yönetim ve ekonomi araştırmaları dergisi
journal of management and economics research
Cilt/Volume: 21 Sayı/Issue: 2 Haziran/June 2023 ss. /pp. 120-133
A. H. Dogan etc. http://dx.doi.org/10.11611/yead.1078847

STATISTICAL ANALYSIS OF COVID-19 OUTBREAK WITH BENFORD'S LAW

Res. Asst. Ali Hasan DOGAN*

Res. Asst. Cemali ALTUNTAS**

Res. Asst. Caneren GUL***

Prof. Nursu TUNALIOGLU (PhD)****

Prof. Bahattin ERDOGAN (PhD)*****

ABSTRACT

The coronavirus disease first identified in mid-December 2019 in Wuhan, China is an ongoing pandemic and the virus has spread around the world. As of 13 March 2020, the number of new cases started to increase significantly in Europe, and Europe was considered as the new center of the Covid-19 pandemic as announced by the WHO. Confirmed case rate (CCR), computed from the numbers of confirmed cases over numbers of tests of the countries can be used to confirm the quality of the numbers, and to detect the manipulation for health surveillance systems of the countries for managing the situation by testing whether or not follow Benford's Law (BL). The main aim of this study is to test CCRs of the countries in Europe by BL to detect the data qualities and to monitor the manipulations, which can help to take precautions for the health surveillance systems of the countries.

Keywords: *Outbreak, Covid-19, Benford's Law, Chi-Squared Test, Kolmogorov-Smirnov Test.*

Jel codes: *C02, C12, C46, C83.*

* Yildiz Technical University, Faculty of Civil Engineering, Department of Geomatic Engineering, Istanbul/Türkiye, E-mail: alihasan@yildiz.edu.tr
** Yildiz Technical University, Faculty of Civil Engineering, Department of Geomatic Engineering, Istanbul/Türkiye, E-mail: cemali@yildiz.edu.tr
*** Yildiz Technical University, Faculty of Civil Engineering, Department of Geomatic Engineering, Istanbul/Türkiye, E-mail: cgul@yildiz.edu.tr
**** Yildiz Technical University, Faculty of Civil Engineering, Department of Geomatic Engineering, Istanbul/Türkiye, E-mail: ntunali@yildiz.edu.tr
***** Yildiz Technical University, Faculty of Civil Engineering, Department of Geomatic Engineering, Istanbul/Türkiye, E-mail: berdogan@yildiz.edu.tr

Makale Geçmişi/Article History

Başvuru Tarihi / Date of Application	: 25 Şubat / February 2022
Düzeltme Tarihi / Revision Date	: 12 Mayıs / May 2023
Kabul Tarihi / Acceptance Date	: 12 Haziran / June 2023

Araştırma Makalesi/Research Article

1. INTRODUCTION

Coronavirus disease (Covid-19) is an infectious disease caused by a newly discovered coronavirus (WHO, 2020). Covid-19 was first identified in mid-December 2019 and has since spread globally, resulting in the ongoing 2020 coronavirus pandemic. All of the countries have tried to control its spread and taken measures for fighting the epidemic. The pandemic has caused the death of more than 1.24 million people and approximately 49.4 million confirmed cases globally as of 7 November 2020. The first situation report released by WHO (2020) announced the updates as; as of 20 January 2020, 282 confirmed cases of 2019-nCoV have been reported from four countries including China (278 cases), Thailand (2 cases), Japan (1 case) and the Republic of Korea (1 case). According to the report declared by WHO (2020), as of the announcement date of the report, the first two confirmed cases of 2019-nCoV acute respiratory disease were reported in Italy; both had travel history to Wuhan City. According to this report, the confirmed cases of 2019-nCoV for the countries reported (20 countries) is 9826 (which 9720 of them were reported in China) (WHO, 2020). Epidemic management is essential to prevent and control the outbreak spread for countries (Idrovo & Manrique-Hernández, 2020). To do that, the data quality should be accurate, and the case numbers should be controlled whether they agree with expectations during events (Gómez-Camponovo et al., 2016). A fast and easy method applied in several fields, especially for fraud-detection in finance, so-called Benford's Law (BL), has recently implemented to check the data qualities reported during pandemics.

BL, also known as the First-digit or Significant-digit law, was first discovered by Simon Newcomb in 1881, and Newcomb (1881) stated that the first pages of the logarithmic tables were worn out faster than the last pages due to much more use and reported that digits did not occur with an equal frequency. However, after this discovery he did not sustain his study (Mir, 2014). Frank Benford then separately rediscovered the phenomenon in 1938 by observing logarithmic tables, which was an evidence of selective use of the digits and he examined the accuracy of BL with large dataset gathered from different fields such as lengths of the rivers, population, atomic weight, cost data, and street address etc. In the study, he proved that the occurrence of any first non-zero digit (even of second or third digit) in a numerical data found in nature depends on a logarithmic distribution, and concluded that the frequency of the first significant digit (i.e., nonzero digit, k) following logarithmic distribution can be computed as (Benford, 1938):

$$P(k) = \log_{10}(k+1) - \log_{10}(k) = \log_{10}\left(1+\frac{1}{k}\right) \tag{1}$$

Where, P is the probability of the number k, and k is the any first digit in the set $\{1,2,3,\dots,9\}$ According to the BL, the occurrences of the non-zero first leading digits (the leftmost digit) from 1 to 9 are 30.1%, 17.6%, 12.5%, 9.7%, 7.9%, 6.7%, 5.8%, 5.1%, and 4.6%, respectively.

In the literature, there have been wide ranges of implementations of BL from social to numerical application fields (Becker 1982; Mir 2012; Nagasaka 1984; Hill 1995; Ausloos et al., 2015; Tunalioglu and Erdogan, 2019). More comprehensive information can be found in "Benford Online Bibliography (Berger et al., 2009)". In addition to these studies, BL was also studied in several studies that have been conducted to confirm the data quality of the confirmed cases/death numbers for outbreaks to detect the surveillance systems using this distribution.

Idrovo et al. (2011) investigated the performance of public health surveillance (PHS) systems whether those properly responded to A(H1N1) pandemic and represented an algorithm based on BL in order to evaluate the data quality in terms of the mortality ratio computed by confirmed deaths over confirmed cases. The main query was that if the incidence follows the distribution described by BL, there is evidence that reporting was satisfactory. In the data analyses, Kuiper, χ^2 and log-likelihood ratio tests were implemented to explore the data reported whether follow or not BL, which can be an evidence to represent the quality of PHS system. According to the results, they concluded that BL-based method could be useful to examine the performance of PHS system during a pandemic. Gómez-Camponovo et al. (2016) aimed to evaluate the performance of the PHS system for dengue epidemic in Paraguay during 2009-2011 with BL with respect to data quality and system sensitivity. They used weekly reported dengue cases, and tested these data with χ^2 and log-likelihood ratio tests, and calculated correlations between socioeconomic (i.e. population, percentage of household with dirt floors, percentage of houses burning or burying trash, houses with no electricity) variables. Although they indicated that first-digit BL analysis did not fit with the related data, BL with second-digit analyses fitted well with the observed data during the pandemic. BL was also implemented by Manrique-Hernandez et al. (2017) during ongoing Zika virus epidemic to explore the global performance of PHS systems. They examined the suspected and confirmed cases reported for 39 countries in America with Kuiper, log-likelihood ratio and Pearson χ^2 tests, and concluded that BL can be a useful and quick tool for the countries to detect the performance of surveillance system. In addition to those, studies investigating the relationship between BL and Covid-19 pandemic confirmed data have been conducted during this epidemic. As considering Covid-19 as a natural event that occurs randomly, it is expected that the epidemic event will follow an exponential distribution in the rate of spread and obey the BL, since the characteristic of the naturally occurring events follows a specified theoretical distribution. Balashov et al. (2020) searched for an answer in their study if manipulation can be done on the pandemic data reported during Covid-19 outbreak by the less developed countries. They implemented BL to cumulative number confirmed cases and the cumulative number of deaths with macroeconomic indicators. Observed proportions were tested with four goodness-of-fit measures; χ^2, modified version of the Kuiper, M-statistic, and D-statistic to examine the fitness of BL. Idrovo & Manrique-Hernández (2020) examined the data quality of Chinese surveillance of Covid-19 with cumulative confirmed case data released by WHO gathered between 1-55 situation reports. They tested the findings to conclude whether or not following BL with χ^2 and log-likelihood ratio, and indicated that use of BL for health surveillance in terms of accuracy of the data quality enhances a quick check as a tool. Zhang (2020) implemented BL to the cumulative case numbers reported during Covid-19 pandemic in China from 15 January 2020 to 10 February 2020, and concluded that any fluctuations or manipulation was detected, which meant that the numbers obey the BL. Although Zhang (2020) tests BL-observed and -expected frequencies with χ^2, more comprehensive statistical tests such as Kolmogorov-Smirnov (KS) or Kuiper were recommended.

In this study, the confirmed case data qualities of the European countries have been investigated by BL, since as of 13 March, 2020 the number of new cases started to increase significantly after China and Europe was considered the new center of the Covid-19 pandemic as announced by WHO. As the data for Covid-19 can be the numbers of new/daily confirmed cases, deaths, and recoveries, we focused on a model that normalize the new daily-confirmed cases with daily-announced test numbers. In short, our goal was to investigate whether the number of normalized confirmed cases announced by the officials followed BL or not, and to test the model against empirical Covid-19 data on the confirmed case rate (CCR).

The coronavirus metrics (number of confirmed cases, deaths and recoveries) can be changed country by country due to several reasons such as measures taken to prevent the spread of the outbreak by social distancing (physical distancing) rules, stay-at-home calls or official lockdowns, testing etc. (Moosa, 2020; Balashov et al., 2020). Depending on these measures, number of daily tests would be different for countries as well. Although, the spread of virus differs due to above-mentioned conditions, the number of confirmed cases results in constant in pro-rata. To examine the spread of the virus independent from testing utility, CCR, defined with the number of daily cases over the number of tests, which will differ country to country, is used as an indicator with BL to detect the data quality released.

2. DATA

WHO continuously shares the numbers of cases and deaths, which are officially reported by the health ministries of the countries (URL-1). Since the number of cases and deaths depends on the population of the countries, two parameters, the level of spread and especially the number of tests, are sufficient to see the tendency of the epidemic. In addition to the number of cases, the number of tests is also needed to determine the rate of spread of the virus and to reveal the situation in countries. The data gathered from “Our World in Data (OWID)”, which has a statistical data sharing web site, was examined (Roser et al., 2020). BL was implemented to the data of the randomly selected European countries from the first case to 31 October 2020. The countries (order of first case appearance) are given in Table 1.

Table 1. The Selected Countries in Europe For Covid-19 Pandemic Analyses

Country	First Reported Case Date	Country	First Reported Case Date
France	25.01.2020	Iceland	29.02.2020
Finland	30.01.2020	Ireland	1.03.2020
Italy	31.01.2020	Luxembourg	1.03.2020
Russia	1.02.2020	Czech Republic	2.03.2020
United Kingdom	1.02.2020	Latvia	3.03.2020
Belgium	4.02.2020	Portugal	3.03.2020
Switzerland	26.02.2020	Ukraine	4.03.2020
Croatia	26.02.2020	Poland	4.03.2020
Austria	26.02.2020	Slovenia	5.03.2020
Norway	27.02.2020	Hungary	5.03.2020
Romania	27.02.2020	Serbia	7.03.2020
Denmark	27.02.2020	Slovakia	7.03.2020
Greece	27.02.2020	Malta	7.03.2020
Estonia	28.02.2020	Bulgaria	8.03.2020
Lithuania	28.02.2020		

3. STATISTICAL ANALYSIS

The Null-Hypothesis (H_0) indicates that the frequencies obtained from observations are the same as expected frequencies basis of BL, is written to test for compliance with BL between the observed and expected first digit distributions. To test whether H_0 hypothesis is accepted or not, different goodness-of-fit measures can be applied as $\mathcal{X}^2$ test or KS test comparing with a critical value. If the observed frequencies are close to corresponding expected frequencies, the $\mathcal{X}^2$ test value or KS test value will be small, indicating a good fit. A good fit leads to acceptance of H_0, whereas a poor fit leads to its rejection.

To test whether H_0 hypothesis is accepted or not, the $\mathcal{X}^2$ goodness-of-fit statistic is calculated using (Walpole et al., 1998);

$$\mathcal{X}^2 = \sum_{k=1}^{9} \frac{(N_O - N_B)^2}{N_B} \qquad (2)$$

Where $\mathcal{X}^2$ is a value of a random variable whose sampling distribution is approximated very closely by the chi-squared distribution with v=9-1 degrees of freedom. N_O is the observed frequency, N_B is the BL frequency. The test is based on how good a fit we have between the frequency of occurrence of observations in an observed sample and the expected frequencies obtained from the hypothesized distribution (Walpole et al., 1998).

In addition to $\mathcal{X}^2$ test, the KS test can be used to serve as a goodness of fit test. The KS test for a given cumulative distribution function $F(x)$ is written as follows (Massey Jr., 1951; Marsaglia et al., 2003):

$$D_n = \text{maximum}\left(\left|F(x_1^O) - F(x_1^B)\right|, \left|F(x_2^O) - F(x_2^B)\right|, \dots, \left|F(x_9^O) - F(x_9^B)\right|\right) \tag{3}$$

where $D_n\sqrt{n}$ is the KS test value, n is the number of the observations, the subindices for $F(x)$ define the first digit, $F(x^O)$ and $F(x^B)$ are the cumulative distribution functions for observed and BL, respectively,

4. RESULTS AND DISCUSSION

To test the compliance with BL, two different statistical tests as $\mathcal{X}^2$ test and KS test were applied. The confidence levels for two tests were selected as 95%. As we have 9 first digits, the degrees of freedom is taken 9-1=8 for $\mathcal{X}^2$ test. The critical value for $\mathcal{X}^2$ test is computed 15.5073. For the KS test the critical value given in Massey Jr. (1951) was used as 1.36. If these values are exceeded in any case by $\mathcal{X}^2$ or KS test values, H_0 hypothesis will be rejected, otherwise, it will be accepted.

The statistical analyses for each country are shown in Tables 2-4. The frequencies and the number of counts for each leading digit are given in Tables. The numbers of the total frequencies are given in the brackets after the country names. In addition to this, the corresponding frequencies (in brackets) for each leading digit as predicted by BL;

$$f_B = \mathrm{N} \log_{10}\left(1 + \frac{1}{k}\right) \tag{4}$$

N is the total number of frequencies and the root mean square error (Δf) calculated from the binomial distribution are shown in Tables (Shao and Ma, 2009).

$$\Delta f = \sqrt{NP(k)(1 - P(k))} \tag{5}$$

The results of the CCR accepted by H_0 hypothesis for both $\mathcal{X}^2$test and KS test are given in Table 2. The number of total frequencies changes between 138 and 250. The maximum test values for $\mathcal{X}^2$test and KS test were calculated for Slovenia and Croatia, respectively. The observed and BL distributions of significant digits for these countries are shown in Figure 1. The minimum test values for $\mathcal{X}^2$ test and KS test were calculated for Estonia and Austria, respectively. The observed and BL distributions of significant digits for these countries are represented in Figure 2. According to the results given in Table 2, the CCR results for fifteen European countries fit the BL proportions based on both $\mathcal{X}^2$test and KS test.

Table 2. The Frequencies and Proportions for the CCR, in Which H_0 Hypothesis Is Accepted for Both $\mathcal{X}^2$ Test and KS Test

First	France (156)		Finland (233)		Italy (250)		Switzerland (233)		Croatia (206)	
Digit	Frequency	Proportion	Frequency	Proportion	Frequency	Proportion	Frequency	Proportion	Frequency	Proportion
1	40 (47.0 ± 5.7)	0.2564	73 (70.1 ± 7.0)	0.3133	81 (75.3 ± 7.3)	0.3240	78 (70.1 ± 7.0)	0.3348	59 (62.0 ± 6.6)	0.2864
2	33 (27.5 ± 4.8)	0.2115	42 (41.0 ± 5.8)	0.1803	43 (44.0 ± 6.0)	0.1720	44 (41.0 ± 5.8)	0.1888	24 (36.3 ± 5.5)	0.1165
3	16 (19.5 ± 4.1)	0.1026	27 (29.1 ± 5.0)	0.1159	24 (31.2 ± 5.2)	0.0960	31 (29.1 ± 5.0)	0.1330	22 (25.7 ± 4.7)	0.1068

4	14 (15.1 ± 3.7)	0.0897	26 (22.6 ± 4.5)	0.1116	26 (24.2 ± 4.7)	0.1040	30 (22.6 ± 4.5)	0.1288	26 (20.0 ± 4.2)	0.1262
5	6 (12.4 ± 3.4)	0.0385	13 (18.4 ± 4.1)	0.0558	18 (19.8 ± 4.3)	0.0720	13 (18.4 ± 4.1)	0.0558	17 (16.3 ± 3.9)	0.0825
6	15 (10.4 ± 3.1)	0.0962	20 (15.6 ± 3.8)	0.0858	22 (16.7 ± 4.0)	0.0880	5 (15.6 ± 3.8)	0.0215	24 (13.8 ± 3.6)	0.1165
7	9 (9.0 ± 2.9)	0.0577	12 (13.5 ± 3.6)	0.0515	14 (14.5 ± 3.7)	0.0560	8 (13.5 ± 3.6)	0.0343	12 (11.9 ± 3.4)	0.0583
8	9 (8.0 ± 2.8)	0.0577	12 (11.9 ± 3.4)	0.0515	10 (12.8 ± 3.5)	0.0400	11(11.9 ± 3.4)	0.0472	12 (10.5 ± 3.2)	0.0583
9	14 (7.1 ± 2.6)	0.0897	8 (10.7 ± 3.2)	0.0343	12 (11.4 ± 3.3)	0.0480	13(10.7 ± 3.2)	0.0558	10 (9.4 ± 3.0)	0.0485
X^2	14.8340		4.4962		4.7378		15.2999		14.4925	
KS	0.9921		0.3367		0.3632		1.3194		1.3255	
First Digit	**Austria (229)**		**Norway (192)**		**Denmark (221)**		**Estonia (223)**		**Lithuania (138)**	
	Frequency	**Proportion**	**Frequency**	**Proportion**	**Frequency**	**Proportion**	**Frequency**	**Proportion**	**Frequency**	**Proportion**
1	74 (68.9 ± 6.9)	0.3231	60 (57.8 ± 6.4)	0.3125	76 (66.5 ± 6.8)	0.3439	68 (67.1 ± 6.8)	0.3049	50 (41.5 ± 5.4)	0.3623
2	39 (40.3 ± 5.8)	0.1703	28 (33.8 ± 5.3)	0.1458	30 (38.9 ± 5.7)	0.1357	32 (39.3 ± 5.7)	0.1435	22 (24.3 ± 4.5)	0.1594
3	22 (28.6 ± 5.0)	0.0961	30 (24.0 ± 4.6)	0.1563	21 (27.6 ± 4.9)	0.0950	25 (27.9 ± 4.9)	0.1121	13 (17.2 ± 3.9)	0.0942
4	21 (22.2 ± 4.5)	0.0917	11 (18.6 ± 4.1)	0.0573	25 (21.4 ± 4.4)	0.1131	21 (21.6 ± 4.4)	0.0942	16 (13.4 ± 3.5)	0.1159
5	19 (18.1 ± 4.1)	0.0830	12 (15.2 ± 3.7)	0.0625	13 (17.5 ± 4.0)	0.0588	18 (17.7 ± 4.0)	0.0807	12 (10.9 ± 3.2)	0.0870
6	23 (15.3 ± 3.8)	0.1004	18 (12.9 ± 3.5)	0.0938	18 (14.8 ± 3.7)	0.0814	20 (14.9 ± 3.7)	0.0897	11 (9.2 ± 2.9)	0.0797
7	7 (13.3 ± 3.5)	0.0306	16 (11.1 ± 3.2)	0.0833	11 (12.8 ± 3.5)	0.0498	16 (12.9 ± 3.5)	0.0717	9 (8.0 ± 2.7)	0.0652
8	12 (11.7 ± 3.3)	0.0524	9 (9.8 ± 3.1)	0.0469	17 (11.3 ± 3.3)	0.0769	11 (11.4 ± 3.3)	0.0493	4 (7.1 ± 2.6)	0.0290
9	12 (10.5 ± 3.2)	0.0524	8 (8.8 ± 2.9)	0.0417	10 (10.1 ± 3.1)	0.0452	12 (10.2 ± 3.1)	0.0538	1 (6.3 ± 2.5)	0.0072
X^2	9.0830		10.6987		10.5528		4.4552		9.8630	
KS	0.3346		0.6066		0.6372		0.6610		0.7200	
First Digit	**Iceland (187)**		**Czech Republic (243)**		**Portugal (241)**		**Slovenia (214)**		**Slovakia (227)**	
	Frequency	**Proportion**	**Frequency**	**Proportion**	**Frequency**	**Proportion**	**Frequency**	**Proportion**	**Frequency**	**Proportion**
1	51 (56.3 ± 6.3)	0.2727	70 (73.2 ± 7.2)	0.2881	74 (72.5 ± 7.1)	0.3071	52 (64.4 ± 6.7)	0.2430	82 (68.3 ± 6.9)	0.3612
2	37 (32.9 ± 5.2)	0.1979	45 (42.8 ± 5.9)	0.1852	54 (42.4 ± 5.9)	0.2241	51 (37.7 ± 5.6)	0.2383	37 (40.0 ± 5.7)	0.1630
3	23 (23.4 ± 4.5)	0.1230	37 (30.4 ± 5.2)	0.1523	25 (30.1 ± 5.1)	0.1037	29 (26.7 ± 4.8)	0.1355	26 (28.4 ± 5.0)	0.1145
4	13 (18.1 ± 4.0)	0.0695	27 (23.5 ± 4.6)	0.1111	22 (23.4 ± 4.6)	0.0913	16 (20.7 ± 4.3)	0.0748	25 (22.0 ± 4.5)	0.1101
5	19 (14.8 ± 3.7)	0.1016	21 (19.2 ± 4.2)	0.0864	20 (19.1 ± 4.2)	0.0830	16 (16.9 ± 4.0)	0.0748	14 (18.0 ± 4.1)	0.0617
6	17 (12.5 ± 3.4)	0.0909	14 (16.3 ± 3.9)	0.0576	14 (16.1 ± 3.9)	0.0581	18 (14.3 ± 3.7)	0.0841	11 (15.2 ± 3.8)	0.0485
7	10 (10.8 ± 3.2)	0.0535	12 (14.1 ± 3.6)	0.0494	11 (14.0 ± 3.6)	0.0456	6 (12.4 ± 3.4)	0.0280	13 (13.2 ± 3.5)	0.0573
8	15 (9.6 ± 3.0)	0.0802	9 (12.4 ± 3.4)	0.0370	8 (12.3 ± 3.4)	0.0332	11 (10.9 ± 3.2)	0.0514	11 (11.6 ± 3.3)	0.0485
9	2 (8.6 ± 2.9)	0.0107	8 (11.1 ± 3.3)	0.0329	13 (11.0 ± 3.2)	0.0539	15 (9.8 ± 3.1)	0.0701	8 (10.4 ± 3.1)	0.0352
X^2	13.4230		4.8167		6.9572		15.4505		6.1808	
KS	0.4905		0.6998		0.8383		0.8490		0.9071	

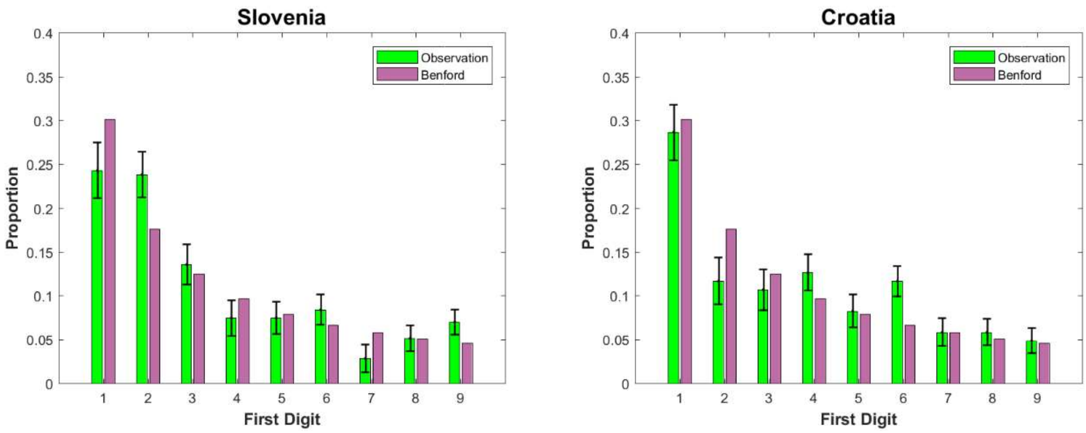

Figure 1. The Observed and BL Distributions of Significant Digits for Slovenia and Croatia

Cilt/Volume: 21 Sayı/Issue: 2 Haziran/June 2023 ss. /pp. 120-133
A. H. Dogan etc. http://dx.doi.org/10.11611/yead.1078847

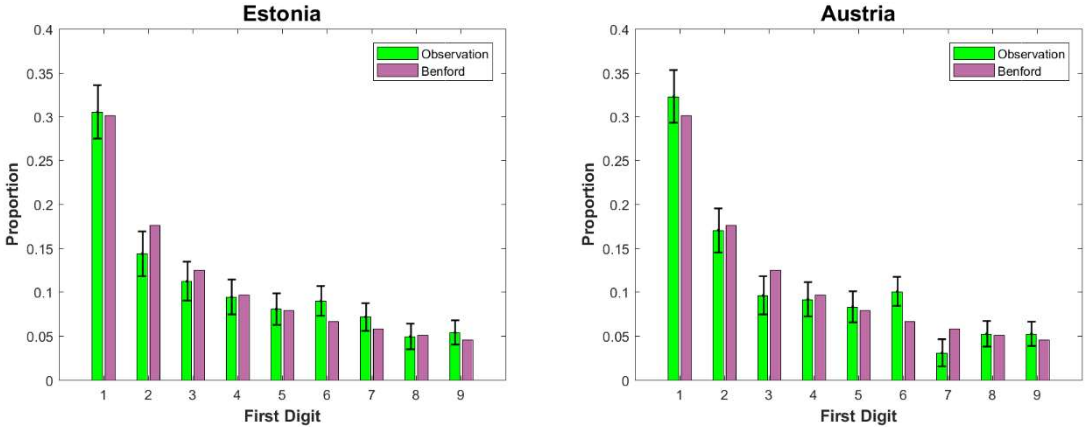

Figure 2. The Observed and BL Distributions of Significant Digits for Estonia and Austria

The results of the CCR, in which H_0 hypothesis is not accepted for both $\mathcal{X}^2$ test and KS test are given in Table 3. The number of total frequencies changes between 142 and 244. Russia has the maximum test values for both $\mathcal{X}^2$test and KS test. The observed and BL distributions of significant digits for Russia are shown in Figure 3. The minimum test values of $\mathcal{X}^2$test and KS test were calculated for Ireland and Hungary, respectively. The observed and BL distributions of significant digits for these countries are given in Figure 4. According to the results given in Table 3, the CCRs for ten European countries do not fit the BL proportions based on both $\mathcal{X}^2$ test and KS test.

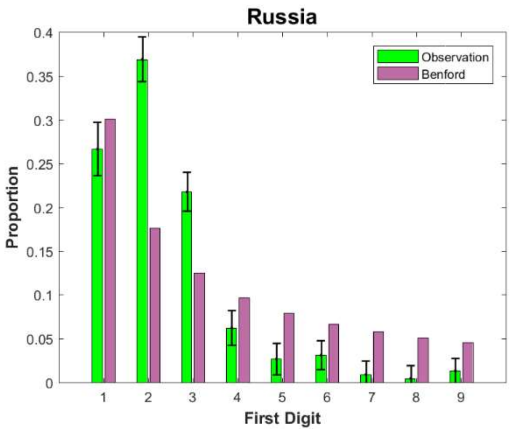

Figure 3. The Observed and BL Distributions of Significant Digits for Russia

Table 3. The Frequencies and Proportions for The Confirmed Case Rate in Which H_0 Hypothesis Is Not Accepted for Both $\mathcal{X}^2$ Test and KS Test

First Digit	Russia (225)		United Kingdom (215)		Belgium (244)		Ireland (227)		Ukraine (159)	
	Frequency	Proportion	Frequency	Proportion	Frequency	Proportion	Frequency	Proportion	Frequency	Proportion
1	60 (67.7 ± 6.9)	0.2667	50 (64.7 ± 6.7)	0.2326	77 (73.5 ± 7.2)	0.3156	73 (68.3 ± 6.9)	0.3216	51 (47.9 ± 5.8)	0.3208
2	83 (39.6 ± 5.7)	0.3689	35 (37.9 ± 5.6)	0.1628	64 (43.0 ± 5.9)	0.2623	59 (40.0 ± 5.7)	0.2599	11 (28.0 ± 4.8)	0.0692
3	49 (28.1 ± 5.0)	0.2178	19 (26.9 ± 4.8)	0.0884	30 (30.5 ± 5.2)	0.1230	21 (28.4 ± 5.0)	0.0925	8 (19.9 ± 4.2)	0.0503
4	14 (21.8 ± 4.4)	0.0622	17 (20.8 ± 4.3)	0.0791	17 (23.6 ± 4.6)	0.0697	19 (22.0 ± 4.5)	0.0837	8 (15.4 ± 3.7)	0.0503
5	6 (17.8 ± 4.1)	0.0267	31 (17.0 ± 4.0)	0.1442	11 (19.3 ± 4.2)	0.0451	11 (18.0 ± 4.1)	0.0485	16 (12.6 ± 3.4)	0.1006
6	7 (15.1 ± 3.7)	0.0311	28 (14.4 ± 3.7)	0.1302	12 (16.3 ± 3.9)	0.0492	11 (15.2 ± 3.8)	0.0485	22 (10.6 ± 3.2)	0.1384
7	2 (13.0 ± 3.5)	0.0089	19 (12.5 ± 3.4)	0.0884	14 (14.2 ± 3.7)	0.0574	14 (13.2 ± 3.5)	0.0617	19 (9.2 ± 2.9)	0.1195
8	1 (11.5 ± 3.3)	0.0044	11 (11.0 ± 3.2)	0.0512	10 (12.5 ± 3.4)	0.0410	13 (11.6 ± 3.3)	0.0573	10 (8.1 ± 2.8)	0.0629
9	3 (10.3 ± 3.1)	0.0133	5 (9.8 ± 3.1)	0.0233	9 (11.2 ± 3.3)	0.0369	6 (10.4 ± 3.1)	0.0264	14 (7.3 ± 2.6)	0.0881
$\mathcal{X}^2$	102.9662		36.7086		17.9921		17.6320		51.2279	
KS	3.7691		1.9968		1.5737		1.5726		2.6279	
First Digit	Poland (183)		Hungary (227)		Serbia (235)		Malta (173)		Bulgaria (142)	
	Frequency	Proportion	Frequency	Proportion	Frequency	Proportion	Frequency	Proportion	Frequency	Proportion
1	60 (55.1 ± 6.2)	0.3279	64 (68.3 ± 6.9)	0.2819	93 (70.7 ± 7.0)	0.3957	78 (52.1 ± 6.0)	0.4509	25 (42.7 ± 5.5)	0.1761
2	67 (32.2 ± 5.2)	0.3661	26 (40.0 ± 5.7)	0.1145	34 (41.4 ± 5.8)	0.1447	26 (30.5 ± 5.0)	0.1503	34 (25.0 ± 4.5)	0.2394
3	25 (22.9 ± 4.5)	0.1366	26 (28.4 ± 5.0)	0.1145	39 (29.4 ± 5.1)	0.1660	18 (21.6 ± 4.3)	0.1040	22 (17.7 ± 3.9)	0.1549
4	16 (17.7 ± 4.0)	0.0874	33 (22.0 ± 4.5)	0.1454	31 (22.8 ± 4.5)	0.1319	9 (16.8 ± 3.9)	0.0520	20 (13.8 ± 3.5)	0.1408
5	5 (14.5 ± 3.7)	0.0273	21 (18.0 ± 4.1)	0.0925	9 (18.6 ± 4.1)	0.0383	6 (13.7 ± 3.6)	0.0347	15 (11.2 ± 3.2)	0.1056
6	1 (12.3 ± 3.4)	0.0055	14 (15.2 ± 3.8)	0.0617	7 (15.7 ± 3.8)	0.0298	11 (11.6 ± 3.3)	0.0636	12 (9.5 ± 3.0)	0.0845
7	5 (10.6 ± 3.2)	0.0273	10 (13.2 ± 3.5)	0.0441	3 (13.6 ± 3.6)	0.0128	9 (10.0 ± 3.1)	0.0520	6 (8.2 ± 2.8)	0.0423
8	3 (9.4 ± 3.0)	0.0164	18 (11.6 ± 3.3)	0.0793	9 (12.0 ± 3.4)	0.0383	8 (8.8 ± 2.9)	0.0462	3 (7.3 ± 2.6)	0.0211
9	1 (8.4 ± 2.8)	0.0055	15 (10.4 ± 3.1)	0.0661	10 (10.8 ± 3.2)	0.0426	8 (7.9 ± 2.7)	0.0462	5 (6.5 ± 2.5)	0.0352
$\mathcal{X}^2$	68.6669		17.7851		33.3640		22.3021		19.8172	
KS	3.0916		1.3718		2.1359		1.9708		1.4892	

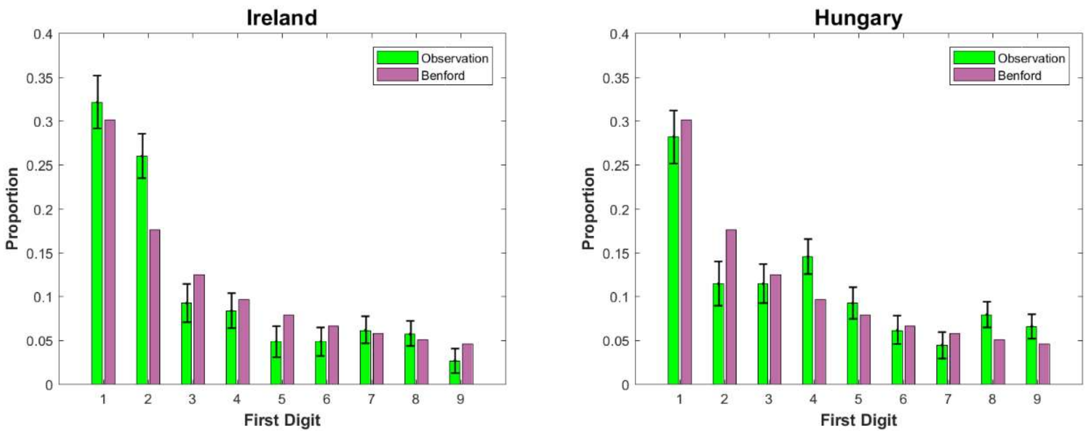

Figure 4. The observed and BL distributions of significant digits for Ireland and Hungary

In this study, two different goodness-of-fit measures were used. The proportions of the four countries had different characteristics when considering the test statistics. The mentioned results were given in Table 4. Although their CCR results fitted BL depending on KS test, they did not fit when $\mathcal{X}^2$ test was applied. This case may be happened since two test statistics have different distribution functions from each other. The number of total frequencies changes between 188 and 217. Greece has the minimum test values for both $\mathcal{X}^2$test and KS test. The observed and BL distributions of significant digits for Greece are shown in Figure 5. The maximum test values for $\mathcal{X}^2$test and KS test were calculated for Luxembourg and Romania, respectively. The observed and BL distributions of significant digits for these countries are given in Figure 6. According to the results given in Table 4, although the only KS test values were smaller than critical value, it can be interpreted as the CCR results for four European countries fit the BL proportions since $\mathcal{X}^2$ test may be subjected to large fluctuations for small data sets (Zhang, 2020).

Table 4. The frequencies and proportions for the CCR in which H_0 hypothesis is accepted for only KS test

First Digit	Romania (210)		Greece (188)		Luxembourg (214)		Latvia (217)	
	Frequency	Proportion	Frequency	Proportion	Frequency	Proportion	Frequency	Proportion
1	56 (63.2 ± 6.6)	0.2667	48 (56.6 ± 6.3)	0.2553	70 (64.4 ± 6.7)	0.3271	58 (65.3 ± 6.8)	0.2673
2	25 (37.0 ± 5.5)	0.1190	38 (33.1 ± 5.2)	0.2021	28 (37.7 ± 5.6)	0.1308	55 (38.2 ± 5.6)	0.2535
3	26 (26.2 ± 4.8)	0.1238	35 (23.5 ± 4.5)	0.1862	14 (26.7 ± 4.8)	0.0654	24 (27.1 ± 4.9)	0.1106
4	26 (20.4 ± 4.3)	0.1238	24 (18.2 ± 4.1)	0.1277	26 (20.7 ± 4.3)	0.1215	29 (21.0 ± 4.4)	0.1336
5	28 (16.6 ± 3.9)	0.1333	12 (14.9 ± 3.7)	0.0638	20 (16.9 ± 4.0)	0.0935	18 (17.2 ± 4.0)	0.0829
6	19 (14.1 ± 3.6)	0.0905	11 (12.6 ± 3.4)	0.0585	13 (14.3 ± 3.7)	0.0607	9 (14.5 ± 3.7)	0.0415
7	7 (12.2 ± 3.4)	0.0333	11 (10.9 ± 3.2)	0.0585	17 (12.4 ± 3.4)	0.0794	9 (12.6 ± 3.4)	0.0415
8	14 (10.7 ± 3.2)	0.0667	7 (9.6 ± 3.0)	0.0372	20 (10.9 ± 3.2)	0.0935	9 (11.1 ± 3.2)	0.0415
9	9 (9.6 ± 3.0)	0.0429	2 (8.6 ± 2.9)	0.0106	6 (9.8 ± 3.1)	0.0280	6 (9.9 ± 3.1)	0.0276
$\mathcal{X}^2$	19.0169		16.0443		21.7012		16.6902	
KS	1.3410		0.9914		1.1512		1.0279	

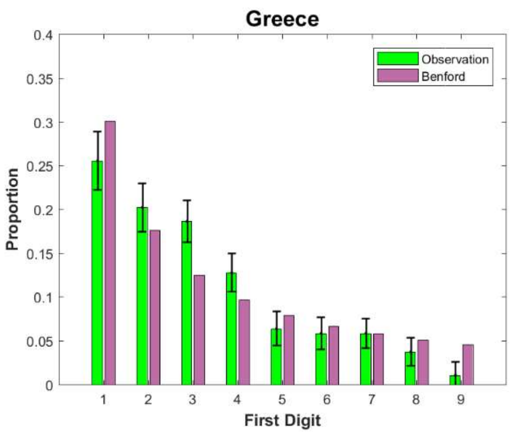

Figure 5. The observed and BL distributions of significant digits for Greece

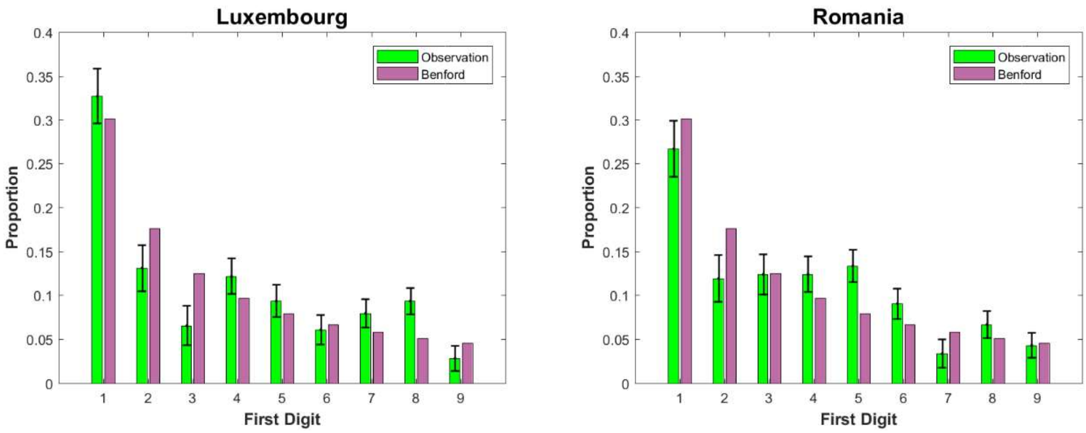

Figure 6. The observed and BL distributions of significant digits for Luxembourg and Romania

5. CONCLUSION

The coronavirus disease has affected all over the world. The outbreak was declared a public health emergency of international concern by the WHO, and then the problem of the management of the outbreak has carried out for all countries. To control the spread of the epidemic, the accuracy of the data quality has important effect. The data may include different values such as number of cases, deaths, recoveries and spread rate. Depending on these metrics, number of daily tests would be different for countries as well, that's why, to examine the spread of the virus independent from testing utility, CCR, defined with the number of daily cases over the number of tests was used as an indicator. Since BL was applied as a useful tool for fraud detection in financial fields, recently it has been implemented during epidemic to detect the data quality reported officially and sensitivity of the health surveillance systems. BL has been implemented successfully to several previous pandemics and epidemics such as A(H1N1), dengue, Zika virus for detecting quality of confirmed data i.e. death, case, test numbers announced. Moreover, BL based fraud detection in health surveillance system has carried out for PHS sensitivity. In this study, BL was implemented to the data of the randomly selected 29 European countries. Two different goodness-of-fit measures have been applied to the CCR results. According to the obtained results, an overall evaluation of the surveillance system performance based on the first digit shows that the data reported by 19 countries fit BL distribution, which can be an indicator of a good data quality and sensitive surveillance system.

REFERENCES

Ausloos, M., Herteliu, C. and Ileanu, B. (2015) "Breakdown of Benford's Law for Birth Data", Physica A. 419:736–745.

Balashov, VS., Yan, Y. and Zhu, X. (2007) "Are Less Developed Countries More Likely to Manipulate Data During Pandemics? Evidence from Newcomb-Benford Law", arXiv, 14841. 2020.

Becker, PW. (1982) "Patterns in Listings of Failure-Rate & MTTF Values and Listings of Other Data", IEEE Trans Reliab, 31:132–134.

Benford, F. (1938) "The Law of Anomalous Numbers. Proceeding of the American Philosophical Society", 78(4):551-573.

Berger, A., Hill, TP. and Rogers, E. (2020) "Benford Online Bibliography. http://www.benfordonline.net. 2009; Accessed March 25, 2020.

Gómez-Camponovo, M., Moreno, J., Idrovo, AJ., Páez, M. and Achkar, M. (2016) "Monitoring the Paraguayan Epidemiological Dengue Surveillance System (2009-2011) Using Benford's Law", Biomédica, 36:583-92.

Hill, TP. (1995) "A Statistical Derivation of The Significant-Digit Law", Stat Sci.,10(4):354–363.

Idrovo, AJ., Fernández-Nino, JA., Bojorquez-Chapela, I. and Moreno-Montoya, J. (2011) "Performance of Public Health Surveillance Systems During the Influenza A(H1N1) Pandemic in The Americas: Testing A New Method Based on Benford's Law", Epidemiol. Infect., 139, 1827–1834.

Idrovo, AJ. and Manrique-Hernández, EF. (2020) "Data Quality of Chinese Surveillance Of COVID-19: Objective Analysis Based on WHO's Situation Reports", Asia Pacific Journal of Public Health, 32(4):165–167.

Manrique-Hernández, EF., Fernández-Nino, JA. and Idrovo, AJ. (2017) "Global Performance of Epidemiologic Surveillance of Zika Virus: Rapid Assessment of An Ongoing Epidemic", Public Health, 143:14-16.

Marsaglia, G., Tsang, WW. and Wang, J. (2003) "Evaluating Kolmogorov's Distribution", Journal of Statistical Software, 8(18):1-4.

Massey, Jr. FJ. (1951) "The Kolmogorov-Smirnov Test for Goodness of Fit", Journal of the American statistical Association, 46(253):68-78.

Mir, TA. (2012) "The Law of The Leading Digits and The World Religions", Physica A., 391:792–798.

Mir, TA. (2014) "The Benford Law Behavior of The Religious Activity Data", Physica A., 408: 1–9.

Moosa, IA. (2020) "The Effectiveness of Social Distancing in Containing Covid-19", Applied Economics., DOI: 10.1080/00036846.2020.1789061.

Nagasaka, K. (1984) "On Benford's Law", Ann Inst Stat Math, 36(Part A):337–352.

Newcomb, S. (1881) "Note on The Frequency of Use of The Different Digits in Natural Numbers", Am J Math, 4(1):39–40.

Roser, M., Ritchie, H., Ortiz-Ospina, E. and Hasell, J. (2020) "Coronavirus Pandemic (COVID-19)", Our World in Data.

Shao, L. and Ma, BQ. (2009) "First Digit Distribution of Hadron Full Width", Mod. Phys. Lett., A24: 3275–3282.

Tunalioglu, N. and Erdogan, B. (2019) "Usability of The Benford's Law for The Results of Least Square Estimation", Acta Geodaetica et Geophysica. 54:315–331.

Walpole, RE., Myers, RH. and Myers, SL. (1998) "Probability and Statistics for Engineers and Scientists", sixth edition, Prentice.

WHO (2020) Situation Reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ Accessed April 30, 2020.

Zhang, J. (2020) "Testing Case Number of Coronavirus Disease 2019 In China with Newcomb-Benford Law", arXiv:2002.05695.

[URL-1] https://covid19.who.int/ (Accessed: 31.10.2020)

KATKI ORANI / ***CONTRIBUTION RATE***	**AÇIKLAMA /** ***EXPLANATION***	**KATKIDA BULUNANLAR /** ***CONTRIBUTORS***
Fikir veya Kavram / *Idea or Notion*	Araştırma hipotezini veya fikrini oluşturmak / *Form the research hypothesis or idea*	Res. Asst. Ali Hasan DOGAN Res. Asst. Cemali ALTUNTAS Res. Asst. Caneren GUL Prof. Nursu TUNALIOGLU (PhD) Prof. Bahattin ERDOGAN (PhD)
Tasarım / *Design*	Yöntemi, ölçeği ve deseni tasarlamak / *Designing method, scale and pattern*	Res. Asst. Ali Hasan DOGAN Res. Asst. Cemali ALTUNTAS Res. Asst. Caneren GUL Prof. Nursu TUNALIOGLU (PhD) Prof. Bahattin ERDOGAN (PhD)
Veri Toplama ve İşleme / *Data Collecting and Processing*	Verileri toplamak, düzenlenmek ve raporlamak / *Collecting, organizing and reporting data*	Res. Asst. Ali Hasan DOGAN Res. Asst. Cemali ALTUNTAS Res. Asst. Caneren GUL Prof. Nursu TUNALIOGLU (PhD) Prof. Bahattin ERDOGAN (PhD)

Tartışma ve Yorum / *Discussion and Interpretation*	Bulguların değerlendirilmesinde ve sonuçlandırılmasında sorumluluk almak / *Taking responsibility in evaluating and finalizing the findings*	Res. Asst. Ali Hasan DOGAN Res. Asst. Cemali ALTUNTAS Res. Asst. Caneren GUL Prof. Nursu TUNALIOGLU (PhD) Prof. Bahattin ERDOGAN (PhD)
Literatür Taraması / *Literature Review*	Çalışma için gerekli literatürü taramak / *Review the literature required for the study*	Res. Asst. Ali Hasan DOGAN Res. Asst. Cemali ALTUNTAS Res. Asst. Caneren GUL Prof. Nursu TUNALIOGLU (PhD) Prof. Bahattin ERDOGAN (PhD)

Hakem Değerlendirmesi: Dış bağımsız.

Çıkar Çatışması: Yazar çıkar çatışması bildirmemiştir.

Finansal Destek: Yazar bu çalışma için finansal destek almadığını beyan etmiştir.

Teşekkür: -

Peer-review: Externally peer-reviewed.

Conflict of Interest: The author has no conflict of interest to declare.

Grant Support: The author declared that this study has received no financial support.

Acknowledgement: -