ECRONICON OPEN ACCESS

EC PULMONOLOGY AND RESPIRATORY MEDICINE

Research Article

Determination of the Reliability of Covid-19 Data in the Republic of North Macedonia Using Benford's law

Dejan Zdraveski[1]*, Margarita Janeska[1] and Petar Avramovski[2]

[1]University "St. Kliment Ohridski" - Bitola, Faculty of Economics-Prilep, Department of Business Informatics, Republic of North Macedonia

[2]University "St. Kliment Ohridski" - Bitola, Clinical Hospital-Bitola, Department of Internal Medicine, Republic of North Macedonia

***Corresponding Author:** Dejan Zdraveski, University "St. Kliment Ohridski" - Bitola, Faculty of Economics-Prilep, Department of Business Informatics, Republic of North Macedonia.

Received: November 26, 2021; **Published:** December 30, 2021

Abstract

Coronavirus or COVID-19 is a highly contagious and pathogenic viral infection caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which has caused a global pandemic that has led to dramatic loss of human life worldwide.

The accuracy and indisputability of Covid-19 data should be one of the basic principles of the Institute of Public Health and the Ministry of Health in every national economy. These attributes are achieved through the use of validated method for collection and processing data, attested procedures, and balanced calculations using all available sources of information.

Benford's law is used to detect anomalies in the large amounts of data. The basic premise is that when the distribution of leading digits in a data set differs significantly from the expected distribution under Benford's law, this anomaly may be an indication of fraud or manipulation. Thus, Benford's Law is an important control mechanism that can contribute to increase of data integrity and information quality.

The purpose of this paper was to understand the realiability of Covid-19 data regarding the number of tests per day, the number of positive cases, the number of deaths and the number of healed persons published by the Institute of Public Health in Republic of North Macedonia. In this paper was examined reliability of data by applying Benford's law and the corresponding tests to assess the compliance of the distribution of the leading digits of this data with the distribution of the same according to Benford's law. The presence of anomalies in the reports from the Center for Public Health will be identified through five tests: z-statistics, MAD-Mean absolute deviation test, χ^2 - test, Kolmogorov-Smirnov test and Kuiper test. Microsoft Excel will be used to analyze the data.

Keywords: *Benford's Law; First Digit Test; Anomalies; Covid-19; Statistical Tests*

Introduction

At the end of 2019, the world began to face a new type of corona virus that began to spread from China andin the next two yearthroughout the world. There is no country in the world that has not registered a case of a new type of coronavirus called SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2).

Like many other countries in Europe, the pandemic in Republic of North Macedonia began in early 2020, with the first case being a woman infected with the virus arriving from Italy. The number of infected people gradually increased and until today the number is over 200,000 infected with SARS-CoV-2, which represents about 10% of the population in North Macedonia.

The fight against this virus in North Macedonia is carried out in different ways through the use of restrictive measures, and especially it is emphasized with the vaccination process that began in February 2021, exactly one year after the appearance of the first case. Today the level of vaccinated in North Macedonia is quite low and is just under 50% of the total population, compared to Western European countries where the vaccination rate exceeds 80%.

The Institute of Public Health publishes daily, weekly and monthly reports on the number of tests performed, SARS-CoV-2 positives, SARS-CoV-2 deaths, and SARS-CoV-2 healed. Very often the citizens of North Macedonia are skeptical about these data in their reliability. Therefore, we used statistical tests and Benford's law, to determine the reliability of these data published by the Institute of Public Health of the Republic of North Macedonia.

The importance of this paper is in overcoming the mistrust that citizens have regarding the published data for the previously mentioned categories.

Materials and Methods

The research used data that are publicly available on the website of the Institute of Public Health in the Republic of North Macedonia, for the daily reported tests performed, cases of Covid-19 patients, deceased and healed persons, from 01.04.2020until 16.10.2021. It should be noted that for several days of the year there was no data at all (for 4 days), and for some days, there was no data on the number of dead and/or healed persons. There were also data with zero value (number of deaths and number of healed persons), which do not participate in the analysis of the data. Therefore, the total number of data on tests and infected persons is the same, and the total number of data on dead and healed persons is smaller and different.

Measuring the quality of the reports underlying Benford's law can be done with a variety of statistical tests. During the testing, two hypotheses are opposed [13]

- H_0 (data follow the Benford distribution) and
- H_1 (data deviate from the Benford distribution).

Depending on the test results, we either accept or reject the null hypothesis. The testing process takes place through the following steps

- Calculation of sample statistics test,
- Selection of the level of significance, on the basis of which the critical P-value is determined,
- If the value of the sample statistics is greater than the critical value, then the null hypothesis for the selected significance level is rejected. Otherwise, the null hypothesis is accepted.

Statistical tests are used to confirm one of the hypotheses. The following tests were used for the research in this paper

- MAD (Mean absolute deviation) test,
- Z -test,
- χ^2 -test,
- Kolmogorov-Smirnov test and
- Kuiper test.

Theoretical framework of benford's law

Benford's law is an observation of the leading digits of numbers found in real data sets. Intuitively, it can be expected that the leading digits of these numbers will be equally distributed so that each of the digits from 1 to 9 will appear with equal probability. In fact, it often happens that the number 1 occurs more often than the number 2, the number 2 more often than the number 3 and so on. This observation is a simplified version of Benford's law. Specifically, Benford's law predicts the frequency of leading digits using logarithms based on 10, which predicts specific frequencies that decrease as the digits increase from 1 to 9. This phenomenon usually occurs in very different cases of real-world data. The most famous version of Benford's law is related to the distribution of the first leading digits, which states that [1],

$$P(D) = \log(1+\frac{1}{D})\ D \in \{1,2,\ldots,9\},$$

Where D is random variable; $D : R \to \{1,2,\ldots,9\}$ and indicates the first significant digit of the randomly selected real number.

Calculating according to the previous formula, the following values are obtained for the probabilities of the first digit by positions.

The probability that the first digit is 1

$$P(D_1 = 1) = \log(1+\frac{1}{1}) = \log 2 = 0{,}301029995$$

The probability that the first digit is 2

$$P(D_1 = 2) = \log(1+\frac{1}{2}) = \log\frac{3}{2} = 0{,}176091259, \text{ etc.}$$

Benford's law of the first digit is that the probability of a given first digit d is proportional to the space between d and d + 1 on a logarithmic scale. That's about 30% for digit 1, but less than 5% for digit 9.

$$P(D_2 = d_2) = \sum_{j=1}^{9}\log(1+\frac{1}{10j+d_2})\ d_2 \in \{0, 1, 2, \ldots, 9\}$$

Thus, the probabilities of all digits of the individual position are calculated and presented in table 1.

Digit	1st place	2nd place	3rd place	4th place
0		0,11968	0,10178	0,10018
1	0,30103	0,11389	0,10138	0,10014
2	0,17609	0,10882	0,10097	0,10010
3	0,12494	0,10433	0,10057	0,10006
4	0,09691	0,10031	0,10018	0,10002
5	0,07918	0,09668	0,09979	0,09998
6	0,06695	0,09337	0,09940	0,09994
7	0,05799	0,09035	0,09902	0,09990
8	0,05115	0,08757	0,09864	0,09986
9	0,04576	0,08500	0,09827	0,09982

Table 1: *First, Second, Third, and Fourth-Digit Proportions of Benford's Law [6].*

According to Benford's law, the first digit in natural numbers is more likely to be a small digit than a large digit. For example, the number 1 is 6.58 times more likely to appear in the first position than the number 9.

The following figure shows the distribution of relative frequencies of the first significant digits according to Benford's law.

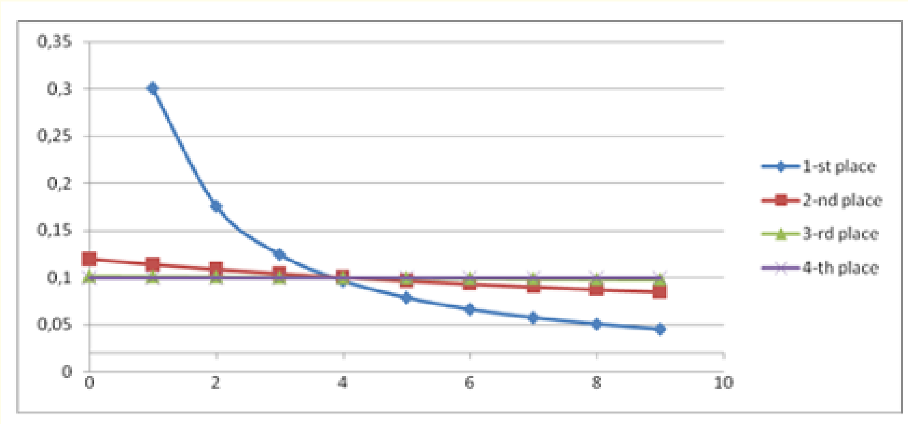

Figure 1: *Distribution of relative frequencies of the first significant digits according to Benford's law.*

It is important to note that Benford's law does not apply to all numerical populations. The data must meet the following criteria for law enforcement [7].

- All data should measure the same phenomena (as in the research all data refer to covid-19),
- There should be no minimum or maximum values embedded in the data set, except for the number zero, which is an acceptable minimum. The presence of a minimum or a maximum will disturb the frequencies of the digits,
- The number of data should have four or more digits for a good fit in Benford's law;
- A large data set is required to fit the Benford distribution. The small sample can not make the expected Benford frequencies, which will cause deviations from the law. As the data set increases, the expected frequencies of Benford's law are shown to be obtained,
- The data should not be composed of data that follow a predefined system.

Results of data analysis for Covid-19 in Republic of North Macedonia

Analysis of data on the number of tests per day (number of tested persons)

The analysis is presented in the following tables

First digit	Benford'Law	Actual relative f	Actual absolute f	Expected absolute f	Difference	Abs. Difference	Z-score	Hi-square
1	0,30103	0,2803571	157	168,5768	-2,0672857	2,0672857	1,0664988	0,7950222
2	0,17609	0,2571429	144	98,6104	8,1052857	8,1052857	5,0356450	20,892480
3	0,12494	0,2035714	114	69,9664	7,8631429	7,8631429	5,6275649	27,712701
4	0,09691	0,0696429	39	54,2696	-2,7267143	2,7267143	2,1811415	4,2963406
5	0,07918	0,0517857	29	44,3408	-2,7394286	2,7394286	2,4008160	5,3075304
6	0,06695	0,0410714	23	37,492	-2,5878571	2,5878571	2,4502284	5,6016767
7	0,05799	0,05	28	32,4744	-0,799	0,799	0,8089776	0,6164935
8	0,05115	0,0214286	12	28,644	-2,9721429	2,9721429	3,1925822	9,6712308
9	0,04576	0,025	14	25,6256	-2,076	2,076	2,35098331	5,2742014
	1	1	560			3,5485397		80,167676

Table 2: *Calculated values for Z - test, χ^2 - test, MAD (Mean absolute deviation) test for the number of tests per day.*

Cumulative Actual	Cumulative Benford'Law	Abs Difference	Dn+	Dn-	Kuiper's test
0,280357143	0,30103	0,020672857	0,020672857	-0,02067286	
0,5375	0,47712	0,06038	-0,08105286	0,081052857	
0,741071429	0,60206	0,139011429	-0,07863143	0,078631429	
0,810714286	0,69897	0,111744286	0,027267143	-0,02726714	
0,8625	0,77815	0,08435	0,027394286	-0,02739429	
0,903571429	0,8451	0,058471429	0,025878571	-0,02587857	
0,953571429	0,90309	0,050481429	0,00799	-0,00799	
0,975	0,95424	0,02076	0,029721429	-0,02972143	
1	1	0	0,02076	-0,02076	
		0,139011429	0,029721429	0,081052857	0,110774286

Table 3: *Calculated values for Kolmogorov-Smirnov test and Kuiper test for the number of tests per day.*

MAD (Mean absolute deviation) test

The critical value of this test for the first digit (for which we are examining) is 0. 015.The calculated value is 3,549 and it is much higher than the critical one, which means that according to this test the null hypothesis is rejected.

Z-test

From the data in table 2 it can be seen that except for digit 1 and digit 7, in all digits there is a greater or lesser deviation (the largest is in digit 2 and digit 3), i.e., the value of z is greater than 1.96, which means that according to this test the distribution of data on the number of daily tests does not follow the Benford distribution

χ^2 - test

The value of the calculated χ^2 - test is 80.17, which is higher than the limit value for this test (15.51), which means that the null hypothesis is rejected for this test.

Kolmogorov-Smirnov test

The critical value for this test is

$$KS = \frac{1{,}148}{\sqrt{n}} = \frac{1{,}148}{\sqrt{560}} = 0{,}048$$

Since the calculated value 0.139 is greater than the critical value, according to this test the null hypothesis is rejected.

Kuiper test

The critical value for this test is

$$K = \frac{1{,}321}{\sqrt{n}} = \frac{1{,}321}{\sqrt{560}} = 0{,}056$$

Because the calculated value of 0.111 is greater than the critical value, the null hypothesis is rejected according to this test.

The distribution of digits, the deviation from Benford's law and the cumulative relative frequencies for the number of tests per day are presented in the following figure.

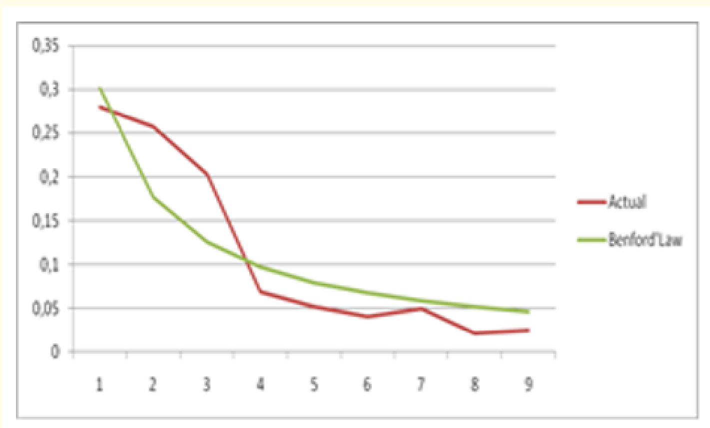

Figure 2: *The distribution of digits and the deviation from Benford's lawfor the number of tests per day.*

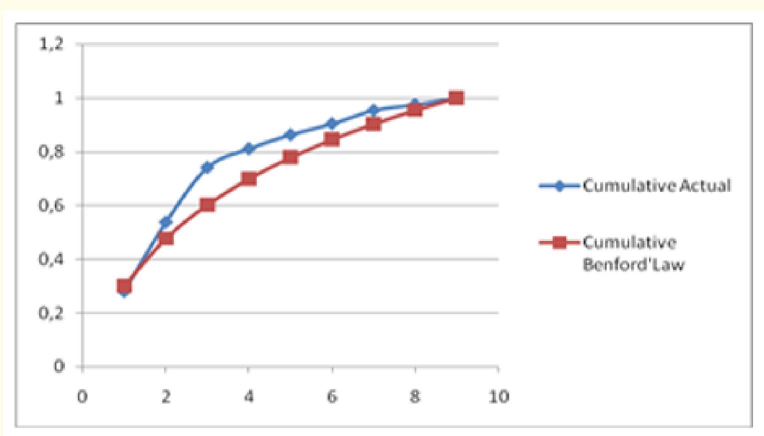

Figure 3: *Cumulative relative frequencies for the number of tests per day.*

Analysis of data on the number of positive cases of Covid-19 per day

The analysis of data for number of positive cases per day is presented in the following tables.

First digit	Benford'Law	Actual relative f	Actual absolute f	Expected absolute f	Difference	Abs. Difference	Z-score	Hi-square
1	0,30103	0,3928571	220	168,5768	9,1827143	9,182714286	4,737300432	15,68629549
2	0,17609	0,0946429	53	98,6104	-8,1447143	8,144714286	5,060141177	21,09623922
3	0,12494	0,1285714	72	69,9664	0,3631429	0,363142857	0,259897351	0,059107357
4	0,09691	0,0910714	51	54,2696	-0,5838571	0,583857143	0,467036473	0,196984761
5	0,07918	0,075	42	44,3408	-0,418	0,418	0,366332268	0,123573428
6	0,06695	0,0553571	31	37,492	-1,1592857	1,159285714	1,097631971	1,124134855
7	0,05799	0,0446429	25	32,4744	-1,3347143	1,334714286	1,351381631	1,720329101
8	0,05115	0,0660714	37	28,644	1,4921429	1,492142857	1,602812858	2,437604245
9	0,04576	0,0517857	29	25,6256	0,6025714	0,602571429	0,682386981	0,444343756
	1	1	560			2,586793651		42,88861222

Table 4: Calculated values for Z - test, χ^2 - test, MAD (Mean absolute deviation) test for the number of positive cases.

Cumulative Actual	Cumulative Benford'Law	Abs Difference	Dn+	Dn-	Kuiper's test
0,392857143	0,30103	0,091827143	-0,09182714	0,091827	
0,4875	0,47712	0,01038	0,081447143	-0,08145	
0,616071429	0,60206	0,014011429	-0,00363143	0,003631	
0,707142857	0,69897	0,008172857	0,005838571	-0,00584	
0,782142857	0,77815	0,003992857	0,00418	-0,00418	
0,8375	0,8451	0,0076	0,011592857	-0,01159	
0,882142857	0,90309	0,020947143	0,013347143	-0,01335	
0,948214286	0,95424	0,006025714	-0,01492143	0,014921	
1	1	0	-0,00602571	0,006026	
		0,091827143	0,081447143	0,091827	0,173274286

Table 5: *Calculated values for Kolmogorov-Smirnov test and Kuiper test for the number of positive cases.*

MAD (Mean absolute deviation) test

The critical value of this test for the first digit (for which we are examining) is 0.015.The calculated value is 2,587 and it is much higher than the critical one, which means that according to this test the null hypothesis is rejected.

Z-test

From the data in the table 4, it can be seen that in digit 1 and digit 2 there is a statistically significant deviation, i.e., the value of z is greater than 1.96, which means that the data does not follow the Benford distribution and the null hypothesis is rejected.

χ^2 - test

The value of the calculated χ^2 - test is 42.89, that is higher than the limit value for this test (15.51), which means that according to this test the null hypothesis is rejected.

Kolmogorov-Smirnov test

The critical value for this test is

$$KS = \frac{1{,}148}{\sqrt{n}} = \frac{1{,}148}{\sqrt{560}} = 0{,}048$$

Since the calculated value 0.092 is greater than the critical value, according to this test the null hypothesis is rejected.

Kuiper test

The critical value for this test is

$$K = \frac{1{,}321}{\sqrt{n}} = \frac{1{,}321}{\sqrt{560}} = 0{,}056$$

Because the calculated value 0.173 is greater than the critical value, the null hypothesis is rejected according to this test.

The distribution of digits, the deviation from Benford's law and the cumulative relative frequencies for the number of positive cases are presented in the following figure.

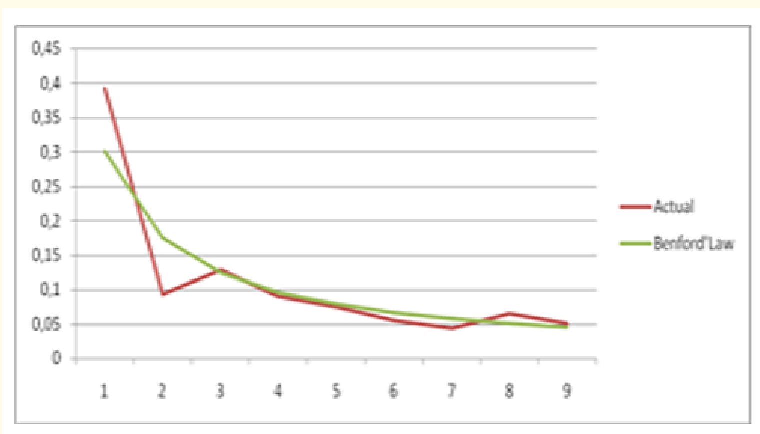

Figure 4: *The distribution of digits and the deviation from Benford's law for the number of positive cases.*

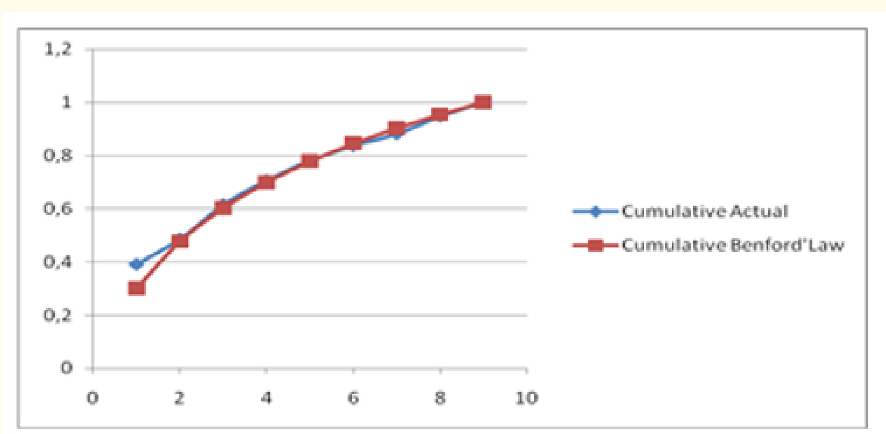

Figure 5: *Cumulative relative frequencies for the number of positive cases.*

Analysis of data on the number of deaths from Covid-19 per day

The analysis is presented in the following tables.

First digit	Benford'Law	Actual relative f	Actual absolute f	Expected absolute f	Difference	Abs. Difference	Z-score	Hi-square
1	0,30103	0,283730159	143	151,71912	-1,729984127	1,729984127	0,846687658	0,501077607
2	0,17609	0,23015873	116	88,74936	5,406873016	5,406873016	3,186795568	8,367354766
3	0,12494	0,158730159	80	62,96976	3,379015873	3,379015873	2,294224077	4,605846909
4	0,09691	0,095238095	48	48,84264	-0,167190476	0,167190476	0,126875273	0,014537342
5	0,07918	0,05952381	30	39,90672	-1,965619048	1,965619048	1,634253786	2,459312646
6	0,06695	0,055555556	28	33,7428	-1,139444444	1,139444444	1,023483096	0,977386341
7	0,05799	0,043650794	22	29,22696	-1,433920635	1,433920635	1,377323863	1,787012773
8	0,05115	0,03968254	20	25,7796	-1,146746032	1,146746032	1,168586633	1,295744548
9	0,04576	0,033730159	17	23,06304	-1,202984127	1,202984127	1,292419043	1,593911906
	1	1	504			1,952419753		21,60218484

Table 6: *Calculated values for Z - test, χ^2 - test, MAD (Mean absolute deviation) test for the number of deaths.*

Cumulative Actual	Cumulative Benford'Law	Abs Difference	Dn+	Dn-	Kuiper's test
0,283730159	0,30103	0,017299841	0,017299841	-0,0173	
0,513888889	0,47712	0,036768889	-0,05406873	0,054069	
0,672619048	0,60206	0,070559048	-0,03379016	0,03379	
0,767857143	0,69897	0,068887143	0,001671905	-0,00167	
0,827380952	0,77815	0,049230952	0,01965619	-0,01966	
0,882936508	0,8451	0,037836508	0,011394444	-0,01139	
0,926587302	0,90309	0,023497302	0,014339206	-0,01434	
0,966269841	0,95424	0,012029841	0,01146746	-0,01147	
1	1	0	0,012029841	-0,01203	
		0,070559048	0,01965619	0,054069	0,073724921

Table7: *Calculated values for Kolmogorov-Smirnov test and Kuiper test for the number of deaths.*

MAD (Mean absolute deviation) test

The critical value of this test for the first digit (for which we are examining) is 0.015.The calculated value is 1.9524 and it is greater than the critical one, which means that according to this test the null hypothesis is rejected.

Z-test

It can be seen in the presented data in table 6 that there is a deviation only in digit 2 and digit 3 (not as high as in the previous research), i.e., the value of z is greater than 1.96, which means that the data do not follows the Benford distribution, so null hypothesis is rejected.

χ^2 - test

The value of the calculated χ^2 - test is 21.6, and it is slightly higher than the limit value for this test (15.51), which means according to this test the null hypothesis is rejected.

Kolmogorov-Smirnov test

The critical value for this test is

$$KS = \frac{1,148}{\sqrt{n}} = \frac{1,148}{\sqrt{504}} = 0,051$$

Since the calculated value 0.071 is very slightly greater (can be said to be approximately equal) than the critical value, according to this test the null hypothesis is accepted.

Kuiper test

The critical value for this test is

$$K = \frac{1,321}{\sqrt{n}} = \frac{1,321}{\sqrt{504}} = 0,059$$

Since the calculated value is 0.074 it is very slightly greater (can be said to be approximately equal) than the critical value so according to this test the null hypothesis is accepted.

The distribution of digits, the deviation from Benford's lawand the cumulative relative frequencies for number of deaths are presented in the following figure

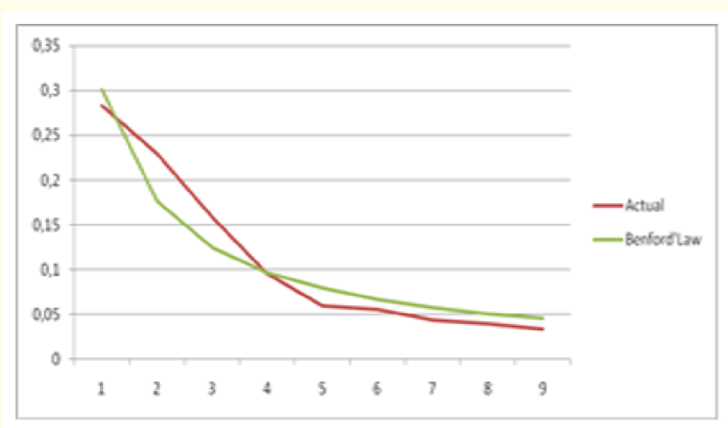

Figure 6: *The distribution of digits and the deviation from Benford's law for the number of deaths.*

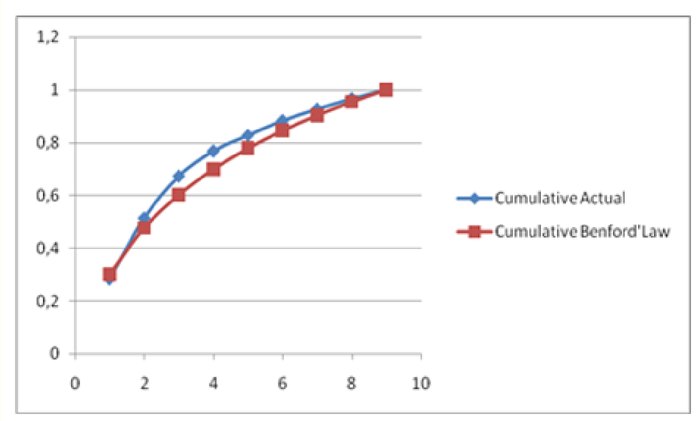

Figure 7: *Cumulative relative frequencies for the number of deaths.*

Analysis of data on the number of healed persons from Covid-19 per day

Data analysis is presented in following table.

First digit	Benford'Law	Actual relative f	Actual absolute f	Expected absolute f	Difference	Abs. Difference	Z-score	Hi-square
1	0,30103	0,288354898	156	162,85723	-1,267510166	1,267510166	0,642711287	0,288728988
2	0,17609	0,125693161	68	95,26469	-5,039683919	5,039683919	3,077476164	7,803135882
3	0,12494	0,118299445	64	67,59254	-0,664055453	0,664055453	0,467125217	0,190943315
4	0,09691	0,11090573	60	52,42831	1,399573013	1,399573013	1,100384283	1,093502527
5	0,07918	0,092421442	50	42,83638	1,324144177	1,324144177	1,140614235	1,197987587
6	0,06695	0,086876155	47	36,21995	1,992615527	1,992615527	1,854361463	3,208438388
7	0,05799	0,068391867	37	31,37259	1,040186691	1,040186691	1,035155549	1,009407999
8	0,05115	0,053604436	29	27,67215	0,245443623	0,245443623	0,259136617	0,063716973
9	0,04576	0,055452865	30	24,75616	0,969286506	0,969286506	1,078894509	1,110748111
	1	1	541			1,549166564		15,96660977

Table 8: *Calculated values for Z - test, χ^2 - test, MAD (Mean absolute deviation) test for healed persons.*

Cumulative Actual	Cumulative Benford'Law	Abs Difference	Dn+	Dn-	Kuiper's test
0,288354898	0,30103	0,012675102	0,012675102	-0,01268	
0,414048059	0,47712	0,063071941	0,050396839	-0,0504	
0,532347505	0,60206	0,069712495	0,006640555	-0,00664	
0,643253235	0,69897	0,055716765	-0,01399573	0,013996	
0,735674677	0,77815	0,042475323	-0,01324144	0,013241	
0,822550832	0,8451	0,022549168	-0,01992616	0,019926	
0,890942699	0,90309	0,012147301	-0,01040187	0,010402	
0,944547135	0,95424	0,009692865	-0,00245444	0,002454	
1	1	0	-0,00969287	0,009693	
		0,069712495	0,050396839	0,019926	0,070322994

Table 9: *Calculated values for Kolmogorov-Smirnov test and Kuiper test for healed persons.*

MAD (Mean absolute deviation) test

The critical value of this test for the first digit is 0.015. The calculated value is 1.5492 and it is greater than the critical one, which means that according to this test the null hypothesis is rejected.

Z-test

It can be seen from the data in table 8 that there is a deviation only in digit 2 which is not so high, i.e., the value of z is greater than 1.96, which means that the data follow the Benford distribution and the null hypothesis can be accepted.

χ^2 - test

The value of the calculated χ^2 - test is 15.97, which is approximately equal to the limit value for this test (15.51). That means, the null hypothesis can be accepted for this test.

Kolmogorov-Smirnov test

The critical value for this test is

Since the calculated value is 0.070, it can be said that it is slightly greater than the critical value, which means that according to this test the null hypothesis can be accepted.

Kuiper test

The critical value for this test is:

$$KS = \frac{1{,}148}{\sqrt{n}} = \frac{1{,}148}{\sqrt{541}} = 0{,}049$$

Since the calculated value is 0.070, which is minimally greater than the critical value (almost equal), according to this test the null hypothesis is accepted.

The distribution of digits, the deviation from Benford's lawand the cumulative relative frequencies for healed are presented in the following figure.

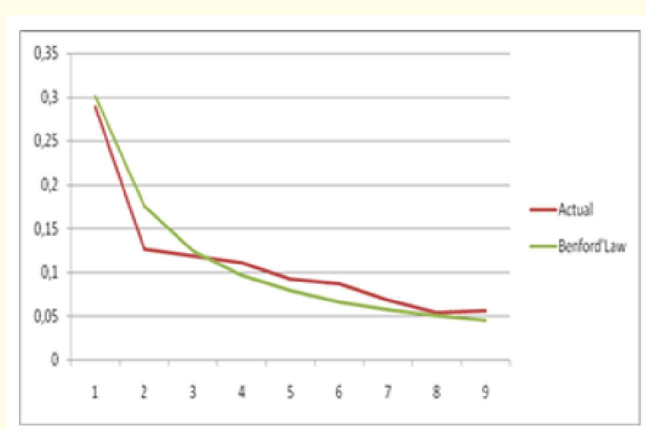

***Figure 8:** The distribution of digits and the deviation from Benford's law for the number of healed persons.*

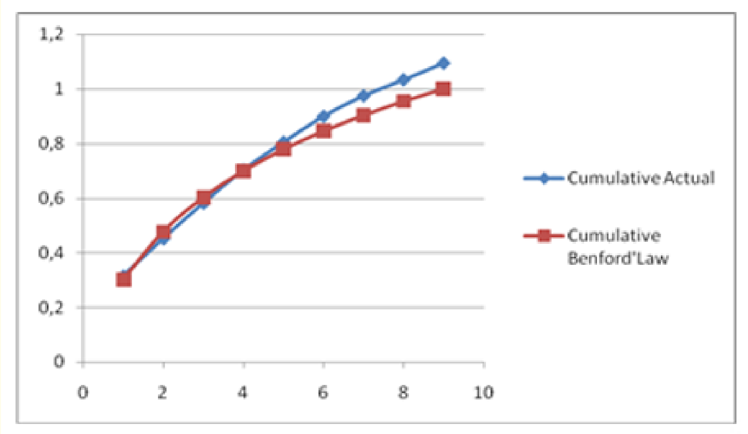

Figure 9: *Cumulative relative frequencies for the number of healed persons.*

Given the fact that all data refers to Covid-19, a summary analysis of the data can be performed. This analysis is presented in the following tables.

First digit	Benford'Law	Actual relative f	Actual absolute f	Expected absolute f	Difference	Abs. Difference	Z-score	Hi-square
1	0,30103	0,312240185	676	651,72995	1,121018476	1,121018476	1,137123	0,903802759
2	0,17609	0,175981524	381	381,23485	-0,010847575	0,010847575	0,013251	0,000144673
3	0,12494	0,152424942	330	270,4951	2,748494226	2,748494226	3,867712	13,09019322
4	0,09691	0,091454965	198	209,81015	-0,545503464	0,545503464	0,857979	0,664789778
5	0,07918	0,069745958	151	171,4247	-0,943404157	0,943404157	1,625667	2,433537116
6	0,06695	0,059584296	129	144,94675	-0,736570439	0,736570439	1,371246	1,754429372
7	0,05799	0,051732102	112	125,54835	-0,625789838	0,625789838	1,245814	1,462048587
8	0,05115	0,045265589	98	110,73975	-0,588441109	0,588441109	1,242826	1,465609504
9	0,04576	0,041570439	90	99,0704	-0,41895612	0,41895612	0,93288	0,830441344
	1	1	2165			0,859891712		22,60499635

Table 10: *Calculated values for Z - test, χ^2 - test, MAD (Mean absolute deviation) test for summary data.*

Cumulative Actual	Cumulative Benford'Law	Abs Difference	Dn+	Dn-	Kuiper's test
0,312240185	0,30103	0,011210185	-0,01121	0,011210185	
0,488221709	0,47712	0,011101709	0,000108	-0,00010848	
0,640646651	0,60206	0,038586651	-0,02748	0,027484942	
0,732101617	0,69897	0,033131617	0,005455	-0,00545503	
0,801847575	0,77815	0,023697575	0,009434	-0,00943404	
0,861431871	0,8451	0,016331871	0,007366	-0,0073657	
0,913163972	0,90309	0,010073972	0,006258	-0,0062579	
0,958429561	0,95424	0,004189561	0,005884	-0,00588441	
1	1	0	0,00419	-0,00418956	
		0,038586651	0,009434	0,027484942	0,036918984

Table 11: *Calculated values for Kolmogorov-Smirnov test and Kuiper test for summary data.*

MAD (Mean absolute deviation) test

The critical value of this test for the first digit is 0.015.The calculated value is 0.86 and it is greater than the critical one, which means that according to this test the null hypothesis is rejected. However, it should be noted that the calculated value of the MAD test has the lowest value in the summary data analysis.

Z-test

It can be seen from the data in table 10 that there is a deviation only in digit 3 i.e., the value of z is greater than 1.96, which means that the data does not follow the Benford distribution, so the null hypothesis is rejected.

χ^2 - test

The value of the calculated χ^2 - test is 22.60, and it is higher than the limit value for this test (15.51), which means that according to this test the null hypothesis is rejected.

Kolmogorov-Smirnov test

The critical value for this test is

$$K = \frac{1{,}321}{\sqrt{n}} = \frac{1{,}321}{\sqrt{541}} = 0{,}057$$

Since the calculated value is 0.039 and it is slightly greater (almost equal) than the critical value, according to this test the null hypothesis is accepted.

Kuiper test

The critical value for this test is

Since the calculated value is 0.0369 and it is slightly greater than the critical value, according to this test the null hypothesis can be accepted.

The distribution of digits, the deviation from Benford's lawand the cumulative relative frequencies for summary data are presented in the following figure

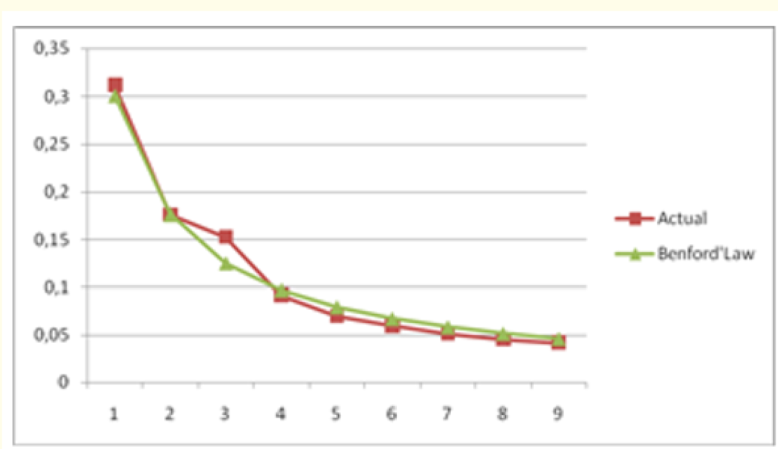

Figure 10: *The distribution of digits and the deviation from Benford's law for the summary data.*

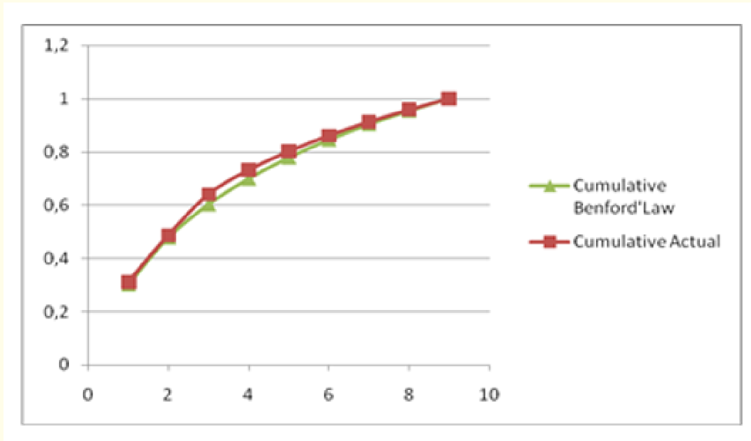

Figure 11: *Cumulative relative frequencies for the number of healed persons.*

Conclusion

The paper is an attempt to create a database of potentially erroneous data relating to Covid-19 in terms of the number of tests, the number of infected people, the number of deaths and the number of healed peoples per day in North Macedonia based on Benford's law (first digit law).Based on the data published by the Institute of Public Health in North Macedonia, the Benford distribution and the tests used, the following general conclusions were reached

- The distribution of the first digit of the data on the number of daily tests for Covid-19 does not follow the Benford distribution according to all tests, because in 7 digits there are statistically significant deviation,
- The distribution of the first digit of the data on the number of positive cases of Covid-19 according to all tests does not follow the Benford distribution (although there is a deviation of only two digits, those deviations are high and statistically significant),
- The distribution of the first digit of the data on the number of deaths per day from Covid-19 is that it follows the Benford distribution according to 2 tests: Kolmogorov and Kuiper, and according to three tests: z-test, χ^2 - test and MAD test does not follow the Benford distribution (only 2 digits have deviations, which are statistically significant, but do not have high values),
- The distribution of the first digit of the data on the number of healed persons per day from Covid-19 follows the Benford distribution according to 3 tests: z-test, Kolmogorov and Kuiper, and according to two tests: χ^2 - test and MAD test does not follow the Benford distribution (only there is a statistically significant deviation in one digit),
- The distribution of the first digit of the total data relating to Covid-19 follows the Benford distribution according to the Kolmogorov and Kuiper tests, and does not follow the z-test, χ^2 - test and MAD test (although only 1 digit has a deviation that is statistically significant, it has a higher value than the deviation in the number of recovered persons).

Given that the χ^2 - test has a small statistical "strength", the Kolmogorov-Smirnov test and the Kuiper test a large statistical "strength" when used to test the compliance of small samples with Benford's law, it can be concluded that the data sets that are published in terms of deaths and healed persons are compliant with Benford's law, i.e., their reliability is on high level. In other words, the data referring to the deaths and healed peoples does not contain anomalies and they are completely real.

The data sets published in terms of the number of tests and the number of people infected with Covid-19 are not compliant with Benford's law, i.e., their reliability is on very low level. This means that this data set, which refers to the number of tests per day and the number of positive cases, contains anomalies that make this data inaccurate. The factors that influence this data set to contain anomalies can be numerous and will be the subject of research in future scientific papers.

This study showed that in order to get a complete picture of the compliance of the Covid-19 data distribution with the Benford distribution, more tests need to be applied. This would be the next phase in which this research would take place.

Bibliography

1. Berger A., *et al.* "A basic theory of Benford's Law". *Probability Surveys* 8 (2011): 1-126.
2. Daniel P Pike "Testing for the Benford Property". *Rochester Institute of technology*.
3. Díaz Cusi J., *et al*. "On the Ability of the Benford's Law to Detect Earthquakes andDiscriminate Seismic Signals". *Seismological Research Letters* 86.1 (2015).
4. Manoochehrnia Pooyan., *et al*. "Benford's Law and Its Application to Lightning Data". Transactions on Electromagnetic Compability 52.4 (2010).
5. Morrow J. "Benford's law, families of distributions and a test basis" (2020).
6. Nigrini MJ. "A Taxpayer Compliance Application of Benford's Law". *Journal of the American Taxation Association* 18.1 (1996): 74.
7. Nigrini MJ. "Digital analysis using Benford's law: tests statistics for auditors". Canada, Vancouer: Global Audit Publications (2000).
8. Nigrini MJ. "Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations". *John Wiley and Sons, New Jersey* (2011).
9. Nigrini MJ. "Benford's Law: Application for Forensic Accounting, Auditing and Fraud Detection". *Wiley and Sons: New Jersey* (2012).
10. Sanches João M and i Marques Jorge S. "Image Reconstruction Using the Benford's Law". *International Conference on Image Processing* (2006).
11. Stefan Engel. "Fact and Fiction in EU-Governmental Economic Date" Catholic University of Eichstätt-Ingolstadt. German Economic Review (2011).
12. Wells Kevin., *et al*. "Quantifying the Partial Volume Effect in PET Using Benford's Law". *Transactions on Nuclear science* 54.5 (2007).
13. ZagorkaLozanov-Crvenkovic, Statistika,Novi Sad (2012).

Volume 11 Issue 1 January 2022