

ANÁLISE COMPARATIVA DE DADOS NOMINAIS E NUMÉRICOS DO SISTEMA NACIONAL DE INFORMAÇÕES SOBRE SANEAMENTO (SNIS)

Afonso Eris Ferreira de Andrade (*), Kaynã Monteiro dos Santos, Édson Aparecido Abdul Nour, Natanael Fernandes Sarmento.

* Faculdade de Engenharia Civil, Arquitetura e Urbanismo (FEC) - Universidade Estadual de Campinas (Unicamp), email: afonsoeris@hotmail.com

RESUMO

O Sistema Nacional de Informações sobre Saneamento (SNIS) é a base de dados mais ampla e abrangente no cenário nacional, sua metodologia é autodeclaratória e voluntária, mas ainda requer de instrumentos de validação. Dessa forma, esse estudo visa comparar, para o ano de 2017, os dados nominais do SNIS com as respostas da Pesquisa Básica de Informações Municipais (MUNIC/IBGE), assim como verificar a distribuição de dados numéricos coletados pelo SNIS com a Lei de Benford (lei dos primeiros dígitos) através do teste estatístico de independência de Qui-quadrado de Pearson. A comparação dos dados nominais do SNIS mostrou que 36% dos municípios brasileiros não declararam informações sobre a existência de uma Política e de um Plano Municipal de Saneamento no Município, as quais foram informadas no mesmo ano na pesquisa realizada pelo IBGE. Outros 9% e 14% dos municípios declararam informações com sentidos opostos em ambas as consultas sobre a existência de uma Política e um de Plano Municipal de Saneamento, respectivamente. Ademais, a análise estatística de dados numéricos de conformidade com a Lei de Benford identificou que os números de 8 dos 16 dados analisados do SNIS não apresentou conformidade com a lei, ou seja, não apresentavam uma distribuição natural, indicando que as séries de dados possam ter sido arredondadas, replicadas, reproduzidas, modificadas propositalmente ou não. Finalmente, ainda que o SNIS tenha representado um marco importante para direcionar a tomada de decisões sobre políticas públicas, regulamentar, avaliar, diagnosticar as questões sanitárias em todos os níveis federativos, o estudo mostrou que a participação dos municípios é insuficiente, como também é preciso criar uma forma consistente de validação dos dados coletados.

PALAVRAS-CHAVE: SNIS, MUNIC, Lei de Benford, Validação de dados, Saneamento

INTRODUÇÃO

O Sistema Nacional de Informações sobre Saneamento (SNIS) é uma base de dados instituída pela Lei Nº 11.445, para coletar e sistematizar dados sobre o saneamento no país, disponibilizando indicadores de desempenho. Com base nesses dados levantados por esse sistema, os entes federativos, municipais e estaduais podem parametrizar suas políticas públicas voltadas ao saneamento e, assim, alavancar sua eficiência. Entretanto, a metodologia usada, de responsabilidade do Ministério do Desenvolvimento Regional, requer ainda de instrumentos de validação, visto que a coleta dessas informações é voluntária e autodeclaratória por parte dos prestadores deste serviço em cada município.

OBJETIVOS

Este estudo visa analisar a consistência de dados nominais e numéricos para formulação dos indicadores do Sistema Nacional de Informações sobre Saneamento (SNIS) de forma comparativa a outra base de dados e uma lei de distribuição numérica.

METODOLOGIA

A avaliação da consistência dos dados coletados pelo Sistema Nacional de Informações sobre Saneamento (SNIS) foi categorizada de duas maneiras, diferindo de acordo com o tipo de dado analisado: nominais ou numéricos. Os dados nominais, os quais compreendiam respostas em forma de palavras, foram comparados com informações encontradas em uma base de dados disponível pelo IBGE, que fornecia exatamente o mesmo dado coletado no mesmo ano. Enquanto, os dados numéricos, foram comparados estatisticamente com uma lei de distribuição numérica com aplicações comprovadas na contabilidade.

DADOS NOMINAIS

Para analisar este tipo de informação, procurou-se uma base de dados com metodologia diferente da adotada pelo SNIS. A Pesquisa de Informações Básicas Municipais (MUNIC) é feita pelo Instituto Brasileiro de Geografia e Estatística (IBGE, 2018) e abrange todos os municípios da federação. A coleta de dados é obrigatória e feita diretamente com as

prefeituras pelo próprio órgão de pesquisa através de entrevistas com gestores da administração pública que detêm informação sobre os órgãos públicos.

Das diversas informações que continham tanto no SNIS como no MUNIC escolheu-se dois dados presentes em ambas bases de dados: a existência de um Plano Municipal de Saneamento Básico e a existência de uma Política Municipal de Saneamento Básico, requisitos obrigatórios segundo a Lei N° 11.445/2007 para todos os municípios brasileiros, indicados respectivamente pelos indicadores PO001 e P028 na base do SNIS, enquanto nos dados do MUNIC eram identificados respectivamente pelos códigos SMSBDG06024 e SMSBDG0603.

Na base do SNIS, os municípios respondiam as informações com “Sim”, “Não” ou simplesmente não declaravam uma resposta, enquanto para a base da MUNIC eram oferecidas três opções de respostas nominais: “Sim”, “Não” e “Em elaboração”. Desse modo, os dados obtidos pelo SNIS e pela MUNIC em 5.565 municípios referente ao ano de 2017 foram comparados e categorizados segundo a concordância ou conflitância das respostas.

DADOS NUMÉRICOS

Para a análise comparativa dos dados numéricos do SNIS utilizou-se a distribuição de Benford, também conhecida como a lei dos primeiros dígitos. Newcomb (1881) observou que nos livros de logaritmos as primeiras páginas eram mais usadas que as últimas e sugeriu que os logaritmos dos números que iniciam com o primeiro dígito de valor baixo eram mais consultados do que os números com primeiro dígito de valor alto. Consequentemente, os algarismos 1 e 2 seriam mais usados do que 8 e 9. A probabilidade da ocorrência dos dígitos em uma lista de números naturais foi apresentada anos mais tarde. Benford (1938) observou dados com mais de 20.000 observações e verificou que a frequência dos dígitos variava em determinadas posições nos números. A frequência dos primeiros dígitos segue uma relação logarítmica como apresentava na equação 1, em que $F(a)$ é a frequência do dígito a ser usado como primeiro algarismo.

$$F(a) = \log\left(\frac{a+1}{a}\right) = \log\left(1 + \frac{1}{a}\right) \quad (\text{Equação 1})$$

A lei de Benford calcula as frequências esperadas de vários dígitos em variadas posições em um número. É igualmente possível calcular a distribuição para dois dígitos, nesse caso, o dígito a varia de 00 a 99. Esse resultado permite o cálculo da frequência de um número específico estar na segunda posição, como por exemplo, a probabilidade do segundo algarismo ser 2 é a soma das probabilidades dos dois primeiros números serem 02, 12, 22, 32, 42, 52, 62, 72, 82 e 92. Dessa forma, pode-se calcular a probabilidade de um dígito a ser o n ésimo-algarismo de um número. A tabela 1 apresenta a probabilidade de um dígito de 0 a 9 aparecer em um número dada sua posição de acordo com a lei de Benford.

Tabela 1. Frequência esperada dos dígitos segundo a posição no algarismo. Fonte: Benford (1938)

Dígito	0	1	2	3	4	5	6	7	8	9
Frequência no primeiro algarismo	-	30,1%	17,6%	12,5%	9,7%	7,9%	6,7%	5,8%	5,1%	4,6%
segundo algarismo	12,0%	11,4%	10,8%	10,4%	10,0%	9,7%	9,3%	9,0%	8,8%	8,5%

Uma importante aplicação da lei foi observada na contabilidade. Em hipótese, quando as receitas de corporações estão pouco abaixo de números redondos, como \$789.000 e \$19,97 milhões, gerentes tenderiam a arredondar esses números para cima, \$800.000 e \$20,00 milhões, respectivamente. Desse modo, o excesso de números zeros e a escassez de nove no segundo dígito sinalizariam a alteração dos dados. Carslaw (1988) ao consultar os dados de receitas de empresas da Nova Zelândia, observou que, de fato, havia mais segundos dígitos zeros e menos segundos dígitos nove do que esperado. Dessa forma, as frequências da lei de Benford podem ser usadas para detectar fraude. Nigrini (1994) utilizou dados de um caso de fraude e identificou que os dígitos fraudulentos se destacavam dos demais. Nos dez anos de dados verificados, os desvios quando comparados à distribuição de Benford eram significativamente maiores nos últimos cinco anos. Isso pode ser justificado porque a lei de Benford é contra intuitiva, portanto, as pessoas não tendem a produzir números conforme essa distribuição. Ademais, ao longo do tempo, indivíduos tendem a repetir suas ações e entrar em uma rotina, podendo evidenciar ainda mais a produção de dados identificados por Nigrini (1994). Das variadas aplicações da lei, um estudo de Freitas et al. (2018) demonstra a aplicação do uso da lei de Benford na averiguação de dados do SNIS.

TESTE ESTATÍSTICO

Para determinar se existe associação entre a frequência dos primeiros dígitos dos dados do SNIS com a frequência teórica determinada pela distribuição de Benford foi utilizado o Teste de Independência Qui-quadrado de Pearson. O teste avalia

a existência de associação entre variáveis por meio da determinação de um valor de dispersão (p-valor) e verifica se a frequência de uma determinada observação em uma amostra está associada ou não com uma frequência esperada ou teórica dado um valor de significância. Portanto, quando o p-valor calculado entre a associação da distribuição dos primeiros dígitos dos dados escolhidos do SNIS e a distribuição de Benford é menor que o valor de significância de 5% (p-valor < 0,05), as informações do SNIS não possuem conformidade com a lei de Benford e podem ser questionadas.

A análise do teste foi realizada para 16 parâmetros das séries de dados de água, esgoto, resíduos sólidos e águas pluviais do SNIS no ano de 2017. Os dados escolhidos da série de água foram: população total atendida com abastecimento de água (AG001), quantidade de economias ativas (AG003), extensão da rede de água (AG005) e volume de água faturado (AG011). Na série de esgoto foram escolhidos os dados: população total atendida com esgotamento sanitário (ES001), extensão da rede de esgotos (ES004), volume de esgotos coletado (ES005) e volume de esgotos tratado (ES006). As informações de resíduos sólidos estudada: população total atendida declarada (Co164), quantidade total de resíduos coletados (Co119), quantidade de resíduos domiciliares coletados (Co111) e quantidade de trabalhadores remunerados alocados no manejo de resíduos sólidos (TB015). Por fim, os dados de águas pluviais analisados foram: extensão de vias públicas em áreas urbanas (IE017), extensão de vias públicas em áreas urbanas com pavimento e meio-fio (IE019), quantidade de bocas de lobo existente (IE021) e total de vias públicas com redes ou canais de águas pluviais subterrâneos (IE024).

RESULTADOS

Realizou-se o cruzamento de informações entre as duas bases de dados para contabilizar como os municípios responderam às perguntas sobre a existência de um Plano Municipal de Saneamento Básico e sobre a existência de uma Política Municipal de Saneamento Básico. Nas tabelas 2 e 3 são apresentadas as matrizes de confusão de todos os 5.565 municípios analisados, onde é possível confrontar as informações declaradas no SNIS e coletadas pelo IBGE no ano de 2017.

Tabela 2. Matriz de confusão relativa aos indicadores com informação da existência de um Plano Municipal de Saneamento Básico. Fonte: Autores

		SMSBDG06024 (MUNIC)		
		Não	Sim	Em Elaboração
PO001 (SNIS)	-	896	558	556
	Não	854	445	509
	Sim	350	1123	274

Tabela 3. Matriz de confusão relativa aos indicadores com informação da existência de uma Política Municipal de Saneamento Básico. Fonte: Autores

		SMSBDG0603 (MUNIC)		
		Não	Sim	Em Elaboração
PO028 (SNIS)	-	838	568	604
	Não	682	280	605
	Sim	222	1465	301

Com relação a existência de um Plano Municipal de Saneamento Básico, 350 municípios declararam “Sim” no SNIS, entretanto responderam “Não” no MUNIC, outros 445 municípios declararam não existir plano Municipal de Saneamento Básico no SNIS, enquanto responderam possuir plano no MUNIC. No que se refere a existência de uma Política Municipal de Saneamento, 222 municípios que declararam possuir uma política no SNIS responderam que não havia política municipal ao IBGE, outros 280 que declararam não possuir política no SNIS afirmaram possuir uma política municipal ao IBGE. Desse modo, a figura 1 reúne em categorias como as informações dos municípios confrontaram nas duas bases de dados.

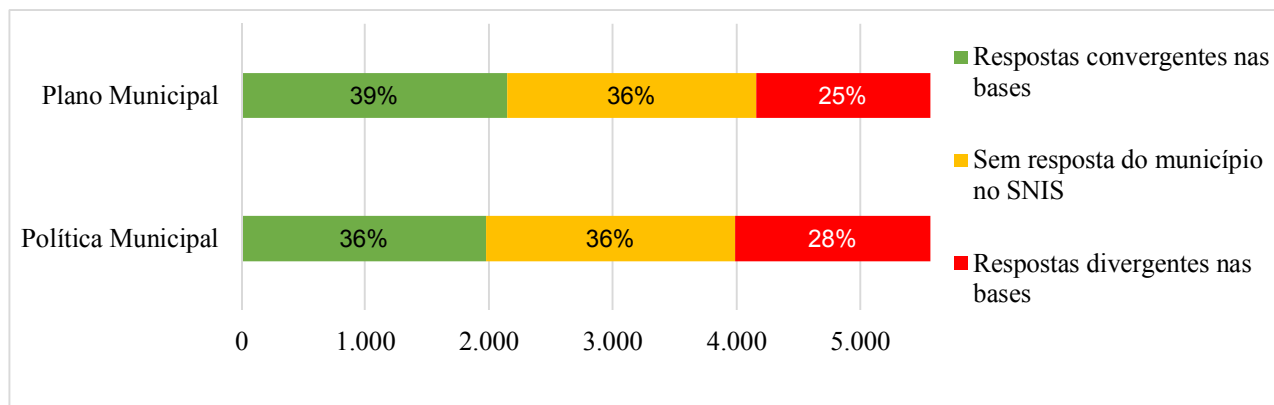


Figura 1. Representação dos dados cruzados entre as bases de dados SNIS vs MUNIC. Fonte: Autores

Para a base de dados do SNIS no ano de 2017, 2.010 municípios não declararam dados sobre os planos ou políticas voltadas ao saneamento de suas cidades, representando 36,1% do total, ilustrado em amarelo na figura 1. O destaque em verde representa a proporção dos municípios que declararam a mesmas informações nas bases, representando 39% e 36% dos municípios no que se refere a existência de um Plano e de uma Política Municipal, respectivamente. As respostas divergentes representam pelo menos um quarto dos 5.565 municípios em ambas as respostas. É importante esclarecer que quando se trata da existência de um Plano Municipal de Saneamento Básico, desses 25% apresentados como respostas divergentes, 64%, 16% do total, responderam “Em elaboração” na pesquisa do IBGE e os outros 36%, 9% do total, declararam, de fato, respostas de sentido opostos em ambas as pesquisas: “Sim” e “Não” ou vice-versa. Quando se trata da existência de uma Política Municipal, dos 28% apresentados como respostas divergentes, os municípios que responderam “Em elaboração” na pesquisa do IBGE representam 50%, 14% do total, enquanto os outros 50%, 14% do total, declararam, de fato, respostas de sentido opostos.

A distribuição numérica do primeiro dígito dos 16 parâmetros escolhidos do SNIS foi quantificada e comparada estatisticamente com a distribuição de Benford. A figura 02 apresenta a distribuição do primeiro algarismo dos dados de população total atendida com esgotamento sanitário (ES001) e total de vias públicas com redes ou canais de águas pluviais subterrâneos (IE024). Esses dois indicadores, ES001 e IE024, correspondem à distribuição de dados que, respectivamente, mais e menos se associaram com a distribuição de Benford pelo Teste de Independência Qui-quadrado de Pearson.

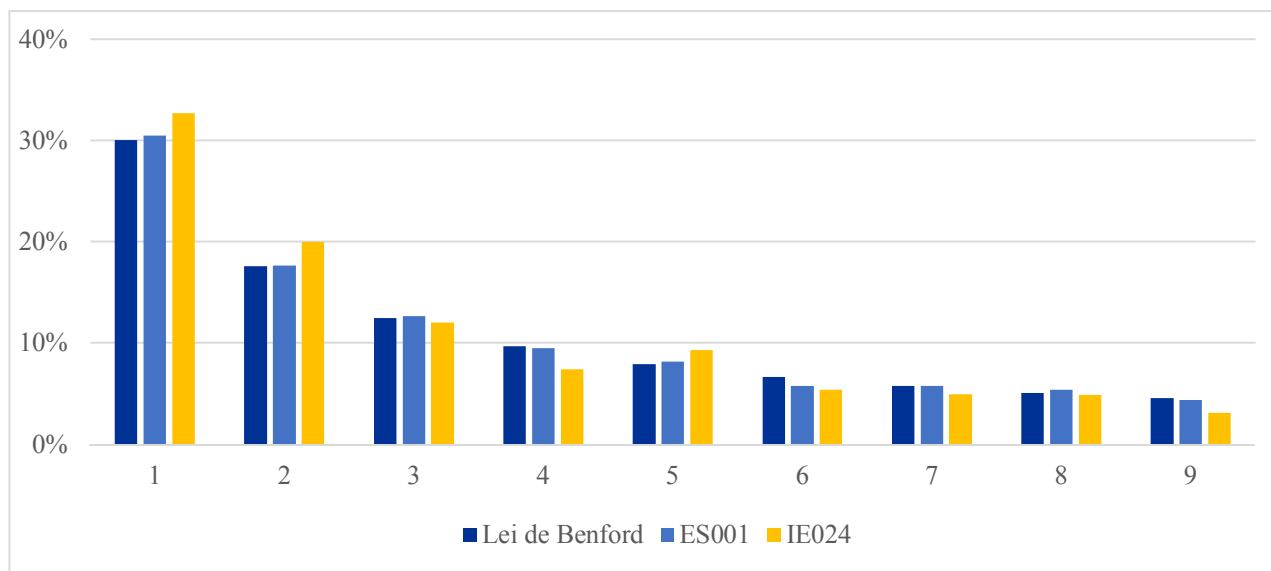


Figura 2. Distribuições observadas do primeiro algarismo dos dados de ES001 e IE024. Fonte: Autor

Para um nível de significância (α) de 5% foi analisado se cada série de dados possui ou não conformidade com a distribuição de Benford. Quando não há associação significativa com a lei, os dados não seguem uma distribuição natural, indicando a presença de arredondamentos, reprodução de dados ou imperfeições geradas ou não propositalmente. Numa condição de auditoria, como afirmado por Nigrini (1996), seriam necessários outros testes como a verificação dos segundos dígitos, cálculo de fatores de distorção ou a verificação de subgrupos da amostra para identificar as alterações na série de dados e se houve amplificação ou atenuação dos valores. Os resultados dos testes estatísticos estão resumidos na tabela 4.

Tabela 4. Parâmetros estatísticos do teste Qui-quadrado e a conformidade com a lei de Benford. Fonte: Autor

Série	Dado	X ² (8)	X ² crítico (8)	p-valor	Qtd. Observações	Resultado $\alpha = 0,05$
Água	AG001	9,706	15,507	0,286	3349	Há conformidade
	AG003	20,310	15,507	0,009	3346	Não há conformidade
	AG005	5,994	15,507	0,648	3376	Há conformidade
	AG011	10,709	15,507	0,219	3192	Há conformidade
Esgoto	ES001	3,938	15,507	0,863	2439	Há conformidade
	ES004	10,711	15,507	0,219	2423	Há conformidade
	ES005	18,828	15,507	0,016	2437	Não há conformidade
	ES006	5,347	15,507	0,720	1834	Há conformidade
Resíduos Sólidos	Co164	10,660	15,507	0,222	3556	Há conformidade
	Co119	20,041	15,507	0,010	3556	Não há conformidade
	Co111	6,642	15,507	0,576	1188	Há conformidade
	TB015	23,983	15,507	0,002	3556	Não há conformidade
Águas Pluviais	IE017	28,906	15,507	$3,29 \cdot 10^{-4}$	3666	Não há conformidade
	IE019	25,353	15,507	0,001	3603	Não há conformidade
	IE021	38,858	15,507	$5,22 \cdot 10^{-6}$	3309	Não há conformidade
	IE024	56,582	15,507	$2,17 \cdot 10^{-9}$	2575	Não há conformidade

Dos dezesseis dados avaliados, oito mostraram conformidade com a distribuição teórica de Benford, enquanto os outros oito não mostraram associação estatística suficiente. Na série de Água e na série de Esgoto, um dado em cada série foi rejeitado no teste estatístico, ao mesmo tempo que na série de Resíduos Sólidos dois dados não estavam em conformidade com a distribuição. Por fim, o teste Qui-quadrado apontou que todas as informações analisadas da série de Águas Pluviais não possuíam conformidade suficiente com a lei.

CONCLUSÕES

O Sistema Nacional de Informações sobre Saneamento é o maior sistema de informações do saneamento básico do país e representou um marco importante no setor, pois promoveu em nível nacional o planejamento, a avaliação, a regulamentação e o diagnóstico das questões sanitárias do país. Todavia, a comparação das informações declaradas no SNIS com as respostas coletadas pelo MUNIC mostrou que os dados nominais foram contrários em pelo menos 9% dos municípios brasileiros no ano de 2017. Além disso, 36% das prefeituras não declararam informações ao SNIS, as quais no mesmo ano foram informadas ao IBGE. Nota-se que ainda há participação insuficiente dos municípios ao declarar informações ao sistema. Ademais, o teste estatístico comparativo das informações numéricas das quatro séries de dados do mesmo ano com a distribuição de Benford apontou que em metade dos dados analisados os números possuíam um indicativo de “artificialidade”. É necessário, pois, criar uma forma consistente de validação dos dados declarados pelos municípios e prestadoras de serviços municipais. O decreto nº 7.217/2010 estabelece diretrizes para criação de políticas públicas municipais para a coleta e sistematização de dados relativos aos serviços de saneamento para avaliação e acompanhamento dos resultados, entretanto o adiamento dos prazos para criação de planos municipais de saneamento básico estabelecidos por outros quatro decretos mostrou que ainda há muito a ser realizado na esfera municipal.

REFERÊNCIAS BIBLIOGRÁFICAS

- BENFORD, F. The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society*, v. 78, n. 4, p. 551–572, 1938.
- BRASIL, **Lei nº 11.445**, de 5 de Janeiro de 2007. Estabelece diretrizes nacionais para o saneamento básico. Diário Oficial da União, Brasília-DF, 2007.
- BRASIL, **Decreto nº 7.217**, de 21 de Junho de 2010. Regulamenta a Lei no 11.445, de 5 de janeiro de 2007, que estabelece diretrizes nacionais para o saneamento básico, e dá outras providências. Diário Oficial da União, Brasília-DF, 2010.
- CARSLAW, C. A. P. N. Anomalies in Income Numbers: Evidence of Goal Oriented Behavior. *The Accounting Review*, v. 63, n. 2, p. 321–327, abr. 1988
- FREITAS, R. Medindo o saneamento: potencialidades e limitações dos bancos de dados brasileiros. **FGV CERI (Centro de Educação em Regulação e Infraestrutura)**. 2018.

6. Instituto Brasileiro de Geografia e Estatística (IBGE). **Perfil dos municípios brasileiros : Saneamento básico : Aspectos gerais da gestão da política de saneamento básico : 2017** / IBGE, Coordenação de População e Indicadores Sociais. - Rio de Janeiro : IBGE, 2018. 39p.
7. 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.
8. NIGRINI, M. Digital analysis and the reduction of auditor litigation risk; p. 14, jan. 1996.
9. NIGRINI, M. Using digital frequencies to detect fraud. The White Paper; p.p.3-6. abril. 1994.