



APLICAÇÃO DA LEI DE BENFORD EM SÉRIES MENSAIS DE PRECIPITAÇÃO

Application of Benford's law in monthly series of precipitation

Aplicación de la ley de Benford en series mensuales de precipitación

Iulo Pessotti Moro  

Universidade Federal do Espírito Santo
iulo.floresta@gmail.com

Sidney Sara Zanetti  

Universidade Federal do Espírito Santo
sidney.zanetti@ufes.br

Roberto Avelino Cecílio  

Universidade Federal do Espírito Santo
roberto.cecilio@ufes.br

Resumo: A Lei de Benford descreve uma distribuição logarítmica dos dígitos iniciais em dados reais, com dígitos menores ocorrendo com maior frequência do que dígitos maiores. Com isso, este estudo investigou a aplicação desta lei em séries mensais de precipitação registradas em nove estações meteorológicas automáticas no estado do Espírito Santo, Brasil, no período de 2006 a 2023. O objetivo foi avaliar a conformidade dos dados com a distribuição teórica preconizada pela lei de Benford, tanto nos dados originais quanto após o preenchimento de falhas das séries. Foram utilizados dois métodos para preenchimento de falhas: a média dos mesmos meses em outros anos com dados e a interpolação linear. A aderência à Lei de Benford foi verificada com o teste Qui-quadrado, comparando a distribuição observada do primeiro e segundo dígitos com o padrão teórico esperado. Os resultados indicaram que a série original de precipitação adere à Lei de Benford, sugerindo confiabilidade dos registros. O preenchimento de falhas com a média alterou a distribuição dos dígitos, especialmente do segundo, comprometendo essa conformidade. Em contraste, a interpolação linear preservou o padrão numérico e a aderência à lei. Conclui-se, portanto, que a Lei de Benford tem potencial para avaliar a qualidade de séries pluviométricas, recomendando-se a interpolação linear para tratar dados ausentes quando é necessário preservar a distribuição natural inicial dos dados.

Palavras-chave: Controle de qualidade de dados. Preenchimento de falhas. Teste de aderência. Distribuição logarítmica.

Abs tract: Benford's Law describes a logarithmic distribution of leading digits in real-world data, with smaller digits occurring more frequently than larger ones. Accordingly, this study investigated the application of this law to monthly precipitation series recorded at nine automatic weather stations in the state of Espírito Santo, Brazil, from 2006 to 2023. The objective was to assess the conformity of the data with the theoretical distribution predicted by Benford's Law, both in the original data and after filling gaps in the series. Two gap-filling methods were used: the average of the same months from other years with data and linear interpolation. Adherence to Benford's Law was verified using the Chi-square test, comparing the observed distribution of the first and second digits with the expected theoretical pattern. The results indicated that the original precipitation series adheres to Benford's Law, suggesting the reliability of the records. Gap-filling with the average altered the digit distribution, particularly for the second digit, compromising this conformity. In contrast, linear interpolation preserved the numerical pattern and adherence to the law. It is concluded, therefore, that Benford's Law has the potential to assess the quality of precipitation series, and linear interpolation is recommended for treating missing data when it is necessary to preserve the initial natural distribution of the data.

Keywords: Data quality control. Gap filling. Goodness-of-fit test. Logarithmic distribution.

Resumen: La Ley de Benford describe una distribución logarítmica de los primeros dígitos en datos reales, donde los dígitos menores ocurren con mayor frecuencia que los dígitos mayores. En este contexto, este estudio investigó la aplicación de dicha ley en series mensuales de precipitación registradas en nueve estaciones meteorológicas automáticas en el estado de Espírito Santo, Brasil, durante el período de 2006 a 2023. El objetivo fue evaluar la conformidad de los datos con la distribución teórica propuesta por la ley de Benford, tanto en los datos originales como después de llenar los datos faltantes de las series. Se utilizaron dos métodos para el relleno de datos: el promedio de los mismos meses de otros años con datos y la interpolación lineal. La adherencia a la Ley de Benford se verificó con la prueba de Chi-cuadrado, comparando la distribución observada del primer y segundo dígito con el patrón teórico esperado. Los resultados indicaron que la serie original de precipitación se adhiere a la Ley de Benford, lo que sugiere la fiabilidad de los registros. El relleno de datos con el promedio alteró la distribución de los dígitos, especialmente del segundo, comprometiendo dicha conformidad. En contraste, la interpolación lineal preservó el patrón numérico y la adherencia a la ley. Se concluye, por lo tanto, que la Ley de Benford tiene potencial para evaluar la calidad de las series pluviométricas, recomendándose la interpolación lineal para tratar datos ausentes cuando es necesario preservar la distribución natural inicial de los datos.

Palabras clave: Control de calidad de datos. Relleno de datos faltantes. Prueba de bondad de ajuste. Distribución logarítmica.

Submetido em: 18/02/2025

Aceito para publicação em: 08/10/2025

Publicado em: 17/10/2025

1. INTRODUÇÃO

A Lei de Benford, uma curiosa descoberta matemática, descreve a distribuição inesperada de dígitos em conjuntos de dados do mundo real. Ao contrário da crença comum de que os dígitos ocorrem com igual probabilidade, a Lei de Benford afirma que dígitos menores como 1, 2 e 3 aparecem com mais frequência como primeiro dígito do que dígitos maiores como 7, 8 e 9 (Nigrini, 2012).

A primeira observação documentada desse fenômeno foi feita pelo matemático e astrônomo americano Simon Newcomb, em 1881. Em seu artigo, Newcomb observou que as primeiras páginas de livros de tabelas de logaritmos, amplamente utilizados na época para cálculos matemáticos, apresentavam maior desgaste do que as páginas posteriores. Ele deduziu que isso se devia ao uso mais frequente de números iniciados por dígitos menores, sugerindo uma distribuição não uniforme dos dígitos iniciais em dados naturais (Newcomb, 1881).

Apesar da relevância de sua observação, a descoberta de Newcomb permaneceu relativamente desconhecida até que Frank Benford, em 1938, a redescobriu e popularizou por meio de uma extensa pesquisa (Nigrini, 2012). Benford analisou 20 conjuntos de dados distintos, incluindo populações de cidades, constantes físicas e estatísticas geográficas, confirmando a prevalência da distribuição não uniforme dos dígitos iniciais e consolidando a lei que leva seu nome (Benford, 1938).

A Lei de Benford, embora contraintuitiva, revela uma ordem subjacente em conjuntos de dados aparentemente aleatórios. Essa propriedade tem sido amplamente utilizada em áreas como auditoria financeira, detecção de fraudes e análise forense de dados, devido à sua capacidade de identificar desvios significativos em relação à distribuição esperada dos dígitos (Durtschi, Hillison e Pacini, 2004; Marzio, Fensore e Passamonti, 2024; Oliveira *et al.*, 2018; Silva, Travassos e Costa, 2017).

Nos últimos anos, a aplicação da Lei de Benford tem se estendido a diversos campos científicos, incluindo estudos ambientais e geocientíficos. Embora a aplicação específica da Lei de Benford em dados meteorológicos seja menos explorada na literatura, alguns pesquisadores têm investigado sua utilidade na validação de dados ambientais e na detecção

de anomalias em séries temporais (Lee e Carvalho, 2019; Li e Fu, 2016; Sambridge, Tkalčić e Jackson, 2010).

Nigrini e Miller (2007) aplicaram a Lei de Benford em dados de vazão de rios e áreas de superfície de lagos, dois importantes conjuntos de dados hidrológicos, concluindo que os dados relacionados aos corpos d'água estavam em conformidade com a Lei de Benford e a não conformidade poderia indicar problemas de integridade dos dados, como conjuntos de dados incompletos ou erros de medição.

Em um estudo focado nas mudanças de regime de precipitação na China e nos EUA no final da década de 1970, Yang e Fu (2017) propõem um novo método para detectar mudanças na intensidade da precipitação, baseando-se na Lei de Benford. Os resultados demonstraram que a avaliação da conformidade com a Lei de Benford permitiu identificar mudanças abruptas nos padrões de precipitação de forma mais direta e clara do que a utilização de variáveis tradicionais como, médias e desvios padrão.

Diante da escassez de estudos sobre a aplicação da Lei de Benford em séries temporais de precipitação, este trabalho tem como objetivo geral demonstrar sua aplicação em dados mensais de precipitação registrados por estações meteorológicas automáticas. O estudo visa analisar a conformidade dos dados com a Lei de Benford utilizando testes estatísticos de aderência, com o intuito de avaliar a consistência dos dados meteorológicos, tanto em sua forma bruta quanto após o preenchimento de falhas nas séries temporais.

2. METODOLOGIA

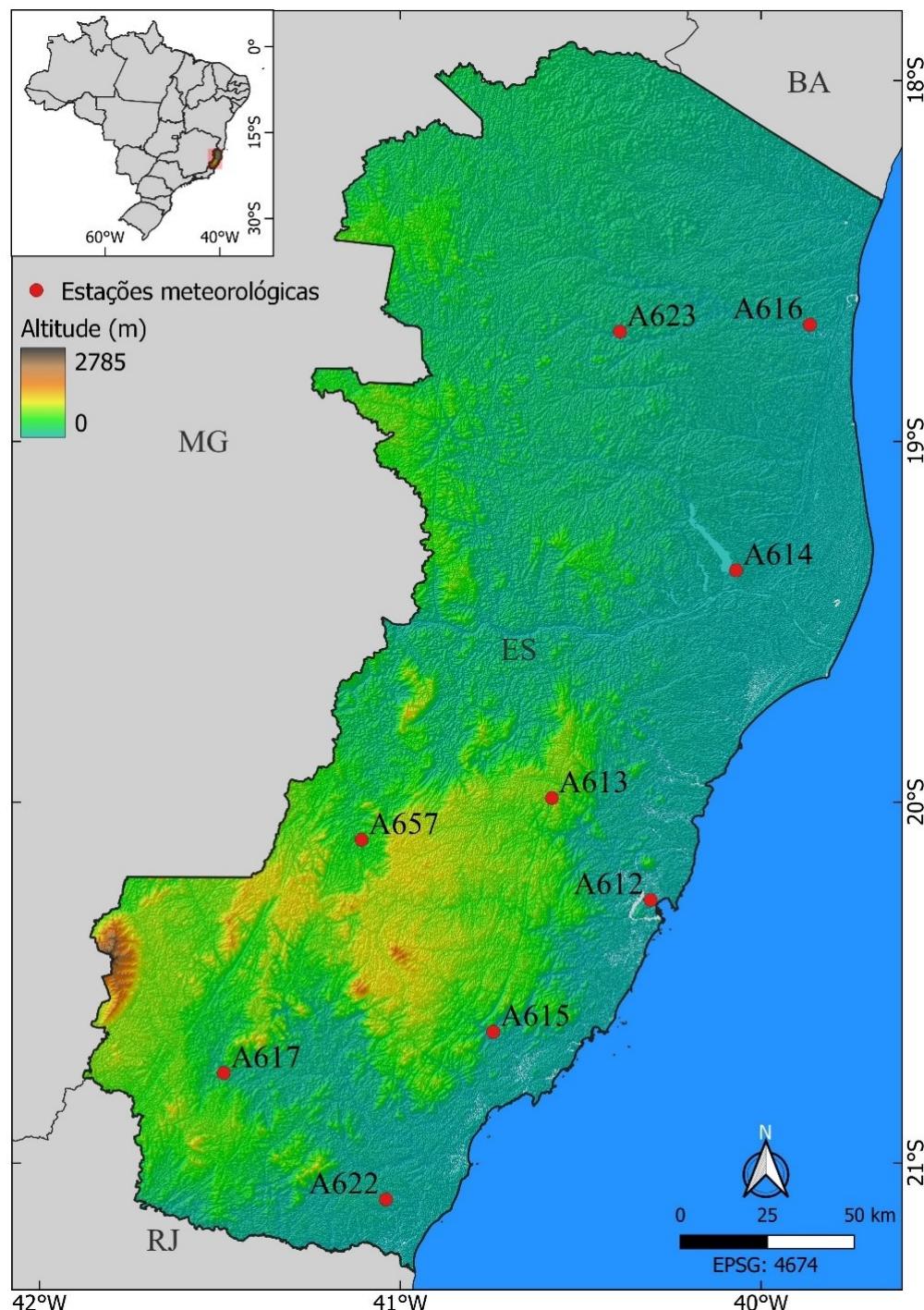
2.1. Área de estudo, aquisição e tratamento dos dados

Elegeu-se, como área em estudo, o estado do Espírito Santo, o qual está localizado na região Sudeste do Brasil, com uma área de 46.095 km². O Espírito Santo apresenta variada tipologia topográfica, com uma variação altimétrica desde o nível do mar até 2.890 m de altitude.

Para o presente estudo, selecionaram-se nove estações meteorológicas automáticas (EMAs) (Figura 1) geridas pelo Instituto Nacional de Meteorologia (INMET), das quais obtiveram-se as séries mensais de precipitação pluviométrica. A descrição das estações e dos dados existentes, bem como a quantidade de falhas estão dispostos na Tabela 1.



Figura 1 -Localização e identificação das estações meteorológicas automáticas utilizadas, do Instituto Nacional de Meteorologia para o estado do Espírito Santo



Fonte: autores (2025).

Tabela 1 - Tamanho das séries de dados utilizadas, das estações meteorológicas automáticas do Instituto Nacional de Meteorologia localizadas no estado do Espírito Santo

ID	Cidade	Data inicial	Data final	Tamanho da série (meses)	Número de meses faltantes	% meses faltantes
A612	Vitória	Out. /2006	Dez. /2023	207	28	13.53
A613	Santa Teresa	Ago. /2007	Nov. /2023	197	55	27.92
A614	Linhares	Out. /2006	Dez. /2023	207	21	10.14
A615	Alfredo Chaves	Nov. /2006	Out. /2023	206	33	16.02
A616	São Mateus	Out. /2006	Dez. /2023	207	16	7.73
A617	Alegre	Out. /2006	Dez. /2023	207	21	10.14
A622	Presidente Kennedy	Jun. /2008	Dez. /2023	187	33	17.65
A623	Nova Venécia	Jun. /2008	Jan. /2023	180	32	17.78
A657	Afonso Claudio	Set. /2011	Nov. /2023	148	30	20.27

Fonte: INMET, organização dos autores (2025).

O controle de qualidade inicial das séries de precipitação foi realizado por meio do teste de homogeneidade normal padrão (SNHT) (Alexandersson, 1986), teste de Pettitt (1979) e teste de Buishand (1982), buscando garantir que as variações presentes fossem decorrentes apenas de fatores climáticos e temporais, conforme demonstrado por Moro et al. (2025).

Utilizaram-se dois métodos para o preenchimento das falhas mensais de dados. O primeiro método consistiu em apenas atribuir, ao mês com falha, o valor da precipitação acumulada média mensal, calculada considerando-se os meses que continham valores disponíveis ao longo da série histórica. Ou seja, para cada mês com dados ausentes, foi calculada a média dos valores registrados naquele mesmo mês em outros anos (McKinney, 2023). Este procedimento assegura que a média mensal histórica registrada não seja alterada.

O segundo método utilizado foi a interpolação linear para estimar os valores ausentes entre pontos de dados existentes. Conforme McKinney (2023), considerando a série temporal de precipitação $\{(t_i, y_i)\}$, em que y_i representa o valor de precipitação no tempo t_i , a interpolação linear para um ponto ausente y_k no tempo t_k entre dois pontos conhecidos (t_i, y_i) e (t_j, y_j) é dada pela Equação 1.

$$(1) \quad y_k = y_i + (t_k - t_i) \cdot \left(\frac{y_j - y_i}{t_j - t_i} \right)$$

Em que: y_i é o valor conhecido imediatamente anterior a y_k ; y_j é o valor conhecido imediatamente posterior a y_k ; t_i e t_j são os instantes de tempo correspondentes aos valores conhecidos.

Após a interpolação linear, podem persistir valores ausentes, especialmente nos extremos da série temporal, onde não há dados suficientes para interpolação. Para tratar esses casos, aplicou-se o preenchimento para trás (*Backward Fill*), em que os valores ausentes são preenchidos utilizando o próximo valor válido subsequente, e o preenchimento para frente (*Forward Fill*). Caso ainda restem valores ausentes após o preenchimento para trás, estes são preenchidos utilizando o último valor anterior válido.

A Organização Meteorológica Mundial desaconselha métodos como a média e a interpolação linear para dados de precipitação, pois eles assumem um comportamento aditivo e suave que contradiz a natureza multiplicativa, de distribuição log-normal e espacialmente complexa da chuva (WMO, 2020, 2021, 2023). No entanto, a utilização desses métodos neste estudo foi deliberada, com o objetivo específico de investigar como diferentes técnicas de imputação de dados afetam a conformidade da série com a Lei de Benford. Dado que tanto a distribuição da precipitação quanto a Lei de Benford possuem uma base logarítmica, a análise foca em avaliar a resiliência desta propriedade estatística fundamental da série completa frente a métodos de preenchimento que, embora não sejam fisicamente ideais, são de aplicação comum.

2.2. Lei de Benford

Para analisar a distribuição dos dígitos nos dados das séries medidas e das séries preenchidas, utilizou-se a Lei de Benford como referência teórica. Benford (1938) derivou as frequências esperadas dos dígitos em listas de números, definindo a Equação 2 para determinar as probabilidades de ocorrência do primeiro dígito. Conforme proposto por Newcomb (1881), o primeiro dígito dos dados segue uma distribuição logarítmica. A probabilidade de um dígito “d” ocorrer como primeiro dígito é expressa pela Equação 2.

$$(2) \quad \text{Prob}(D_1 = d_1) = \log\left(1 + \frac{1}{d_1}\right), \quad d_1 \in \{1, 2, \dots, 9\}$$

Para estender a análise aos segundos dígitos, aplicou-se a Equação 3, que calcula a probabilidade de ocorrência de cada dígito na segunda posição. Essa equação considera todas as combinações possíveis de primeiros dígitos (d_1) e segundos dígitos (d_2).

$$(3) \quad \text{Prob}(D_2 = d_2) = \sum_{d_1=1}^9 \log\left(1 + \frac{1}{d_1 d_2}\right), \quad d_2 \in \{0, 1, \dots, 9\}$$

Essas equações permitem prever a frequência esperada de cada dígito em suas respectivas posições. Com base nelas, a distribuição teórica da Lei de Benford (Tabela 2) foi comparada com a distribuição observada nos dados coletados, utilizando um teste estatístico de aderência para avaliar a conformidade da distribuição dos dados com a distribuição teórica de Benford.

Tabela 2 - Probabilidades esperadas de ocorrência do primeiro e segundo dígitos, de acordo com a lei de Benford

Dígito	Probabilidade de ocorrência do 1º dígito	Probabilidade de ocorrência do 2º dígito
0	-	0,11968
1	0,30103	0,11389
2	0,17609	0,10882
3	0,12494	0,10433
4	0,09691	0,10031
5	0,07918	0,09668
6	0,06695	0,09337
7	0,05799	0,09035
8	0,05115	0,08757
9	0,04576	0,08500
Total	1,00000	1,00000

Fonte: autores (2025).

2.3. Avaliando a conformidade com a lei de Benford

Foi aplicado o teste de aderência Qui-quadrado para verificar a conformidade da distribuição dos dados de pluviosidade com a distribuição de Benford. Neste teste, a hipótese nula (H_0) é que não há diferença significativa entre as frequências observadas e as frequências esperadas, ou seja, os dados aderem à distribuição de Benford (valor-p > 0,05). A hipótese alternativa (H_a) sugere que os dados não estão em conformidade com a lei de Benford (Costa Neto, 2002).

O teste Qui-quadrado é usado para comparar um conjunto de resultados reais com os resultados esperados, sendo a estatística do teste Qui-quadrado (χ^2) calculada conforme a Equação 4 (Costa Neto, 2002).

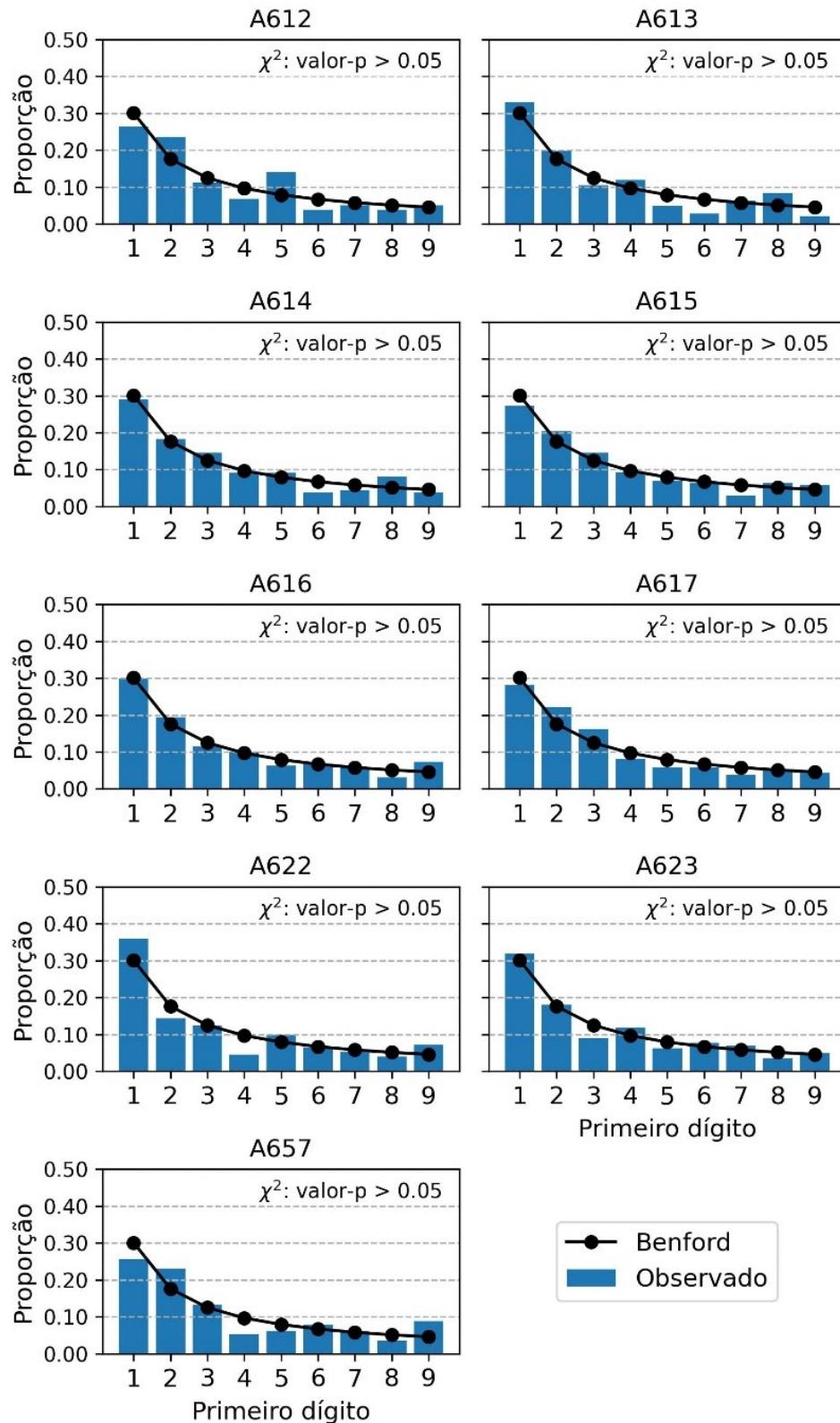
$$(4) \quad \chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

Em que: χ^2 é a estatística do teste Qui-quadrado; O_i é a contagem real da ocorrência de cada dígito; E_i é a contagem esperada pela lei de Benford para o respectivo dígito; e n representa o número de compartimentos (no caso do primeiro dígito é 9, pois varia de 1 a 9 e para o segundo dígito é 10, pois varia de 0 a 9).

3. RESULTADOS E DISCUSSÃO

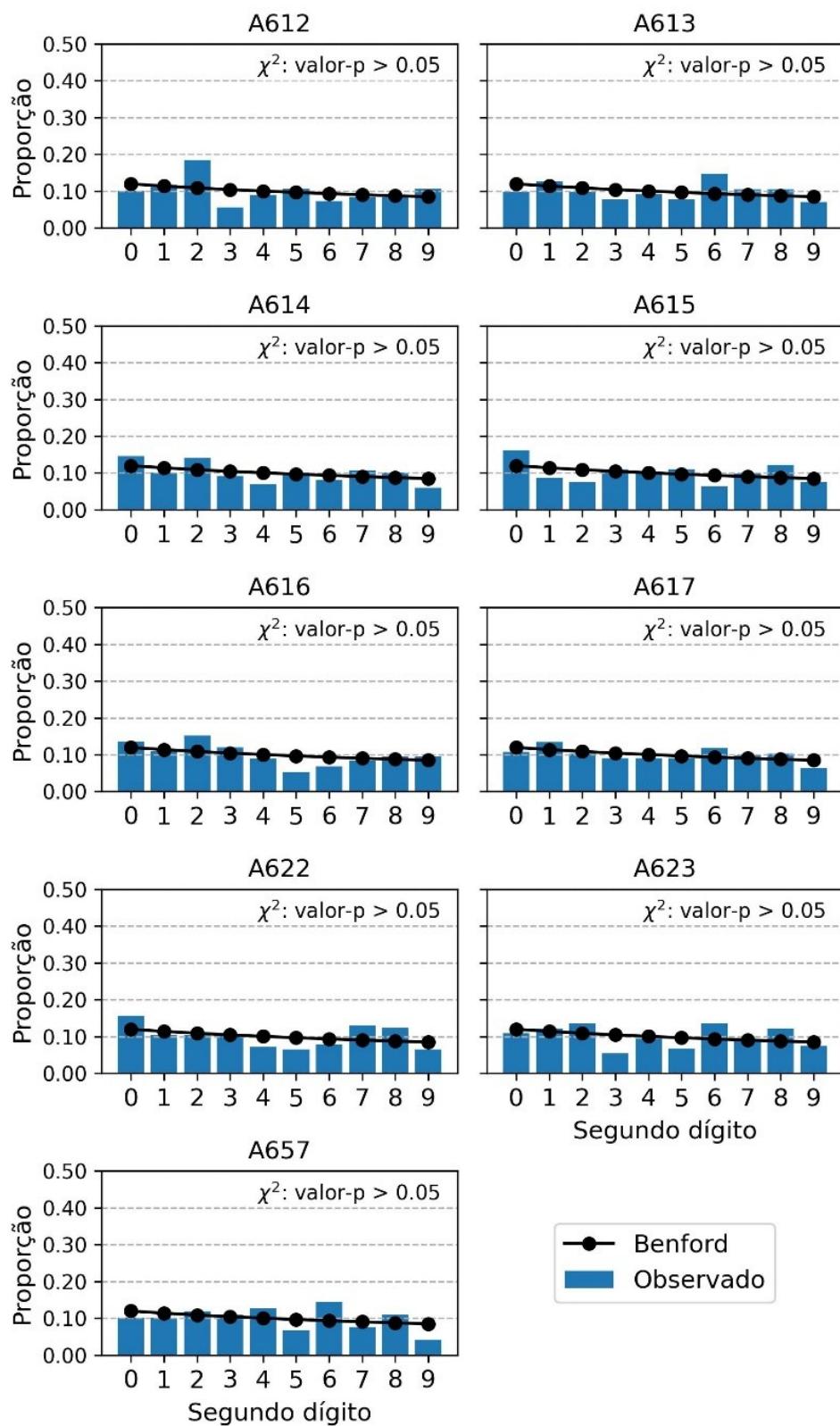
A análise das séries temporais mensais de precipitação nas nove estações meteorológicas automáticas do Espírito Santo revelou uma tendência de conformidade com a Lei de Benford, tanto para o primeiro dígito (Figura 2), quanto para o segundo dígito (Figura 3). O teste estatístico de aderência χ^2 não rejeitou a hipótese nula (H_0), pois todas as estações apresentaram valor- $p > 0,05$, indicando que não há diferenças estatisticamente significativas entre as frequências observadas e as esperadas.

Figura 2 – Proporções de ocorrência do primeiro dígito da lei de Benford do acumulado mensal de chuva observada, para cada estação meteorológica avaliada



Fonte: autores (2025).

Figura 3 – Proporções de ocorrência do segundo dígito da lei de Benford do acumulado mensal de chuva observada, para cada estação meteorológica avaliada



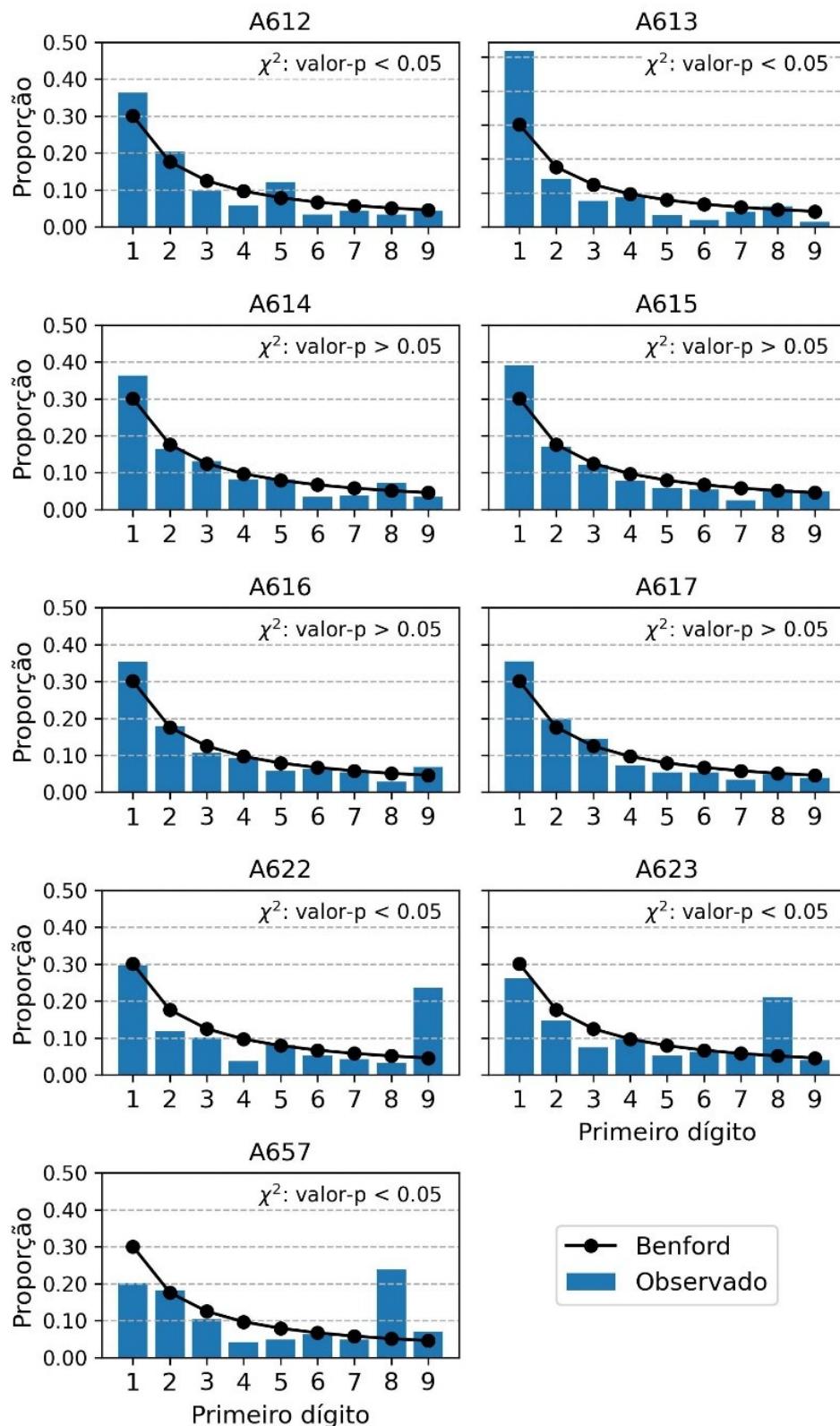
Fonte: autores (2025).

Essa conformidade inicial indica que os dados brutos de precipitação seguem um padrão compatível com a distribuição logarítmica da Lei de Benford. Isso sugere que, quando os dados abrangem uma ampla faixa de valores, tendem a seguir naturalmente a distribuição estatística descrita pela Lei de Benford.

Embora algumas estações, como A612 e A613, apresentem maior desvio para alguns dígitos, de modo geral, a conformidade foi estatisticamente significativa para o primeiro e segundo dígitos em todas as estações meteorológicas.

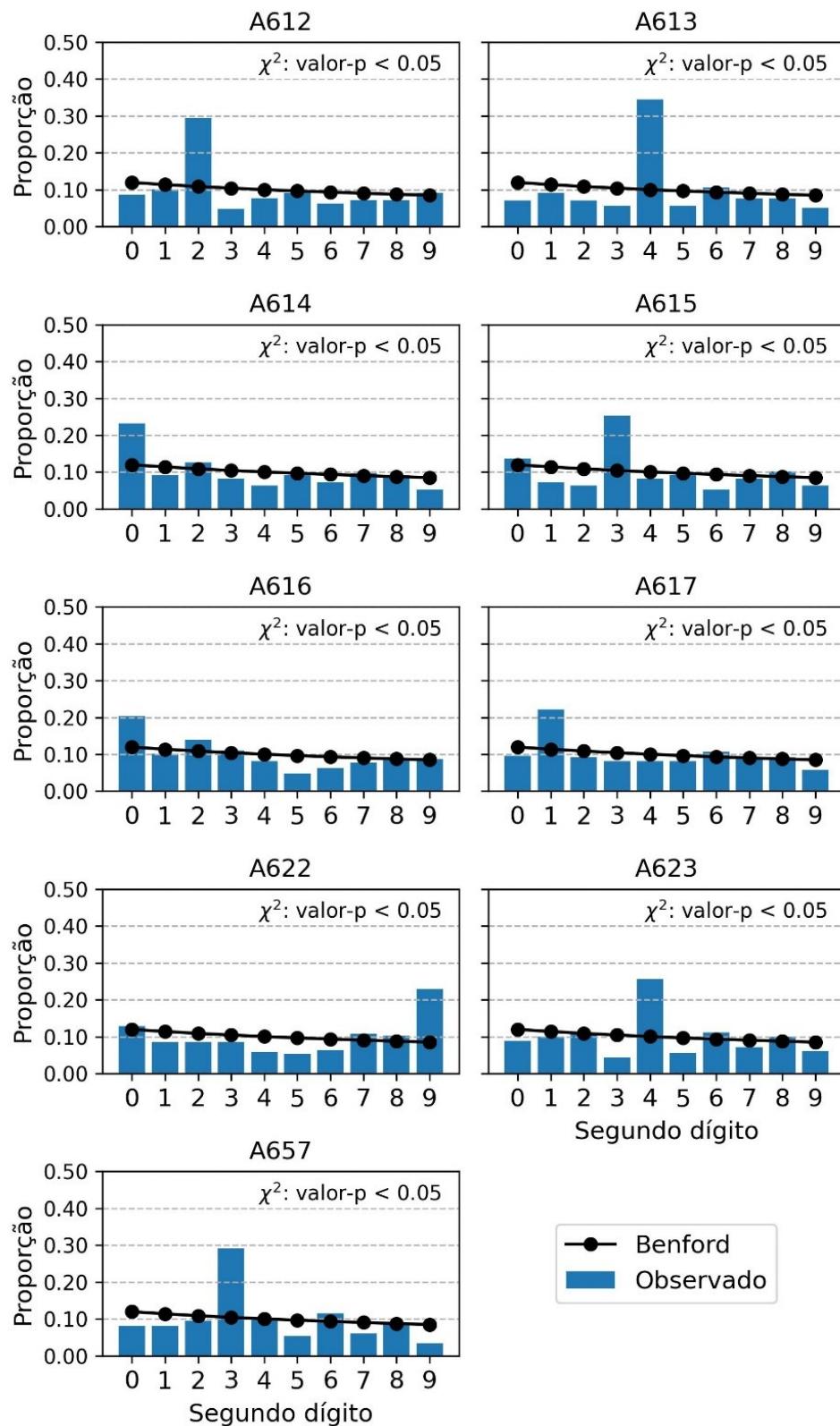
Após o preenchimento das falhas nas séries temporais pelo método da média, a análise da distribuição dos dígitos foi repetida, conforme a Figura 4 para o primeiro dígito e a Figura 5 para o segundo dígito.

Figura 4 – Proporções de ocorrência do primeiro dígito da lei de Benford do acumulado mensal de chuva observada após o preenchimento de falhas com a média, para cada estação meteorológica avaliada



Fonte: autores (2025).

Figura 5 – Proporções de ocorrência do segundo dígito da lei de Benford do acumulado mensal de chuva observada após o preenchimento de falhas com a média, para cada estação meteorológica avaliada



Fonte: autores (2025).

Observou-se uma leve mudança nas frequências dos primeiros dígitos nas estações A614, A615, A616 e A617 (Figura 4), porém não foram estatisticamente significativas, indicando que tais estações permaneceram seguindo a distribuição da Lei de Benford mesmo após o preenchimento de falhas com a média, considerando apenas a distribuição do primeiro dígito.

Contudo, após o preenchimento de falhas, as estações A612, A613, A622, A623 e A657 apresentaram uma não conformidade com a Lei de Benford, ou seja, rejeitando-se H_0 no teste de aderência de χ^2 , pois o valor-p do teste foi menor que 0,05 nestas estações meteorológicas. Isso indica que o método de preenchimento de falhas com a média alterou a distribuição de ocorrência natural dos dígitos, resultando nas anomalias detectadas pelo teste de Benford.

As estações meteorológicas A614, A615, A616 e A617 foram as que apresentaram as menores porcentagens de falhas (Tabela 1), o que pode ter sido a causa para a não detecção da alteração do padrão de Benford no primeiro dígito dessas estações. Nas demais estações, com maiores números de falhas, foram detectados desvios na distribuição no primeiro dígito.

Dessa forma, a quantidade de dados faltantes (preenchidos) influenciou significativamente a distribuição dos dígitos após o preenchimento. Esta situação corrobora o exposto trabalho de Cembrera-Conde, Silva-Ramírez e Cembrera (2024), que afirmam que séries com maior proporção de falhas preenchidas tendem a sofrer alterações mais expressivas na distribuição natural dos dígitos devido ao método de preenchimento, o que pode comprometer a aderência à Lei de Benford.

Para o segundo dígito (Figura 5), os resultados indicam que o preenchimento de falhas afetou de forma significativa a conformidade dos dados com a Lei de Benford. No geral, o teste de aderência indicara que em todas as séries temporais de precipitação, após o preenchimento de falhas com a média, tiveram a estrutura da distribuição teórica dos dígitos alteradas em todas as estações meteorológicas, independentemente do número de falhas nas séries.

A discrepância mais acentuada observada no segundo dígito, após o preenchimento das falhas (Figura 5), reforça a ideia de que o método de imputação afetou os aspectos mais sutis da distribuição numérica dos dados. Enquanto o primeiro dígito está associado a ordens de magnitude, o segundo dígito captura variações mais finas. Portanto, alterações nesse nível

podem indicar que o método de preenchimento está inserindo padrões artificiais que não refletem a variabilidade natural dos dados de precipitação.

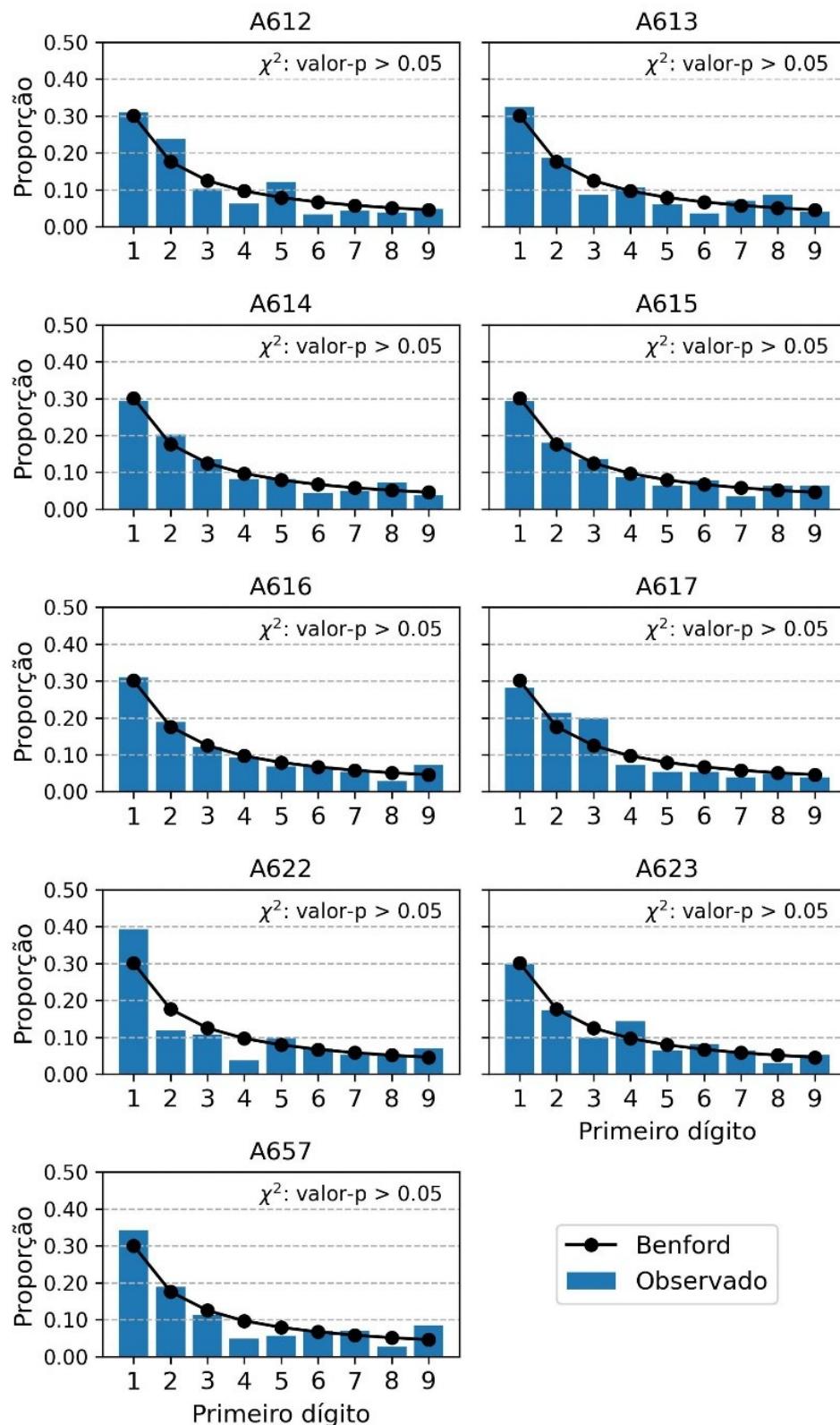
A maior sensibilidade do segundo dígito ao preenchimento de falhas pode ser atribuída à distribuição teórica de Benford para o segundo dígito, que é consideravelmente mais uniforme do que a do primeiro dígito, com probabilidades que variam de 11,968% (para o dígito 0) a 8,5% (para o dígito 9). Essa distribuição quase plana significa que não há uma dominância natural de um dígito sobre os outros. Consequentemente, a introdução repetida de valores artificiais, como uma média mensal constante, cria picos de frequência anômalos para dígitos específicos que se desviam de forma estatisticamente significativa da distribuição esperada (Marzio, Fensore e Passamonti, 2024; Nigrini, 2012). Portanto, o segundo dígito funciona como um detector mais sensível a padrões não naturais e à perda de variabilidade introduzidos por métodos de preenchimento mais simples.

Além disso, a variação na conformidade entre diferentes séries sugere que a quantidade e o padrão das falhas originais podem influenciar o impacto do método de preenchimento. Estações com maior quantidade de dados faltantes, ou com padrões de precipitação mais complexos, podem gerar ruído e serem mais suscetíveis a desvios na distribuição dos dígitos após o preenchimento de falhas (Marzio, Fensore e Passamonti, 2024).

Esses resultados têm implicações importantes para a análise de séries temporais hidrológicas (Nigrini e Miller, 2007). Primeiro, destacam a necessidade de utilizar métodos de preenchimento que preservem as propriedades estatísticas dos dados originais, especialmente quando análises baseadas na distribuição dos dígitos são previstas. Segundo, sugerem que a Lei de Benford pode ser uma ferramenta útil para avaliar a qualidade dos dados após procedimentos de preenchimento de falhas, funcionando como um indicador de possíveis anomalias introduzidas (Pain e Croset, 2023).

No caso do preenchimento utilizando interpolação linear, a Figura 6 apresenta a distribuição do primeiro dígito para os dados de precipitação mensal. Após a interpolação, observa-se que as estações estão em conformidade com a Lei de Benford (Figura 6). O teste de aderência χ^2 resultou em valores-p maiores que 0,05 nestes casos, indicando não rejeição da hipótese de conformidade com a Lei de Benford.

Figura 6 – Proporções de ocorrência do primeiro dígito da lei de Benford do acumulado mensal de chuva observada após o preenchimento de falhas com interpolação linear, para cada estação meteorológica avaliada

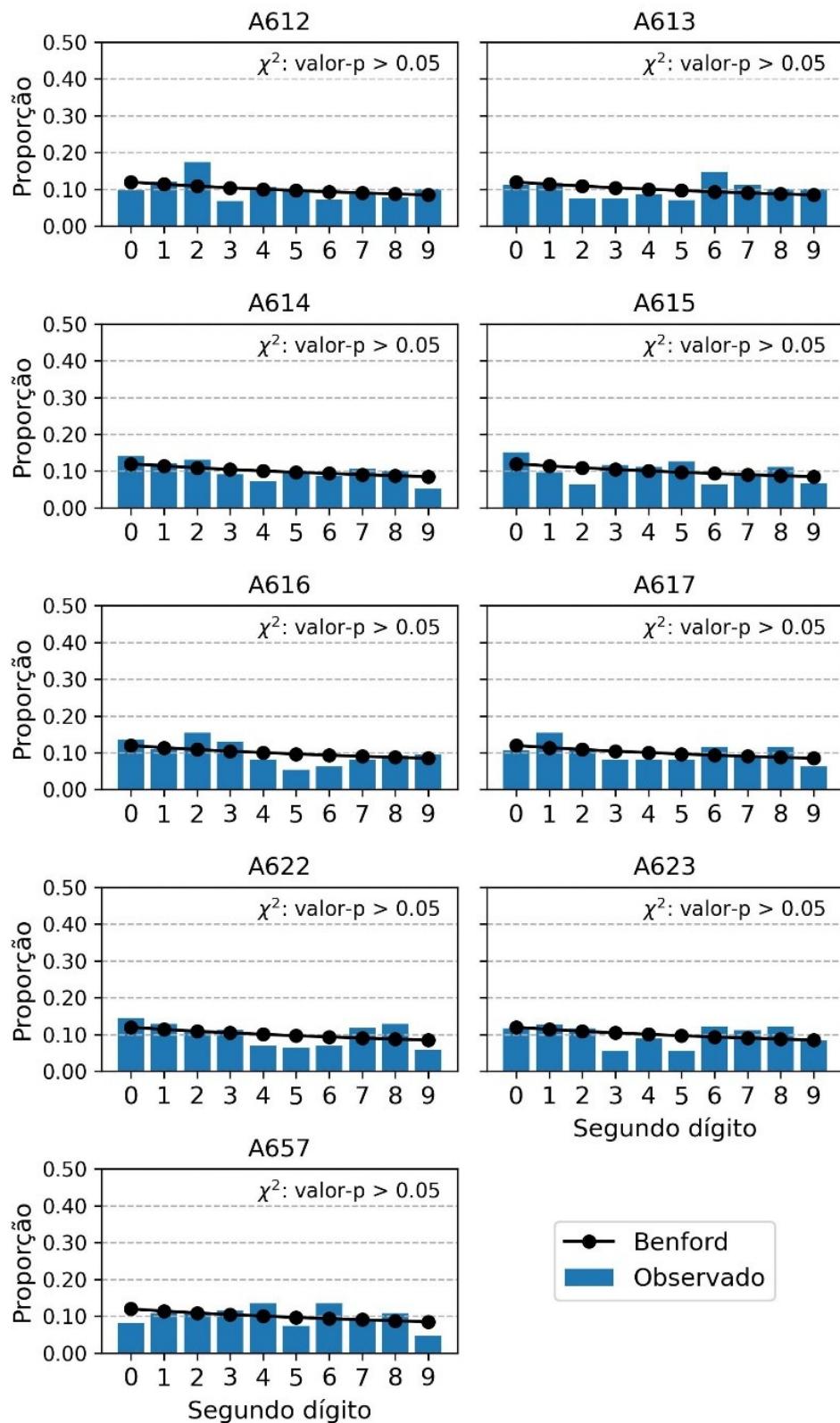


Fonte: autores (2025).

A ausência de discrepâncias entre as distribuições dos dados interpolados e de Benford, indicam que a interpolação linear não introduziu mudanças significativa na distribuição dos primeiros dígitos, especialmente nas estações que tinham uma maior quantidade de dados faltantes ou padrões de precipitação mais variáveis. Esses resultados sugerem que o método de interpolação linear pode preencher as séries com valores que mantêm as características estatísticas naturais dos dados, mantendo a adequação à Lei de Benford.

Adicionalmente, na distribuição do segundo dígito nas séries, após o preenchimento com interpolação linear (Figura 7), também se observa uma tendência de conformidade com a Lei de Benford, ou seja, as estações seguiram a distribuição teórica dos segundos dígitos esperada pela Lei de Benford. Isso sugere que os padrões naturais dos segundos dígitos não foram significativamente alterados pelo processo de interpolação linear.

Figura 7 – Proporções de ocorrência do segundo dígito da lei de Benford do acumulado mensal de chuva observada após o preenchimento de falhas com interpolação linear, para cada estação meteorológica avaliada



Fonte: autores (2025).

É importante considerar que o preenchimento de falhas em séries temporais é um procedimento comum para lidar com dados incompletos, porém a escolha do método pode impactar significativamente as análises subsequentes dos dados. Métodos de preenchimento que não preservam as características estatísticas originais dos dados podem levar a distorções nas distribuições dos dígitos. Por exemplo, se o método tende a atribuir valores médios (constantes) ou seguir um padrão menos variável, pode reduzir a diversidade numérica dos dígitos iniciais, desviando a distribuição observada daquela prevista pela Lei de Benford.

Portanto, a análise da aderência dos dados meteorológicos à Lei de Benford tem potenciais aplicações práticas no contexto da avaliação da qualidade dos dados (Nigrini e Miller, 2007), servindo como um indicador e não como prova de qualidade (Noleto-Filho *et al.*, 2022). Ao examinar se a distribuição dos primeiros dígitos dos dados segue o padrão esperado pela Lei de Benford, é possível detectar anomalias ou irregularidades que possam indicar problemas como erros de medição ou efeitos indesejados de métodos de preenchimento de falhas.

Dessa forma, incorporar a análise da Lei de Benford na fase inicial de avaliação das séries meteorológicas pode servir de complementação aos métodos tradicionais de verificação de dados, podendo contribuir para a análise de consistência das séries temporais de dados de precipitação.

4. CONSIDERAÇÕES FINAIS

Os resultados obtidos indicam que a Lei de Benford pode ser uma ferramenta útil para a análise de consistência de séries mensais de precipitação, especialmente para a identificação de anomalias e possíveis problemas nos dados. A conformidade observada nos dados brutos das estações sugere que as séries temporais de precipitação seguem o padrão de distribuição de dígitos previsto por Benford, o que reforça a confiabilidade dos dados meteorológicos fornecidos pelas estações do INMET.

O método de preenchimento das falhas com a média introduziu desvios na conformidade com a Lei de Benford, indicando que este método não é o mais adequado para o preenchimento de falhas mensais de precipitação.

A alteração na distribuição dos dígitos, após o preenchimento de falhas com a média, indica que o método pode ter introduzido vieses ou reduzido a variabilidade numérica, comprometendo a integridade dos dados para análises que dependem da distribuição natural dos dígitos, como é o caso da Lei de Benford.

A interpolação linear não alterou a conformidade dos dados de precipitação com a Lei de Benford, tanto para o primeiro quanto para o segundo dígito. Isso sugere que este método preservou as propriedades numéricas e estatísticas características das séries de dados originais.

O estudo apresenta limitações que devem ser consideradas. A análise foi baseada em dados de nove estações meteorológicas restritas ao estado do Espírito Santo, o que reduz a representatividade dos resultados para outras regiões com diferentes condições climáticas. Além disso, a utilização intencional de métodos de preenchimento de falhas não recomendados pela Organização Meteorológica Mundial (média e interpolação linear), embora adequada ao objetivo de avaliar seu efeito sobre a Lei de Benford, constitui uma limitação metodológica. Recomenda-se que pesquisas futuras ampliem a cobertura espacial e testem técnicas de preenchimento mais robustas para confirmar e generalizar os resultados obtidos.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior -Brasil (CAPES) - Código de Financiamento 001.

REFERÊNCIAS

- ALEXANDERSSON, H. A homogeneity test applied to precipitation data. *Journal of Climatology*, v. 6, n. 6, p. 661–675, 10 jan. 1986.
- BENFORD, F. The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society*, v. 78, n. 4, p. 551–572, 1938.
- BUISHAND, T. A. Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology*, v. 58, n. 1–2, p. 11–27, ago. 1982.
- COSTA NETO, P. L. DE O. **Estatística**. 3. ed. São Paulo, SP: Blucher, 2002.

CUMBRERA-CONDE, I.; SILVA-RAMÍREZ, E.-L.; CUMBRERA, F. L. Signal processing analysis for detection of anomalies in numerical series. **Expert Systems with Applications**, v. 255, p. 124708, 1 dez. 2024.

DURTSCHI, C.; HILLISON, W.; PACINI, C. The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data. **Journal of Forensic Accounting**, v. 5, n. 1, p. 17–34, 2004.

LEE, J.; CARVALHO, M. DE. Technological improvements or climate change? Bayesian modeling of time-varying conformance to Benford's Law. **PLOS ONE**, v. 14, n. 4, p. e0213300, 12 abr. 2019.

LI, Q.; FU, Z. Quantifying non-stationarity effects on organization of atmospheric turbulent eddy motion by Benford's law. **Communications in Nonlinear Science and Numerical Simulation**, v. 33, p. 91–98, 1 abr. 2016.

MARZIO, M. DI; FENSORE, S.; PASSAMONTI, C. Validating Benfordness on contaminated data. **Socio-Economic Planning Sciences**, v. 95, p. 102008, 1 out. 2024.

MCKINNEY, W. **Python para análise de dados: tratamento de dados com pandas, NumPy e Jupyter**. 3. ed. São Paulo, SP: Novatec Editora Ltda, 2023.

MORO, I. P.; CECÍLIO, R. A.; ZANETTI, S. S.; XAVIER, A. C.; PEZZOPANE, J. E. M. Tendências em séries temporais meteorológicas e seus impactos na evapotranspiração de referência diária. **Revista Brasileira de Climatologia**, v. 36, p. 202–224, 4 fev. 2025.

NEWCOMB, S. Note on the Frequency of Use of the Different Digits in Natural Numbers. **American Journal of Mathematics**, v. 4, n. 1/4, p. 39, 1881.

NIGRINI, M. J. **Benford's law: applications for forensic accounting, auditing, and fraud detection**. Hoboken: John Wiley & Sons, Inc., 2012.

NIGRINI, M. J.; MILLER, S. J. Benford's Law Applied to Hydrology Data—Results and Relevance to Other Geophysical Data. **Mathematical Geology**, v. 39, n. 5, p. 469–490, 12 set. 2007.

NOLETO-FILHO, E. M.; CARVALHO, A. R.; THOMÉ-SOUZA, M. J. F.; ANGELINI, R. Reporting the accuracy of small-scale fishing data by simply applying Benford's law. **Frontiers in Marine Science**, v. 9, 2 set. 2022.

OLIVEIRA, J. A. J. DE; MILANI FILHO, M. A. F.; FRANCISCHETTI, C. E.; OLIVEIRA JR, S. V. DE; PADOVEZE, C. L. Uma análise nos gastos da união: aplicabilidade da lei de Newcomb-Benford. **CONTABILOMETRIA - Brazilian Journal of Quantitative Methods Applied to Accounting**, v. 5, n. 2, p. 123–141, 2018.

PAIN, J.-C.; CROSET, P. Checking the reliability of opacity databases. **The European Physical Journal D**, v. 77, n. 4, p. 60, 13 abr. 2023.

PETTITT, A. N. A Non-Parametric Approach to the Change-Point Problem. **Applied Statistics**, v. 28, n. 2, p. 126, 1979.



SAMBRIDGE, M.; TKALČIĆ, H.; JACKSON, A. Benford's law in the natural sciences. **Geophysical Research Letters**, v. 37, n. 22, 17 nov. 2010.

SILVA, W. B. DA; TRAVASSOS, S. K. DE M.; COSTA, J. I. DE F. Using the Newcomb-Benford Law as a Deviation Identification Method in Continuous Auditing Environments: A Proposal for Detecting Deviations over Time. **Revista Contabilidade & Finanças**, v. 28, n. 73, p. 11–26, 1 abr. 2017.

WMO - WORLD METEOROLOGICAL ORGANIZATION. **Guidelines on Homogenization (WMO-No. 1245)**. Geneva: WMO, 2020.

WMO - WORLD METEOROLOGICAL ORGANIZATION. **Guidelines on Surface Station Data Quality Control and Quality Assurance for Climate Applications (WMO-No. 1269)**. Geneva: WMO, 2021.

WMO - WORLD METEOROLOGICAL ORGANIZATION. **Guide to Climatological Practices (WMO-No. 100)**. Geneva: WMO, 2023.

YANG, L.; FU, Z. Out-phased decadal precipitation regime shift in China and the United States. **Theoretical and Applied Climatology**, v. 130, n. 1–2, p. 535–544, 24 out. 2017.

