

Aplicação da Lei de Benford nos números de casos confirmados de COVID-19 em diferentes países

Benford's Law applied to COVID-19 confirmed case numbers in different countries

Rhômulo Oliveira Menezes

Universidade Federal do Pará (UFPA), Instituto de Educação Matemática e Científica (IEMCI), Programa de Pós-Graduação em Educação em Ciências e Matemáticas (PPGECM), Belém, PA, Brasil

 <https://orcid.org/0000-0001-9042-8323>, rhomulo.menezes4542@escola.seduc.pa.gov.br

Informações do Artigo

Como citar este artigo

MENEZES, Rhômulo Oliveira. Aplicação da Lei de Benford nos números de casos confirmados de COVID-19 em diferentes países. **REMAT: Revista Eletrônica da Matemática**, Bento Gonçalves, RS, v. 7, n. 1, p. e3005, 5 fev. 2021. DOI: <https://doi.org/10.35819/remat2021v7i1id4586>



Histórico do Artigo

Submissão: 17 de setembro de 2020.
Aceite: 6 de dezembro de 2020.

Palavras-chave

COVID-19
Números de Casos Confirmados
Lei de Benford
Teste de Comparação de Frequências

Keywords

COVID-19
Numbers of Confirmed Cases
Benford's Law
Frequency Comparison Test

Resumo

A Lei de Benford afirma que, em conjuntos de números aleatórios, a probabilidade de o primeiro dígito desses números ser 1 é maior do que a dos dígitos seguintes. Dessa forma, a distribuição proposta por essa lei mostra que o 1 tem aproximadamente 30,1% de chances de ser o primeiro; em seguida, o dígito 2, com 17,6%; o 3, com 12,5% e assim por diante, até chegar ao dígito 9, com 4,6% de chance. Nesse contexto, assume-se, como objetivo, verificar se a Lei de Benford se aplica aos números de casos confirmados da doença COVID-19 em diferentes países. A pesquisa, do tipo quantitativa, tratou e analisou dados coletados no site da Organização Mundial de Saúde (OMS) dos seguintes países, escolhidos aleatoriamente, conforme notoriedade nas mídias nacionais e internacionais: China, Itália, Nova Zelândia, Brasil e Estados Unidos da América (EUA). Para avaliar a discrepância entre as frequências relativas observadas e esperadas, utilizou-se o teste de comparação de frequências. Os resultados encontrados mostraram que os números da China e da Nova Zelândia tiveram X^2 calculado menor que o X^2 crítico, enquanto a Itália, o Brasil e os EUA tiveram X^2 calculado maior que o X^2 crítico, todos em um nível de significância de 5%. Assim, concluiu-se que a Lei de Benford aplicou-se aos números de casos confirmados da doença COVID-19 na China e na Nova Zelândia, sendo rejeitada pelos números de casos confirmados da doença COVID-19 na Itália, no Brasil e nos EUA.

Abstract

Benford's Law states that in sets of random numbers the prospect of the first digit of these numbers being 1 is greater than that of the following digits. Thus, the distribution proposed by Benford's law shows that number 1 has approximately a 30.1% chance of being the first digit then the digit 2, with 17.6%, digit 3, with 12.5%, and so on, until number 9 is reached, having a 4.6% chance. In this context, the objective of the study is to verify whether this law applies to the numbers of confirmed cases of COVID-19 in different countries. The research, employing a quantitative method, treated and analyzed data collected on the World Health Organization (WHO) website, selecting the following countries, chosen randomly, according to the notoriety received in national and international media: China, Italy, New Zealand, Brazil and the United States of America (USA). To assess the discrepancy between the observed and expected relative frequencies, the frequency comparison test was used. The results found demonstrated that the numbers for China and New Zealand had calculated X^2 smaller than the critical X^2 , while Italy, Brazil and the USA had calculated X^2 greater than the critical X^2 , all at a 5% significance level. Consequently, it was concluded that Benford's Law applied to the numbers of confirmed cases of COVID-19

disease in China and New Zealand, being rejected by the numbers of confirmed cases of COVID-19 disease in Italy, Brazil and the USA.

1. Introdução

A Lei de Benford – também conhecida como Lei do Primeiro Dígito, Lei de Newcomb-Benford ou Lei dos Números Anômalos – foi observada por Newcomb (1881), astrônomo e matemático do século XIX, que, ao manusear livros de tabelas logarítmicas, percebeu que as primeiras páginas apresentavam maior desgaste em relação ao uso do que as seguintes, fazendo-o inferir que mais números usados começavam pelo dígito 1, ou seja, mantissas de menor valor ocorriam com maior frequência que mantissas de maior valor. No entanto, a observação de Newcomb (1881) foi demonstrada e difundida por Benford (1938), que a analisou em conjuntos de números oriundos de diferentes contextos não relacionados, indicando-a como uma lei de probabilidade geral de ampla aplicação.

Muitos foram os estudos realizados com o propósito de identificar dados fabricados a partir da distribuição da Lei de Benford. Cunha (2013), em seu estudo da arte, destacou vários trabalhos que evidenciam essas aplicações. Dentre eles, os estudos de Nigrini (1992, 2000, 2012) e suas descobertas a partir da hipótese de que dados contábeis seguiam a distribuição da Lei de Benford detectaram dados fraudulentos aleatórios em sete companhias de New York, nos quais poucos valores começavam com o número 1 e muitos começavam com o 6. Para Cunha (2013, p. 52), este autor “mostrou a aplicabilidade generalizada da lei e seus usos práticos para detectar fraudes, erros e outras anomalias”.

Outra aplicação da Lei de Benford foi em dados macroeconômicos, como demonstrado no artigo de Rauch *et al.* (2011). Ao utilizar a Lei de Benford nos dados macroeconômicos dos países membros da União Europeia, os autores construíram um ranking dos 27 países membros, conforme a extensão do desvio encontrado, indicando a Grécia como o país que teve maior desvio. Essa suspeita de manipulação dos dados foi posteriormente confirmada pela própria Comissão Europeia. Assim, esse estudo “evidenciou a efetividade da Lei de Benford na detecção de irregularidades e manipulações de dados macroeconômicos” (CUNHA, 2013, p. 40).

Além de dados contábeis e macroeconômicos, outra aplicação realizada foi em dados de eleições. Mebane (2006, 2009) analisou dados das eleições iranianas de 2009. Diferentemente das aplicações supracitadas, o estatístico descobriu que a contagem de votos tendia a seguir a Lei de Benford no segundo dígito, e não no primeiro, como comumente usada, fazendo-o constatar anomalias que indicavam fortemente a ocorrência de fraude na vitória de Ahmadinejad no Iran.

Assim, neste trabalho, considerando a singularidade da Lei de Benford de mostrar probabilidades distintas para a ocorrência do primeiro dígito em um universo aleatório de números, e considerando o conjunto de números derivados da pandemia do novo coronavírus

(SARS-COV-2), causador da doença COVID-19, assumimos como objetivo verificar se a Lei de Benford se aplica aos números de casos confirmados da doença COVID-19 em diferentes países.

O artigo foi organizado em quatro seções, a contar da introdução. Na segunda seção, descreveremos características da pesquisa realizada, evidenciando o tipo de pesquisa, a fonte de dados, bem como instrumentos de coleta e de análise dos dados. Na terceira seção, apresentaremos os resultados encontrados. Encerraremos o artigo na quarta seção, na qual apresentaremos conclusões.

2. Metodologia

Nesta seção, apresentamos aspectos metodológicos da pesquisa relatada neste artigo, destacando o tipo de pesquisa, a fonte de coleta dos dados analisados, bem como características da Lei de Benford e do método estatístico utilizado na validação dos dados analisados.

2.1. Tipo de pesquisa e fonte de coleta dos dados analisados

Pesquisas de cunho quantitativo “consistem essencialmente em encontrar relações entre variáveis, fazer descrições recorrendo ao tratamento estatístico de dados recolhidos, testar teorias” (CARMO; FERREIRA, 2008, p. 196). Dessa forma, entendemos a pesquisa relatada neste artigo como sendo do tipo quantitativo, na medida em que buscamos verificar a Lei de Benford nos números confirmados de COVID-19 em diferentes países.

Os países foram selecionados de forma arbitrária, de acordo com a notoriedade que ganhavam com a proliferação da doença COVID-19 em seus territórios, como especificado a seguir: China, primeiros casos confirmados no mundo; Itália, um dos países da Europa mais afetados; Nova Zelândia, país referência no enfretamento e no controle de casos; Brasil, país mais afetado na América do Sul; e EUA, país que lidera os rankings como um dos mais afetados no mundo.

A coleta aconteceu entre os dias 12 e 16 de setembro de 2020, considerando o intervalo de meses janeiro-agosto, no *site* da Organização Mundial de Saúde (OMS¹), no qual acessamos os números diários de casos confirmados de COVID-19 na China, na Itália, na Nova Zelândia, no Brasil e nos EUA.

2.2. A Lei de Benford

A Lei de Benford é dada por

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

¹ Disponível em: <https://covid19.who.int/>. Acesso em: 12 e 16 de set. de 2020.

em que $n \in [1,9]$ e $P(n)$ referem-se à probabilidade de ocorrência do primeiro dígito n em um universo de números aleatórios, como assinalado na Tabela 1.

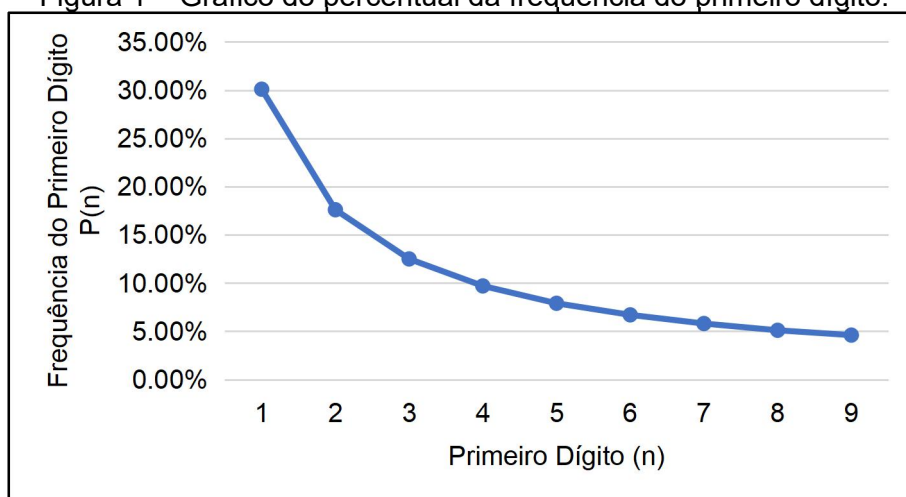
Tabela 1 – Lei de Benford para o primeiro dígito.

<i>Dígito (n)</i>	<i>Probabilidade do Primeiro Dígito P(n)</i>	<i>Percentual</i>
1	0,301	30,1%
2	0,176	17,6%
3	0,125	12,5%
4	0,097	9,7%
5	0,079	7,9%
6	0,067	6,7%
7	0,058	5,8%
8	0,051	5,1%
9	0,046	4,6%

Fonte: Adaptado de Benford (1938, p. 556).

Na Figura 1 é possível observar o comportamento dos dados gerados com o modelo matemático da Lei de Benford (Tabela 1) no gráfico:

Figura 1 – Gráfico do percentual da frequência do primeiro dígito.



Fonte: Elaboração do autor (2020).

2.3. Teste de comparação de frequências (χ^2)

Para o estudo deste artigo, considerando que foram investigadas amostras de números derivadas dos casos confirmados da doença COVID-19 em alguns países (Frequências observadas), segundo a Lei de Benford, com os percentuais para a frequência do primeiro dígito desses números (Frequências esperadas), utilizamos o teste de comparação de frequências. De acordo com Campos, Wodewotzki e Jacobini (2013), os resultados obtidos por meio de amostras nem sempre coincidem com os resultados teóricos esperados. Por exemplo, ao lançar uma moeda honesta 100 vezes, teoricamente estima-se que se pode esperar 50 vezes resultando cara e 50 vezes resultando coroa; no entanto, muitas vezes esse resultado não é obtido.

Com o teste de comparação de frequências seria possível comparar, por exemplo, se o total de números de casos de COVID-19 de um determinado país (Frequências observadas) que começam com o dígito 1 difere de modo significativo do percentual ajustado pela Lei de Benford (Frequências esperadas), que no caso de o primeiro dígito ser 1, é de 30,1% de chance. Essa medida de discrepância existente entre as frequências observadas e as frequências esperadas é dada pela estatística X^2 , que é expressa por:

$$X^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \dots + \frac{(o_k - e_k)^2}{e_k} = \sum \frac{(o_j - e_j)^2}{e_j}. \quad (2)$$

De acordo com Campos, Wodewotzki e Jacobini (2013, p. 102), o processo do teste de comparação de frequências acontece da seguinte forma:

As frequências esperadas são calculadas com base em uma hipótese H_0 . Se, para essa hipótese, o valor de X^2 calculado for maior que o valor crítico (dado em tabela), concluímos que as frequências observadas diferem de modo significativo das esperadas e rejeitamos H_0 ao nível de significância correspondente. No caso contrário, devemos aceitá-la ou, pelo menos, não rejeitá-la.

Para este estudo, consideramos, para o X^2 calculado, o somatório das nove possibilidades da Lei de Benford. Dessa forma, para o primeiro dígito temos $j \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, com grau de liberdade igual a 8 ($gl = 9 - 1 = 8$), expresso por

$$X^2 = \sum_{j=1}^9 \frac{(o_j - e_j)^2}{e_j}. \quad (3)$$

E, para o X^2 crítico, fixamos o nível de significância em 5%, delimitando para este estudo os valores utilizados para o teste de comparação de frequências:

Tabela 2 – Valores considerados para o teste de comparação de frequências.

X^2 calculado	Grau de liberdade (gl)	X^2 crítico ²
$X^2 = \sum_{j=1}^9 \frac{(o_j - e_j)^2}{e_j}$	8	15,507

Fonte: Elaboração do autor (2020).

Assim, para avaliar se a hipótese H_0 deste estudo (Lei de Benford aplicando-se aos números de casos confirmados de COVID-19 em diferentes países) foi rejeitada ou não, comparamos os valores de X^2 calculado e X^2 crítico, permitindo inferir duas opções:

$$X^2 \text{ calculado} \geq X^2 \text{ crítico} \rightarrow \text{rejeitar } H_0, \quad (4)$$

$$X^2 \text{ calculado} < X^2 \text{ crítico} \rightarrow \text{não rejeitar } H_0. \quad (5)$$

² Disponível em: <https://www.ime.unicamp.br/~cnaber/Tabela%20da%20Qui-quadrado.pdf>. Acesso em: 16 de set. de 2020.

3. Resultados

Nesta seção, apresentamos os resultados obtidos, ao analisar as frequências relativas observadas no primeiro dígito dos números de casos confirmados de COVID-19 na China, na Itália, na Nova Zelândia, no Brasil e nos EUA. Os dados investigados referem-se aos números de casos confirmados de janeiro até agosto, de acordo com o surgimento dos primeiros casos em cada país.

3.1. China

Segundo o site de monitoramento da OMS, na China, de 4 de janeiro a 31 de agosto de 2020, houve 90.383 casos confirmados de COVID-19, conforme organizado na Tabela 3.

Tabela 3 – Números de casos diários confirmados de COVID-19 na China.

<i>Janeiro</i>		<i>Fevereiro</i>		<i>Março</i>		<i>Abril</i>		<i>Mai</i>		<i>Junho</i>		<i>Julho</i>		<i>Agosto</i>	
<i>Dia</i>	<i>n.º</i>	<i>Dia</i>	<i>n.º</i>	<i>Dia</i>	<i>n.º</i>	<i>Dia</i>	<i>n.º</i>	<i>Dia</i>	<i>n.º</i>	<i>Dia</i>	<i>n.º</i>	<i>Dia</i>	<i>n.º</i>	<i>Dia</i>	<i>n.º</i>
1	-	1	2101	1	579	1	93	1	12	1	18	1	5	1	166
2	-	2	2590	2	206	2	86	2	6	2	9	2	31	2	180
3	-	3	2827	3	128	3	78	3	2	3	5	3	15	3	157
4	1	4	3233	4	120	4	73	4	8	4	1	4	9	4	114
5	0	5	3892	5	143	5	55	5	3	5	11	5	19	5	109
6	3	6	3697	6	146	6	75	6	2	6	6	6	14	6	122
7	0	7	3151	7	102	7	66	7	3	7	9	7	25	7	133
8	0	8	3387	8	46	8	90	8	6	8	5	8	21	8	120
9	0	9	2653	9	45	9	88	9	1	9	4	9	33	9	92
10	0	10	2985	10	20	10	56	10	14	10	3	10	46	10	121
11	41	11	2472	11	31	11	78	11	20	11	11	11	42	11	113
12	0	12	2022	12	26	12	102	12	1	12	7	12	35	12	61
13	0	13	15152	13	10	13	112	13	7	13	12	13	46	13	82
14	0	14	4050	14	30	14	99	14	6	14	58	14	55	14	99
15	0	15	2644	15	27	15	51	15	5	15	49	15	54	15	70
16	0	16	2008	16	29	16	50	16	9	16	45	16	20	16	66
17	4	17	2051	17	39	17	352	17	6	17	44	17	78	17	98
18	17	18	1893	18	58	18	31	18	10	18	36	18	82	18	67
19	59	19	1751	19	122	19	43	19	6	19	37	19	80	19	54
20	77	20	396	20	116	20	16	20	5	20	30	20	131	20	33
21	93	21	892	21	82	21	14	21	2	21	27	21	84	21	40
22	152	22	825	22	1	22	35	22	13	22	21	22	74	22	50
23	131	23	649	23	104	23	15	23	2	23	52	23	135	23	38
24	261	24	221	24	164	24	9	24	3	24	28	24	139	24	41
25	462	25	517	25	102	25	12	25	11	25	21	25	160	25	23
26	688	26	411	26	92	26	14	26	7	26	29	26	179	26	34
27	776	27	440	27	132	27	3	27	1	27	24	27	189	27	32
28	1772	28	329	28	120	28	6	28	3	28	18	28	217	28	30
29	1462	29	430	29	128	29	22	29	0	29	14	29	212	29	22
30	1741			30	114	30	4	30	18	30	23	30	223	30	28
31	1984			31	90			31	5			31	276	31	32

Fonte: OMS (2020).

Com esses números, foi possível encontrar o X^2 calculado para comparar as frequências relativas observadas do primeiro dígito desses números com as frequências relativas esperadas, de acordo com a Lei de Benford, como indica a Tabela 4.

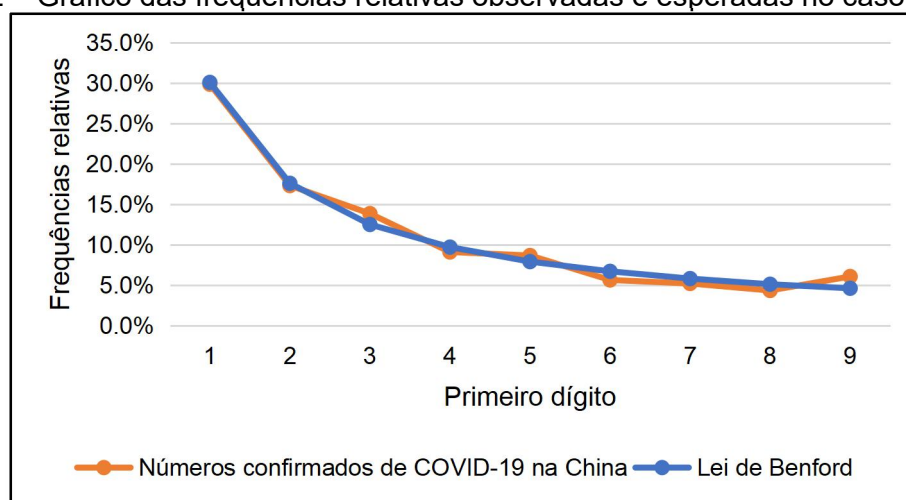
Tabela 4 – Teste de comparação de frequências para o primeiro dígito dos números de casos confirmados de COVID-19 na China.

Dígito	o_j	Frequência relativa observada	e_j	Frequência relativa esperada	$(o_j - e_j)$	$(o_j - e_j)^2$	$\frac{(o_j - e_j)^2}{e_j}$
1	69	29,9%	69,531	30,1%	-0,531	0,282	0,004
2	40	17,3%	40,656	17,6%	-0,656	0,430	0,011
3	32	13,9%	28,875	12,5%	3,125	9,766	0,338
4	21	9,1%	22,407	9,7%	-1,407	1,980	0,088
5	20	8,7%	18,249	7,9%	1,751	3,066	0,168
6	13	5,6%	15,477	6,7%	-2,477	6,136	0,396
7	12	5,2%	13,398	5,8%	-1,398	1,954	0,146
8	10	4,3%	11,781	5,1%	-1,781	3,172	0,269
9	14	6,1%	10,626	4,6%	3,374	11,384	1,071
Total	231					X^2 calculado =	2,492

Fonte: Elaboração do autor (2020).

Como o X^2 calculado é inferior ao X^2 crítico (15,507), aceitamos a hipótese que o número de casos confirmados de COVID-19 na China segue a Lei de Benford. A Figura 2 ilustra as frequências relativas observadas e esperadas, permitindo visualmente observar como a medida de discrepância entre as frequências é pequena.

Figura 2 – Gráfico das frequências relativas observadas e esperadas no caso da China.



Fonte: Elaboração do autor (2020).

3.2. Itália

De acordo com a OMS, na Itália, de 29 de janeiro a 31 de agosto de 2020, houve 268.366 casos confirmados de COVID-19. Não consideramos o número do dia 20 de junho, pois foi computado erroneamente com um número negativo (-148). Apresentamos esses números organizados pelos dias do mês na Tabela 5.

Tabela 5 – Números de casos diários confirmados de COVID-19 na Itália.

Janeiro		Fevereiro		Março		Abril		Maió		Junho		Julho		Agosto	
Dia	n.º	Dia	n.º	Dia	n.º	Dia	n.º	Dia	n.º	Dia	n.º	Dia	n.º	Dia	n.º
1	-	1	0	1	561	1	4053	1	1872	1	355	1	142	1	379
2	-	2	0	2	347	2	4782	2	1965	2	178	2	182	2	295
3	-	3	0	3	466	3	4668	3	1900	3	318	3	201	3	238
4	-	4	0	4	587	4	4585	4	1389	4	321	4	223	4	159
5	-	5	0	5	769	5	4805	5	1221	5	177	5	235	5	190
6	-	6	0	6	778	6	4316	6	1075	6	518	6	192	6	384
7	-	7	1	7	1247	7	3599	7	1444	7	270	7	208	7	401
8	-	8	0	8	1492	8	3039	8	1401	8	197	8	137	8	552
9	-	9	0	9	1797	9	3836	9	1327	9	280	9	193	9	347
10	-	10	0	10	977	10	4204	10	1083	10	283	10	214	10	463
11	-	11	0	11	2313	11	3951	11	802	11	202	11	276	11	259
12	-	12	0	12	2651	12	4694	12	744	12	379	12	188	12	412
13	-	13	0	13	2547	13	4092	13	1402	13	163	13	234	13	476
14	-	14	0	14	3497	14	3153	14	888	14	346	14	169	14	522
15	-	15	0	15	3590	15	2972	15	992	15	338	15	114	15	574
16	-	16	0	16	3233	16	2667	16	789	16	301	16	162	16	629
17	-	17	0	17	3526	17	3786	17	875	17	210	17	230	17	477
18	-	18	0	18	4207	18	3493	18	675	18	328	18	231	18	320
19	-	19	0	19	5318	19	3491	19	451	19	331	19	249	19	401
20	-	20	0	20	5986	20	3047	20	813	20	-	20	218	20	642
21	-	21	1	21	6557	21	2256	21	665	21	264	21	190	21	840
22	-	22	7	22	0	22	2729	22	642	22	224	22	128	22	947
23	-	23	113	23	5560	23	3370	23	652	23	221	23	280	23	1071
24	-	24	105	24	4789	24	2646	24	669	24	113	24	306	24	1209
25	-	25	93	25	5249	25	3021	25	531	25	577	25	252	25	953
26	-	26	78	26	5210	26	2357	26	300	26	296	26	274	26	876
27	-	27	250	27	6153	27	2324	27	397	27	255	27	254	27	1366
28	-	28	238	28	5959	28	1739	28	584	28	175	28	168	28	1409
29	6	29	240	29	5974	29	2091	29	593	29	174	29	202	29	1460
30	0			30	5217	30	2086	30	516	30	126	30	288	30	1444
31	0			31	4050			31	416			31	382	31	1365

Fonte: Fonte: OMS (2020).

Com os números diários de casos confirmados de COVID-19 na Itália tabelados, foi possível encontrar o X^2 calculado, conforme demonstra a Tabela 6.

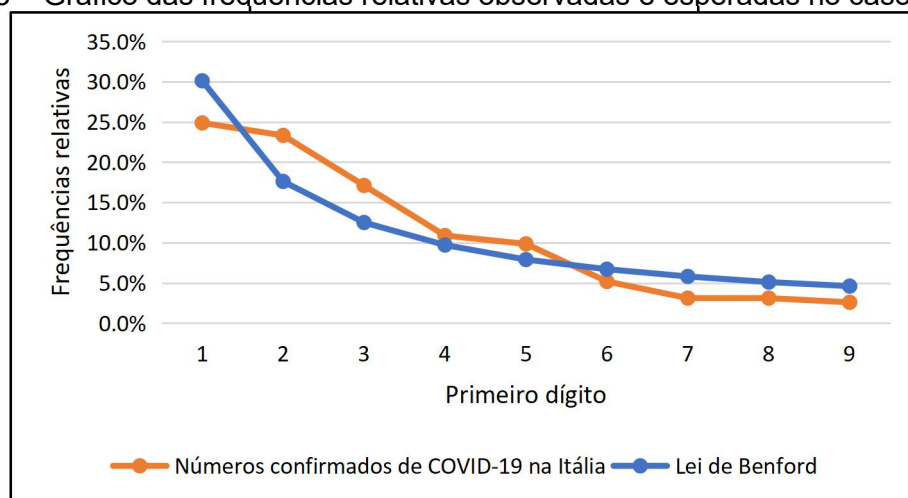
Tabela 6 – Teste de comparação de frequências para o primeiro dígito dos números de casos confirmados de COVID-19 na Itália.

Dígito	o_j	Frequência relativa observada	e_j	Frequência relativa esperada	$(o_j - e_j)$	$(o_j - e_j)^2$	$\frac{(o_j - e_j)^2}{e_j}$
1	48	24,9%	58,093	30,1%	-10,093	101,869	1,754
2	45	23,3%	33,968	17,6%	11,032	121,705	3,583
3	33	17,1%	24,125	12,5%	8,875	78,766	3,265
4	21	10,9%	18,721	9,7%	2,279	5,194	0,277
5	19	9,8%	15,247	7,9%	3,753	14,085	0,924
6	10	5,2%	12,931	6,7%	-2,931	8,591	0,664
7	6	3,1%	11,194	5,8%	-5,194	26,978	2,410
8	6	3,1%	9,843	5,1%	-3,843	14,769	1,500
9	5	2,6%	8,878	4,6%	-3,878	15,039	1,694
Total	193					X^2 calculado =	16,071

Fonte: Elaboração do autor (2020).

Para os números de casos confirmados de COVID-19 na Itália, o X^2 calculado foi maior que o X^2 crítico (15,507), indicando, em um nível de significância de 5%, a rejeição da hipótese H_0 . A Figura 3 mostra o comportamento das frequências relativas observadas e esperadas, que confirma essa rejeição.

Figura 3 – Gráfico das frequências relativas observadas e esperadas no caso da Itália.



Fonte: Elaboração do autor (2020).

3.3. Nova Zelândia

Diferentemente da China e da Itália, que computaram os primeiros casos confirmados de COVID-19 em janeiro, na Nova Zelândia, de acordo com a OMS, de 28 de fevereiro a 31 de agosto, houve 1.387 casos confirmados de COVID-19, conforme apresentado na Tabela 7.

Tabela 7 – Números de casos diários confirmados de COVID-19 na Nova Zelândia.

Fevereiro		Março		Abril		Maio		Junho		Julho		Agosto	
Dia	n.º	Dia	n.º	Dia	n.º	Dia	n.º	Dia	n.º	Dia	n.º	Dia	n.º
1	-	1	0	1	47	1	3	1	0	1	0	1	2
2	-	2	0	2	76	2	2	2	0	2	2	2	3
3	-	3	1	3	51	3	2	3	0	3	0	3	2

4	-	4	0	4	50	4	1	4	0	4	0	4	0
5	-	5	0	5	48	5	0	5	0	5	3	5	2
6	-	6	2	6	39	6	1	6	0	6	1	6	0
7	-	7	1	7	32	7	1	7	0	7	2	7	0
8	-	8	0	8	26	8	2	8	0	8	1	8	0
9	-	9	0	9	23	9	1	9	0	9	3	9	0
10	-	10	0	10	23	10	2	10	0	10	2	10	0
11	-	11	0	11	20	11	3	11	0	11	1	11	1
12	-	12	0	12	14	12	0	12	0	12	1	12	4
13	-	13	0	13	15	13	0	13	0	13	0	13	14
14	-	14	1	14	8	14	0	14	0	14	1	14	13
15	-	15	0	15	6	15	1	15	0	15	2	15	7
16	-	16	0	16	6	16	0	16	2	16	1	16	13
17	-	17	5	17	2	17	1	17	0	17	1	17	9
18	-	18	9	18	8	18	0	18	1	18	1	18	13
19	-	19	15	19	4	19	4	19	0	19	3	19	6
20	-	20	14	20	7	20	0	20	2	20	1	20	5
21	-	21	13	21	2	21	0	21	0	21	1	21	11
22	-	22	0	22	5	22	1	22	4	22	0	22	6
23	-	23	36	23	0	23	0	23	2	23	0	23	3
24	-	24	40	24	2	24	0	24	1	24	1	24	8
25	-	25	47	25	3	25	0	25	3	25	0	25	7
26	-	26	73	26	4	26	0	26	1	26	0	26	5
27	-	27	76	27	1	27	0	27	2	27	0	27	7
28	5	28	78	28	2	28	0	28	4	28	1	28	12
29	0	29	60	29	2	29	0	29	2	29	1	29	13
		30	76	30	3	30	0	30	0	30	2	30	2
		31	48	31		31	0	31		31	0	31	9

Fonte: OMS (2020).

Com os números de casos confirmados de COVID-19 na Nova Zelândia, foi possível encontrar o χ^2 calculado, conforme apresentado na Tabela 8.

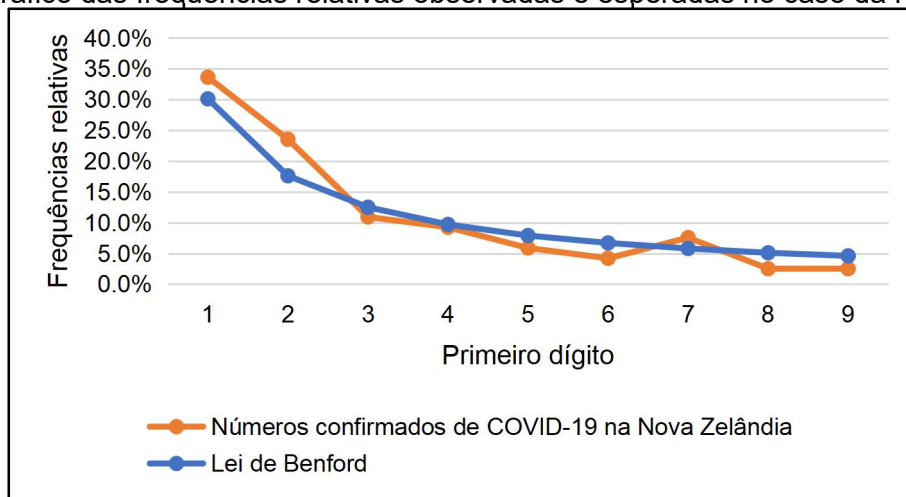
Tabela 8 – Teste de comparação de frequências para o primeiro dígito dos números de casos confirmados de COVID-19 na Nova Zelândia.

Dígito	o_j	Frequência relativa observada	e_j	Frequência relativa esperada	$(o_j - e_j)$	$(o_j - e_j)^2$	$\frac{(o_j - e_j)^2}{e_j}$
1	40	33,6%	35,819	30,1%	4,181	17,481	0,488
2	28	23,5%	20,944	17,6%	7,056	49,787	2,377
3	13	10,9%	14,875	12,5%	-1,875	3,516	0,236
4	11	9,2%	11,543	9,7%	-0,543	0,295	0,026
5	7	5,9%	9,401	7,9%	-2,401	5,765	0,613
6	5	4,2%	7,973	6,7%	-2,973	8,839	1,109
7	9	7,6%	6,902	5,8%	2,098	4,402	0,638
8	3	2,5%	6,069	5,1%	-3,069	9,419	1,552
9	3	2,5%	5,474	4,6%	-2,474	6,121	1,118
Total	119					χ^2 calculado =	8,157

Fonte: Elaboração do autor (2020).

Para os números de casos confirmados de COVID-19 na Nova Zelândia, o X^2 calculado foi menor que o valor de 15,507 do X^2 crítico, indicando, em um nível de significância de 5%, o aceite da hipótese H_0 . Na Figura 4, é possível observar o comportamento das frequências relativas observadas e esperadas para os dígitos considerados pela Lei de Benford.

Figura 4 – Gráfico das frequências relativas observadas e esperadas no caso da Nova Zelândia.



Fonte: Elaboração do autor (2020).

3.4. Brasil

Os números de casos confirmados de COVID-19 no Brasil, assim como na Nova Zelândia, também começaram a ser computados no final da segunda quinzena do mês de fevereiro. E, de acordo com a OMS, de 26 de fevereiro a 31 de agosto, houve 3.846.153 casos confirmados de COVID-19, como apresentado na Tabela 9.

Tabela 9 – Números de casos diários confirmados de COVID-19 no Brasil.

Fevereiro		Março		Abril		Maio		Junho		Julho		Agosto	
Dia	n.º	Dia	n.º	Dia	n.º	Dia	n.º	Dia	n.º	Dia	n.º	Dia	n.º
1	-	1	1	1	323	1	6276	1	33274	1	24052	1	57837
2	-	2	0	2	1138	2	7218	2	16409	2	33846	2	52383
3	-	3	0	3	1119	3	6209	3	11598	3	46712	3	45392
4	-	4	1	4	1074	4	4970	4	28936	4	48105	4	25800
5	-	5	4	5	1146	5	4588	5	28633	5	42223	5	16641
6	-	6	1	6	1222	6	6633	6	30925	6	37923	6	51603
7	-	7	11	7	852	7	6935	7	30830	7	26051	7	57152
8	-	8	0	8	926	8	10503	8	27075	8	20229	8	53139
9	-	9	6	9	1661	9	9888	9	18912	9	45305	9	50230
10	-	10	0	10	2210	10	10222	10	15654	10	44571	10	49970
11	-	11	9	11	1930	11	10611	11	32091	11	42619	11	23010
12	-	12	18	12	1781	12	6760	12	32913	12	45048	12	22048
13	-	13	46	13	1089	13	5632	13	30412	13	39023	13	52160
14	-	14	23	14	1442	14	9258	14	25982	14	24831	14	55155
15	-	15	0	15	1261	15	11385	15	21704	15	20286	15	60091
16	-	16	79	16	1832	16	13944	16	17110	16	41857	16	50644
17	-	17	34	17	3058	17	15305	17	20647	17	39924	17	41576
18	-	18	57	18	2105	18	14919	18	34918	18	45403	18	23101

19	-	19	133	19	3257	19	7938	19	32188	19	34177	19	19373
20	-	20	193	20	2917	20	13140	20	22765	20	28532	20	47784
21	-	21	283	21	2055	21	17408	21	54771	21	23529	21	49298
22	-	22	0	22	1927	22	19951	22	34666	22	20257	22	45323
23	-	23	0	23	2498	23	18508	23	17459	23	41008	23	30355
24	-	24	1297	24	2678	24	20803	24	21432	24	67860	24	50032
25	-	25	0	25	3735	25	16508	25	39436	25	59961	25	23421
26	5	26	232	26	3503	26	15813	26	42725	26	55891	26	17078
27	0	27	0	27	5514	27	11687	27	39483	27	51147	27	47134
28	0	28	482	28	3379	28	16324	28	46860	28	24578	28	47161
29	0	29	502	29	4613	29	20599	29	38693	29	23284	29	44235
		30	487	30	5385	30	26417	30	30476	30	40816	30	43412
		31	352			31	26928			31	69074	31	41350

Fonte: OMS (2020).

Considerando os números de casos confirmados de COVID-19 no Brasil, encontramos o valor de X^2 calculado, como apontado na Tabela 10.

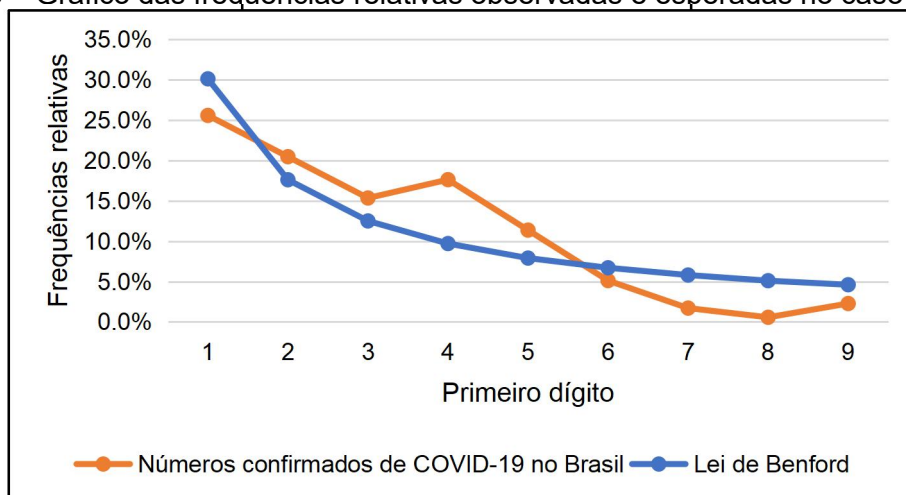
Tabela 10 – Teste de comparação de frequências para o primeiro dígito dos números de casos confirmados de COVID-19 no Brasil.

Dígito	o_j	Frequência relativa observada	e_j	Frequência relativa esperada	$(o_j - e_j)$	$(o_j - e_j)^2$	$\frac{(o_j - e_j)^2}{e_j}$
1	45	25,6%	52,976	30,1%	-7,976	63,617	1,201
2	36	20,5%	30,976	17,6%	5,024	25,241	0,815
3	27	15,3%	22,000	12,5%	5,000	25,000	1,136
4	31	17,6%	17,072	9,7%	13,928	193,989	11,363
5	20	11,4%	13,904	7,9%	6,096	37,161	2,673
6	9	5,1%	11,792	6,7%	-2,792	7,795	0,661
7	3	1,7%	10,208	5,8%	-7,208	51,955	5,090
8	1	0,6%	8,976	5,1%	-7,976	63,617	7,087
9	4	2,3%	8,096	4,6%	-4,096	16,777	2,072
Total	176					X^2 calculado =	32,098

Fonte: Elaboração do autor (2020).

Os números de casos confirmados de COVID-19 no Brasil geraram um X^2 calculado maior que o X^2 crítico (15,507). Indicaram uma grande discrepância entre as frequências relativas observadas e as frequências relativas esperadas, fazendo com que a hipótese H_0 fosse rejeitada para esse conjunto de números, em um nível de significância de 5%. Visualmente, na Figura 5, é possível perceber como as frequências relativas observadas não coincidem com as frequências relativas esperadas em nenhum dígito.

Figura 5 – Gráfico das frequências relativas observadas e esperadas no caso do Brasil.



Fonte: Elaboração do autor (2020).

3.5. Estados Unidos da América (EUA)

Os primeiros casos confirmados de COVID-19 nos EUA começaram em janeiro. Segundo a OMS, de 20 de janeiro a 31 de agosto de 2020, houve 5.899.504 casos confirmados de COVID-19 no país. Os números diários foram organizados na Tabela 11.

Tabela 11 – Números de casos diários confirmados de COVID-19 nos EUA.

Janeiro		Fevereiro		Março		Abril		Maió		Junho		Julho		Agosto	
<i>Dia</i>	<i>n.º</i>	<i>Dia</i>	<i>n.º</i>	<i>Dia</i>	<i>n.º</i>	<i>Dia</i>	<i>n.º</i>	<i>Dia</i>	<i>n.º</i>	<i>Dia</i>	<i>n.º</i>	<i>Dia</i>	<i>n.º</i>	<i>Dia</i>	<i>n.º</i>
1	-	1	1	1	0	1	22559	1	31379	1	23482	1	35757	1	67823
2	-	2	0	2	2	2	24103	2	31774	2	26116	2	43556	2	67499
3	-	3	3	3	44	3	26298	3	26753	3	14692	3	54271	3	58388
4	-	4	0	4	21	4	28103	4	31839	4	24890	4	53213	4	47183
5	-	5	0	5	19	5	32105	5	29266	5	14583	5	51933	5	49151
6	-	6	1	6	65	6	33510	6	16200	6	20069	6	57186	6	49629
7	-	7	0	7	0	7	26493	7	22267	7	28922	7	43686	7	53373
8	-	8	0	8	0	8	29510	8	22119	8	28918	8	46194	8	55318
9	-	9	0	9	259	9	31709	9	30204	9	17848	9	50263	9	61028
10	-	10	0	10	224	10	30859	10	25870	10	17536	10	64630	10	53893
11	-	11	1	11	0	11	35386	11	26642	11	17235	11	58975	11	47964
12	-	12	0	12	291	12	31606	12	23767	12	20315	12	66281	12	39894
13	-	13	2	13	277	13	31633	13	18044	13	21745	13	62369	13	54791
14	-	14	0	14	414	14	29308	14	21424	14	22133	14	60113	14	55907
15	-	15	0	15	36	15	24446	15	20840	15	25314	15	58720	15	52799
16	-	16	0	16	0	16	25802	16	27090	16	21754	16	60711	16	55359
17	-	17	0	17	1822	17	28711	17	22813	17	18514	17	67165	17	54375
18	-	18	0	18	3551	18	32549	18	31967	18	27921	18	71484	18	41073
19	-	19	0	19	3351	19	30023	19	13227	19	23139	19	74354	19	39125
20	5	20	0	20	4777	20	28252	20	24417	20	23046	20	66963	20	37908
21	0	21	19	21	0	21	27668	21	23310	21	36617	21	62788	21	46259
22	0	22	1	22	0	22	25634	22	22787	22	32349	22	57276	22	43952
23	0	23	0	23	16354	23	24019	23	20475	23	27575	23	62929	23	45960
24	1	24	18	24	20341	24	29127	24	24151	24	26519	24	69641	24	44946
25	0	25	0	25	0	25	30719	25	26158	25	34191	25	71714	25	37765

26	3	26	6	26	16420	26	38509	26	15253	26	37601	26	74235	26	32883
27	0	27	0	27	0	27	32417	27	24886	27	40526	27	63968	27	37030
28	0	28	3	28	16894	28	29218	28	16362	28	44458	28	61498	28	45484
29	0	29	0	29	18093	29	22541	29	19606	29	44580	29	54022	29	46194
30	1			30	19332	30	20517	30	21214	30	41008	30	59629	30	44002
31	1			31	17987			31	17962			31	65406	31	43983

Fonte: OMS (2020).

Com os números diários de casos confirmados de COVID-19 nos EUA, foi possível encontrar o X^2 calculado, como apresentado na Tabela 12.

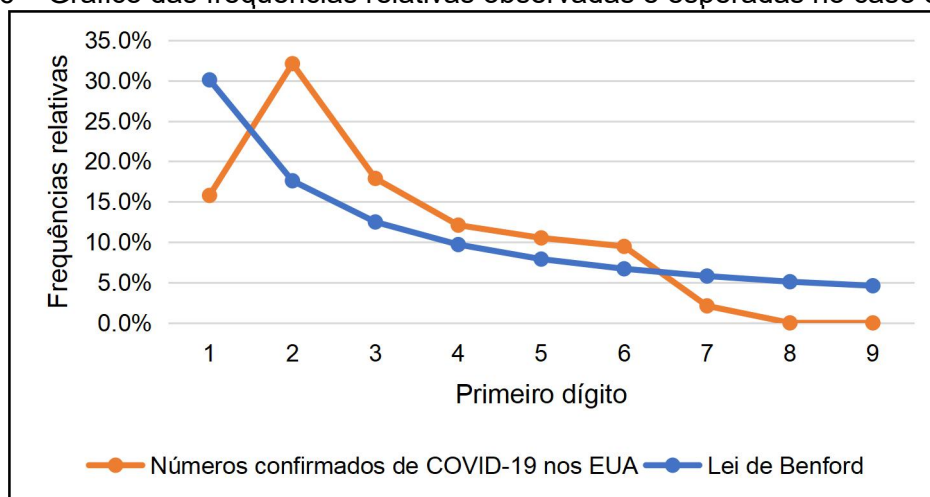
Tabela 12 – Teste de comparação de frequências para o primeiro dígito dos números de casos confirmados de COVID-19 nos EUA.

Dígito	o_j	Frequência relativa observada	e_j	Frequência relativa esperada	$(o_j - e_j)$	$(o_j - e_j)^2$	$\frac{(o_j - e_j)^2}{e_j}$
1	30	15,8%	57,190	30,1%	-27,190	739,296	12,927
2	61	32,1%	33,440	17,6%	27,560	759,554	22,714
3	34	17,9%	23,750	12,5%	10,250	105,063	4,424
4	23	12,1%	18,430	9,7%	4,570	20,885	1,133
5	20	10,5%	15,010	7,9%	4,990	24,900	1,659
6	18	9,5%	12,730	6,7%	5,270	27,773	2,182
7	4	2,1%	11,020	5,8%	-7,020	49,280	4,472
8	0	0,0%	9,690	5,1%	-9,690	93,896	9,690
9	0	0,0%	8,740	4,6%	-8,740	76,388	8,740
Total	190					X^2 calculado =	67,940

Fonte: Elaboração do autor (2020).

O X^2 calculado foi de 67,940, enquanto o X^2 crítico é de 15,507. Dessa forma, é possível concluir que o X^2 calculado foi maior que o X^2 crítico, o que indica, em um nível de 5% de significância, que a hipótese H_0 para esse conjunto de números foi rejeitada. Na Figura 6, é possível observar visualmente a discrepância entre as frequências relativas observadas e esperadas, encontrada pelo X^2 calculado.

Figura 6 – Gráfico das frequências relativas observadas e esperadas no caso dos EUA.








Fonte: Elaboração do autor para este texto (2020).

4. Conclusão

O objetivo elencado neste artigo foi verificar se a Lei de Benford aplicava-se aos números de casos confirmados de COVID-19 em diferentes países (hipótese H_0). Ao utilizar o teste de comparação de frequências nos números de casos confirmados de COVID-19 na China, na Itália, na Nova Zelândia, no Brasil e nos EUA, foi possível avaliar a discrepância entre as frequências relativas observadas e as esperadas, para o aceite ou a rejeição da H_0 , conforme organizado no Quadro 1.

Quadro 1 – Avaliação da hipótese H_0 para os números de casos confirmados de COVID-19.

Países	X^2 calculado	X^2 crítico	Hipótese H_0
	2,492	15,507	Aceita
	8,157	15,507	Aceita
	16,071	15,507	Rejeitada
	32,098	15,507	Rejeitada
	67,940	15,507	Rejeitada

Fonte: Elaboração do autor (2020).

Assim, a hipótese H_0 de que a Lei de Benford se aplicava aos números de casos confirmados de COVID-19 foi aceita pelos números da China e da Nova Zelândia e rejeitada pelos números da Itália, do Brasil e dos EUA. A China, em particular, foi o país com os números que apresentaram a menor discrepância, seguida da Nova Zelândia. Por outro lado, a Itália passou por pouco do valor do X^2 crítico, enquanto o Brasil e os EUA foram os países com as maiores discrepâncias.

Em outros contextos, a Lei de Benford foi aplicada na auditoria de dados contábeis (NIGRINI, 1992, 2000, 2012), na auditoria de dados macroeconômicos (RAUCH *et al.*, 2011) e até mesmo na busca de fraudes eleitorais (MEBANE, 2006, 2009). Com este estudo, não tivemos a intenção de auditar os números informados pelos países e contabilizados pela OMS. Segundo Campos, Wodewotzki e Jacobini (2013), é raro que frequências observadas concordem muito bem com as esperadas e, dessa forma, as discrepâncias encontradas nos números da Itália, do Brasil e dos EUA mostram que nem sempre a distribuição proposta pela Lei de Benford é a que melhor se ajusta para descrever o primeiro dígito. No entanto, com exceção dos números dos EUA, os outros países mostraram que o dígito 1 apareceu com mais frequência do que os demais.

Para futuras pesquisas, propomos investigar mais a fundo motivos que levaram os números de casos confirmados de COVID-19 a serem ajustados pela Lei de Benford, enquanto

outros não. Talvez se deva adotar outras opções viáveis para tratar esses dados, estendendo a investigação para o segundo, o terceiro, o quarto dígito e assim por diante.

Referências

- BENFORD, F. The law of anomalous numbers. **Proceedings of The American Philosophical Society**, v. 78, p. 551-572, 1938. Disponível em: <https://www.jstor.org/stable/984802>. Acesso em: 01 jul. 2020.
- CAMPOS, R. C; WODEWOTZKI, M. L. L.; JACOBINI, O. R. **Educação Estatística: teoria e prática em ambientes de modelagem matemática**. Coleção Tendências em Educação Matemática. 2. ed. Belo Horizonte: Autêntica, 2013.
- CARMO, H.; FERREIRA, M. M. **Metodologia da investigação: guia para auto-aprendizagem**. 2. ed. Lisboa-PT: Universidade Aberta, 2008.
- CUNHA, F. C. R. da. **Aplicações da lei Newcomb-Benford à auditoria de obras públicas**. Orientador: Maurício Soares Bugarin. 2013. 486f. Dissertação (Mestrado em Regulação e Gestão de Negócios) – Universidade de Brasília, Brasília, 2013. Disponível em: <https://repositorio.unb.br/handle/10482/16379>. Acesso em: 20 dez. 2020.
- MEBANE, W. R. Election Forensics: vote counts and Benford's Law. **Papers, Posters and Syllabi**, The Society Political Methodology, n. 620, 2006.
- MEBANE, W. R. **Note on the presidential election in Iran**. Michigan: University of Michigan, 2009.
- NEWCOMB, S. Note on the frequency of use of the different digits in natural Numbers. **American Journal of Mathematics**, v. 4, n. 1, p. 39-40, 1881. DOI: <https://doi.org/10.2307/2369148>.
- NIGRINI, M. J. **The detection of income tax evasion through an analysis of digital frequencies**. Tese (Ph.D.) – University of Cincinnati, Cincinnati, OH. 1992.
- NIGRINI, M. J. **Digital analysis using Benford's Law: Tests statistics for auditors**. Global Audit Publication. Berlim, Heidelberg: Springer, 2000.
- NIGRINI, M. J. **Benford's law. Applications for Forensic Accounting, Auditing and Fraud Detection**. Hoboken, New Jersey: John Wiley & Sons, 2012.
- OMS. **World Health Organization**. Disponível em: <https://covid19.who.int/>. Acesso em: 12 e 16 set. de 2020.
- RAUCH, B.; BRÄHLER, G.; GÖTTSCHE, M.; ENGEL, S. Fact and fiction in EU-Governmental Economic Data. **German Economic Review**, v. 12, n. 3, p. 243-255, 2011. DOI: <https://doi.org/10.1111/j.1468-0475.2011.00542.x>.