

LEI DE NEWCOMB-BENFORD: UMA APLICAÇÃO COM DADOS DE EXECUÇÃO ORÇAMENTÁRIA DE DESPESAS COM A PANDEMIA DE COVID-19 NO ESTADO DO PARANÁ

Vitória Eduarda Bello (Universidade Federal do Paraná - UFPR) E-mail: vitoria.bello@ufpr.br

Anselmo Chaves Neto (Universidade Federal do Paraná - UFPR) E-mail: anselmo@ufpr.br

Resumo: O presente estudo buscou verificar se a Lei de Newcomb-Benford pode ser utilizada para identificar irregularidades na execução orçamentária de despesas com a pandemia do novo Coronavírus (COVID-19). Para tanto, utilizou-se a formulação da Lei de Newcomb-Benford para a análise do primeiro e dos dois primeiros dígitos dos gastos com a pandemia de COVID-19 nos anos de 2020 a 2022 obtidos do Portal da Transparência do Governo do Estado do Paraná. Estatisticamente, utilizou-se dos testes de Qui-quadrado de Pearson, Teste do Arco Mantissa e Desvio Absoluto Médio. No que se refere aos resultados, indica-se que os dados de execução orçamentária de despesas com a pandemia de COVID-19 não seguem a distribuição esperada com as probabilidades dos dígitos formuladas pela Lei de Newcomb-Benford, o que sugere maiores investigações voltadas para a confirmação da regularidade dessas despesas.

Palavras-chave: Lei de Newcomb-Benford, COVID-19, Execução Orçamentária do Governo do Estado do Paraná.

NEWCOMB-BENFORD LAW: AN APPLICATION WITH DATA ON BUDGET EXECUTION OF EXPENSES WITH THE COVID-19 PANDEMIC IN THE STATE OF PARANÁ

Abstract: The present study sought to verify whether the Newcomb-Benford Law can be used to identify irregularities in the budget execution of expenses with the new Coronavirus (COVID-19) pandemic. To this end, the formulation of the Newcomb-Benford Law was used to analyze the first and first two digits of expenses with the COVID-19 pandemic in the years 2020 to 2022 obtained from the Transparency Portal of the Government of the State of Paraná. Statistically, Pearson's Chi-square test, Mantissa Arc Test and Mean Absolute Deviation were used. Regarding the results, it is indicated that the budget execution data of expenses with the COVID-19 pandemic do not follow the expected distribution with the probabilities of the digits formulated by the Newcomb-Benford Law, which suggests further investigations aimed at confirmation of the regularity of these expenses.

Keywords: Benford's Law, COVID-19, Budget Execution of the Government of the State of Paraná.

1. Introdução

A população atual vive cercada por enormes quantidades de dados que aparecem tanto em canais de comunicação ou redes sociais e, tais dados, são transmitidos rapidamente por satélites que orbitam a Terra. Dessa forma, cientistas e analistas de dados precisam, de forma ágil, selecionar e observar comportamentos padronizados nesses grandes conjuntos (MILLER, 2015, p. 03).

A Lei dos Números Anômalos, Lei dos Dígitos Significativos, Lei de Newcomb-Benford ou Lei de Benford, está relacionada com identificação de padrões numéricos em um conjunto de dados. Contudo, Benford não foi o primeiro a observar o padrão numérico.

Segundo Miller (2015, p. 04), o astrônomo e matemático do século XIX Simon Newcomb (1835 - 1909), observou este comportamento mais de 50 anos antes de Benford, já que em 1881 publicou um artigo que discutia que os dez dígitos não apareciam com igual frequência, inclusive quando se observava tabelas logarítmicas.

Em 1912, por meio do *Calcul des Probabilités*, Henri Poincaré (1854 - 1912) traz uma formalização da Lei de Newcomb-Benford, quando observa uma lista de logaritmos onde em dada posição significativa, dados os dígitos de zero a nove, a ocorrência de números pares ou ímpares são eventos equiprováveis. Franel (1917) apresenta algumas correções ao trabalho de Poincaré e confirma os cálculos de probabilidade. Neste sentido, a Lei de Newcomb-Benford continua a ser discutida.

Anteriormente a Franel, em 1916, estudava-se a Lei de Newcomb-Benford de forma a observar a distribuição dos números em módulo (distribuição igual de números módulo um), que foi particularmente relevante no desenvolvimento de sequências de baixa discrepância.

Em 1938, o estudo avançou, e o artigo *Lei dos Números Anômalos* foi publicado na revista *Proceedings da American Philosophical Society* por Frank Benford (1883-1948). Benford estudou a distribuição dos dígitos iniciais de 20 conjuntos de dados, incluindo rios, áreas, populações, constantes físicas, sequências matemáticas, dentre outros. Ele chegou a mesma conclusão que Newcomb (1881), de que nos livros que continham tabelas logarítmicas, as primeiras páginas apresentavam maior desgaste em relação às últimas, o que indicaria maior consulta em páginas que constavam os logaritmos dos números com primeiros dígitos pequenos.

O artigo de Benford teve um destino muito melhor do que o artigo de Newcomb, possivelmente em parte porque precedeu imediatamente a um artigo de física de Bethe, Rose e Smith no espalhamento múltiplo de elétrons. Considerando que foi publicado décadas antes de haver outro artigo sobre o trabalho de Newcomb. E, o próximo artigo após o artigo de Benford foi publicado seis anos depois por S. A. Goutsmit e W. H. Furry (1944), e a partir daí os artigos começaram a ocorrer cada vez com maior frequência (MILLER, 2015, p. 7). Contudo, S. A. Goutsmit e W. H. Furry (1944), acreditavam que a Lei de Newcomb-Benford era resultado natural do sistema numérico usado (BOWMAN, 2016, p. 3).

A Lei de Newcomb-Benford ficou conhecida por sua enorme aplicação e por dar suporte nas investigações relacionadas a fraudes. As pessoas tendem a acreditar, erroneamente, que os números se distribuem uniformemente na natureza e dessa forma, ao se alterar valores de uma planilha em benefício próprio, não se questionaria as frequências com as quais os dígitos aparecem.

Sua formulação pode ser entendida como, dado um número qualquer, a probabilidade de que o primeiro dígito deste número seja o número 1 é maior que a probabilidade de que seja o número 2 que é maior do que seja o número 3 e assim sucessivamente, de forma que a frequência diminui até chegar no número 9 (NEWCOMB, 1881).

Este artigo está relacionado com o exame das planilhas orçamentárias disponíveis no Portal da Transparência do Governo do Estado do Paraná (2022), de modo a verificar se a execução orçamentária de despesas utilizadas para o combate da pandemia de COVID-19, nos anos de 2020 a 2022, segue ou não a formulação da Lei de Newcomb-Benford.

Para tanto, utilizou-se de todo o conjunto de dados disponível nesse período e sabendo que de acordo com Nigrini (2012, p. 71) há uma variedade de testes estatísticos que podem ser usados para determinar o melhor ajuste dos dados às respectivas distribuições de probabilidades que a Lei de Newcomb-Benford estabelece. Esses testes são categorizados como testes primários, testes avançados e testes associados.

Dessa forma, o presente trabalho tem por objetivo apresentar os cálculos e análises gráficas obtidas para o primeiro e os dois primeiros dígitos de cada observação das

despesas públicas em questão, de acordo com alguns métodos estatísticos. Os resultados sugerem que os dados não seguem a Lei de Newcomb-Benford e que por esse motivo pode ter havido irregularidades que devem ser investigadas.

1.1. Problema motivador

Durante o período de coleta de dados, de 2020 a 2022, o mundo passou pela pandemia do novo coronavírus. Segundo a Organização Mundial da Saúde (OMS):

A doença do coronavírus (COVID-19) é uma doença infecciosa causada pelo vírus SARS-CoV-2. A maioria das pessoas infectadas com o vírus apresentará doença respiratória leve a moderada e se recuperará sem precisar de tratamento especial. No entanto, alguns ficarão gravemente doentes e exigirão atenção médica. Os idosos e aqueles com condições médicas subjacentes, como doenças cardiovasculares, diabetes, doenças respiratórias crônicas ou câncer, são mais propensos a desenvolver doenças graves. Qualquer pessoa pode ficar doente com COVID-19 e ficar gravemente doente ou morrer em qualquer idade (OMS, 2022).

Dessa forma, diversos protocolos foram elaborados pelo Governo Federal e pelo Governo do Estado do Paraná, visando a contenção do vírus. Os orçamentos públicos passaram por adaptações e a verba foi destinada principalmente para a área da saúde pública. De modo a garantir o acesso aos dados orçamentários, no Portal da Transparência do Governo do Estado do Paraná (2022), criou-se um histórico de planilhas relacionadas ao combate à COVID-19 e que tem por objetivo veicular dados e informações detalhados sobre a execução financeira e orçamentária. É possível visualizar e baixar os dados abertos, de modo a desenvolver técnicas que corroborem com a participação por parte do cidadão como uma ferramenta de acesso à informação e que os pesquisadores trabalhem colaborativamente pela sociedade.

2. Desenvolvimento

A formulação da Lei de Newcomb-Benford segue o modelo matemático da Equação (1) a seguir:

$$P(D_1 = d_1) = \log\left(1 + \frac{1}{d_1}\right) \quad (1)$$

onde P é a probabilidade de ocorrência do primeiro dígito significativo e entende-se por 'dígito' o primeiro algarismo que compõe qualquer número que se queira analisar. Além disso, D_1 representa o primeiro dígito e $d_1 = 1, 2, \dots, 9$ é o valor do dígito significativo avaliado. De acordo com Nigrini (2012, p. 5), exemplifica-se que a probabilidade de D_1 assumir o valor 1 é avaliada substituindo este valor na Equação (1) e o resultado está exibido na Equação (2).

$$P(D_1 = 1) = \log\left(1 + \frac{1}{1}\right) = \log(2) = 0,3010. \quad (2)$$

Os cálculos para as demais probabilidades seguem a mesma fórmula e se obtém a Figura 1 a seguir, que mostra os valores já calculados em um gráfico em barras para melhor entendimento.

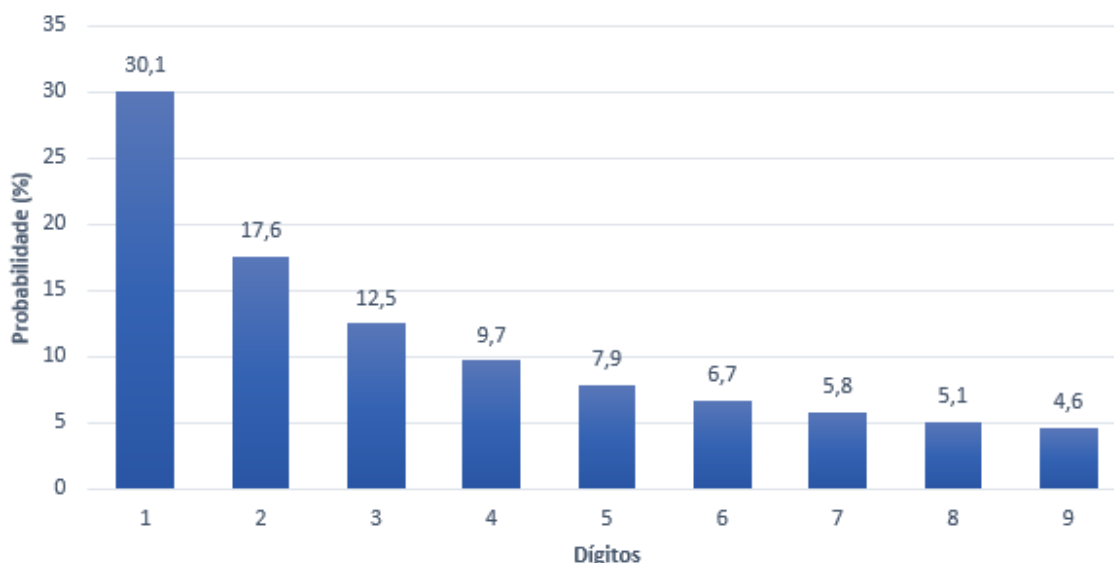


Figura 1 – Distribuição de probabilidades para o primeiro dígito de cada observação segundo a Lei de Newcomb-Benford

Ainda, de acordo com Nigrini (2012), existem diversos testes estatísticos que podem ser usados para verificar se os dados seguem ou não a Lei de Newcomb-Benford. Nesse sentido, no presente trabalho se aborda o Teste Qui-quadrado de Pearson, Desvio Absoluto Médio (DAM) e Teste do Arco da Mantissa (AM). Além disso, para melhor entendimento, já feitos os cálculos, a Tabela 1 mostra as frequências esperadas para todos os dígitos de 0 a 9 das primeiras quatro posições em qualquer número.

Tabela 1 – Probabilidade dos dígitos de 0 a 9 das primeiras quatro posições em qualquer número

Dígitos	Varição	Segunda Posição	Terceira Posição	Quarta Posição
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	0,10331	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

Fonte: Adaptado de Nigrini (1996, p. 74)

Assim, sintetizando os principais estudiosos do assunto e os primeiros a desenvolverem a teoria, Santos (2018) cita que:

Por muitos anos, a Lei de Benford foi reconhecida como uma curiosidade matemática. Vários pesquisadores contribuíram para o desenvolvimento da teoria da Lei de Benford e forneceram explicações para a teoria, de Simon Newcomb sendo o primeiro pesquisador, a Frank Benford, de quem a Lei tomou seu nome, Theodore Hill, que forneceu a primeira prova matemática da Lei e Mark Nigrini, que é a força motriz por trás da aplicação da lei na detecção de fraude (SANTOS, 2018, p. 26).

Por meio dos estudos de Mark Nigrini, incluindo dados que auxiliam no entendimento de quando a Lei está sendo rigorosamente aceita ou não, tornou-se parte essencial e é um dos principais referenciais utilizados atualmente. Segundo Nigrini (2012, p. 13), a fórmula a seguir na Equação (3) pode ser usada para calcular o primeiro, os dois primeiros, os três primeiros e os quatro primeiros dígitos. Essa expressão também pode ser usada para calcular as probabilidades do segundo, terceiro e quarto dígitos, embora possa ser visto na Tabela 1 que as probabilidades são iguais para a maioria dos propósitos práticos a partir do quarto dígito.

$$P(D_1 = d_1, \dots, D_k = d_k) = \log \left[1 + \left(\frac{1}{\sum_{i=1}^k d_i \cdot 10^{k-i}} \right) \right] \quad (3)$$

2.1. Teste Qui-quadrado de Pearson

O teste Qui-quadrado avalia as evidências contra alguma suposição. Assim, faz comparação entre valores observados e valores esperados de acordo com uma suposição. A estatística do teste, expressão adiante, tem distribuição Qui-quadrado com $k-1$ graus de liberdade sendo k o número de dígitos possíveis. E, a decisão de se aceitar, ou não, a suposição como verdadeira é feita a um determinado nível de significância, α . A estatística do teste é exibida na Equação (4) a seguir:

$$\chi^2 = \sum_{i=1}^k \frac{(CR - CE)^2}{CE} \quad (4)$$

onde CR representa a contagem real e CE representa a contagem esperada dos dígitos. No presente estudo, o teste Qui-quadrado foi aplicado na análise dos dois primeiros dígitos, a um nível de significância de 5%. E, considerando que existem 90 dígitos possíveis, o número de graus de liberdade é $k - 1 = 89$. Mas, quando se considera apenas o primeiro dígito existem 9 dígitos possíveis e o número de graus de liberdade é 8.

No estudo de caso em questão, encontrou-se o valor da estatística de 8337 para o primeiro dígito, o que configura não conformidade com a Lei de Newcomb-Benford, pois o valor-p correspondente é menor que 5%. O mesmo ocorre com o teste dos dois primeiros dígitos, a estatística de teste foi de 110068, também indicando não conformidade, já que seu valor-p é menor que 5%.

2.2. Desvio Absoluto Médio (DAM)

O Desvio Absoluto Médio (DAM), é uma medida de conformidade com a Lei de Newcomb-Benford que ignora o número de registros, isto é, que independe do número de observações. Sua fórmula está de acordo com a Equação (5):

$$DAM = \frac{\sum_{i=1}^k |PR - PE|}{k} \quad (5)$$

onde PE denota a proporção esperada de acordo com a Lei de Newcomb-Benford, PR denota a proporção observada e k representa o número de séries dos dígitos analisados (09 para o teste do primeiro dígito e 90 para o dos dois primeiros dígitos). Além disso, um importante fato sobre o DAM é que ele não sofre influência de eventuais outliers, isto

é, observações que se destacam das demais. Ainda, de acordo com Nigrini (2012), que estabelece parâmetros para avaliar se o DAM segue a Lei de Newcomb-Benford ou não, tem-se a Tabela 2 a seguir.

Tabela 2 – Valores críticos e conclusões para valores de DAM para o primeiro dígito e para os dois primeiros dígitos

Dígitos	Variação	Conclusões
Primeiro Dígito	0,000 a 0,006	Conformidade
	0,006 a 0,012	Conformidade aceitável
	0,012 a 0,015	Conformidade Marginalmente aceitável
	Acima de 0,015	Não conformidade
Primeiros Dois Dígitos	0,0000 a 0,0012	Conformidade
	0,0012 a 0,0018	Conformidade aceitável
	0,0018 a 0,0022	Conformidade Marginalmente aceitável
	Acima de 0,0022	Não conformidade

Fonte: Adaptado de Nigrini (2012, p. 160)

Neste trabalho, encontrou-se o valor de 0,02041361 para o primeiro dígito. Isto indica não conformidade com a Lei de Newcomb-Benford segundo a Tabela 2. Da mesma forma, ocorre com o teste dos dois primeiros dígitos, o valor encontrado para o DAM foi de 0,003832284, também indicando não conformidade.

2.3. Teste do Arco da Mantissa (AM)

Alexander (2009) propõe um teste rigoroso baseado nas mantissas, que se denomina de teste do Arco da Mantissa (AM). Segundo Nigrini (2012, p. 165), o valor numérico da mantissa é plotado no eixo y . No teste AM, cada valor numérico é plotado no círculo unitário e produz um conjunto de pontos uniformemente distribuídos no círculo de raio 1 e centrados na origem (0). O centro de gravidade é chamado de vetor médio (MV) e suas coordenadas x e y são calculadas conforme mostrado as Equações (6) e (7) a seguir:

$$\text{Coordenada } x = \frac{\sum_{i=1}^N \cos(2\pi(\log(x_i) \bmod 1))}{N} \quad (6)$$

$$\text{Coordenada } y = \frac{\sum_{i=1}^N \text{sen}(2\pi(\log(x_i) \bmod 1))}{N} \quad (7)$$

O comprimento médio é calculado como $L^2 = (\text{coordenada } x)^2 + (\text{coordenada } y)^2$ e o valor-p correspondente é $1 - e^{-L^2 \cdot N}$. Na aplicação deste trabalho rejeitou-se a hipótese de que os dados seguissem a Lei de Newcomb-Benford, devido ao valor de 0,0055018 para L^2 que foi encontrado.

3. Metodologia

3.1 Coleta de Dados

Este trabalho é um estudo de caso com dados do Portal da Transparência do Governo do Estado do Paraná (2022), correspondentes à execução orçamentária para a contingência da pandemia de COVID-19 durante o período de junho a dezembro de 2020, fevereiro a

dezembro de 2021 e fevereiro a outubro de 2022. Dessa forma, para uma investigação completa, os dados em questão foram agrupados em uma única planilha. Com relação à coleta de dados, utilizou-se as planilhas de despesas que estão disponíveis nos dados abertos. Além disso, realizou-se um processo denominado aqui de ‘tratamento de dados’ em que se filtrou somente valores de orçamento que tenham números maiores ou iguais a 10, por conta da utilização também dos dois primeiros dígitos de cada observação para a análise, já que valores menores que 10 possuem um único dígito. Portanto, trabalhou-se com 195418 observações após a filtragem e agrupamento dos dados, já que havia somente uma observação menor do que 10, de modo que a quantidade total antes da filtragem foi de 195419 observações. A Tabela 3 a seguir ilustra a quantidade de dados brutos, anterior ao processo de tratamento dos dados, e quantas observações foram coletadas.

Tabela 3 – Número de observações por ano antes do processo de tratamento dos dados

	Ano	Número de Observações
1	2020	57807
2	2021	100798
3	2022	36814

Fonte: Elaborada pelos autores com dados extraídos do Portal da Transparência do Estado do Paraná (2022)

3.2. Análise dos Dados

Os dados disponibilizados foram analisados por meio do programa R (R CORE TEAM, 2019), que permite utilizar diversos mecanismos de análise gráfica e tabelas de resumos estatísticos. Mais especificamente, foi utilizado o pacote *benford.analysis* (CINELLI, 2018), que permite utilizar testes estatísticos para verificar o ajuste com as probabilidades dos primeiros dígitos, conforme a Lei de Newcomb-Benford determina.

4. Resultados e Discussão

Dados financeiros tendem a seguir a Lei de Newcomb-Benford e tentativas de manipulação podem acabar sendo identificadas. Dessa forma, dados que forem criados, fraudados ou de alguma forma, alterados, tendem a não estar em conformidade com a Lei de Newcomb-Benford. Além disso, dados viesados ou incompletos podem apresentar problemas ao se aplicar essa Lei, de modo que ela fornece uma forma de investigação orçamentária e de gastos contábeis governamentais.

Neste trabalho, procurou-se verificar se há alguma discrepância no primeiro e nos dois primeiros dígitos dos lançamentos da execução orçamentária no Paraná com relação aos gastos com o novo Coronavírus. O primeiro gráfico, exibido na Figura 2, mostra a contagem de observações com relação ao seu primeiro dígito, comparando-a com o valor esperado pela Lei de Newcomb-Benford (linha em vermelho). Percebe-se que há um salto claro no dígito 8. Contudo, o dígito 1 é o mais frequente, o que coincide com o que é indicado pela Lei em questão.

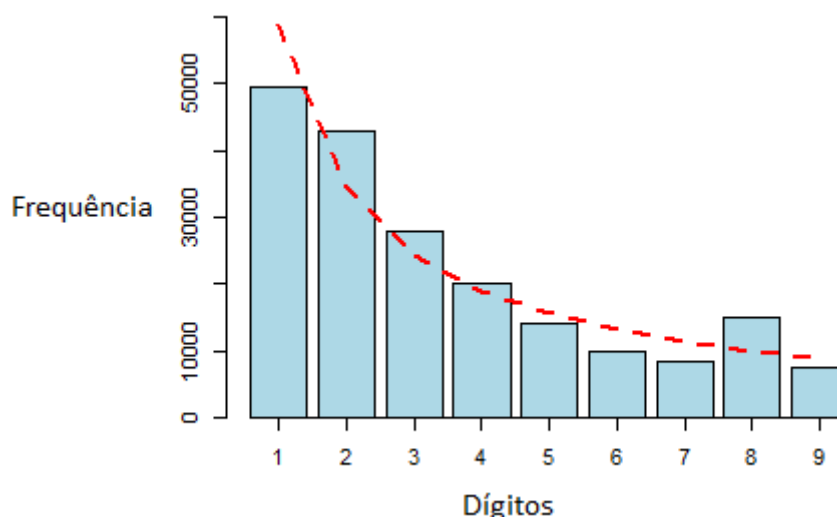


Figura 2 – Contagem de observações com relação ao primeiro dígito de cada observação das despesas públicas do Governo do Paraná com COVID-19 em comparação com o valor esperado pela Lei de Newcomb-Benford (linha vermelha)

Isso pode ser observado por meio da Tabela 4, em que as 5 maiores diferenças absolutas são mostradas. Veja que o primeiro grupo é do dígito 1 seguido pelo dígito 2, que está de acordo com a Lei de Newcomb-Benford. Contudo, no terceiro grupo, o que tem maior frequência, é o dígito 8, que deveria estar entre os últimos, devido a sua menor probabilidade de aparecer em planilhas como o primeiro dígito de observações.

Tabela 4 – As 5 maiores diferenças absolutas com relação ao primeiro dígito de cada observação das despesas públicas do Governo do Paraná com COVID-19

	Dígitos	Diferença Absoluta
1	1	9211,08
2	2	8366,46
3	8	4957,00
4	3	3469,46
5	6	3138,35

Fonte: Elaborada pelos autores com dados extraídos do Portal da Transparência do Estado do Paraná (2022)

O mesmo ocorre quando se observa o gráfico da Figura 3, que indica a contagem de observações com relação aos dois primeiros dígitos de cada observação em comparação com o valor esperado pela Lei de Newcomb-Benford (linha em vermelho).

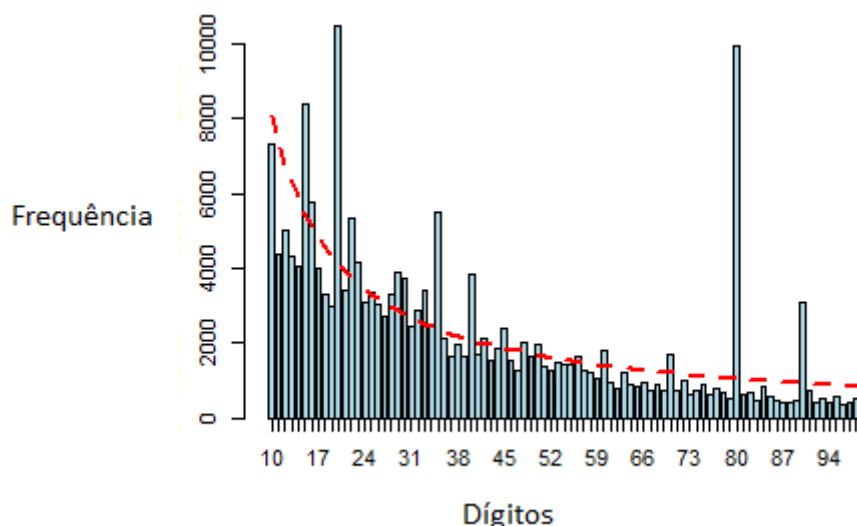


Figura 3 – Contagem de observações com relação aos dois primeiros dígitos de cada observação das despesas públicas do Governo do Paraná com COVID-19 em comparação com o valor esperado pela Lei de Newcomb-Benford (linha vermelha)

Percebe-se que há um salto claro no dígito 80, bem como alguns desajustes no dígito 20. Isso pode ser observado na Tabela 5, onde as cinco maiores diferenças absolutas são mostradas. Veja que o primeiro grupo é o dos números que começam com o dígito 80, como observado graficamente na Figura 3 e o segundo grupo é o dos números que começam com 20. Esses registros indicam a necessidade de serem analisados com mais atenção.

Tabela 5 – As 5 maiores diferenças absolutas com relação ao primeiro dígito de cada observação das despesas públicas do Governo do Paraná com COVID-19

	Dígitos	Diferença Absoluta
1	80	8892,78
2	20	6333,43
3	35	3091,59
4	11	3004,08
5	15	2908,23

Fonte: Elaborada pelos autores com dados extraídos do Portal da Transparência do Estado do Paraná (2022)

Foi feito, ainda, uma análise mais específica, dado a quantidade de observações. Assim, trabalhou-se com uma nova tabela contendo somente os dados suspeitos, que comecem com o primeiro dígito 80 e 20. Esse novo conjunto de dados foi formado com aproximadamente 10% das observações (20416 observações), permitindo restringir a análise.

Ainda, o valor DAM, segundo o teste feito, indicou 0,003832284, que segundo Nigrini (2012), indicaria não conformidade com a Lei de Newcomb-Benford conforme citado anteriormente. Novamente, uma análise mais detalhada é indicada, a partir dos dados que se tornaram suspeitos, bem como o conhecimento específico do conjunto de dados que está sendo estudado.

5. Considerações Finais

De acordo com os resultados, a conclusão que se tem é que os dados, que formam 195418

observações do orçamento líquido para combate à pandemia de COVID-19 no Estado do Paraná não seguem a Lei de Newcomb-Benford. Foram realizadas três estatísticas de testes, que possuem rigor nessa detecção e todas indicaram rejeitar a hipótese que os dados seguem a Lei. Duas delas indicaram valores pequenos e próximos de zero para o valor-p (Qui-quadrado e Teste do Arco da Mantissa com valor-p ambos iguais a $2,2e^{-16}$ que é menor do 5%). Além disso, o DAM também indicou não conformidade com a Lei de Newcomb-Benford de acordo com as estatísticas estabelecidas por Nigrini (2012, p. 160). Por conseguinte, sabendo que tais dados não seguem a Lei, é indicado uma análise dos valores considerados suspeitos. Nesse estudo observou-se que os valores com maior frequência iniciavam com os primeiros dígitos iguais a 80 e 20. Dessa forma, uma nova planilha foi criada, somente com os dados suspeitos. Essa planilha contém 20416, isto é, 10,4473% dos dados originais, restringindo a análise.

Os valores iniciados por 80, tomando uma análise mais minuciosa, indicam na maior parte dos dados, o pagamento de bolsas para estudantes de medicina, farmácia e áreas da saúde afins, visando o combate à pandemia. Nesse sentido, pode haver explicações para os dados com maior desajuste, principalmente gráficos, conforme foi analisado na Figura 3 quando havia uma frequência aumentada para o dígito iniciado por 80.

O segundo dígito de maior frequência, conforme a Tabela 5, é o dígito 20 que pode ter variações em pagamentos de despesas. Não pode ser especificamente determinado por conta de indicar valores de 200 ou 2000 em reais, por exemplo. Por conta disso, maiores esclarecimentos devem ser tomados junto ao Governo do Estado, para que se possa esclarecer qual o destino comum de orçamentos iniciados com estes dígitos. Segundo Nigrini (2012), há diversos detalhes que precisam ser ponderados pelos pesquisadores quando se refere à dados de aplicação que sejam reais:

Ao trabalhar com dados do mundo real, geralmente temos grandes conjuntos de dados, e as estatísticas de teste usuais geralmente indicam não conformidade. Com dados do mundo real, não estamos realmente em um cenário em que nossos dados são uma amostra extraída de uma população. Nossos dados geralmente são a população (NIGRINI, 2012, p. 169).

O teste Qui-quadrado, por exemplo, avalia todos os dígitos ao mesmo tempo. Assim, leva em consideração o número de registros no cálculo da conformidade. Dessa forma, tolera-se apenas pequenos desvios da Lei de Newcomb-Benford para N grande, tornando-os úteis apenas para uma comparação de conjuntos de dados de tamanho semelhante ao longo do tempo.

Referências

ALEXANDER, J. Remarks on the use of Benford's Law. Working paper, Case Western Reserve University, Department of Mathematics and Cognitive Science, 2009.

BENFORD, F. The law of anomalous numbers. Proceedings of the American Philosophical Society, v. 78, n. 4, p. 551-572, 1938.

BRASIL. Governo do Estado do Paraná. Portal da Transparência: Execução Orçamentária - Dados Abertos, 2022. Disponível em: <https://www.coronavirus.pr.gov.br/execucao-orcamentaria-dados-abertos>. Acesso em: 28 out. 2022.

BOWMAN, A. Contributions to the Testing of Benford's Law. Masters Thesis, McMaster University, Hamilton, Canada, 2016.

- CINELLI, C.** benford.analysis, 2018. Disponível em: <http://github.com/carloscinelli/benford.analysis>. Acesso em: 28 out. 2022.
- FRANEL, J.** A propos des tables de logarithmes. Lectronic Research Archive for Mathematics. p. 286-295, 1917.
- GOUDSMIT, S. A., FURRY, W. H.** Significant figures of numbers in statistical tables. Nature, v. 154, n. 3921, p. 800-801, 1944.
- MILLER, S. J.** Benford's Law. Princeton University Press, 2015.
- 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.
- NIGRINI, M. J.** A Taxpayer Compliance Application of Benford's Law. Journal of the American Taxation Association, p. 72-91, 1996.
- NIGRINI, M. J.** Benford's Law: Applications for forensic accounting, auditing, and fraud detection. John Wiley & Sons, 2012.
- ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS).** Doença de coronavírus (COVID-19). 2022. Disponível em: https://www.who.int/health-topics/coronavirus#tab=tab_1. Acesso em: 28 out. 2022.
- R DEVELOPMENT CORE TEAM.** R: A language and environment for statistical Computing, 2019. R Foundation for Statistical Computing, Vienna. Disponível em: <https://www.rproject.org/>. Acesso em: 28 out. 2022.
- SANTOS, R. F. R.** Avaliação da conformidade de dados financeiros utilizando a Lei de Benford. 2018. 54 p. Monografia (Especialista em MBA, Auditoria Integral) – Departamento de Ciências Contábeis, do Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, Curitiba, 2018.