

DEBATE (/ENSAIO/DEBATE/)

A covid-19 e a inexorável Lei dos Números Anômalos

FOTO: ADRIANO MACHADO/REUTERS - 04.OUT.2020

Maurício Bugarin

06 de Outubro de 2020

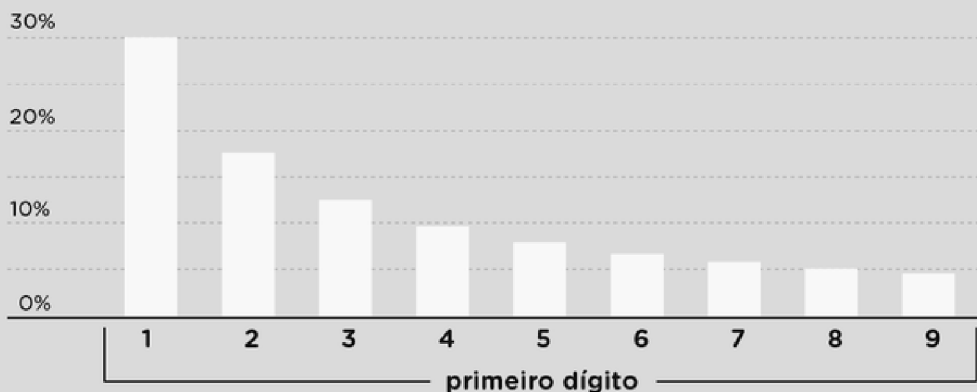
Para verificar a relação entre a pandemia e a Lei de Benford, foi construída uma base de dados de casos de coronavírus totais reportados pelos estados brasileiros

Em fins do século 19, ao consultar um livro de tabelas de logaritmos, o astrônomo canadense-americano Simon Newcomb percebeu que as páginas iniciais, que correspondiam aos números com primeiro dígito mais baixo (1, 2, 3) pareciam muito mais gastas que as páginas finais, correspondendo aos números com primeiro dígito mais elevado (7, 8, 9). Em 1881, Newcomb publicou um artigo ("Note on the frequency of the different digits in natural numbers") em que sugere que coleções de números que aparecem espontaneamente na natureza tendem a seguir uma regra que atribui maior frequência aos primeiros dígitos mais baixos e postulou uma fórmula logarítmica para descrever a frequência decrescente de cada um dos possíveis primeiros dígitos 1 a 9.

Distribuição teórica da frequência do primeiro dígito 1 a 9

Em bases de dados numéricos segundo Newcomb (1881)

40% de frequência esperada (Lei de Benford)



Fonte: Newcomb (1881)

NEXO

Esse trabalho precursor caiu no ostracismo até que, em 1938, o físico americano Frank Benford também identificou essa regularidade (“The law of anomalous numbers”). Ele coletou 20 bancos de dados numéricos “naturais” das mais variadas fontes, incluindo desde comprimentos de rios e populações de cidades a quadrados de números naturais e constantes físicas, e comprovou estatisticamente que os primeiros números nessas bases também seguiam a distribuição decrescente. Trata-se de uma característica intrigante, uma vez que em uma primeira abordagem leiga se esperaria uma distribuição uniforme em que cada um desses primeiros dígitos aparecesse com a mesma frequência. Ademais, Benford e estudos sucessores determinaram distribuições teóricas para as frequências do segundo dígito (que, ainda que também decrescente, decresce mais lentamente) e também dos dois primeiros dígitos (que se assemelha a uma distribuição expandida daquela do primeiro dígito).

Desde o final do ano de 2019, defrontamos-nos com o grande desafio da pandemia do novo coronavírus, que causa a covid-19. Para verificar a relação entre a pandemia e a Lei dos Números Anômalos, mais conhecida atualmente como Lei de Benford, foi construída uma base de dados de casos de covid-19 totais reportados pelos estados brasileiros de 25 de fevereiro de 2020, data em que o primeiro caso foi reportado, até 15 de setembro deste ano, data em que atingimos um total de 4.384.299 brasileiros contaminados pelo vírus.

AS EVIDÊNCIAS VISUAL E ESTATÍSTICA PRELIMINARES SUGEREM QUE OS ESTADOS TÊM REPORTADO COM SERIEDADE OS NÚMEROS DE CASOS DE COVID-19

NESTA SÉRIE

DEBATE (/ENSAIO/DEBATE/?

MODULO=DEBATE&TEMA=TAG&EDICAO=AS%20CRISES%20DO%20NOVO%20CORONAV%3%ADRUS&POSICAO=1)

A covid-19 e a inexorável Lei dos Números Anômalos (/ensaio/debate/2020/A-covid-19-e-a-inexorável-Lei-dos-Números-Anômalos?

modulo=Debate&tema=tag&edicao=As%20crises%20do%20novo%20coronav%3%ADRus&posicao=1)

DEBATE (/ENSAIO/DEBATE/?

MODULO=DEBATE&TEMA=TAG&EDICAO=AS%20CRISES%20DO%20NOVO%20CORONAV%3%ADRUS&POSICAO=2)

Uma recuperação verde da América Latina é necessária (/ensaio/debate/2020/Uma-recuperação-verde-da-América-Latina-é-necessária?

modulo=Debate&tema=tag&edicao=As%20crises%20do%20novo%20coronav%3%ADRus&posicao=2)

DEBATE (/ENSAIO/DEBATE/?

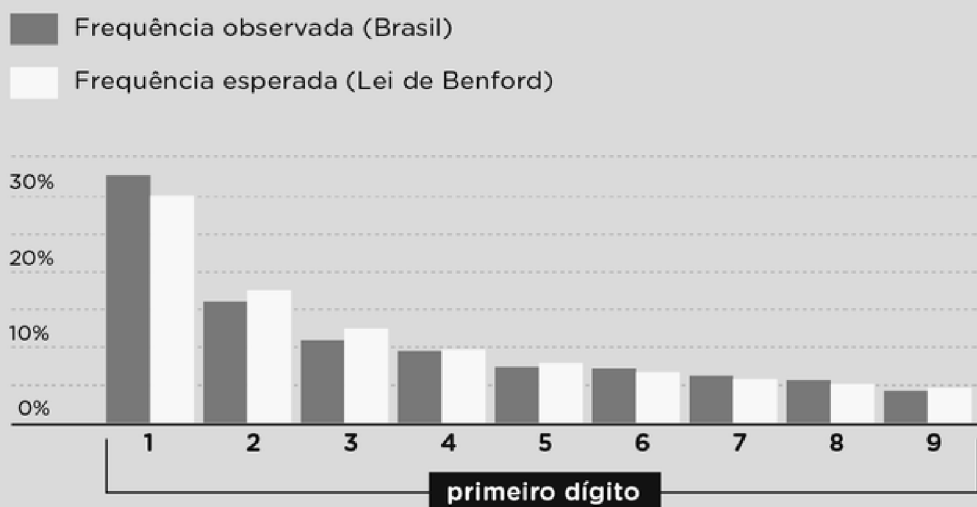
MODULO=DEBATE&TEMA=TAG&EDICAO=AS%20CRISES%20DO%20NOVO%20CORONAV%3%ADRUS&POSICAO=3)

As mortes ‘ocultas’ da pandemia e o Setembro Amarelo (/ensaio/debate/2020/As-mortes-‘ocultas’-da-pandemia-e-o-Setembro-Amarelo?

modulo=Debate&tema=tag&edicao=As%20crises%20do%20novo%20coronav%3%ADRus&posicao=3)

As figuras a seguir apresentam as frequências do primeiro dígito, do segundo dígito e dos dois primeiros dígitos nesse banco de dados, bem como as frequências correspondentes previstas pela Lei de Benford. A base de dados usada é aquela dos números totais acumulados de casos reportados, de acordo com o Ministério da Saúde do Brasil.

Frequências observadas e esperadas do primeiro dígito de casos diários acumulados de covid-19 nos estados brasileiros



Fonte: Ministério da Saúde para os casos de covid-19

NEXO

Frequências observadas e esperadas do segundo dígito de casos diários acumulados de covid-19 nos estados brasileiros

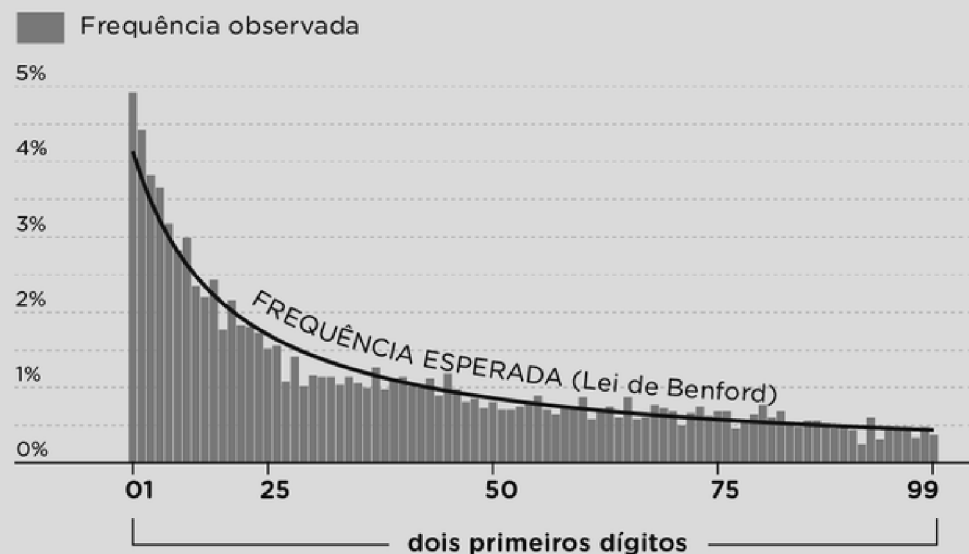


Fonte: Ministério da Saúde para os casos de covid-19

NEXO

As figuras apresentam uma impressionante conformação geral dos dados à distribuição teórica, sugerindo fortemente que a propagação do vírus no país também segue a Lei dos Números Anômalos. No caso do segundo dígito, chega a haver uma total identificação entre a frequência esperada e aquela realizada para o primeiro dígito 1. Essa identificação perfeita também ocorre nos dois primeiros dígitos 58, 59, 62 e 83.

Frequências observadas e esperadas dos dois primeiros dígitos de casos diários acumulados de covid-19 nos estados brasileiros



Fonte: Ministério da Saúde para os casos de covid-19

NEXO

O artigo de 1938 de Frank Benford desencadeou uma série de pesquisas mais recentes sobre como se usar essa regularidade para detectar manipulação de dados nas mais diversas áreas, como análise contábil, auditoria de obras públicas e até mesmo análise de fraudes em requerimentos de seguro-desemprego, dentre muitas outras aplicações.

Diversos testes estatísticos têm sido desenvolvidos e considera-se que quanto mais distante uma base de dados numéricos estiver da distribuição teórica, mais indícios deve haver de manipulações.

As evidências visual e estatística preliminares sugerem que os estados têm reportado com seriedade os números de casos de covid-19. Por exemplo, o teste do chi-quadrado para a frequência dos dois primeiros dígitos leva a um valor $\chi^2 = 100,96$, enquanto o limite para se considerar a conformidade aceitável é de 112,02. À medida que mais dados se tornam disponíveis, estudos mais aprofundados do ponto de vista estatístico deverão colocar à prova esses resultados favoráveis preliminares.

Maurício Bugarin é PhD em economia pela UiuC (University of Illinois at Urbana-Champaign), professor titular do Departamento de Economia da UnB (Universidade de Brasília) e líder do grupo de pesquisas do CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) "Economics and Politics Research Group". Atualmente é pesquisador visitante no Departamento de Ciência Política da Vanderbilt University, nos EUA. Publica no site: www.bugarinmauricio.com (<http://www.bugarinmauricio.com>).

Os artigos publicados no nexo ensaio são de autoria de colaboradores eventuais do jornal e não representam as ideias ou opiniões do Nexo. O Nexo Ensaio é um espaço que tem como objetivo garantir a pluralidade do debate sobre temas relevantes para a agenda pública nacional e internacional. Para participar, entre em contato por meio de ensaio@nexojornal.com.br informando seu nome, telefone e email.

The Trust Project (<https://thetrustproject.org/>) **SAIBA MAIS**