A Surpreendente Lei de Benford

Vitor Amorim - IFSP; Aroldo Rodrigues - UFOPA

Na história da Matemática, por vez ou outra nos deparamos com descobertas surpreendentes, seja por seu caráter contraintuitivo, pela forma como a novidade foi revelada ou pelas inusitadas aplicações que elas proporcionam. Entre essas descobertas, o achado que ficou conhecido como Lei de Benford certamente ganha destaque.

Além de sua descoberta ter contado com elementos do acaso e de seu desenvolvimento ao longo de mais de um século ter continuado a revelar propriedades surpreendentes, esta inesperada lei acabou se transformando em uma poderosa ferramenta de combate a fraudes.

Este texto pretende explorar de forma breve algumas dessas características e mostrar como a sua combinação torna este tema rico para exploração didática em diversos segmentos do ensino.

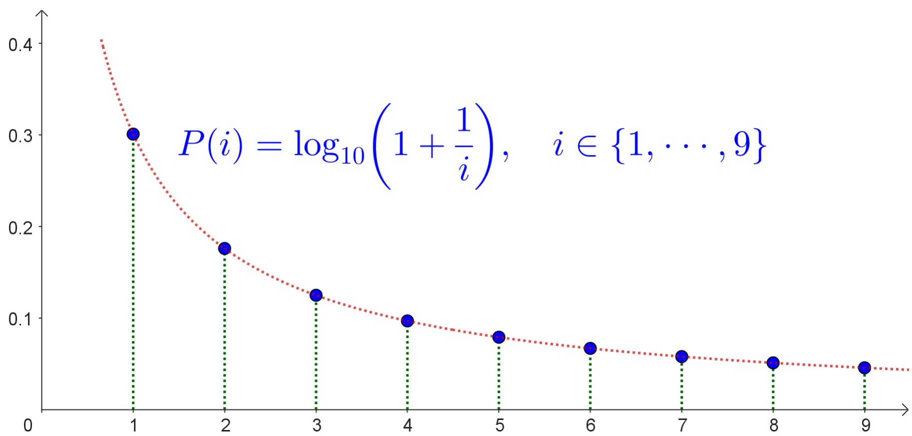
A Descoberta

Imagine que você dispõe de uma grande lista de dados de uma natureza arbitrária, como os comprimentos de todos os rios do planeta, o balancete de uma grande empresa, as populações de todas as cidades do Brasil ou os boletins de urna de todas as seções eleitorais na última eleição. Ao extrair o primeiro algarismo significativo de cada um desses números, com que frequência poderíamos esperar a aparição de cada um dos algarismos de 1 a 9 nesta posição?

A resposta mais comum e intuitiva para esta pergunta é $1/9 \cong 11{,}11\%$, já que a princípio não há motivo algum para acreditarmos que algum dos algarismos teria mais predisposição para aparecer na primeira posição do que outros. E é aí que está o primeiro ponto contraintuitivo da Lei de Benford.

Em meados do século XIX, o astrônomo Simon Newcomb (1835-1909) foi o primeiro a observar que, em grandes conjuntos de números das mais variadas naturezas, o algarismo 1 aparece na primeira posição significativa com maior frequência que o 2, que por sua vez é mais frequente que o 3 e assim sucessivamente.

A parte mais curiosa desta descoberta foi a forma como ela ocorreu. Antigamente, quando não existiam calculadoras, as tábuas de logaritmos eram imprescindíveis para a tarefa de realizar longos cálculos numéricos. Foi observando as extensas tábuas de logaritmo utilizadas na época que Newcomb percebeu que as páginas correspondentes aos algarismos menores estavam mais escuras e desgastadas do que as dos algarismos maiores, o que o levou a desconfiar que a distribuição de frequências do primeiro algarismo não era uniforme. Ao se debruçar sobre o fenômeno observando conjuntos de dados, ele propôs, como modelo, a seguinte lei de distribuição de probabilidades para o primeiro algarismo significativo:



Apesar de surpreendente e intrigante, a descoberta de Newcomb ficou esquecida no tempo e, por volta de 1938, ganhou novo destaque sendo redescoberta pelo físico Frank Benford (1883-1948).

Além de promover sua divulgação, Benford desenvolveu generalizações e realizou testes do funcionamento da Lei em imensos conjuntos numéricos de diversas origens. Hoje sabemos que este modelo descreve com bastante precisão a distribuição de frequências do primeiro algarismo em variados tipos de conjuntos de dados, como listas de comprimentos de rios, populações, extensões territoriais, balancetes contábeis entre outros.

Os Avanços

Ao longo do século XX, novas e surpreendentes descobertas sobre lei de Benford foram feitas, entre as quais destacamos:

- Foi obtida uma lei geral de distribuição aplicável aos algarismos das posições seguintes após a primeira, cuja fórmula para a probabilidade do algarismo i aparecer na posição j é dada por:

$$P_j(i) = \sum_{k=10^{j-2}}^{10^{j-1}-1} \log_{10}\left(1 + \frac{1}{10k+i}\right).$$

- A partir da lei acima, podemos mostrar que a distribuição de probabilidades dos algarismos converge para a distribuição uniforme, com 10% de probabilidade para cada algarismo (note que a partir do segundo algarismo o zero passa a ser considerado).

- A lei de Benford é invariante por mudança de escala. Assim, um conjunto de extensões de rios dadas em quilômetros, continuará aderindo à lei se os dados forem convertidos em milhas;

- Provou-se também que, se um conjunto de números adere à lei na base decimal, então este conjunto, quando convertido à uma base $b \geq 2$, irá satisfazer à lei análoga com a mudança da base do logaritmo para b;

- Sequências matemáticas conhecidas, como a sequência das potências de 2 ou a sequência de Fibonacci, satisfazem a lei de Benford. Os dois casos são consequências do seguinte resultado: se $\log_{10}(a)$ é irracional, então qualquer conjunto do tipo $\{a^n; n \geq 1\}$ satisfaz esta lei.

Vale observar que nem todo conjunto de dados adere a lei de Benford. Um conjunto de dados limitado em um intervalo restrito, como o conjunto das alturas de um grupo de humanos, claramente não vai satisfazer esta lei. Da mesma forma, a lei não se aplica a um conjunto de dados provenientes de uma distribuição uniforme, como os resultados de sorteios de uma loteria.

Apesar desses e outros casos em que a lei não funciona, descobertas recentes têm mostrado condições sobre um conjunto de dados para que a lei funcione e, de certa forma, justificado o fato de ser bastante amplo o seu campo de validade.

Atualmente, a lei de Benford é amplamente utilizada desde o combate a fraudes em orçamentos de obras públicas à confiabilidade de sistemas eleitorais. Afinal, embora não seja uma prova de fraude, quando um conjunto de dados não obedece a lei de Benford, ele merece no mínimo ser alvo de um escrutínio.

Modelagem Matemática e Consequências

As ferramentas matemáticas utilizadas para modelar fenômenos aleatórios em geral, como variáveis aleatórias e suas distribuições de probabilidades, podem ser encontradas da referência [1]. Para o desenvolvimento das propriedades específicas da Lei de Benford, utilizamos principalmente a referência [4].

O primeiro algarismo significativo de um número real tomado ao acaso, pode ser modelado por uma variável aleatória discreta $X \in \{1, \cdots, 9\}$. Assim, a lei apresentada anteriormente pode ser vista como a distribuição de probabilidades de X: $P(X = i) = \log(1 + 1/i)$ para $i = 1, \cdots, 9$, onde log sempre representará, a partir deste ponto, o logaritmo de base decimal.

Entretanto, a modelagem da lei de Benford e o estudo de suas propriedades ganham em praticidade e profundidade quando modelados por uma variável aleatória contínua, conforme detalharemos a seguir, de modo que a distribuição desta nova variável poderá ser associada à distribuição de X.

Sabemos que todo número $z \in \mathbb{R}$ pode ser escrito na forma $z = m \cdot 10^k$ onde $k \in \mathbb{Z}$ e $m \in [1, 10)$, chamado de *mantissa* de z. Dessa forma, a qualquer variável aleatória $Z \in \mathbb{R}$, pode-se associar a variável aleatória contínua $M \in [1, 10)$ que descreve a mantissa de Z. Note que, para $i \in \{1, \cdots, 9\}$, tem-se $X = i$ se, e somente se, $M \in [i, i + 1)$, estabelecendo, dessa forma, uma relação entre X e M.

Assim, podemos estudar as propriedades de X estabelecendo uma função densidade de probabilidade para M, que seja condizente com a lei de Benford. Esta função pode ser dada por:

$$f(x) := \begin{cases} \frac{1}{x \ln(10)}, & \text{se } x \in [1, 10) \\ 0, & \text{caso contrário} \end{cases}.$$

É fácil verificar que $f(x)$ define uma densidade de probabilidade, pois $\int_1^{10} f(x)dx = 1$. Além disso, podemos verificar a equivalência com a lei de Benford, observando que, para $i \in \{1, \cdots, 9\}$, temos

$$\begin{aligned} P(X = i) = P(i \leq M < i + 1) &= \int_i^{i+1} f(x)dx \\ &= \frac{1}{\ln(10)}(\ln(i + 1) - \ln(i)) \\ &= \frac{\ln\left(\frac{i+1}{i}\right)}{\ln(10)} \\ &= \log\left(1 + \frac{1}{i}\right). \end{aligned}$$

Dessa forma, propriedades matemáticas da lei de Benford podem ser obtidas pelo estudo da variável aleatória M, seja por meio de sua função densidade f ou de sua função de distribuição acumulada, dada por $F_M(x) = P(M \leq x)$. Neste ponto, lembramos que uma variável aleatória qualquer pode ser determinada univocamente por sua lei de distribuição acumulada. Assim, vale a pena ter em mãos a lei da função $F_M(x)$, dada por:

$$F_M(x) = P(M \leq x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0, & \text{se } x < 1 \\ \log(x), & \text{se } 1 \leq x < 10 \\ 1, & \text{se } x \geq 10 \end{cases}.$$

Com esta ferramenta em mãos, podemos obter, entre outras coisas, as demonstrações das propriedades mencionadas na seção anterior. Faremos aqui a prova da propriedade da invariância da lei por mudança de escala. Ou seja, se uma variável M que descreve a mantissa de um conjunto de números tem a lei de Benford como distribuição, então, dada uma constante $c > 0$, a variável cM também adere a esta lei.

Dado que o nosso interesse está na mantissa de um número real, podemos considerar, sem perda de generalidade, $1 \leq c < 10$. Isso se justifica porque uma constante fora deste intervalo pode ser escrita na forma $c = c' \cdot 10^k$ com $1 \leq c' < 10$ e $k \in \mathbb{Z}$, de modo que apenas o valor de c' terá impacto na mantissa de um número multiplicado por c.

Assim, considere um conjunto de números que satisfaz a lei de Benford e seja $M \in [1, 10)$ a variável aleatória que descreve a mantissa deste conjunto. Ao aplicarmos ao conjunto um fator de escala $c \in [1, 10)$, obteremos números do tipo $cM \cdot 10^k$, onde $k \in \mathbb{Z}$ e $1 \leq cM < 100$. Assim, se N descreve a mantissa do novo conjunto, teremos:

$$N = \begin{cases} cM, & \text{se } 1 \leq M < \frac{10}{c} \\ \frac{cM}{10}, & \text{se } \frac{10}{c} \leq M < 10 \end{cases}.$$

Portanto, a função de distribuição acumulada de N deve satisfazer, para $1 \leq x < 10$:

$$\begin{aligned} F_N(x) = P(N \leq x) = P(1 \leq N \leq x) &= P\left(1 \leq cM \leq x\right) + P\left(1 \leq \frac{cM}{10} \leq x\right) \\ &= P\left(\frac{1}{c} \leq M \leq \frac{x}{c}\right) + P\left(\frac{10}{c} \leq M \leq \frac{10x}{c}\right) \\ &= P\left(1 \leq M \leq \frac{x}{c}\right) + P\left(\frac{10}{c} \leq M \leq \frac{10x}{c}\right). \end{aligned}$$

Consideremos então os seguintes casos: 1) Se $1 \leq x < c$, então o primeiro termo da soma acima é igual a zero e, aplicando a lei de distribuição de M, o segundo termo resulta em $\log(x)$; 2) Se $c \leq x < 10$, obtemos:

$$\begin{aligned} F_N(x) &= P\left(1 \leq M \leq \frac{x}{c}\right) + P\left(\frac{10}{c} \leq M \leq \frac{10x}{c}\right) \\ &= P\left(1 \leq M \leq \frac{x}{c}\right) + P\left(\frac{10}{c} \leq M \leq 10\right) \\ &= \log\left(\frac{x}{c}\right) + \log(c) = \log(x). \end{aligned}$$

Logo, temos $F_N(x) = \log(x) = F_M(x)$ para todo $1 \leq x < 10$. Como as funções são constantes iguais a 0 e 1 para os casos $x < 1$ e $x \geq 10$, respectivamente, concluímos que as variáveis N e M possuem a mesma distribuição de probabilidades, ou seja, ambas satisfazem a lei de Benford.

Lei de Benford e Triângulo de Sierpinski

Com o passar do tempo foram descobertos diversos contextos de aparição do padrão da lei de Benford. Em particular, sequências numéricas bastante conhecidas na Matemática como a sequência das potências de 2 e a sequência de Fibonacci obedecem a este padrão. Podemos mostrar que esses dois casos são consequências de um resultado mais geral, que comentaremos mais abaixo, e que também dá origem ao seguinte resultado:

Dado um triângulo equilátero qualquer de área A_0, a sequência das áreas (pintadas em preto) das figuras produzidas pelas iterações que transformam o triângulo original no triângulo de Sierpinski satisfaz a lei de Benford.

Antes de apresentar uma justificativa para esta surpreendente afirmação, vamos demonstrar, de forma parcial, o seguinte resultado: se $\log(a)$ é irracional, então o conjunto $A = \{a^n, n \in \mathbb{N}\}$ satisfaz a lei de Benford.

Afirmamos inicialmente que a variável aleatória $M \in [1, 10)$ que descreve a mantissa de um conjunto segue a lei de Benford se, e somente se, a variável $Y = \log(M)$ tem distribuição uniforme no intervalo $[0, 1)$. Isso pode ser verificado obtendo as funções de distribuição acumulada de M e Y sob cada uma das condições, suficiente e necessária. Faremos um caso e deixaremos o outro por conta do leitor. Se assumirmos que M satisfaz a lei de Benford, temos, para $0 \leq x < 1$:

$$F_Y(x) = P(Y \leq x) = P\left(\log(M) \leq x\right) = P\left(M \leq 10^x\right) = \log\left(10^x\right) = x,$$

que é exatamente a lei da distribuição (acumulada) uniforme quando $0 \leq x < 1$. Os casos $x < 0$ e $x \geq 1$ seguem de imediato.

De volta aos elementos do conjunto A, queremos mostrar que sua respectiva mantissa M segue a lei de Benford ou, equivalentemente, que $Y = \log(M)$ tem distribuição uniforme no intervalo $[0, 1)$.

Note que se um número real está escrito na forma $z = m \cdot 10^k$, podemos obter o logaritmo de sua mantissa fazendo $\log(m) = \log(z) - k = \log(z) - \lfloor \log(z) \rfloor$, onde $\lfloor x \rfloor$ indica a função piso de x, que descreve o maior número inteiro menor ou igual a x. Com isso em mente, definimos a função $d(x) = x - \lfloor x \rfloor$ e obtemos $\log(m) = d(\log(z))$.

Assim, estamos interessados na distribuição dos números $d\left(\log\left(a^n\right)\right) = d\left(n \log(a)\right)$, para $n \in \mathbb{N}$. Neste ponto, aplicaremos um resultado conhecido como Teorema de Weyl (cuja demonstração pode ser encontrada em [2]), que afirma que se ψ é irracional, então os elementos da sequência $(d(n\psi))_{n\in\mathbb{N}}$ têm distribuição uniforme no intervalo $[0, 1)$.

Logo, se $\log(a)$ irracional, os elementos da sequência $(d\left(n \log(a)\right))_{n\in\mathbb{N}}$ (e, portanto, os logaritmos das mantissas das potências a^n) têm distribuição uniforme em $[0, 1)$. Ou seja, os elementos do conjunto A aderem à lei de Benford.

Este poderoso resultado pode ser aplicado para mostrar a aderência das potências de 2 e dos números de Fibonacci à lei de Benford. Mas neste texto vamos nos ater a demonstrar o resultado enunciado sobre a sequência que gera o triângulo de Sierpinski.

Inicialmente, note que a sequência das áreas pintadas em preto forma uma progressão geométrica de razão $\frac{3}{4}$, pois a cada iteração a área em questão corresponde a três quartos da área anterior. Assim, a área do n-ésimo triângulo pode ser dada pelo termo geral $A_n = A_0 \cdot \left(\frac{3}{4}\right)^n$.

Pela propriedade da mudança de escala, temos que a sequência $(A_n)_{n\in\mathbb{N}}$ adere à lei de Benford se, e somente se, a sequência dada por $\frac{1}{A_0} \cdot A_n = \left(\frac{3}{4}\right)^n$ também o faz. Logo, aplicando o resultado anterior, nosso argumento estará finalizado se mostrarmos que $\log\left(\frac{3}{4}\right)$ é irracional.

De fato, se $\log\left(\frac{3}{4}\right)$ fosse racional existiriam inteiros $p < 0$ e $q > 0$ tais que:

$$\log\left(\frac{3}{4}\right) = \frac{p}{q} \Leftrightarrow \frac{3}{4} = 10^{\frac{p}{q}} \Leftrightarrow 10^{-p} \cdot 3^q = 4^q \Leftrightarrow 2^{-p} \cdot 5^{-p} \cdot 3^q = 2^{2q},$$

o que contradiz o Teorema Fundamental da Aritmética. Portanto, somos obrigados a concluir que $\log\left(\frac{3}{4}\right)$ é irracional, o que encerra a demonstração.

Possibilidades de Uso Didático

Os diversos pontos explorados neste texto, entre outras possibilidades não descritas aqui, fazem da lei de Benford um tema rico para exploração em sala de aula em múltiplos contextos de profundidade, etapas de ensino e objetivos de aprendizagem.

Existem inesgotáveis fontes que podem embasar a construção de uma sequência didática para a exploração deste tema. Em particular, sugerimos as referências [3] e [5] para trazer algumas inspirações, sendo que a segunda se trata de um episódio de uma excelente série da Netflix que trata de diversos temas ligados à ciência, tecnologia e sociedade.

Sem a pretensão de esgotar as possibilidades, apresentaremos a seguir mais algumas ideias gerais para uso didático do tema, que podem ser destrinchadas e melhoradas de acordo com o contexto em que serão trabalhadas.

- Características do desenvolvimento histórico de novas ideias matemáticas podem ser discutidas a partir da forma inusitada como a lei de Benford foi descoberta, a qual evidencia que nem todo desenvolvimento é linear nem tampouco sistemático, a partir de um encadeamento lógico de ideais.

- Outra característica histórica com elementos de interesse está no uso das tábuas de logaritmos, prática que era necessária para a realização de cálculos numéricos extensos, mas tem pouca ou nenhuma utilidade nos dias de hoje. Tanto o contexto da criação do conceito de logaritmo e sua aplicação em cálculos numéricos, quanto a obsolescência deste e de outros métodos matemáticos podem ser explorados nesta discussão.

- O desenvolvimento histórico da lei de Benford ilustra também de forma precisa o modo de operação da chamada Modelagem Matemática. A observação do fenômeno, seu registro e organização, a seleção de ferramentas matemáticas abstratas que o aproximam, a comparação entre a ferramenta e o fenômeno observado e o aprofundamento matemático proporcionando novas descobertas são etapas que compõem o cerne desta prática científica. Evidenciar todo este processo em sala de aula promoverá maior compreensão deste mecanismo que aparece de forma recorrente nos contextos escolar, acadêmico, profissional e da vida em sociedade como um todo.

- Ainda no contexto da Modelagem Matemática, as aplicações e limitações do modelo precisam ser bem compreendidas. Por exemplo, a lei de Benford pode ser aplicada para análise de documentos contábeis de obras públicas ou de boletins de urnas das eleições com objetivo de verificar a aderência ao modelo e, possivelmente, encontrar um indicativo de fraude (caso não haja a aderência). Entretanto, é igualmente importante ressaltar duas limitações deste modelo: 1) A aderência à lei de Benford **não é uma prova** de que não houve fraude, visto que o fraudador pode conhecer e forjar a obediência ao modelo; 2) A não aderência **não é uma prova** de fraude, pois outros diversos fatores, como alguns que citamos na segunda seção deste texto, podem alterar a distribuição dos algarismos, questão esta que pode gerar mais uma rica discussão sobre o tema.

- A questão dos espaços amostrais equiprováveis e não equiprováveis pode ser explorada em conjunto com o caráter contraintuitivo da lei. Que tipo de distribuição esperaríamos? Por quê? Em quais casos é intuitivo ter uma distribuição equiprovável ou não?

- As propriedades matemáticas da lei de Benford citadas ao longo do texto geram inúmeros contextos de exploração em sala de aula. Ao final do texto, mostraremos uma proposta mais específica de trabalho com essas propriedades. Mas é importante nomear alguns dos inúmeros temas matemáticos que podem ser desenvolvidos tomando a lei de Benford como pano de fundo, quais sejam: distribuição de frequências e sua relação com probabilidades, distribuições uniformes e não uniformes, logaritmo e função logarítmica, representações gráficas, escalas de medidas e conversões, notação científica, ordem de grandeza e mantissa, sistemas numéricos em diferentes bases, potências, progressões, recorrências e sequência de Fibonacci, irracionalidade de números reais, Teorema Fundamental da Aritmética, propriedades dos triângulos equiláteros, fractais e construção do triângulo de Sierpinski e, para o contexto do ensino superior, variáveis aleatórias discretas e contínuas e suas distribuições de probabilidade.

Encerraremos o texto apresentando de forma detalhada uma proposta de abordagem didática para a lei de Benford. Além de abordar vários aspectos já mencionados, esta proposta ainda inclui o trabalho com planilhas eletrônicas e o consequente desenvolvimento de habilidades matemáticas e computacionais.

O primeiro passo da proposta envolve a busca e a tabulação de alguns dados, por exemplo, as listas das populações de todos os países do mundo e de suas respectivas extensões territoriais (em km^2). Uma lista com dados atualizados pode ser obtida, entre outras fontes, em `https://www.worldometers.info/world-population/population-by-country/`.

Com os dados copiados e organizados em uma planilha, o questionamento feito no início deste texto pode ser repetido: se extrairmos o primeiro algarismo de cada número, com qual frequência devemos esperar a aparição de cada algarismo?

Em seguida, podemos usar a função ESQUERDA do Excel ou comando equivalente em outra planilha eletrônica para tabular os primeiros algarismos significativos de cada lista, como mostrado parcialmente na figura abaixo.

Com a função CONT.SE, podemos construir uma tabela de distribuição de frequências dos algarismos para apresentar o resultado e confrontar com a primeira resposta dada pelos alunos. A função CONT.SE conta a quantidade de células de um certo intervalo que contém um valor específico. Assim, podemos contar quantas vezes aparece cada algarismo e, dividindo pelo total, obter as frequências relativas.

	A	B	C	D	E
1		Populações	Primeiro Algarismo	Extensões Territoriais (km^2)	Primeiro Algarismo
2	**1**	1.439.323.776	1	9.388.211	9
3	**2**	1.380.004.385	1	2.973.190	2
4	**3**	331.002.651	3	9.147.420	9
5	**4**	273.523.615	2	1.811.570	1
6	**5**	220.892.340	2	770.880	7
7	**6**	212.559.417	2	8.358.140	8
8	**7**	206.139.589	2	910.770	9
9	**8**	164.689.383	1	130.170	1
10	**9**	145.934.462	1	16.376.870	1
11	**10**	128.932.753	1	1.943.950	1

Algarismo	Frequência Populações	Frequência Extensões T.
1	29,49%	28,63%
2	15,81%	19,23%
3	12,82%	13,25%
4	8,97%	8,97%
5	11,54%	7,26%
6	7,69%	5,56%
7	3,42%	5,98%
8	5,56%	5,13%
9	4,70%	5,98%
Total	100,00%	100,00%

Para mostrar que não se trata de uma coincidência e aumentar o convencimento dos alunos de que existe um padrão e de que este padrão não é a distribuição uniforme, uma ideia é colocar mais duas colunas de dados. Aproveitando os dados já existentes, podemos construir uma coluna de densidade populacional, obtida pelas razões entre as populações e suas respectivas extensões territoriais.

Uma dica importante para este ponto: se os dados forem retirados de um *site* de um país que adota a vírgula como separador de classes dos números (como é o caso da sugestão que demos acima), então o Excel em língua portuguesa entenderá que a entrada "206,139,589" se trata de um texto e não de um número. Nesse caso, o *software* não realiza a divisão para obter a densidade populacional. Isso pode ser resolvido selecionando todos das células da planilha que contém números, utilizando o comando "Ctrl L" no teclado e, no menu Substituir, solicitar a troca das vírgulas por pontos.

Outra questão que se deve observar é se algum dado utilizado ou gerado por um cálculo (como é o caso das densidades populacionais) resultou em um número positivo menor que 1. Neste caso, a função ESQUERDA retornaria o algarismo 0. Mas há uma forma de contornar isso. Note que se $Z = M \cdot 10^k$, onde $M \in [1, 10)$ e $k \in \mathbb{Z}$, então o primeiro dígito significativo de Z pode ser calculado pela fórmula

$$\left\lfloor \frac{Z}{10^{\lfloor \log(Z) \rfloor}} \right\rfloor = \left\lfloor \frac{M \cdot 10^k}{10^k} \right\rfloor = \lfloor M \rfloor,$$

pois $\lfloor \log(Z) \rfloor = \lfloor k + \log(M) \rfloor = k$. Esta fórmula pode ser inserida em uma planilha eletrônica utilizando a função INT para representar a função piso ($\lfloor x \rfloor$) e LOG10 para calcular o logaritmo decimal. Apesar de ser uma forma mais trabalhosa de inserir na planilha, esta fórmula não depende de que o número seja maior que 1, como é o caso da função ESQUERDA.

Outra atividade interessante seria inserir uma coluna com dados totalmente desconectados dos primeiros, utilizando, por exemplo, a sequência de Fibonacci, cuja construção pode ser feita automaticamente na planilha. Em seguida, tabulamos os primeiros algarismos das duas novas colunas e construímos suas respectivas distribuições de frequências.

Também podemos explicitar a propriedade de mudança de escala. As extensões territoriais podem ser convertidas em mi^2 (milhas quadradas) construindo uma nova coluna na planilha. O momento pode ser aproveitado inclusive para a discussão sobre os sistemas de medidas e a revisão sobre conversões de unidades. No caso em questão, basta dividir os valores em km^2 por 2,59 (cálculo aproximado). Em seguida, acrescentamos esses dados às distribuições de frequências.

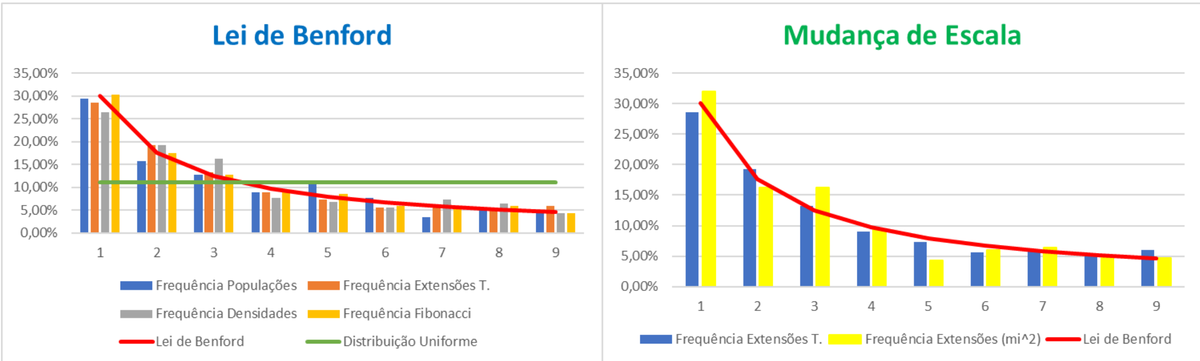
Por fim, caso o tempo permita, poderíamos ampliar o número de colunas pedindo para que os alunos pesquisassem e trouxessem na aula seguinte listas de dados para pesquisar quais seriam os padrões das distribuições de frequências dos primeiros algarismos. Inclusive, os dados reunidos por eles que não satisfizessem a lei de Benford seriam um ótimo objeto de análise, servindo de base para explorar os limites de validade da lei.

A esta altura, os alunos já terão percebido a presença de um padrão aproximado, o padrão da Lei de Benford. Podemos acrescentar então às tabelas de distribuições de frequências, duas colunas com a distribuição uniforme (11,11% para todos) e com a própria Lei de Benford, utilizando o comando

=LOG10(1+1/XX), onde XX é a célula correspondente ao algarismo em questão. Obtemos então o seguinte resultado:

Algarismo	Frequência Populações	Frequência Extensões T.	Frequência Densidades	Frequência Fibonacci	Frequência Extensões (mi^2)	Lei de Benford	Distribuição Uniforme
1	29,49%	28,63%	26,50%	30,34%	32,05%	30,10%	11,11%
2	15,81%	19,23%	19,23%	17,52%	16,24%	17,61%	11,11%
3	12,82%	13,25%	16,24%	12,82%	16,24%	12,49%	11,11%
4	8,97%	8,97%	7,69%	8,97%	9,40%	9,69%	11,11%
5	11,54%	7,26%	6,84%	8,55%	4,27%	7,92%	11,11%
6	7,69%	5,56%	5,56%	5,98%	5,98%	6,69%	11,11%
7	3,42%	5,98%	7,26%	5,56%	6,41%	5,80%	11,11%
8	5,56%	5,13%	6,41%	5,98%	4,70%	5,12%	11,11%
9	4,70%	5,98%	4,27%	4,27%	4,70%	4,58%	11,11%
Total	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Construir representações gráficas para essas distribuições ajudará na visualização a na compreensão dos conceitos. Mostramos alguns exemplos abaixo.



Como última sugestão, podemos construir também colunas de erros para explicitar a diferença entre os conjuntos que parecem aderir à Lei de Benford e o modelo teórico. Esta ação deixará claro um dos aspectos mais importantes da Modelagem Matemática: modelos teóricos são sempre aproximações da realidade, cujo objetivo é ter o máximo de acurácia possível, mas dificilmente sendo um retrato exato do fenômeno observado.

Algarismo	Lei de Benford	Frequência Populações	Erro (em módulo) em relação ao modelo	Frequência Extensões T.	Erro (em módulo) em relação ao modelo
1	30,10%	29,49%	0,62%	28,63%	1,47%
2	17,61%	15,81%	1,80%	19,23%	1,62%
3	12,49%	12,82%	0,33%	13,25%	0,75%
4	9,69%	8,97%	0,72%	8,97%	0,72%
5	7,92%	11,54%	3,62%	7,26%	0,65%
6	6,69%	7,69%	1,00%	5,56%	1,14%
7	5,80%	3,42%	2,38%	5,98%	0,18%
8	5,12%	5,56%	0,44%	5,13%	0,01%
9	4,58%	4,70%	0,13%	5,98%	1,41%
Total	100,00%	100,00%	--------------	100,00%	--------------

Outras propriedades como a invariância por mudança de base (com a lei de Benford adaptada para o logaritmo de base 8) podem ser exploradas neste tipo de atividade com planilhas eletrônicas. E, de forma mais ampla, certamente existe um leque de incontáveis desdobramentos que podem ser feitos a partir desta proposta, ou então de formas diferentes de organizá-la e complementá-la. O objetivo desta seção final do texto era trazer algumas ideias iniciais, entre tantas outras possibilidades, para a inspiração de trabalho em sala de aula com este rico tema que é a lei de Benford.

Referências

[1] DANTAS, C. Probabilidade: um curso introdutório. 3. ed. rev. São Paulo: EdUSP, 2008.

[2] KUIPERS, L. and NIEDERREITER H. Uniform Distribution of Sequences. Wiley, New York, 1974.

[3] MATOS, J. Lei de Benford: conceito, experimentos e aplicações. Dissertação (Profmat). Universidade Estadual do Ceará, Quixadá, 2020.

[4] ROUSSEAU, C.; SAINT-AUBIN, Y. Matemática e Atualidade. v. 2. Rio de Janeiro: SBM, 2015.

[5] DÍGITOS (temporada 1, ep. 4). A Era dos Dados [Seriado]. Direção: Latif Nasser. Produção: Latif Nasser et al.. Reino Unido: Netflix, 2020.