.
7. Government of Pakistan. Coronavirus in Pakistan. <http://covid.gov.pk/>.
8. Goodman W. The promises and pitfalls of Benford's law. *Significance*. 2016;13(3):38-41.
9. Idrovo AJ, Manrique-Hernández EF. Data Quality of Chinese Surveillance of COVID-19: Objective Analysis Based on WHO's Situation Reports. *Asia Pac J Public Health*. 2020;32(4):165-7.
10. Jade Scipioni, 2019. CNBC. June 5. <https://www.cnbc.com/jade-scipioni/>.
11. Koch C, Okamura K. Benford's Law and COVID-19 Reporting. SSRN. 2020.
12. Lee K-B, Han S, Jeong Y. COVID-19, flattening the curve, and Benford's law. *Physica A: Statistical Mechanics and its Applications*. 2020;559:125090.
13. Mebane Jr WR, editor *Election forensics: Vote counts and Benford's law*. Summer Meeting of the Political Methodology Society, UC-Davis, July; 2006
14. Newcomb S. Note on the Frequency of Use of the Different Digits in Natural Numbers. *American Journal of Mathematics*. 1881;4(1):39-40.
15. Raul I. How Valid are the Reported Cases of People Infected with Covid-19 in the World? *International Journal of Coronaviruses*. 2020;1(2):53-6.

16. Sambridge M, Jackson A. National COVID numbers - Benford's law looks for errors. Nature. 2020;581(7809):384. The New York Times, 2020. Print for Tuesday. The New York Times Retrieved April 19, 2020, from. <https://www.nytimes.com/issue/todayspaper/2020/02/25/todays-newyork-times> (Feb. 25).
17. UNESCO, 2020. Half of World's Student Population Not Attending School: UNESCO Launches Global Coalition to Accelerate Deployment of Remote Learning Solutions. UNESCO March 19. <https://en.unesco.org/news/half-worlds-student-populationnot-attending-school-unesco-launches-global-coalition-accelerate>.
18. World Health Organization, 2020a. Considerations for Quarantine of Individuals in the Context of Containment for Coronavirus Disease (COVID-19): Interim Guidance, 19 March 2020 (No.
19. WHO/2019-nCoV/IHR Quarantine/2020.2). World Health Organization. World Health Organization, 2020b. Critical Preparedness, Readiness and Response Actions for COVID-19-7 March 2020.
20. Winter C, Schneider M, Yannikos Y, editors. Detecting Fraud Using Modified Benford Analysis. Advances in Digital Forensics VII; 2011 2011//; Berlin, Heidelberg: Springer Berlin Heidelberg.