TCSA
https://dergipark.org.tr/tr/pub/tcsa
Vol.1 No.1
May 2024: 1-1
Accepted 29 April 2024
Published 2 June 2024

Transactions on Computer Science and Applications

Galatasaray University
TCSA
Transactions on Computer Science and Applications

Coronavirus record fraud estimation by Benford's law analytics

[*]Günce Keziban ORMAN [*,a,b], Sena ATAKAN[a], Elif Ece ERDEM[a], Şükrü Demir İnan ÖZER[a], Timoteos Onur ÖZÇELİK[a]
[a] Galatasaray University, Computer Engineering, Istanbul, Turkiye, 34000,
[b] ORCID: 0000-0003-0402-8417

*** Corresponding author email address**:korman@gsu.edu.tr, **phone:** 00902122274480

Abstract

The COVID-19 pandemic has generated vast amounts of data, including daily case and death counts by country. Analyzing the reliability of this data is crucial, and Benford's Law, a statistical principle that predicts the frequency of leading digits in naturally occurring datasets, can serve as a valuable tool. This study explores Benford's Law applications to these COVID-19 data, departing from previous work in two key ways. First, we leverage the most comprehensive dataset to date, spanning nearly three years of the pandemic, offering a broader and more robust picture. Second, we introduce a novel analysis technique – monotony checking – to assess Benford compliance by examining the decreasing frequency of leading digits. We employ a multi-pronged approach, encompassing chi-square tests, expected frequency calculations, mean absolute distance scores and exponential smoothing. Strikingly, these analyses converge in showcasing significant deviations from Benford's Law in numerous countries across diverse regions. Furthermore, our monotony analysis reinforces these findings, suggesting potential anomalies in data reporting. This research showcases the potential of Benford's Law for scrutinizing health-related data, much like its applications in financial and network domains. The observed discrepancies warrant further investigation to ensure data transparency and reliability in the ongoing fight against COVID-19.

Keywords: Benford's Law, Coronavirus, Statistically Significance

1. Introduction

The novel coronavirus, referred to as COVID-19, initially manifested on January 13, 2020, subsequent to a comprehensive investigation conducted on a cohesive cohort of pat ients who exhibited respiratory manifestations, namely fever, cough, and dyspnea, in the region of Wuhan, located in China, towards the culmination of December. The repercussi ons of the COVID-19 pandemic have been of colossal prop ortions, resulting in a far-reaching impact across the globe, which has subsequently necessitated prodigious scientific e ndeavors within this domain. Nations have perpetually enga ged in the exchange of daily data pertaining to the incidenc e and fatality rates associated with the disease. However, u ncertainties surrounding the precision and veracity of such data, prompted by multifarious factors, have engendered an extensive discourse and deliberation.

In this study, we explore the application of Benford's Law as a mathematical testing environment to assess the reliability of COVID-19 data. Benford's Law provides a mathematical test to verify the reliability of data series (Berger & Hill, 2021; Formann, 2010; Michalski & Stoltz, 2013) and has previously been used to evaluate the accuracy of COVI D-19 data collection (Campolieti, 2022; Kolias, 2022). Ho wever, our study differs in several aspects. Firstly, we analy se data from multiple countries worldwide, rather than focu sing on a single country or region. Secondly, we examine d ata from the entire COVID-19 period, from March 1, 2020, to March 21, 2023, providing a comprehensive scope. Lastl y, we go beyond statistical significance tests and propose to examine the monotony of the first digits in the data. The article proceeds by discussing related works, expla ining the technical details of our methodology, and presenti ng the results.

2. Related Work

Since its inception, Benford's Law has found application across diverse domains, spanning law, finance, and health (Alsunaidi et al., 2021; Balashov et al. 2021; Caffarini et al. 2022; Kilani, 2021; Lee et al. 2020; Rahimi et al. 2023; Zhao et al. 2020). This section offers an overview of notable s tudies. Li et al. (2019) explore the utility of Benford's Law as a tool for discerning data quality and detecting anomalous data in various domains. data in various domains.

The article summarizes the law's application in natural and social sciences, investigating data conditions and influencing factors affecting detection accur acy. The study proposes enhancements to the model, includ ing broadening the detection scope, combining multiple me thods, and reinforcing the explication of detection outcome s. The authors advocate for comprehensive research across disciplines to advance Benford's Law, emphasizing the need to delve into its fundamental principles, integrate it with o ther data processing technologies, and extend its applications. They further recommend refining the foundational theor y of the law, intensifying its collaborative use with other da ta processing techniques, and enhancing the explanation of its outcomes.

Campolieti (2022) focuses on assessing the accuracy of reported COVID-19 deaths in the United States through the application of Benford's Law. The United States has faced severe repercussions from the COVID-19 pandemic, with a high number of deaths, particularly among the elderly and patients with pre-existing medical conditions. An investigation into the accuracy of reported numbers ensued following the revelation of under-reported COVID-19 deaths in nursi ng homes by a report from the New York State Attorney G eneral. Benford's Law is employed as a statistical tool to id entify potential misrepresentations. The study utilizes Benf ord's Law to scrutinize reported COVID-19 deaths on a stat e-by-state basis, revealing deviations from Benford's Law f or the majority of states. Mean absolute deviation (MAD) fi ndings indicate disparities from Benford's Law in New Yor k, while distortion factors point to under-reporting of death s in most states. The study asserts a statistically significant under-reporting of COVID-19 deaths in the United States, with only a select few states adhering to Benford's Law.

A new analytical examination procedure, as outlined by Busta & Weinberg (1998), measures the disparity between the digit distribution of a data set and the Benford digit distribution. This discrepancy may highlight potential data mani pulation, signalling the need for additional audit testing. Th e application is specifically tailored to financial data analysi s with an artificial neural network-based modelling approac h. The primary objective is to discriminate between "norma l" financial data and data subjected to manipulation. The fin dings show that Benford's analytical review procedure detected contamination in data, at a level of 10 percent or more, 68 percent of the time. Conversely, when the data was uncontaminated, the test indicated that the data was "clean" with a 67 percent accuracy. Notably, the results of Benford's ana lytical examination procedure compare favorably to conven tional analytical examination methods. Its distinctive analys is lends itself to complementing these traditional approache s. It is important to note, however, that the study relied on s imulated data, rather than real data, which represents a limit ation. Future research could enhance the study's robustness by incorporating real-world data.

In another work which is dedicated to the analysis of CO VID-19 data with Benford's Law (Kolias, 2022), the researchers focus on detecting variations in the leading digits of COVID-19 data across European Union (EU) countries. Th eir investigation explores the correlation of these disparities with four development indices: GDP percapita, health spen ding relative to GDP, the universal health coverage (UHC) index, and the full vaccination rate. The study reveals subst antial deviations from the Benford distribution in a majority of EU countries. Specifically, Denmark, Ireland, and Greec e exhibit the highest gaps in daily cases, while Cyprus, Italy and Greece demonstrate the highest gaps in deaths. Regres sion analysis results indicate a positive association between the complete vaccination rate and deviations from Benfor d's Law in mortality. This study highlights that applying Be nford's Law to COVID-19 data in the EU can provide insig hts into the relevance of the reported figures and call into q uestioning the relationship between the full vaccination rate and the reported figures.

Arshadi & Jahangir (2014) advocate employing Benford's Law to identify anomalies in Internet traffic, specifically deliberate attacks on TCP streams. Conversely, Formann (2 010) employs simulation to explore the association between the distribution of significant figures and the distribution of the observed variable. As Benford's Law is a generic mathe matical tool, it is adapted to several different domains. In re cent years, it has also been used for COVID-19 data checki ng, as we aimed to do in this work. Our work can be seen a s an extension of prior efforts, as we employ the same statis tical tests. However, our study takes a broader approach, an alysing the data set covering all nations and capturing the e ntirety of the COVID-19 pandemic timeline. Moreover, we propose a novel test that has not been previously applied in conjunction with Benford's Law and the COVID-19 data set.

3. Method

A. Benford's Law

Benford's Law, also known as the law of prime digits, reveals that the distribution of numbers in natural data sets is not uniform and that the probability of encountering different digits in first digits differs. There is neither a homogeneo us distribution nor a probability of 1/9. To explain better, in a series of numerical data, one might expect to see the digit s 1 to 9 appear about as frequently as the first significant di git, i.e. with a frequency of 1/9 for each. But the series very often follows Benford's Law: for almost a third of the data, the first most frequent significant digit is 1 (Berger, A. et a l., 2011).

Benford's Law is used to detect fraud or flaws in data collection based on the distribution of the first digits of the ob served data. A Benford distribution of the first digits arises naturally for exponential processes with multiple changes i n magnitude Michalski, T. et al., 2013. More clearly, the fre quency with which the first digit is 1 is 0.301, the first digit is 2 is 0.176, etc., decreasing to the first digit being 9 only 0. 046 of the time. In Eq. 1, the probability that a number starts with the digit n, for a distribution satisfying

Benford's L aw is represented.

$$P(n)=\log_{10}\left(\frac{n+1}{n}\right) \quad (1)$$

Based on this formula, the distribution of the first digits of nine natural numbers is shown in pie-chart given in Fig. 1.

B. Statistical Significance

Statistical significance is a concept that is based on hypothesis tests. It is usually used for checking if two (or more) different data sets (a.k.a., samples) are coming from the sa me population statistically or not. These tests use the basic hypothesis of having the same mean (μ). Although artificial learning algorithms are proposing more advanced analysis t echniques, they need larger data sets. Hence, statistical anal ysis can still give good insight, especially in cases of lack o f observation. There are several statistical significance tests for different scenarios. Here, we use the chi-squared test.

A. Chi-squared Test

The chi-square (2) test is a statistical method used to determine whether observed data follows an expected distribu tion or whether there is a significant difference between obs erved and expected frequencies. Differently from many stat istical significance tests, it does not compare the sample av erage and infer the population mean, but it uses the frequen cies. Thus, the 2 test offers a good environment for categori cal data sets. The formula of 2 tes

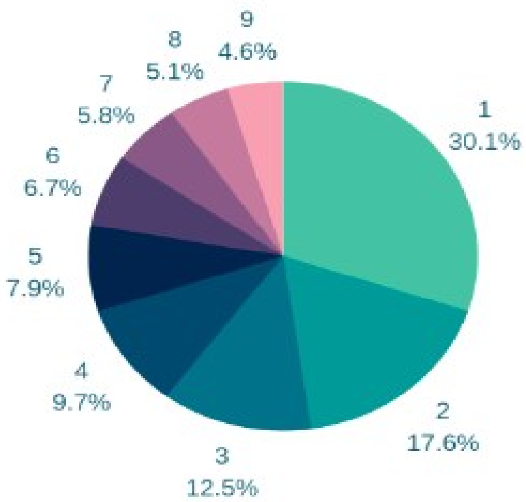

Fig. 1. Pie chart of the expected probabilities of the natural numbers' appe arance at the first digit according to Benford's Law.

$$X^2=\sum\frac{(O_i-E_i)^2}{E_i} \quad (2)$$

Here, $O_i$is the observed frequency of the ith object, and E_i is its expected value. In our case, for each country, the frequency of Covid cases occurrence of the digits 1 to 9 in the f irst digit is compared to the expected value, which is the rat e of occurrence of the first digits according to Benford's La w, according to this formula. The result is compared to the calculated critical value, and an estimate is made to determi ne whether fraud has occurred.

B. Mean Absolute Difference

The Mean Absolute Distance (MAD) is a statistical metric used to measure the average absolute difference betwee n observed values and expected values. It is commonly use d in forecasting and time series analysis to evaluate the acc uracy of predictions. In our Benford Law's analysis case, M AD formula is given by Eq. 3.

$$MAD=\frac{1}{9}\sum_{1}^{9}\left|P(O_i)-P(E_i)\right| \quad (3)$$

C. Our Analysis Approach: Monotony Check

According to Benford's Law, the higher the value of the first digit, the lower the incidence should be. The statistical significance tests check for each digit whether the observed value is different from its expected value according to Benford's Law. However, they do not check the consecutive dec rease in the digits. More clearly, the statistical significance tests do not consider the sequential order of the digit decrea se. In this case, if the appearance of the next digit is larger t han the previous one but the amount is statistically not signi ficant, it can pass the statistical tests. However, according to Benford's Law, consecutive digits should appear fewer ti mes (in logarithmic decreasing order). We check if the obse rved numbers obey the monotone decrease besides the stati stical significance tests.

Algorithm 1 Monotony Check

Input: countries, covid cases, ϵ
Output: monotony state

∀ country ∈ countries, monotony state[country] = TRUE
for country in countries **do**
  digit appearance = $digitfrequencies$(covid cases [country])
  for $i = 1$ to 8 **do**
    if (digit appearance[$i+1$] $\not<$ appearance[i]+ϵ) **then**
      monotony state[country] = FALSE
    end if
  end for
end for
return monotony state

In algorithm 1, we represent the simple procedure of our monotony check approach. This algorithm takes a list of countries and the daily COVID case number of each country. It also takes an error parameter, ε, which is used as the gap between the difference of consecutive digit occurrences. Th e algorithm first sets the monotony state of all countries as TRUE and starts checking each country's monotone decreas e, considering also the gap ε. If any consecutive digit does not obey the consecutive decrease, then it sets the related c ountries' state as false. As a result, it returns the states of all the countries.

In this work, a margin of error (ε) of 10 is set after an ex perimental optimization analysis since this value provides t he best results. It is tested for countries to see whether a mo notonic decrease occurs as the digit values increase.

Table 1
COVID Data Set Sample

Date reported	Country	WHO region	New cases	Cumul. cases	New deaths	Cumul. deaths
12/23/2022	China	WPRO	6966046	50447985	894	36318
12/22/2022	China	WPRO	6434648	43481939	836	35424
12/24/2022	China	WPRO	6327801	56775786	1308	37626
12/21/2022	China	WPRO	5905312	37047291	628	34588
12/25/2022	China	WPRO	5669864	62445650	1369	38995

D. Exponential Smoothing Analyses

The Exponential Smoothing technique has also been employed as another method for time series analysis. The co nsistency of forecasts made with this method is reflected in the graphs. Simultaneously, an anomaly score has been calc ulated for each country. The differences between the model predictions for each day and the actual values were squared, and then the square root of their average was computed as the Exponential Smoothing Root Mean Square Error (RMS E) score.

4. Results

The *Case and death count data* data set, which shows in detail the daily number of COVID cases and deaths betwee n the dates of March 1, 2020, and March 21, 2023, of all countries shared by the World Health Organization (WHO), was used*.

The data set consists of the following features: *Date reported, Country code, Country, WHO region, New cases, Cumulative cases, New deaths, and Cumulative deaths*. There are 275891 total lines corresponding to daily observed case s in all countries. The data set contains a uniform daily time series; that is, it contains data for every day. No cleaning w as necessary, as the data set was clean and suitable for the s tudy. A small data sample of the COVID cases is represent ed in Table 1.

For each country, the distribution of the first digits of the number of cases in terms of numbers and percentages was studied to check if they obey Benford' Law. Results for some countries are shown in Table 2.

A. Analyses of the Chi Square Goodness-of-fit Test

The frequency of occurrence of the numbers 1 to 9 in the first digit was calculated for each country in the first step. Then, for each country, expected values were calculated by u sing Benford's Law formula given in Eq. 1. For instance, if the total number of cases in a country is 100,000, then 0.30 1*100,000 cases starting with 1 are expected. The χ^2 test r esult is calculated by using the formula given in Eq. 2.

Table 2
Digit Occurrence and Frequencies for Turkey, United States of America, and United Kingdom

	Turkey		USA		UK	
Digit	Occur.	Freq.	Occur.	Freq.	Occur.	Freq.
1	218	0.262	325	0.322	234	0.207
2	178	0.214	168	0.167	206	0.182
3	57	0.069	105	0.104	218	0.193
4	37	0.044	111	0.11	155	0.137
5	51	0.061	92	0.091	96	0.085
6	48	0.058	65	0.064	68	0.060
7	70	0.084	55	0.055	64	0.057
8	97	0.117	42	0.042	53	0.047
9	76	0.091	46	0.046	35	0.031

Since the value of degrees of freedom for this test is 8, t he critical value becomes 15.507 when the significance level, α, is set at 0.05. Therefore, there is an anomaly according to Benford's Law in countries whose χ^2 values are greater than 15.507. The more the χ^2 value increases, the greater the anomaly. In total, 154 countries exceeding the critical v alue were identified. Afghanistan (451.47), India (514.16), Indonesia (538.1), and the Maldives (607.5) can be cited as examples of countries whose degree of χ^2 values are the hi ghest ones. In Fig. 2., a bar chart of result χ^2 values for some countries is represented. Accordingly, Turkey, Belarus, Brazil, Mexico, the Maldives, and Indonesia's results are much higher than expected critical values of χ^2.

* https://covid19.who.int/data

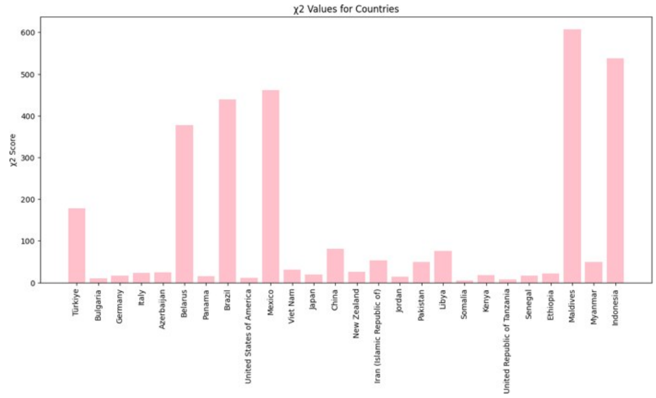

Fig. 2. Result χ^2 values of some countries

B. Analyses of the Mean Absolute Distance Test

To calculate the MAD value, we used the formula given in Eq. 3. We calculate the absolute values of the difference between the actual percentages of the number of cases previously calculated for each figure and the percentages deter mined according to Benford's Law. The values obtained for each number from 1 to 9 are added and divided by 9 to calc ulate a MAD score for each country. The countries with the highest MAD scores are Tajikistan (0.083), Saba (0.087), a nd Belarus (0.074). In Fig. 3., the bar chart of MAD values for some countries is presented. Accordingly, as the results of the χ^2 tests show, Turkey, Belarus, Brazil, Mexico, the Maldives, and Indonesia's have noticeably high deviations. Besides, China and Libya exhibit high MAD values as well.

We examine χ^2 tests and MAD results together by also considering the regions of the countries. The regions are cla ssified by the WHO. These regions are the African Region (AFRO), Eastern Mediterranean Region (EMRO), Southeast Asia Region (SEARO), Americas Region (AMRO), West ern Pacific Region (WPRO), and European Region (EUR O). In Fig. 4., χ^2 and MAD results are represented together The color difference in the figure is related to the degree of deviation. Here, five countries have low MAD values and χ^2 values below the critical value of 15.507: Somalia, Tanzania, Bulgaria, the United States, and Panama. We therefor e see that the number of cases in these countries is consiste nt with Benford's Law. In general, the results of these two t ests indicate the same thing. Moreover, we do not observe a ny homogeneity of deviation among the regions. Hence, the difference is due to the countries themselves. Since during t he period of COVID lockdown travel was also not permitte d, it is reasonable not to observe any similarity between the neighbouring countries or among the countries from the sa me region.

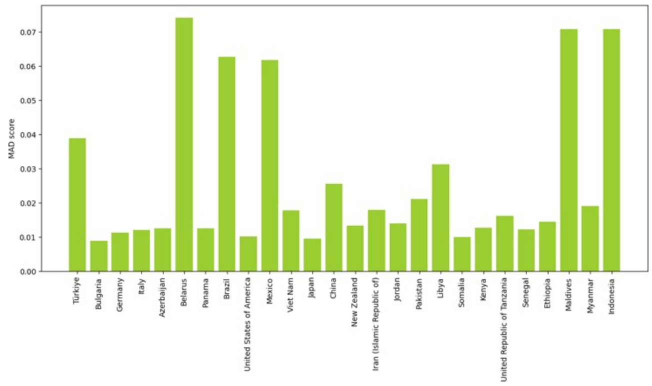

Fig. 3. Result MAD values of some countries

	WHO region	χ2	MAD
Somalia	EMRO	5.43	0.010
Tanzania	AFRO	6.91	0.016
Bulgaria	EURO	9.76	0.009
USA	AMRO	11.25	0.010
Panama	AMRO	15.03	0.013
Senegal	AFRO	16.25	0.012
Germany	EURO	17.03	0.011
Kenya	AFRO	18.53	0.013
Japan	WPRO	19.46	0.009
France	EURO	21.80	0.013
Ethiopia	AFRO	22.43	0.015
Italy	EURO	22.68	0.012
Azerbaijan	EURO	23.88	0.013
New Zealand	WPRO	25.78	0.013
Viet Nam	WPRO	31.04	0.018
Myanmar	SEARO	49.20	0.019
Pakistan	EMRO	49.50	0.021
Iran	EMRO	52.94	0.018
Libya	EMRO	75.50	0.031
China	WPRO	80.78	0.026
Türkiye	EURO	177.45	0.039
Belarus	EURO	377.12	0.074
Brazil	AMRO	439.52	0.063
Mexico	AMRO	462.15	0.062
Indonesia	SEARO	538.10	0.071
Maldives	SEARO	607.51	0.071

Fig. 4. Result MAD values of some countries

C. Analysis of the Monotony Check

According to Benford's Law, the higher the value of the first digit, the lower the incidence should be. This monotonic reduction was tested with a margin of error, ϵ, of 10. 153 countries followed this monotonic decrease, while 82 countries did not follow it. The rate of the number of countries obeying the monotone decrease is 0.65. The countries which do not follow the monotone decrease are Afghanistan, Angola, Antigua and Barbuda, Aruba, Australia, Bahrain, Belarus, Belize, Bhutan, Bolivia (Plurinational State of), Bonaire, Botswana, Brazil, Cabo Verde, Canada, China, Colombia, Cuba, Curaçao, Denmark, Dominican Republic , Ecuador, Egypt, Equatorial Guinea, Finland, Germany, Gu atemala, Guyana, Honduras, Hungary, India, Indonesia, Ira n (Islamic Republic of), Kazakhstan, Kosovo (Alsunaidi et al., 2021), Kyrgyzstan, Latvia, Lesotho, Liberia, Libya, Lie chtenstein, Lithuania, Malawi, Maldives, Martinique, Maur itania, Mauritius, Mayotte, Mexico, Montenegro, Myanmar Nepal, New Zealand, Norway, Oman, Paraguay, Peru, Port ugal, Qatar, Russian Federation, Rwanda, Saba, Saint Vinc ent and the Grenadines, Saudi Arabia, Senegal, Sint Eustatius, Sint Maarten, Slovenia, South Africa, Sri Lanka, Suriname, Sweden, Syrian Arab Republic, Tajikistan, Thailand, The United Kingdom, Turkey, Uganda, United Arab Emirate , United States Virgin Islands, Uruguay, Venezuela (Bolivarian Republic of).

We observe that the countries whose χ^2 and MAD results indicate that they obey Benford's Law are also obeying the monotone decrease as well and as expected. However, we see that the monotony check is giving more sensitive results There are more countries that do not obey it.

The graphical representaion of the RMSE of the exponential smooting fitting to the data set and original data for Br azil, Ethiopia and France is provided in Fig. 5, 6 and 7 resp ectively. We observe a highly low errors specifically for Et hiopia since the number of recorded cases are also low for t his country.

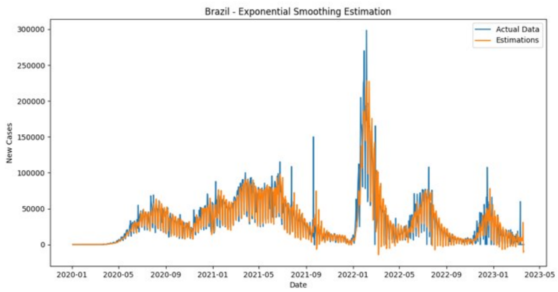

Fig. 5. RMSE of Exponential Smoothing in Brazil: 15162.605

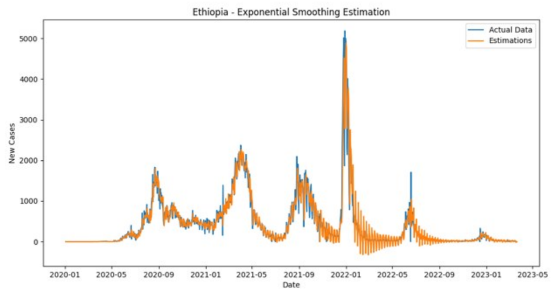

Fig. 6. RMSE of Exponential Smoothing in Ethiopia:220,045

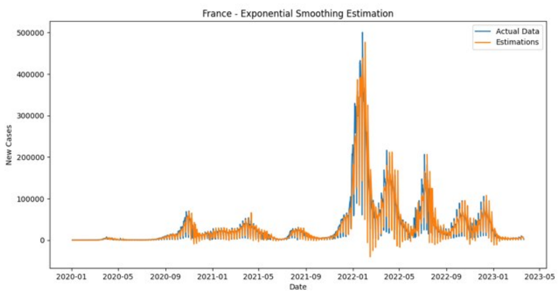

Fig. 7 RMSE of Exponential Smoothing in France: 16645,693

5. Conclusion

In this work, an analytical study was carried out on the relationship of COVID-19 cases with Benford's law. For the analysis, the dataset titled "Daily cases and deaths by date reported to WHO", which contains daily cases and deaths fo r all countries between 03/01/2020 and 03/21/2023, was us ed. The distribution of the first digits of the number of case s was calculated for each country in the dataset. According to Benford's law, the higher the value of the first digit, the lower the expected incidence.

To check whether Benford's law is followed, we first tested the downward trend in cases with a margin of error (10). 153 countries followed this downward trend, but 82 countr ies did not. The rate of compliance with Benford's law is 0. 65.

A chi-square test (X2) was then performed to calculate the differences between the actual frequencies of the first di gits of the number of cases and the frequencies predicted by Benford's law. This test identified 154 countries that exceed ed the critical value of 15.507. These include countries such as Afghanistan, India, Indonesia and the Maldives.

In addition, Exponential Smoothing methods were used toxamine the relationship between time series analyses an d Benford's law analyses. Root Mean Square Error (RMSE) scores were obtained by calculating the differences betwee n predictions made with Exponential Smoothing and actual values.

Finally, the relationship between the anomaly scores obtained by each time series analysis method and Benford's La w was evaluated using Pearson's correlation coefficient. Ho wever, no direct correlation was found between the time ser ies analysis methods and Benford's law. The most likely rea son for this is that time series analysis methods and Benfor d's law analyses approach the data from different perspectives.

In summary, this project investigated the possible detecti on of fraud in COVID-19 cases with Benford's Law and evaluated the relationship between the two using different ana lytical methods. Benford's Law analyses determined that th ere could be fraud in the case count data for Indonesia, Maldives, Afghanistan, Colombia, Finland, Mexico, Montenegro and Portugal. Meanwhile, the number of cases in the follo wing countries: Monaco, Singapore, Spain, Niger and Puert Rico showed a uniform distribution in accordance with Benford's law.

In the future, this work can be expanded by analysing common features between countries with similar estimates of fraud analysis using Benford's Law. These estimates made with Benford's Law can be compared with the analysis of fr aud in data disclosed by countries on different subjects. The correlation between results obtained using different techniq ues other than time series analysis can be examined. The rel iability of Benford's Law in detecting fraud can be analysed in greater details.

References

Alsunaidi, S. J., Almuhaideb, A. M., Ibrahim, N. M., Shaikh, F. S., Alqudaihi, K. S., Alhaidari, F. A., Khan, I. U., Aslam, N., & Alshahrani, M. S. (2021). Applications of Big Data Analytics to Control COVID-19 Pandemic. Sensors (Basel, Switzerland), 21(7), 2282. https://doi.org/10.3390/s21072282

Arshadi, L., & Jahangir, A. H. (2014). Benford's law behavior of I nternet traffic. Journal of Network and Computer Applications, 40, 194-205.

Balashov, V. S., Yan, Y., & Zhu, X. (2021). Using the Newcomb-Benford law to study the association between a country's COVID-19 reporting accuracy and its development. Scientific reports, 11(1), 22914. https://doi.org/10.1038/s41598-021-02367-z

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.

Busta, B., & Weinberg, R. (1998). Using Benford's law and neural networks as a review procedure. Managerial Auditing Journal, 13(6), 356-366.

Caffarini, J., Gjini, K., Sevak, B., Waleffe, R., Kalkach-Aparicio, M., Boly, M., & Struck, A. F. (2022). Engineering nonlinear epileptic biomarkers using deep learning and Benford's law. Scientific reports, 12(1), 5397. https://doi.org/10.1038/s41598-022-09429-w

Campolieti, M. (2022). COVID-19 deaths in the USA: Benford's law and under-reporting. Journal of Public Health, 44(2), e268-e271.

Formann, A. K. (2010). The Newcomb-Benford law in its relation to some common distributions. PloS one, 5(5), e10541.

Kilani, A. (2021). Authoritarian regimes' propensity to manipulate Covid-19 data: a statistical analysis using Benford's Law. Commonwealth & Comparative Politics, 59(3), 319-333.

Kolias, P. (2022). Applying Benford's law to COVID-19 data: the case of the European Union. Journal of Public Health, 44(2), e221-e226.

Lee, K. B., Han, S., & Jeong, Y. (2020). COVID-19, flattening the curve, and Benford's law. Physica A, 559, 125090. https://doi.org/10.1016/j.physa.2020.125090

Li, F., Han, S., Zhang, H., Ding, J., Zhang, J., & Wu, J. (2019, February). Application of Benford's law in Data Analysis. In Journal of Physics: Conference Series (Vol. 1168, No. 3, p. 032133). IOP Publishing.

Michalski, T., & Stoltz, G. (2013). Do countries falsify economic data strategically? Some evidence that they might. Review of Economics and Statistics, 95(2), 591-616.

Rahimi, I., Chen, F., & Gandomi, A. H. (2023). A review on COVID-19 forecasting models. Neural Computing and Applications, 35(33), 23671-23681.

Zhao, S., Gao, D., Zhuang, Z., Chong, M. K., Cai, Y., Ran, J., Cao, P., Wang, K., Lou, Y., Wang, W., Yang, L., He, D., Wang, M H. (2020). Estimating the serial interval of the novel coronavirus disease (COVID-19): a statistical analysis using the public data in Hong Kong from January 16 to February 15, 2020. Frontiers in Physics, 8, 347.