

# The Full Database of Countries with Potential COVID-19 Data Misreport based on Benford's Law

<sup>1</sup>Ahmad Kilani and <sup>2</sup>Georgios P. Georgiou

<sup>1</sup>Department of Comparative Politics, RUDN University, Moscow, Russia

<sup>2</sup>Department of Languages and Literature, University of Nicosia, Nicosia, Cyprus

## Abstract

The aim of this database is to provide researchers and scholars a unified database for potential data misreport by 171 countries regarding their COVID-19 daily reported cases. The analysis employs three different tests (chi-square, Kuiper, and Mean Absolute Deviation) to determine if the data given by each country in the world fit Benford's Law.

## 1. Benford's law

Benford's Law (BL) is extensively used to test the legitimacy of data in numerous applications, such as in finance, economics, and politics. BL indicates that the first digit of a naturally transpire decimal number is more likely to be equal to 1, and the possibilities of the first digit to be equal to the subsequent numbers, decrease progressively.

The formula of BL is as follows, where  $D$  represents the digits between 1 and 9 and  $P$  is the probability:

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

Using this formula, the expected frequencies for digits in first position are illustrated below:

Digit	1	2	3	4	5	6	7	8	9
Expected Frequency	.30103	.17609	.12494	.09691	.07918	.06695	.05799	.05115	.04576

## 2. Tests

To measure the difference between the observed and expected first digit BL distribution for a country we used a *chi-square* test, which is based on the formula below:

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

Where  $O_i$  is the observed and  $E_i$  is the expected absolute frequencies for digit  $i$ .

The *Kuiper* test (a modified Kolmogorov–Smirnov test) (Kuiper 1960) ignores sample size and it is based on the following formula:

$$V = (D_n^+ + D_n^-)(\sqrt{n} + 0.155 + 0,24\sqrt{n})$$

where  $D_n^+ = \sup [F(O_i) - F(E_i)]$  and  $D_n^- = \sup [F(E_i) - F(O_i)]$ , and  $F(\cdot)$  stands for cumulated relative frequencies. The discrepancy statistics  $D^+$  and  $D^-$  characterize the absolute sizes of the differences between the two distributions being compared: the absolute and the observed distribution.

The *Mean Absolute Deviation (MAD)* test that also ignores sample size yet if the data sample size is small it may be inclined to false positives errors, when the results conclude nonconformity from unbiased data (Nigrini 2012). Henceforth, Nigrini's MAD can only be applied to large sample size i.e. the total daily reported cases in the world. Johnson and Weggenmann (2013) have adjusted the MAD calculation method to better solve the false positive problem; hence, an adjusted MAD can be used for country's data set. The MAD statistic is calculated as follows:

$$\text{Mean Absolute Deviation} = \frac{1}{N} \sum_1^N f_i |X_i - \bar{X}|$$

Where  $X_i$  is the difference between actual occurrence rate and Benford occurrence rate and  $\bar{X}$  is the mean of the difference between actual occurrence rate and Benford occurrence rate. Whereas in Nigrini's MAD,  $X_i$  is the actual occurrence rate and  $\bar{X}$  is the Benford occurrence rate. The absolute symbol means that the deviation is given a positive sign irrespective of whether it is positive or negative. Individual differences are then totaled and divided by 9 (the number of leading digits) to yield the mean absolute deviation (Johnson and Weggenmann 2013).

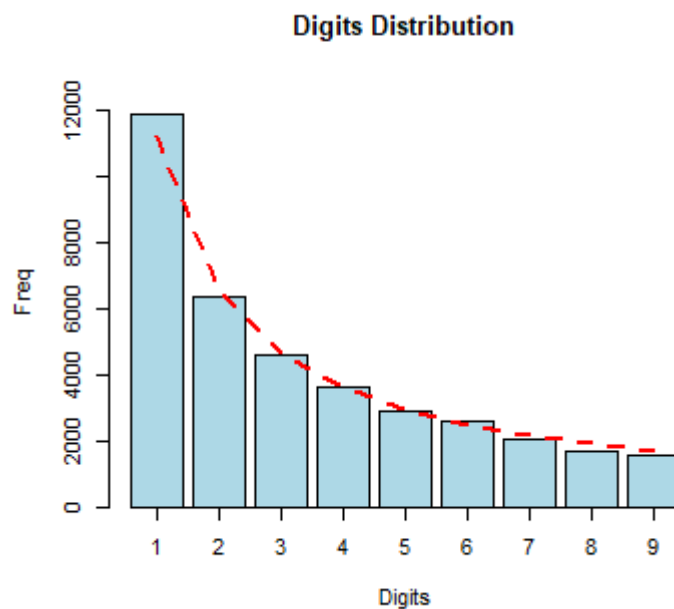
### 3. Data collection

We used publicly available data from the European Centre for Disease Prevention and Control website<sup>1</sup> on reported daily cases of Covid-19 in the world starting from the first case of Covid-19 in each country to the 12<sup>nd</sup> of November 2020.

The below Figure shows the Benfordness of the World's reported daily cases where the blue bars represent the world's observed daily reported cases and the intermittent red line represents BL.

---

<sup>1</sup> <https://www.ecdc.europa.eu/en>



#### 4. The database

The ranges of the tests are shown below. If the value of the chi-square and the Kuiper tests are more than 20.09 and 2 respectively, it is considered that a country's distribution curve does not fit BL. Similarly, in the MAD test, if the value is more than 0.015, the curve does not conform to BL.

$\alpha$ level	chi-square	Kuiper	MAD	MAD conformity ranges
0.01	>15.51	>1.65	>0.006	Close conformity
0.05	15.51 - 17.54	1.65-1.75	0.006-0.012	Acceptable conformity
0.1	17.54 - 20.09	1.75-2.0	0.012-0.015	Marginally acceptable conformity
none	<20.09	<2	<0.015	Nonconformity

<b>Countries</b>	<b>Chi-square</b>	<b>Kuiper</b>	<b>MAD<sup>2</sup></b>
Afghanistan	8.00	1.49	0.013
Albania	19.02	2.98	0.035
Algeria	12.05	1.84	0.021
Andorra	7.99	1.46	0.016
Angola	6.51	0.95	0.014
Argentina	21.77	3.31	0.039
Armenia	8.80	1.65	0.013
Australia	4.94	1.26	0.014
Austria	6.29	1.52	0.014
Azerbaijan	13.51	1.31	0.023
Bahrain	30.62	3.97	0.041
Bangladesh	26.13	3.83	0.040
Belarus	66.84	3.43	0.037
Belgium	8.72	2.04	0.019
Belize	12.78	1.25	0.021
Benin	11.56	1.09	0.015
Bhutan	23.15	2.33	0.033
Bolivia	14.22	2.41	0.026
Bosnia and Herzegovina	9.65	1.87	0.023
Botswana	13.26	1.05	0.021
Brazil	17.72	2.83	0.026
Brunei Darussalam	43.49	1.50	0.051
Bulgaria	8.35	1.97	0.020
Burkina Faso	9.58	1.67	0.023

Burundi	7.37	1.01	0.015
Cambodia	37.80	2.38	0.048
Cameroon	5.96	1.06	0.017
Canada	7.65	1.85	0.013
Cape Verde	9.77	1.53	0.015
Central African Republic	4.85	0.87	0.015
Chad	5.87	1.18	0.012
Chile	23.92	3.00	0.032
China	6.39	1.65	0.013
Colombia	42.10	3.89	0.041
Comoros	21.38	1.29	0.030
Congo	6.55	0.69	0.017
Costa Rica	14.48	2.59	0.031
Croatia	5.71	1.01	0.016
Cuba	17.37	2.89	0.031
Cyprus	6.16	1.65	0.014
Czech Republic	1.46	0.61	0.009
Denmark	6.18	1.13	0.014
Djibouti	5.07	1.09	0.011
Dominican Republic	8.31	1.77	0.020
East Timor	126.28	2.04	0.096
Ecuador	13.27	2.14	0.019
Egypt	41.88	4.62	0.051
El Salvador	9.94	0.99	0.019
Equatorial Guinea	18.09	1.41	0.027

Eritrea	11.04	1.17	0.021
Estonia	4.50	0.70	0.008
Ethiopia	7.46	1.68	0.013
Finland	9.48	2.12	0.020
France	6.95	1.65	0.013
Gabon	16.28	2.03	0.030
Gambia	12.91	1.43	0.021
Georgia	5.90	0.72	0.013
Germany	2.42	0.52	0.009
Ghana	4.44	1.00	0.011
Greece	11.87	2.07	0.018
Guatemala	10.31	1.91	0.015
Guinea	20.05	2.78	0.030
Guinea-Bissau	19.47	1.67	0.031
Guyana	3.80	0.96	0.011
Haiti	4.12	1.21	0.009
Honduras	18.37	3.16	0.031
Hungary	8.73	1.20	0.016
Iceland	4.03	1.21	0.010
India	26.49	3.58	0.035
Indonesia	26.39	2.93	0.030
Iran	56.76	4.51	0.046
Iraq	36.47	4.25	0.044
Ireland	2.20	1.02	0.011
Israel	3.78	1.23	0.011

Italy	7.54	1.67	0.013
Ivory Coast	15.22	2.55	0.028
Jamaica	4.14	0.84	0.008
Japan	15.21	2.70	0.031
Jordan	1.25	0.45	0.009
Kazakhstan	4.35	1.06	0.010
Kenya	4.52	1.15	0.011
Kosovo	5.50	1.35	0.018
Kuwait	90.41	6.18	0.074
Kyrgyzstan	6.46	1.53	0.015
Lao People's Democratic Republic	56.16	1.29	0.057
Latvia	3.28	0.83	0.008
Lebanon	8.18	1.78	0.021
Lesotho	5.69	0.68	0.012
Liberia	8.23	1.45	0.017
Libya	16.55	2.11	0.024
Liechtenstein	27.92	2.18	0.041
Lithuania	3.34	0.87	0.014
Luxembourg	9.98	1.62	0.018
Madagascar	11.31	1.41	0.021
Malawi	2.61	0.75	0.010
Malaysia	7.20	1.86	0.019
Maldives	1.66	0.61	0.005
Mali	8.02	1.85	0.016

Malta	5.30	1.38	0.015
Mauritania	5.45	1.09	0.014
Mauritius	20.34	1.34	0.034
Mexico	33.06	3.34	0.036
Moldova	3.16	0.67	0.013
Mongolia	36.55	2.35	0.045
Montenegro	5.79	1.00	0.011
Morocco	8.30	1.96	0.018
Mozambique	6.63	1.52	0.018
Myanmar	13.83	2.18	0.026
Namibia	3.89	1.09	0.015
Nepal	7.89	1.75	0.017
Netherlands	12.66	2.39	0.022
New Zealand	7.56	1.44	0.016
Nicaragua	43.15	1.94	0.049
Niger	7.65	1.25	0.015
Nigeria	7.96	1.37	0.010
North Macedonia	17.89	2.68	0.029
Norway	2.96	0.97	0.012
Oman	5.83	1.52	0.020
Pakistan	11.13	2.12	0.018
Palestine	33.47	3.87	0.045
Panama	44.52	4.04	0.041
Papua New Guinea	17.16	1.26	0.029
Paraguay	11.35	2.21	0.019



Peru	28.98	3.39	0.041
Philippines	21.62	3.04	0.030
Poland	23.74	3.25	0.036
Portugal	14.25	2.78	0.029
Qatar	40.65	3.68	0.039
Romania	11.81	2.35	0.022
Russia	60.91	4.79	0.057
Rwanda	2.00	0.65	0.008
Saudi Arabia	23.32	2.69	0.028
Senegal	14.81	2.61	0.029
Serbia	7.45	1.52	0.018
Sierra Leone	4.11	1.19	0.013
Singapore	4.70	1.50	0.016
Slovakia	5.06	1.50	0.015
Slovenia	7.27	1.62	0.016
Somalia	3.99	0.91	0.014
South Africa	11.45	2.20	0.024
South Korea	1.99	0.73	0.008
South Sudan	4.11	0.71	0.012
Spain	5.44	0.92	0.012
Sri Lanka	2.69	0.67	0.007
Sudan	5.48	0.97	0.012
Suriname	8.05	1.18	0.012
Swaziland	1.64	0.50	0.006
Sweden	11.71	2.05	0.018

Switzerland	1.42	0.48	0.008
Syria	33.29	2.83	0.034
Taiwan	32.18	2.93	0.040
Tajikistan	163.73	6.43	0.095
Tanzania	24.03	0.69	0.032
Thailand	8.89	1.24	0.015
The Democratic Republic Of The Congo	4.80	1.22	0.015
Togo	8.13	1.77	0.017
Trinidad and Tobago	2.34	0.67	0.005
Tunisia	7.94	1.61	0.017
Turkey	72.28	5.78	0.067
Uganda	15.25	2.43	0.027
Ukraine	18.84	2.82	0.029
United Arab Emirates	5.72	1.61	0.016
United Kingdom	4.82	1.07	0.011
United States	19.84	3.41	0.033
Uruguay	3.92	1.02	0.012
Uzbekistan	21.29	3.15	0.033
Venezuela	11.51	1.59	0.017
Vietnam	6.91	1.35	0.014
Yemen	15.11	1.42	0.018
Zambia	3.60	0.64	0.013
Zimbabwe	8.20	1.42	0.017

## References

- Johnson, G.G., and J. Weggenmann. 2013. "Exploratory research applying Benford's Law to selected balances in the financial statements of state governments." *Academy of Accounting and Financial Studies Journal* 17(3): 31-44.
- Kuiper, N. H. 1960. "Tests concerning random points on a circle. P, ." *roceedings of the Koninklijke Nederlandse Akademie van Wetenschappen Series A*. 63 (pp. 38–47).
- Nigrini, Mark J. 2012. *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. New Jersey: Wiley Corporate F&A.