NOTA TÉCNICA

Diretoria de Estatística e Informações (Direi)

NOTA TÉCNICA:
UMA NOTA SOBRE A LEI DE ZIPF NOS MUNICÍPIOS BRASILEIROS

Número 7/2024

Belo Horizonte

2024

GOVERNO DO ESTADO DE MINAS GERAIS

Governador
Romeu Zema Neto
Vice-Governador
Mateus Simões

SECRETARIA DE ESTADO DE PLANEJAMENTO E GESTÃO

Secretária de Estado de Planejamento e Gestão
Luísa Cardoso Barreto

FUNDAÇÃO JOÃO PINHEIRO
Presidente
Luciana Lopes Nominato Braga
Vice-Presidente
Mônica Moreira Esteves Bernardi
Diretoria de Estatística e Informações (Direi)
Claudio Djissey Shikida

Equipe Técnica
Elaboração
Claudio Djissey Shikida

Revisão
Marielle Durães Ferreira

Normalização
Ana Paula da Silva

SUMÁRIO

1 INTRODUÇÃO[1]

Imagine uma lista de municípios ordenada de forma decrescente, conforme sua população. Mais do que isto, imagine que, nesta lista, o primeiro município tem o dobro da população do segundo que, por sua vez, é três vezes maior que o terceiro e assim por diante. Em outras palavras, a probabilidade de que um município tenha população maior do que um valor S é dada por: $P(tamanho > S) = A/S^b$, com $b \approx 1$.

Este padrão é denominado lei de Zipf, em homenagem ao seu descobridor, o linguista George Kingsley Zipf, que descobriu esta regularidade empírica em seus estudos de linguística. Como observou Monasterio (2004), a lei não é uma "Lei" justamente por ser uma aproximação estatística. Algebricamente, considere a seguinte relação entre o *ranking* dos municípios e sua população:

$$R = AS^{-b} \qquad (1)$$

Nessa relação, R é a posição do município no *ranking* (ordenado de forma decrescente), A é uma constante, S é a população urbana do município e b é um parâmetro que, se observada a lei de Zipf, assume valor unitário. A mesma relação pode ser mais facilmente observada em escala logarítmica: $\ln(R_i) = A - bln(S_i)$, já acrescentando "i" para denotar o município (i =1,..., n).

Estatisticamente falando, estima-se:

$$\ln(R_i) = A + bln(S_i) + e_i \qquad (2)$$

A relação (2) é uma regressão simples, que pode ser estimada por Mínimos Quadrados Ordinários (MQO). Como já dito, se a lei de Zipf é observada, espera-se que o coeficiente estimado de b seja não nulo e próximo a -1[2]. Gabaix (1999) refere-se a esta distribuição como o *steady state* de outra "lei", a lei de Gibrat. Em suas palavras:

> Consider a situation where there are a fixed number of cities […], and that, over time, their sizes grow (and possibly shrink) stochastically. Assume only that, at least for a certain range of (normalized) sizes, the cities follow similar processes; i.e., their growth processes have a common mean (equal to the mean city growth rate) and a common variance. This homogeneity of growth processes is often referred to as Gibrat's law, after Gibrat [1931]. Then, automatically, in the steady state, the distribution of cities in that range will follow Zipf's law with a power exponent of 1. (Gabaix, 1999, p.741)

[1] Agradeço a Caio Gonçalves e Priscila Garcia Maia, ambos da Coordenação de Indicadores Sociais sediada na Diretoria de Estatística e Informações (CIS-DIREI) desta fundação, pela ajuda com a construção da base de dados.

[2] O leitor interessado em desdobramentos da lei de Zipf, com ênfase em estudos de economia regional e urbana, encontrará uma ótima resenha em Orellana Aragón (2009).

Assim, embora pareça uma curiosidade estatística[3], há evidências de que a lei seja relacionada com a atividade econômica das regiões. Outro autor, Orellana Aragón (2009), oferece uma possível explicação para o fenômeno:

> […] as distribuições do tamanho das cidades apresentam peculiaridades próprias quando se trata de distribuições fortemente assimétricas. Em efeito, quando existem muitas cidades pequenas e poucas grandes, o que acontece é que o número de cidades grandes em cada classe decresce conforme aumenta a dimensão que caracteriza aquela classe.
>
> A configuração espacial do equilíbrio que define o número e tamanho dos núcleos urbanos pode [*sic*] ser entendida como o resultado de um processo que participam dois tipos de forças opostas, como se explicou anteriormente: forças centrípetas ou de aglomeração e forças centrifugas ou de dispersão. O equilíbrio espacial correspondente é o resultado de uma complicada interação de forças que puxam e que tiram de consumidores e empresas, até que possam procurar uma localização melhor (Orellana Aragón, 2009, p.43).

Além disso, no modelo econômico inicial elaborado por Gabaix (1999), a convergência, no tempo, para o estado estacionário da lei de Gibrat, ou seja, para a lei de Zipf, baseia-se nas escolhas de indivíduos que, nascendo em uma cidade, buscam maximizar seu gasto com comodidades (*amenities*) que, supostamente, são distribuídos de forma idêntica e independente (IID). Supondo que a função de produção em cada cidade tenha retornos constantes de escala, é possível mostrar que o processo de crescimento será similar entre as cidades, independentemente do tamanho (população) inicial dessas.

Apesar desses *insights*, não existe, na literatura, uma explicação definitiva para o comportamento das populações municipais em "cumprimento" da lei de Zipf. Nesta nota, avalia-se a aderência da lei à população urbana dos municípios brasileiros. A inovação do trabalho é a população das áreas urbanas, construídas experimentalmente pela Coordenação de Indicadores Sociais da Diretoria de Estatística e Informações da Fundação João Pinheiro (CIS-Direi/FJP), com base no Censo de 2022 e nas áreas urbanizadas elaboradas pelo Instituto Brasileiro de Geografia e Estatística (IBGE[4]). O uso destas áreas busca corrigir o problema da classificação tradicional de população urbana/rural de um município[5].

[3] Em trabalhos anteriores, Shikida, Nobre Fernandez e Carraro (2019) e Shikida e Comitti (2020) investigaram a aderência dos dados à lei de Zipf, respectivamente, em dados do campeonato brasileiro e de casos reportados de Covid-19 no Brasil.

[4] Ver: IBGE, [2019?].

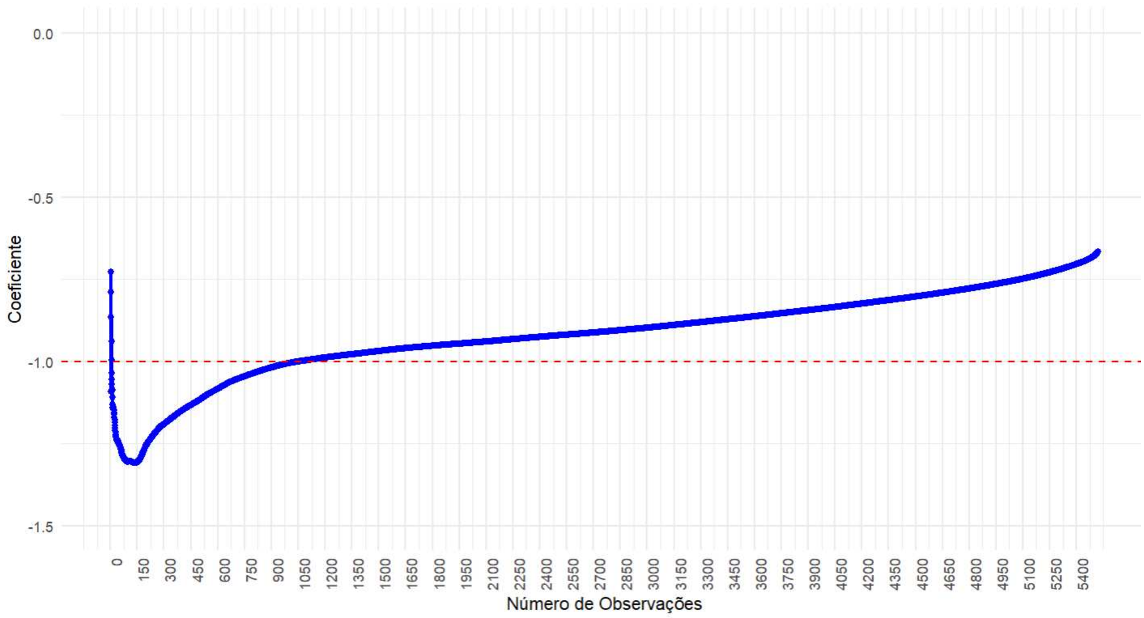
[5] A classificação tradicional, estabelecida pelo primeiro governo Vargas, definiu que a população da sede de um município seria considerada como urbana, independentemente de suas características.

2 INDÍCIOS SOBRE A LEI DE ZIPF

A base de dados contém 5.520 municípios com população urbana que assumem valores não nulos[6]. Partindo dos municípios ordenados por sua população urbana (de forma decrescente), o exercício desta nota consiste em calcular sucessivas regressões em que, a cada nova estimação, inclui-se um município na amostra — desse modo, foram estimadas 5.519 regressões. Isso permite verificar, aproximadamente, até que ponto da distribuição dos municípios a lei de Zipf é verificada.

Nesse sentido, a Gráfico 1 a seguir mostra o comportamento do coeficiente "b"[7].

Gráfico 1: Estimativa da lei de Zipf – Coeficiente x Número de observações



Fonte: Cálculos do autor.

Observa-se que, a despeito de uma queda inicial, o coeficiente converge para -1 quando o número de municípios usados nas regressões aumenta. A convergência se dá em municípios com áreas urbanas em torno de 19 mil habitantes. A partir daí, as estimativas mostram um afastamento da lei de Zipf. Isso significa que esta parece operar para, aproximadamente, os 1.050 ou 1.170 maiores municípios — quando se considera o b estimado em torno de -0.99. Daí em diante, a inclusão de outros municípios lentamente leva o coeficiente para valores próximos de -0.66[8].

Lembrando o modelo de Gabaix (1999) mencionado anteriormente, talvez seja possível afirmar que, no Brasil, ele tenha boa aderência para municípios de população urbana próxima ou maior

6 A base de dados contém mais 49 municípios, mas com população urbana igual a zero.

7 As estimações foram feitas no programa R.

8 Segundo Gabaix (1999), a elevada variância das pequenas cidades seria uma explicação para o desvio dos coeficientes estimados em relação ao *steady state* representado pela lei de Zipf.

do que 19 mil habitantes. Ou seja, o crescimento da população urbana leva ao estado estacionário neste nível. Esta “conclusão”, é bom lembrar, limita-se à amostra analisada.

Dito isso, vale pensar também nas possibilidades de extensão deste estudo. Existem variadas possibilidades de investigação. Por exemplo, pode-se perguntar se existe maior homogeneidade na dinâmica econômica de municípios com população urbana maior do que 19 mil habitantes. Outra questão é analisar a base industrial das cidades nos desvios observados da lei de Zipf. A estimativa por corte regional ou estadual também é outra possibilidade.

3 À GUISA DE CONCLUSÃO

Esta nota usou os dados mais recentes do Censo do IBGE combinados com as áreas urbanas calculadas pela FJP para verificar a adequação das populações urbanas dos municípios brasileiros à lei de Zipf. Mostrou-se que a curiosa relação pode ser pensada como o estado estacionário de um processo econômico-demográfico baseado na lei de Gibrat, tal como exposto por Gabaix (1999). Os resultados indicam que a lei de Zipf é observada quando a regressão envolve os primeiros 1.050 a 1.170 municípios.

O resultado não foge muito ao que se encontra na literatura, mas é desejável – recomendável até – que sejam feitos mais testes (inclusive, com outros métodos de ajuste). Além disso, pode-se experimentar a checagem com áreas urbanas construídas com outros critérios. Finalmente, estimações considerando o corte estadual ou regional também pode ser um caminho para novas descobertas acerca das dinâmicas regionais (o que possibilita, inclusive, como já fizeram alguns autores mencionados, a construção de modelos teóricos que considerem a regularidade de Zipf nos sistemas econômicos). São todas sugestões para aprofundamentos sobre os resultados encontrados aqui.

REFERÊNCIAS

GABAIX, Xavier. Zipf's law for cities: an explanation. **Quarterly Journal of Economics**, Oxford, v. 114, n.3, p.739-767, 1999. Disponível em: https://www.jstor.org/stable/2586883. Acesso em: 8 nov. 2024.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Áreas urbanizadas do Brasil 2019**. Rio de Janeiro, [2019?]. Disponível em: https://biblioteca.ibge.gov.br/visualizacao/livros/liv101973_informativo.pdf. Acesso em: 8 nov. 2011.

MONASTERIO, Leonardo Monteiro. A lei de Zipf: Rio Grande do Sul (1940-2000). **Redes**, Santa Cruz do Sul-RS, v.9, n.2, p.181-190, 2004. Disponível em: https://online.unisc.br/seer/index.php/redes/article/view/10998. Acesso em: 8 nov. 2024.

ORELLANA ARAGÓN, Jorge Alberto. **A lei de Zipf e os efeitos de um tratado de livre comércio:** caso da Guatemala. 2009. 114 f. Dissertação (Mestrado em Economia) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009. Disponível em: https://lume.ufrgs.br/bitstream/handle/10183/16417/000701569.pdf?sequence=1. Acesso em: 3 out. 2024.

SHIKIDA, Claudio Djissey; COMITTIT, Victor. Days of Zipf and Covid? Looking for evidence of Zipf's Law in the infected Brazil. **Covid Economics**, Paris, issue 43, 21 Aug. 2020. Disponível em: https://cepr.org/publications/covid-economics-issue-43#392514_392919_390593. Acesso em: 8 nov. 2024.

SHIKIDA, Claudio Djissey; NOBRE FERNANDEZ, Rodrigo; CARRARO, André. A distribuição do *ranking* de clubes brasileiros regido por uma lei universal: uma aplicação a Lei de Zipf. **Podium Sport, Leisure and Tourism Review**, São Paulo, v.8, n.2, p.230-240, 2019. Disponível em: https://periodicos.uninove.br/podium/article/view/322. Acesso em: 8 nov. 2024.