mathematics

MDPI

(Article)

" Applying Benford's Law to monitor death registration data: A management tool for the COVID-19 pandemic”

Abstract: In Spain, the COVID-19 pandemic has impacted the various regions of the country differently. The availability of reliable and up-to-date information has proved to be fundamental for the management of this health crisis. However, especially during the first wave of the pandemic (February-August 2020), the disparity in the recording criteria and in the timing of providing these figures to the central government created controversy and confusion regarding the real dimension of the pandemic. It is therefore necessary to have objective and homogeneous criteria at the national level to guide health managers in the correct recording and evaluation of the magnitude of the pandemic. Within this context, we propose using Benford’s Law as an auditing tool to monitor the reliability of the number of daily COVID-related deaths to identify possible deviations from the expected trend.

Keywords: Covid-19 deaths; Benford’s Law; health management tool; reliability data; auditing tool.

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Mathematics* **2021**, *9*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Lastname

Received: date
Accepted: date
Published: date

1. Introduction

It is clear that in order to face a health crisis of the magnitude of Covid-19, public health managers and administrators need to have the necessary tools to collect data correctly and predict the latest trends of the different pandemic indicators.

The Covid-19 health crisis has filled the health management literature with countless publications about methodologies for the prediction of incidence, transmission dynamics or number of cases [1]–[8]. Most of these models are based on mathematical and statistical tools such as for instance machine learning [9] linear generalized models [10] or logistic growth models [11].

Indeed, it is desirable that such tools are as operational and simple as possible, so that they can be implemented at all levels of the health and policy organizational hierarchy [14]. Therefore, in this paper we propose using Benford’s Law (BL) [15] as a guide to monitor the correct recording of Covid-related deaths.

In many real-life data sets, the frequency of the first digit does not follow a uniform distribution. In addition, the first digit tends to be small, and so the probability of occurrence of the number 1 in the first position is 30.1%, while the probability of that number being 9 is 4.5% [16]. Then, BL empirically discovered the pattern for the frequency distribution of first digits for many collections of numbers. Therefore, a good approach to analyse a potential manipulation of data recording is to check for the validity of BL.

In this line, BL has been used to detect fraud or errors in data recording in a wide variety of areas. For instance,[17][18][19][20] use BL as a tool for fraud detection in the insurance industry or in other commercial trades. Other authors have also applied BL to test the veracity of scientific data [21] or data with public health relevance, as is the case of the work of Stoerk [22] who focuses on the recording of air quality data in Beijing.

A very recent research already uses BL as a tool to assess the effectiveness of the control interventions in flattening the curve and the spread of COVID [16]. Results from this work suggest that BL it is a suitable approach to analyses COVID-related trends and potential manipulations or registration errors in the number of cases and deaths, because of the data characteristics. When numbers follow an exponential distribution, as is the case of the number of COVID infections or deaths, it has been demonstrated that they follow Benford's Law. This audit methodology has also already been used for other infectious diseases. For instance, in Uruguay [23], authors use the BL to evaluate the dengue case reporting system.

As a result, BL may be a useful tool for testing the reliability of data provided by different countries or indeed regions within the same country. As shows by [24]. Authors applies BL to test the reliability of COVID-19 death-case reporting in countries with authoritarian regimes. They concluded that countries with democratic regimes do conform better to the BL than the authoritarian ones regarding COVID-19 death-case figures reported.

In the case of Spain, the decentralization of some competences, such as health, has caused some difficulties for the central government to collect homogeneous information on the number of COVID-related cases and deaths. Especially during the first months of the pandemic, where there was a continuous readjustment of COVID figures provided by the different regions. Within this context, it is essential to set a common tool to assess the validity of COVID data recorded across the country.

A good clinical recording system is at the core of good health planning at all levels of management, from a single hospital to a nation-wide level. Likewise, in the case of COVID crisis management, the correct recording of cases, fatality or incidence rates is essential to reduce inefficiencies in the field of health management [25].

An incorrect registration or updating of data can cause important inefficiencies both in the allocation of resources and in enactment of control measures. Within this context, authors such as Koch et al. 2020[26], already use BL to assess the veracity of COVID data in China.

Italy was the first European country where the pandemic had a strong impact. Within the Italian context, some authors have stressed the importance of correctly interpreting fatality rate data and discussing the correct recording of deaths to optimize a health policy [27] as well as to analyse the different impacts of the pandemic across the regions of a country [28]

In Spain, the basic providers of health information and data are the Ministry of Health, the Health Departments (Consejerías de Salud) and the Public Health Departments of the regions (known as Autonomous Communities (ACs)). Previously, health care beneficiaries and standards were defined centrally, but since 2002, when the decentralization process for health care responsibilities concluded, the responsibility for services delivery and funding has been devolved to the 17 ACs [29]. This organizational model has sometimes led to a lack of homogeneity in the registration of some health phenomena, as in the case of Covid-19. Specifically, during the first wave of the pandemic, there has been great controversy over the lack of homogeneity of criteria when counting COVID cases and deaths in the different ACs. The Ministry of Health itself did not always had updated data at any given time for all regions~~.~~ The Ministerial Order BOE-A-2020-3953 from March 21, 2020 [30], established that the ACs must provide the central government with the aggregated COVID data on a weekly basis. All the ACs provided their data by filling the related template which included information such as number of confirmed cases, number of hospitalised cases and number of deaths. However, the protocol for recording this information may vary across ACs. Specially, the recording of the number of COVID-related deaths is particularly sensitive, as the cause of death is not always clear, especially among the elderly population or those with chronic comorbidities. In addition, many deaths may be recorded as “unknown cause”. All these elements can lead to divergences from the actual situation.

The issue of both the updating of tools and protocols for recording health information, as well as the establishment of homogeneous health information systems among ACs are topics already addressed in the literature [31][32] This discussion takes on particular relevance in a public health crisis such as the current one.

It is essential that health administrations base their policy and management decisions on reliable data and objective criteria in order to avoid inefficiencies. Therefore, the establishment of common mechanisms to detect possible errors and data deviations is basic to help provide a map of the pandemic situation at the national level, as well as identify the specificities of regional figures.

In this line, this paper proposes using BL as a methodological approach in COVID crisis management to monitor the registration of deaths. Specifically, the aim of this paper is to provide an objective tool capable of detecting possible errors or deviations from the expected trend in the recording of the number of COVID deaths per day in Spain. We focus our analysis on the Spanish case, so the methodology proposed in this paper can be used as a guide to monitor the reliability of COVID-related figures for the health administration both at a regional and central level. In addition, this analysis allows us to look into the pandemic’s impact on the different regions in terms of number of deaths.

This paper is organized as follow: Section 2 describes the methods and empirical procedures. In section 3 data and sources are presented. Section 4, captures the results. And finally, in section 5 we discuss our conclusions.

2. Methods and Empirical Procedure

2.1. Description of Benford’s Law.

Benford’s Law is a mathematical rule conjecture that most sets of numbers verify. It is more frequent for an arbitrary set to verify BL than not. In other words, it is easier to enumerate the set of data that does not verify BL than the set of data that verifies the property[33]. This mathematical law has been used in different scientific fields such as physics [34] or economics [35]. One of its most used applications is to detect tax fraud[36][37].
Thus, BL establishes the (hypothetical) distribution of the digits of the same sequence of numbers. The distribution depends on the position of the digit or digits considered. Therefore, according to BL, the significant digit distribution does not follow a uniform distribution, they are skewed toward the smaller numbers.
The expression of BL states that the probability that the first digit of a magnitude is a specific figure "n" is provided by equation 1:

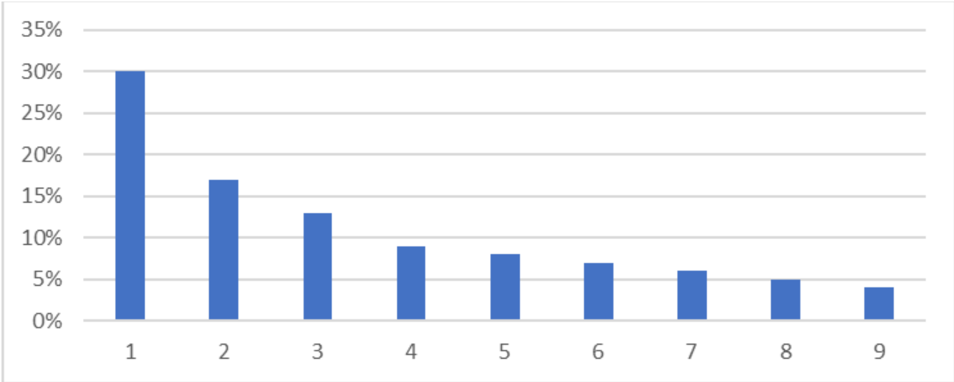
$P(n) = log10n + 1 - log10\ \ n = log10n + 1n = log101 + 1n$ (1)

with n= k

Where $P_{(n)}$ is the probability of a number having the first non-zero digit n.

According with expression 1, BL provides the theoretical proportion for each of the digits from 1 to 9 to be first significant digit. Figure 1 shows the distribution of first significant digit predicted by BL.

Figure 1. Frequency distribution of the first digit according the BL.



Digit	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Proportion	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

An extension of the formula and generalized to any set of "n" first digits is provided by equation 2:

$$Pn1n2 \dots nk = log10\ \ 1 + \ \ 1n1\ \ n2 \dots .nk \qquad (2)$$

Where $P(n1n2 \dots nk)$ is the probability of a set of numbers having the first position.

Therefore, we can obtain the probability of occurrence of each digit according to its position. Thus, for instance, the probability of the first digit is 1 is:

$$log101 + 11 = 0{,}301 * 100\% = 30{,}1\%$$

The probability of the first two digits of the pair 37, is:

$$log101 + 137 = 0{,}0116 * 100\% = 1{,}16\%$$

The probability of the first three digits being the triad 280 is:

$$log101 + 137 = 0{,}0015 * 100\% = 0{,}15\%\ \ [18]$$

2.2. Chi-square test

As a goodness of fit of the analysis, we used the χ^2 (Chi-square) test. Through the χ^2 test we tested whether the n entries in a set of data are compatible with the BL (equation 2). That is to say, we test the null hypothesis for the first digit probabilities, $pi = Pr\ \ (D1 = i)$. Therefore, we are testing the hypothesis specified below [38].

Considering $F \equiv q1, q2, \dots q9$ as a discrete distribution of probability, and this probability is $qi = log101 + 1i$. Also, qi verify that $qi\ \ \geq 0\ \ for\ \ i = 1,2, \dots .9;\ \ \ i = 19qi = 1$. Then, we are testing the following hypothesis:

$$H0 : \{pi\}i = 19 \ follows \ F,$$

$$H1 : \{pi\}i = 19 \ do \ not \ follows \ F$$

The chi-square statistics provide a measure of the distance between real data and Benford distribution. Therefore, the highest chi-square value is, the larger the deviation between real data and Benford distribution [16].

Then, with the chi-square test we are testing the null hypothesis (H_0) that the first digit is the same as expected on BL basis. Hence, the chi-square test points to those sets of numbers in which we must look into the possible causes of noncompliance with BL, and are those for which we can reject the H_0.

2.3. Sensitivity analysis steps:

In order to very results provided by the χ^2 test, we designed a sensitivity analysis following the steps detailed below. As the observed figures are random and that randomization depends on chance, we run this sensitivity analysis to validate results.

Step 1. First, the series of observed values were modifying by random perturbations assuming that:

(i) such a disturbance is unintentional;

(ii) the applied perturbations are independent of each other;

(iii) the perturbation size varies over a 20% range, and within that range any possible outcome is equally likely. This assumption implies to consider the uniform probability distribution taking values within the interval [-0.1, +0.1]. Denoted as U [-0.1 ; +0.1].

Step 2. From the observed fatality rate of a specific AC, an arbitrarily large set of alternative series with a generated perturbation is obtained through a Montecarlo simulation. Specifically, we generated 1000 replications for each series. Therefore, given the observed series $\{x_1^{obs}, x_2^{obs}, x_3^{obs}, \ldots x_n^{obs}\}$, we obtain the ith series modified as $x_k^i = x_k^{obs} \cdot \left(1 + u_k^i\right),\ k = 1, \ldots, n$ where $i = 1, \ldots 1000,\ \{u_k^i\}_{k=1}^n$ are n values obtained by simulation from the distribution U [-0.1 ; +0.1].

Step 3. The BL test is applied to each series $i_0,\ \{u_k^{i_0}\}_{k=1}^n$ generated synthetically, by calculating the statistics distance of χ^2 and the p-value test for that series. Then, we obtain 1000 synthetic series, with their 1000 p-values $\{p_1^i, p_2^i, \ldots, p_n^i\}_{i=1}^{1000}$ and their 1000 χ^2 distances.

Summarizing, as a result of the previous steps, given a observed series, $\{xkobs\}k = 1n$, we can generate the 1000 synthetic simulations $\{\{x^i{}_k\}k = 1n\ \}\}i = 11000$ and the 1000 p-values $\{p_1^i, p_2^i, \ldots, p_n^i\}_{i=1}^{1000}$, then, we calculate both the average p-value, p ,and the average distance χ^2, in addition we calculate quantiles of α-order for those p-values, $q\alpha$.

Step 4. From $q\alpha$ it possible to get the equivalent to a confidence interval that allow validation about the decision of the BL fulfillment with the observed data. That is to say, we goal is to check if the decision for observed data it can be keep for data with perturbations. Then, we set a $q1 - \alpha$ value an take a decision according the scheme displayed in Table 2:

Table 2: Criteria to validate decision about the BL fulfillment

Decision for observed data	*If $q_{0.95} > \alpha$ for data with perturbations*	*If $q_{0.95} < \alpha$ for data with perturbations*
H_0 Fail to reject	We can keep the decision	We cannot keep the decision
H_0 Reject	We cannot keep the decision	We can keep the decision

3. Data and source

For the analysis carried out in this paper, we consider data from Spanish population in the period from March to June 2020, the first important period of the pandemic (first wave). Even if we can use others periods, as the aim of this work is to compare the impact of the pandemic through BL by region in Spain, this period keeps a homogeneous characteristic for all ACs in terms of lockdown measures. In sequel periods, known as second and third waves, the regions had adopted different control measures that can affect the pandemic trend and therefore, the resulting figures wouldn't be so comparable. In addition, the analysed period matched the exponential growth phase of the pandemic.

The data analysed in this work is the number of deaths per day recorded by the different ACs during the period under study. The data are downloaded from "datadista Git-Hub repository" [40] and the information was contrasted with the data available in the official website of the Spanish Government's Department of Health (Ministerio de Sanidad).

There are a few errors in the transcription of the government data. Some errors are just changes in the figures due to transcription errors or changes in the cause of death for some of the deceased. These types of errors in the information were verified and treated by Datadista.

In addition to testing those regions that follow BL in recording the number of daily deaths, we will use a ranking of the fatality rate for the different Spanish regions as a reference. That is to say, we observe whether those regions that deviate from BL are in the first or last positions of the fatality rate ranking. This comparison suggests the potential causes of the deviation from BL. Such causes may relate to errors in the records or low quality data recording, a stronger impact of the pandemic than in other regions, or better real data than in other regions, i.e. regions with few or no deaths per day.

As discussed in the introduction, BL has been used for a wide variety of phenomena due to its versatility. For BL to be applied, the following recommendations must be met: The data must follow a geometrical sequence and it must not contain a theoretical maximum or minimum. In addition, BL is independent of the scale of measurement on which the data are being processed.

Therefore, according to the data characteristics described above, BL is a suitable methodology to achieve the objective proposed in this paper; that is to say, we are going to test if the number of daily COVID-related deaths registered by the different ACs, follows or not BL. Hence, it is necessary to focus the analysis on those regions that show a deviation from what is expected according to BL in the daily death register, in order to identify the possible causes of this deviation. As explained before, we propose to use BL as an auditing guide to identify possible errors or deviations in the COVID figures recorded by the ACs, in order to set a common tool of reliability data assessment at any level of the health administration hierarchy.

4.Results

The summary of the main results of our analysis are displayed in Table 1, where we compare the goodness of fit to BL with the fatality rate in order to identify recording errors or deviations from the expected trend in the daily death figures. Table 1 includes both results of compliance with BL (χ^2 test) and the COVID fatality rate by ACs. Among those regions for which we reject the hypothesis that BL is fulfilled, two kinds of interpretation can be offered. The majority of the ACs for which we reject the H_0, are ranked at the top of the fatality rate ranking (above the Spanish rate). Then, in these cases, the explanation for the deviation from the BL relate to mistakes in the registration of the daily number of deaths, or an uncontrolled pandemic crisis providing skyrocketing figures. The region with the largest χ^2 value is Catalonia, and that is to say, the one with the largest deviation from what was expected according to BL. In fact, Catalonia rectified up to about 20% of the data initially supplied to the Ministry, confirming that there had been errors in the registry or in the counting of cases.

The exception within this group of regions are the cases of Galicia and Extremadura, for which we reject the null hypothesis but they both show low fatality rates. In these cases, the number of daily deaths does not comply with BL probably because of the low number of daily deaths recorded; here the number of deaths recorded daily is only 1 or 2.

Therefore, the fact that the number of daily deaths does not follow BL may be due to either the incorrect recording of cases (daily deaths in this case), or to a favorable evolution of the pandemic within the region. Thus, in some cases, a region may show a good outcome (a number of daily deaths that remains low over time). Thus, as it records few deaths per day, the phenomenon does not follow the exponential trend described by BL. However, in other cases, non-compliance with BL indicates those regions that may have extremely high daily death rates.

For those ACs for which we cannot reject the null hypothesis, in other words, where the number of daily deaths follows BL, lower fatality rates are observed. That is, although the figures follow an exponential trend, they follow the expected trend, and so this can be considered as an indicator that the data have been correctly recorded. However, we can find one exception. This is the case of País Vasco, which is in the middle of the fatality rate ranking but the number of daily deaths follows BL. In this case, although the COVID fatality rate is high, slightly higher than the Spanish rate, we can say that this information is reliable and the daily number of deaths is well recorded.

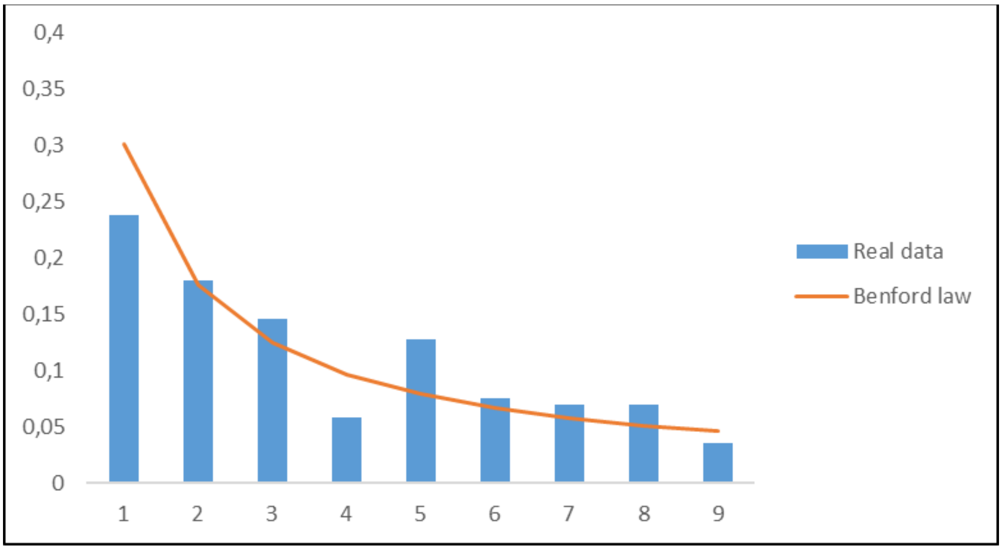
Table 1. Regions ranked by χ2 test.

ACs code	Autonomous communities (ACs)	χ^2 value Estimator	χ^2 Test p-value	Fatality rate (x10[5])	
9	Cataluña	291.947	0.000293***	74.5	7
15	Navarra	217.510	0.005398***	81.5	6
12	**Galicia**	**214.966**	**0.005938*** **	**23.4**	**14**
13	Madrid	195.582	0.012143**	127.4	2
17	La Rioja	178.412	0.022448**	116.2	4
7	Castilla y León	177.992	0.022782**	117.2	3
11	**Extremadura**	**169.880**	**0.030233***	**49.2**	**10**
8	Castilla La Mancha	164.449	0.036437*	143.4	1
2	Aragón	139.270	0.083687*	82.2	5
0	*Spain*	*128.710*	*0.116364*	*60.9*	*9*
1	Andalucía	118.706	0.157069	17.3	16
6	Cantabria	114.593	0.177005	36.0	11
10	C. Valenciana	98.378	0.276588	29.0	13
16	**País Vasco**	**93.121**	**0.316654**	**70.9**	**8**
4	Baleares	55.347	0.699181	19.8	15
3	Asturias	54.368	0.710025	32.8	12
14	Murcia	34.197	0.905324	10.0	17

Reject H_0 at levels *5%,3%, *** 1%.**

In Figure 2, we can graphically observe the frequencies of recording actual daily deaths versus the trend defined by BL for the figures at a national level. Specifically, bars represent the frequency distribution of the first digit of the number of COVID deaths per day in Spain during the period under study (from March to June 2020) while the line represent the theoretical distribution of the BL.

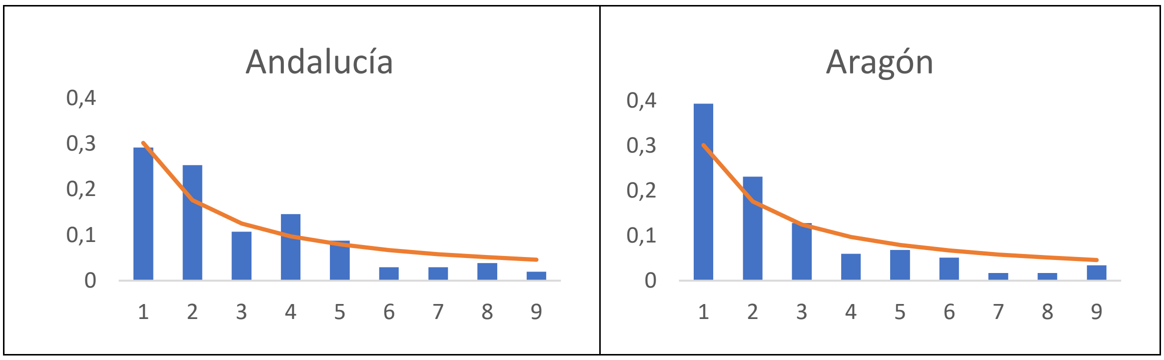
Figure 2. Frequency distribution of the first digit of the number of deaths per day by COVID in Spain.

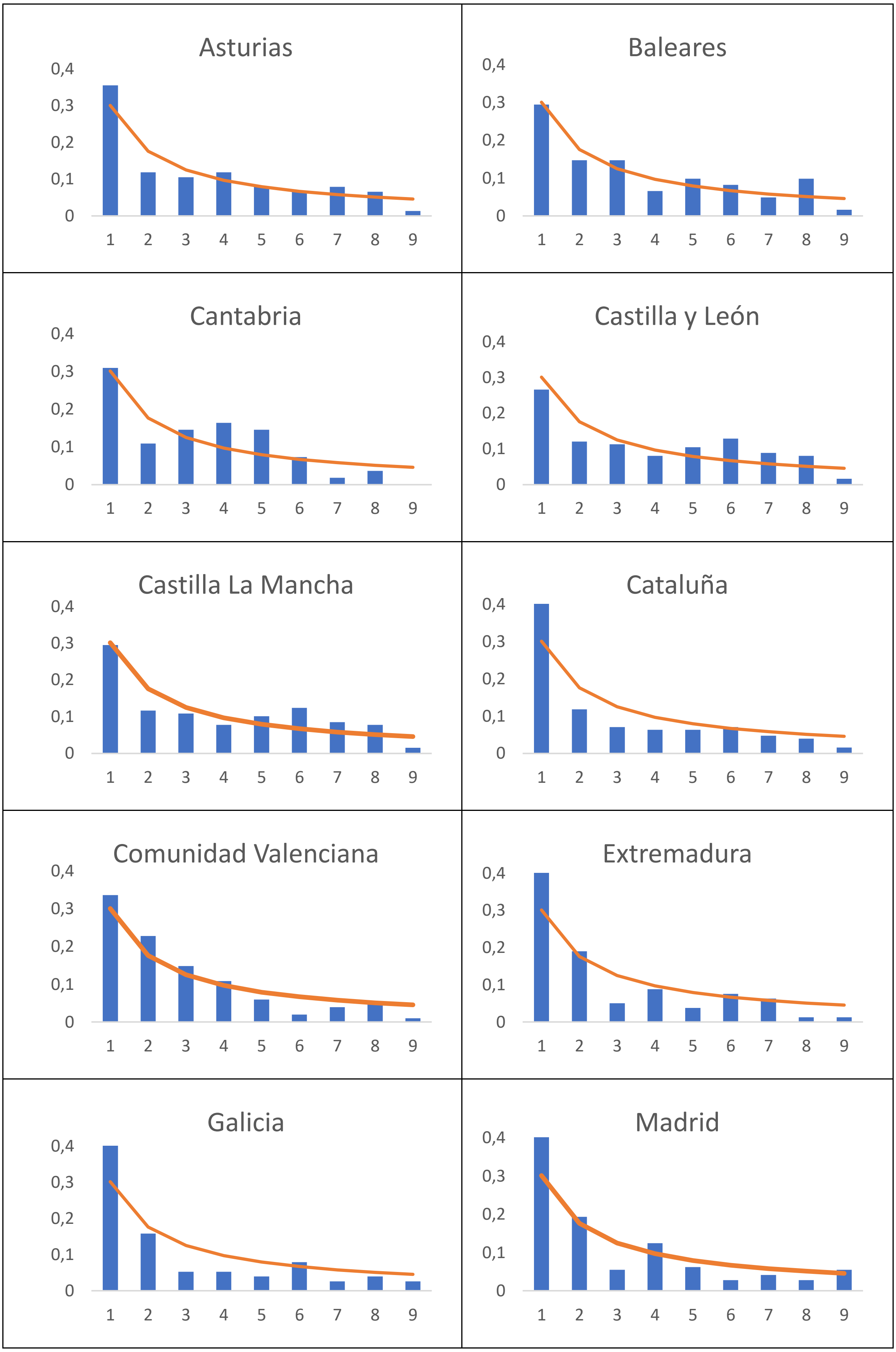


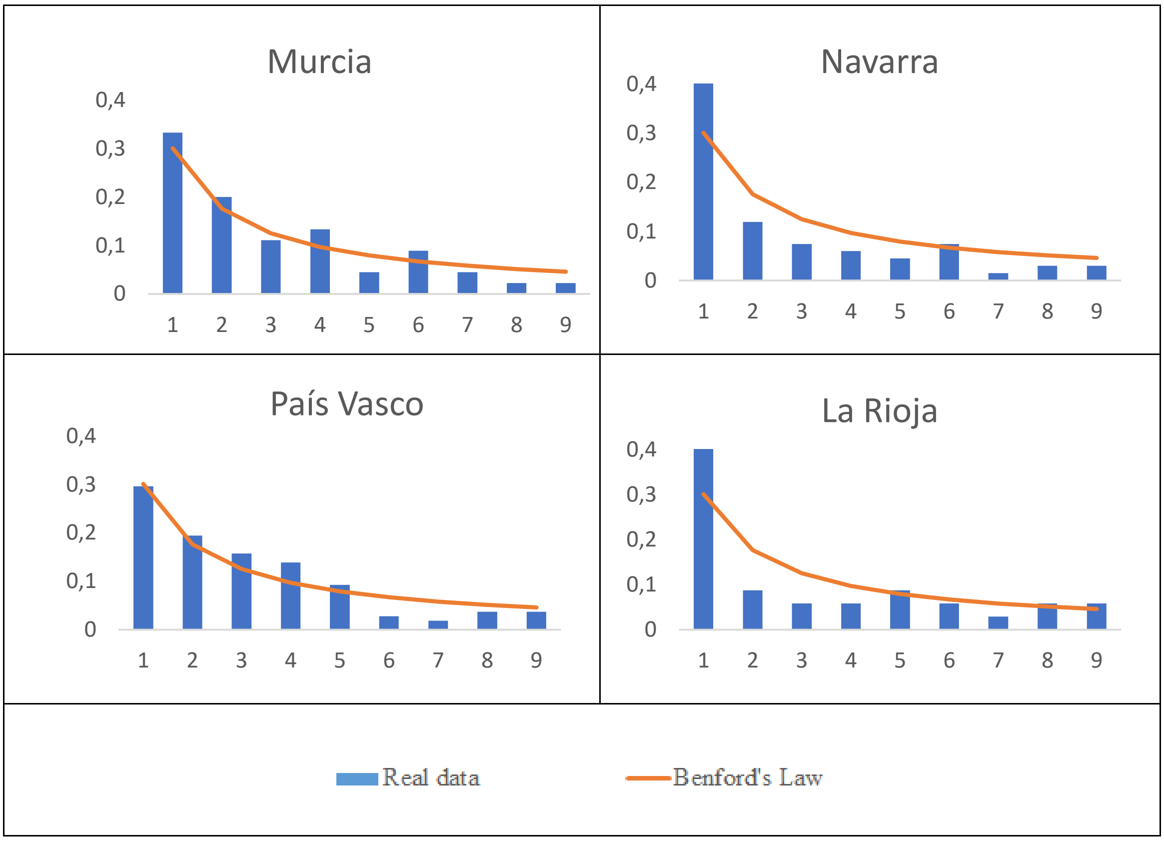
As we can observe in Figure 2, the frequency of daily deaths recorded at a national level, are quite in line with the trend described by BL.

As well, Figure 3 shows the graphic comparison for each AC, and as we can observe that regions such as Cataluña, Navarra or La Rioja, show a divergence in the frequency of the first digits from the trend defined by BL. As shown in Table 1, any of this regions fits the BL. While regions such as Comunidad Valenciana or Murcia, have a frequency of the first figures that is very similar to that described by BL. These are regions for which the number of daily deaths fulfil BL**.**

Figure 3. Frequency distribution of the first digit of the number of COVID deaths per day by AA.CC.

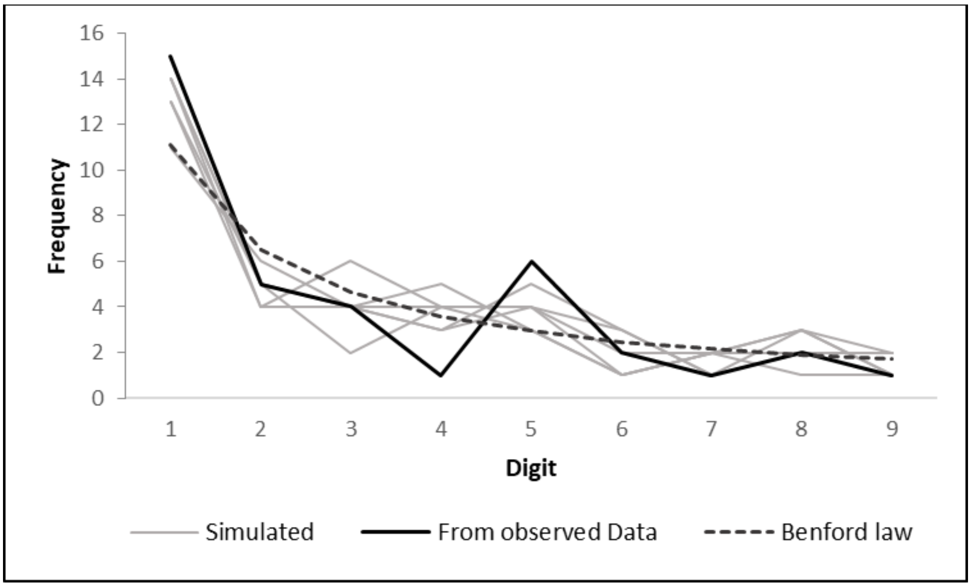



Asturias
Baleares
Cantabria
Castilla y León
Castilla La Mancha
Cataluña
Comunidad Valenciana
Extremadura
Galicia
Madrid
0,4
0,3
0,2
0,1
0
1 2 3 4 5 6 7 8 9



Then, as described within the section 2, in order to verify the reliability of results showing in Table 1 about whether or not the ACs data follows the BL, we run a the sensitivity analysis. Figure 4 graphically show a set of simulations modifying the observed values by random perturbations. Specifically, the figure display the Benford count (coloured lines) of the Spain observed data series (black line) and the Benford curve taking as a reference (dotted line), for the considered period.

Figure 4. Benford's counting. Observed series and simulations.



The sensitivity analysis allowed us to verify results provided by the chi-square test. Therefore, following the criteria detailed in Table 2, results of the sensitivity analysis for each AC are shown in Table 3[4].

Table 3 shown the sensitivity analysis results for each AC.

AC	*Inicial decison (for observed data)*	$q_{95\%}$	*Final decision (for data with perturbations)*
Cataluña	Rejection	0,00013	Rejection
Navarra	Rejection	0,00610	Rejection
Madrid	Rejection	0,00058	Rejection
La Rioja	Rejection	0,00633	Rejection
Galicia	Rejection	0,00583	Rejection
Castilla León	Rejection	0,04743	Rejection
Spain	Fail to reject	0,67138	Fail to reject
C. Valenciana	Fail to reject	0,34756	Fail to reject
Andalucía	Fail to reject	0,47747	Fail to reject
Cantabria	Fail to reject	0,28642	Fail to reject
Baleares	Fail to reject	0,33385	Fail to reject

As display in Table 3, we keep the initial decision about BL fulfilment for all ACs. The sensitivity analysis therefore, verifies and confirm the initial decision taken using the χ^2 test.

[4] Table 3 display a summary of the sensitivity analysis results. For all ACs that are not shown in the table, the decision was keep for data with perturbations.

In summary, we can use BL fit test as an indicator of the reliability of the data recorded in terms of daily COVID deaths for the different regions. Furthermore, comparing this result with the fatality rate can help to interpret the specificities of each AC and identify whether we are dealing with an error in the data recording or a particularity in the trend of the pandemic for that specific region.

Actually, deviations and errors in the recording of data provided by the different ACs, occurred above all in the first months of the pandemic. Subsequently, the records have improved and refined. In fact, if we apply the chi-square test, for the second, third and fourth waves, the number of ACs that meet BL increases.

Our analyses reveal diversity in the profile across the different ACs, and points out those cases with greater deviations and which, therefore, require special attention as to the possible causes of such divergence.

In a country with the characteristics of Spain in terms of health organization, it is crucial to set common tools for the verification of data, in order to have reliable and homogeneous information available throughout the Spanish territory to serve as a basis for public health decisions.

5.Discussion

BL has already been used with different purposes within the COVID-19 pandemic context [41][16][26]. However, the proposal of this paper is to use BL as a health crisis management tool to audit the correctness of the recording of COVID figures for the different ACs and to identify deviations from the expected trend. Specifically, we use the BL for detecting possible errors in the accounting for COVID deaths within the context of the Spanish ACs.

The first wave of the COVID pandemic has generated a great deal of controversy in Spain due to the discrepancy in the figures regarding the number of daily deaths provided by the different ACs. Differences both in the magnitudes of figures and the rate at which they are updated, have created doubts about their reliability.

To understand the statistical deficiencies detected in the records of deaths due to COVID, several situations that have influenced this problem must be taken into account. In the first place, since it was an emerging and unknown disease until 2019, public administrations have had to face a great problem such as the lack of homogenization and consensus at the time of registering COVID-related deaths. Likewise, the death registration system used in each region may have presented logical deficiencies in its operational dynamics in the face of a totally unforeseen situation. Thus, the clinical and diagnostic criteria to confirm a death due to COVID could differ from one administration to another. For example, especially at the beginning of the pandemic, it was very difficult to establish whether a death was due to COVID-19 or not since it was not possible to carry out confirmatory diagnostic tests on all suspected cases, nor was it possible to carry out diagnostic tests that would allow the evolution of the disease to be followed. This was especially evident in the social health centers, where, due to the lack of healthcare resources, it was very difficult to carry out an optimal follow-up of the disease. Furthermore, both under-registration and over-registration of deaths generate a distortion in the information that prevents a correct planning and management of the health resources to contain and control the pandemic. In addition, it also creates a situation of mistrust and misinformation among public opinion.

However, having reliable and homogenous information on the state of the pandemic throughout the country is crucial for good pandemic management. For this reason, the development of models and indicators that serve as guidelines for the correct recording of data, is a key ally for health administrations at all management levels.

This article shows how BL can be taken as a reference for the control of the registration of the number of daily COVID-related deaths. Specifically, we propose to test the hypothesis that the frequency of the daily number of deaths follows BL through a chi-square test as an audit test for the data reliability. Non-compliance with the BL will point to those regions that may have errors in the recording of COVID deaths. Then, it tells us where the focus should be placed to analyse the possible causes of these deviations from the expected trend. While accepting the hypothesis of BL, compliance is a good indicator of the reliability of the data.

In addition, in order to validate our results, we run a sensitivity analysis that allow us to confirm the decisions about the hypothesis of the BL fulfilment. In fact, the sensitivity test yields the same results as using the chi-square test, hence we keep the decision on those ACs that do not follow the BL.

As already mentioned, we focused our analysis on the Spanish case, where very significant differences in COVID figures have been found during the current pandemic among the different ACs. According to our results, while we accept the hypothesis for the aggregate data for Spain, in the case of the different ACs we observe some discrepancies, since not all the ACs fulfil BL.

Once we have identified those ACs whose recording of daily deaths does not comply with BL, we compare it with the ranking of the fatality rate as a reference to find possible causes for the deviation from the expected trend. In fact, the majority of ACs for which we reject the H_0, are ranked at the top of the fatality rate ranking (above the Spanish rate). In these cases, the explanation for the deviation from BL relate to mistakes in the registration of the daily number of deaths (this is the case of Cataluña, Navarra or Madrid among others). These mistakes in the data registration may be due to delays in the information reporting (events on a specific day may be recorded afterwards), human errors, differences on counting or recording criteria, among others. In fact, anomalous figures such as the case of Catalonia has been often reported in the press. In this AC, one of the main recording errors was the delay in reporting and recording information (sometimes attributing to a single day death cases from previous days) [39].

However, there are two ACs (Galicia and Extremadura) that do not fulfil BL but present a fatality rate below the average. In these cases, the non-accomplishment of BL is due to the low number of daily deaths. When on the majority of the days there are just 1, 2 or 0 deaths, this set of numbers doesn't grow exponentially and therefore, the probability distribution for the leading digit doesn't follow BL.

For those ACs for which we cannot reject the null hypothesis, in other words, for which the number of daily deaths follows BL, this can be considered as an indicator of data reliability.

Summarizing, by comparing results of BL hypothesis test with the fatality rate we can better interpret the results. Thus, we obtain two possible explanations for those ACs that do not conform to BL, either there is an error in the recording of the data, or the pandemic is following a positive evolution and the number of deaths per day is very low, therefore, the phenomenon does not follow the exponential trend described by BL.

In this way, the BL can be used as an auditing tool in the recording of COVID data, specifically, for the number of daily deaths, and therefore can help to provide reliable data to health administrations in their different management levels. Thus, BL can be used as an epidemiological tool to generate information on the precision in the registration of notified cases and number of daily deaths for the evaluation of different intervention strategies [25]. Specially in context where health competencies are decentralized, as is the case of Spain, the coordination among CAs and the provision of homogenous information are crucial for public health matters management. Hence, this coordination implies to set common tools and procedures for data auditing at a national level.

As mentioned above, the recording of the number of daily deaths is particularly sensitive as it presents medical and administrative difficulties. Moreover, in general terms, few administrations were ready to deal with a pandemic of the magnitude of COVID-19. However, professionals in the sector have reacted quickly and efficiently and have adapted processes and protocols to the new health reality, improving and refining the correct recording of data. In fact, if we were to carry out the same analysis presented in this paper with data from the third or fourth wave, we would obtain that the number of daily deaths for practically all the ACs comply with BL. Then, as the recording of the number of daily deaths has improved and the data has become more reliable, the more they follow BL.

We believe that this paper can be useful to set common tools for the verification of data, in order to have reliable and homogeneous information available throughout the Spanish territory to serve as a basis for public health decisions.

References

[1] A. Anirudh, "Mathematical modeling and the transmission dynamics in predicting the Covid-19 - What next in combating the pandemic," *Infect. Dis. Model.*, vol. 5, pp. 366–374, Jan. 2020, doi: 10.1016/j.idm.2020.06.002.

[2] Y. Mohamadou, A. Halidou, and P. T. Kapen, "A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19," *Appl. Intell.*, vol. 50, no. 11, pp. 3913–3925, 2020, doi: 10.1007/s10489-020-01770-9.

[3] F. Ahmad, S. N. Almuayqil, M. Humayun, S. Naseem, W. A. Khan, and K. Junaid, "Prediction of COVID-19 Cases Using Machine Learning for Effective Public Health Management," *Computers, Materials \& Continua* , vol. 66, no. 3. 2021, doi: 10.32604/cmc.2021.013067.

[4] D. Yadav, H. Maheshwari, and U. Chandra, "Outbreak prediction of covid-19 in most susceptible countries," *Glob. J. Environ. Sci. Manag.*, vol. 6, no. Special Issue (Covid-19), pp. 11–20, 2020, doi: 10.22034/GJESM.2019.06.SI.02.

[5] S. Li *et al.*, "Development and external evaluation of predictions models for mortality of COVID-19 patients using machine learning method," *Neural Comput. Appl.*, 2021, doi: 10.1007/s00521-020-05592-1.

[6] D. Giuliani, M. M. Dickson, G. Espa, and F. Santi, "Modelling and predicting the spatio-temporal spread of COVID-19 in Italy," *BMC Infect. Dis.*, vol. 20, no. 1, p. 700, 2020, doi: 10.1186/s12879-020-05415-7.

[7] D. Alboaneen, B. Pranggono, D. Alshammari, N. Alqahtani, and R. Alyaffer, "Predicting the Epidemiological Outbreak of the Coronavirus Disease 2019 (COVID-19) in Saudi Arabia," *Int. J. Environ. Res. Public Health*, vol. 17, p. 4568, Jun. 2020, doi: 10.3390/ijerph17124568.

[8] A. Alsayed, H. Sadir, R. Kamil, and H. Sari, "Prediction of Epidemic Peak and Infected Cases for COVID-19 Disease in Malaysia, 2020," *International Journal of Environmental Research and Public Health* , vol. 17, no. 11. 2020, doi: 10.3390/ijerph17114076.

[9] L. Jia, K. Li, Y. Jiang, X. Guo, and T. Zhao, "Prediction and analysis of Coronavirus Disease 2019."

[10] L. Qin *et al.*, "Prediction of Number of Cases of 2019 Novel Coronavirus (COVID-19) Using Social Media Search Index.," *Int. J. Environ. Res. Public Health*, vol. 17, no. 7, Mar. 2020, doi: 10.3390/ijerph17072365.

[11] S. M. Ayyoubzadeh, S. M. Ayyoubzadeh, H. Zahedi, M. Ahmadi, and S. R Niakan Kalhori, "Predicting COVID-19 Incidence Through Analysis of Google Trends Data in Iran: Data Mining and Deep Learning Pilot Study," *JMIR public Heal. Surveill.*, vol. 6, no. 2, pp. e18828–e18828, Apr. 2020, doi: 10.2196/18828.

[12] K. Ganasegeran, A. S. H. Ch'ng, and I. Looi, "What Is the Estimated COVID-19 Reproduction Number and the Proportion of the Population That Needs to Be Immunized to Achieve Herd Immunity in Malaysia? A Mathematical Epidemiology Synthesis," *COVID*, vol. 1, no. 1, pp. 13–19, 2021, doi: 10.3390/covid1010003.

[13] S. W. Park, D. Champredon, J. S. Weitz, and J. Dushoff, "A practical generation-interval-based approach to inferring the strength of epidemics from their speed.," *Epidemics*, vol. 27, pp. 12–18, Jun. 2019, doi: 10.1016/j.epidem.2018.12.002.

[14] M. Caballer-Tarazona, I. Moya-Clemente, D. Vivas-Consuelo, and I. Barrachina-Martínez, "A model to measure the efficiency of hospital performance," *Math. Comput. Model.*, vol. 52, no. 7, pp. 1095–1102, 2010, doi: https://doi.org/10.1016/j.mcm.2010.03.006.

[15] F. Benford, "The Law of Anomalous Numbers," *Proc. Am. Philos. Soc.*, vol. 78, no. 4, pp. 551–572, Jun. 1938, [Online]. Available: http://www.jstor.org/stable/984802.

[16] K.-B. Lee, S. Han, and Y. Jeong, "COVID-19, flattening the curve, and Benford's law," *Physica A*, vol. 559, p. 125090, Dec. 2020, doi: 10.1016/j.physa.2020.125090.

[17] M. Maher and M. Akers, "Using Benfords Law To Detect Fraud In The Insurance Industry," *Account. Fac. Res. Publ.*, vol. 1, Mar. 2011, doi: 10.19030/iber.v1i7.3951.

[18] P. M. Cabeza García, "Aplicación de la ley de Benford en la detección de fraudes ," *Revista Universidad y Sociedad* , vol. 11.

scielocu , pp. 421–427, 2019.

[19] A. Cerioli, L. Barabesi, A. Cerasa, M. Menegatti, and D. Perrotta, "Newcomb{\textendash}Benford law and the detection of frauds in international trade," *Proc. Natl. Acad. Sci.*, vol. 116, no. 1, pp. 106–115, 2019, doi: 10.1073/pnas.1806617115.

[20] L. Barabesi, A. Cerasa, A. Cerioli, and D. Perrotta, "Goodness-of-Fit Testing for the Newcomb-Benford Law With Application to the Detection of Customs Fraud," *J. Bus. Econ. Stat.*, vol. 36, no. 2, pp. 346–358, Apr. 2018, doi: 10.1080/07350015.2016.1172014.

[21] A. Diekmann, "Not the First Digit! Using Benford's Law to Detect Fraudulent Scientif ic Data," *J. Appl. Stat.*, vol. 34, no. 3, pp. 321–329, 2007, doi: 10.1080/02664760601004940.

[22] T. Stoerk, "Statistical corruption in Beijing's air quality data has likely ended in 2012," *Atmos. Environ.*, vol. 127, pp. 365–371, 2016, doi: https://doi.org/10.1016/j.atmosenv.2015.12.055.

[23] M. Gómez-Camponovo, J. Moreno, Á. J. Idrovo, M. Páez, and M. Achkar, "[Monitoring the Paraguayan epidemiological dengue surveillance system (2009-2011) using Benford's law].," *Biomedica*, vol. 36, no. 4, pp. 583–592, Dec. 2016, doi: 10.7705/biomedica.v36i4.2731.

[24] L. Burlac and N. Giannakis, "Benford's Law: Analysis of the trustworthiness of COVID-19 reporting in the context of different political regimes," School of Education, Culture and Communication, Mälardalen University, 2021.

[25] Y. D. Tasri and E. S. Tasri, "Improving clinical records: their role in decision-making and healthcare management – COVID-19 perspectives," *Int. J. Healthc. Manag.*, vol. 13, no. 4, pp. 325–336, Oct. 2020, doi: 10.1080/20479700.2020.1803623.

[26] C. Koch and K. Okamura, "Benford's Law and COVID-19 reporting," *Econ. Lett.*, vol. 196, p. 109573, 2020, doi: https://doi.org/10.1016/j.econlet.2020.109573.

[27] M. Iosa, S. Paolucci, and G. Morone, "Covid-19: A Dynamic Analysis of Fatality Risk in Italy," *Front. Med.*, vol. 7, p. 185, 2020, doi: 10.3389/fmed.2020.00185.

[28] B. Armocida, B. Formenti, S. Ussai, F. Palestra, and E. Missoni, "The Italian health system and the COVID-19 challenge," *Lancet Public Heal.*, vol. 5, no. 5, p. e253, May 2020, doi: 10.1016/S2468-2667(20)30074-8.

[29] M. Caballer-Tarazona, A. Clemente-Collado, and D. Vivas-Consuelo, "A cost and performance comparison of Public Private Partnership and public hospitals in Spain," *Health Econ. Rev.*, vol. 6, no. 1, p. 17, Dec. 2016, doi: 10.1186/s13561-016-0095-5.

[30] M. de Sanidad, "BOE-A-2020-3953. «BOE» núm. 78, de 21 de marzo de 2020, páginas 26505 a 26510 (6 págs.)," 2020. https://www.boe.es/diario_boe/txt.php?id=BOE-A-2020-3953.

[31] B. Moreno Küstner, "La información sanitaria se enreda en la informática," *Gac. Sanit.*, vol. 25, no. 4, pp. 343–344, 2011, doi: 10.1016/j.gaceta.2011.02.014.

[32] Suay, "Los Sistemas de información sanitaria en el marco de un Sistema Nacional de Salud descentralizado," *Arbor*, vol. 180, no. 710, pp. 327-342., 2005.

[33] J. Kazemitabar and J. Kazemitabar, "Measuring the conformity of distributions to Benford's law," *Commun. Stat. - Theory Methods*, vol. 49, no. 14, pp. 3530–3536, Jul. 2020, doi: 10.1080/03610926.2019.1590599.

[34] M. Sambridge, H. Tkalčić, and A. Jackson, "Benford's law in the natural sciences," *Geophys. Res. Lett.*, vol. 37, no. 22, Nov. 2010, doi: https://doi.org/10.1029/2010GL044830.

[35] D. E. Giles, "Benford's law and naturally occurring prices in certain ebaY auctions," *Appl. Econ. Lett.*, vol. 14, no. 3, pp. 157–161, Feb. 2007, doi: 10.1080/13504850500425667.

[36] M. J. Nigrini, "A taxpayer compliance application of Benford's Law," *J. Am. Tax. Assoc.*, vol. 18, no. 1, p. 72, 1996, [Online]. Available: https://www.proquest.com/scholarly-journals/taxpayer-compliance-application-benfords-law/docview/211023799/se-2?accountid=14777.

[37] C. Goh, "Applying visual analytics to fraud detection using Benford's law," *The journal of corporate accounting & finance.*, vol. 31, no. 4. John Wiley & Sons, [New York, N.Y.] :, pp. 202–208, 2020, doi: 10.1002/jcaf.22440.

[38] M. Lesperance, W. J. Reed, M. A. Stephens, C. Tsao, and B. Wilton, "Assessing Conformance with Benford's Law: Goodness-Of-Fit Tests and Simultaneous Confidence Intervals," *PLoS One*, vol. 11, no. 3, p. e0151235, Mar. 2016, [Online]. Available: https://doi.org/10.1371/journal.pone.0151235.
[39] G. Whyman, E. Shulzinger, and E. Bormashenko, "Intuitive considerations clarifying the origin and applicability of the Benford law," *Results Phys.*, vol. 6, pp. 3–6, 2016, doi: https://doi.org/10.1016/j.rinp.2015.11.010.
[40] DATADISTA, "Coronavirus Disease 2019 (COVID-19) in Spain." Harvard Dataverse, 2020, doi: doi:10.7910/DVN/GPFFAQ.
[41] L. Silva and D. Figueiredo Filho, "Using Benford's law to assess the quality of COVID-19 register data in Brazil.," *J. Public Health (Oxf).*, vol. 43, no. 1, pp. 107–110, Apr. 2021, doi: 10.1093/pubmed/fdaa193.
[42] D. Roy and S. P. Mukherjee, "A Note on Characterisations of the Weibull Distribution," *Sankhyā Indian J. Stat. Ser. A*, vol. 48, no. 2, pp. 250–253, Dec. 1986, [Online]. Available: http://www.jstor.org/stable/25050594.
[43] A. Vazquez, "Exact solution of infection dynamics with gamma distribution of generation intervals," *Phys. Rev. E*, vol. 103, no. 4, p. 42306, Apr. 2021, doi: 10.1103/PhysRevE.103.042306.
[44] E. Sevillano and P. Linde, "El desbarajuste de las cifras del coronavirus: Sanidad rebaja en casi 2.000 las muertes desde que empezó la pandemia," *El País*, May 25, 2020.