

A LEI DE ZIPF: RIO GRANDE DO SUL (1940-2000)

*Leonardo Monteiro Monastério**

Resumo

O trabalho faz um teste empírico da aplicabilidade da lei de Zipf no Rio Grande do Sul, entre 1940 e 2000. Essa lei afirma que existe uma relação linear, negativa e aproximadamente unitária entre os logaritmos da ordem das maiores cidades de uma região e suas respectivas populações. Após uma breve panorâmica das diversas explicações teóricas para tal lei, apresentam-se os resultados para o estado. Mostra-se que, feitas algumas ressalvas, a lei de Zipf representa adequadamente o perfil urbano do Rio Grande do Sul.

Palavras-chave: lei de Zipf, regra de ordem-tamanho, hierarquia urbana.

Abstract

ZIPF LAW: RIO GRANDE DO SUL (1940-2000) - This paper is an empirical test of Zipf's law's applicability to Rio Grande do Sul (1940-2000). It states that the log-linear relation between the rank and the population of the cities of a region is close to -1. After a brief overview of theories, the results are presented. It is shown that, with some remarks, Zipf's law is an adequate representation of the urban hierarchy of Rio Grande do Sul.

Keywords: Zipf's law, rank-size rule, urban hierarchy.

INTRODUÇÃO

Ordene de forma decrescente os dados sobre população das cidades de sua região. Divida o número de habitantes da maior cidade da região por dois. Anote. Siga repetindo a operação acrescentando uma unidade ao denominador. Compare a lista de valores observados com a dos calculados. É provável que ambas sejam bastante semelhantes.

* Doutor em Desenvolvimento Econômico (UFPR). Professor do Departamento de Geografia e Economia da Universidade Federal de Pelotas.

Essa estranha “lei”, que mais parece um truque de salão, retrata a relação aproximada entre a ordem de tamanho das cidades e suas respectivas populações pelo mundo afora, hoje e no passado. Conforme se verá adiante, ela equivale a sustentar que a relação entre os logaritmos do número de ordem das cidades (da maior para a menor) e o número de habitantes pode ser aproximada por uma linha reta com inclinação igual a -1. Apontada pelo lingüista George Zipf em 1949¹, essa regularidade continua sendo válida em diversos países e regiões. O recente estudo de Soo (2002) estimou que a média da reta supracitada estimada é de -1,1 para os 75 países estudados, sendo que em 71 destes a inclinação estava entre -0,8 e -1,5². Apesar da regularidade enunciada por Zipf não ser tão precisa quanto às constantes da Física, ela é um dos padrões mais robustos que pode ser encontrado nas Ciências Sociais. Levando isso em conta, optou-se por chamá-la de lei - e não Lei - de Zipf.

Por que essa lei existe? A polêmica das razões profundas que sustentam essa lei pode ser encontrada no clássico livro de Walter Isard (1956) e até na recente obra de Fujita et alli (2002). O problema chamou a atenção até mesmo de autores não identificados com a ciência regional, como o ganhador do prêmio Nobel, Herbert Simon. Neste trabalho tem-se apenas uma panorâmica do desafio teórico.

O objetivo desse trabalho é testar se a lei em questão se aplica ao caso do Rio Grande do Sul no período 1940-2000. A bibliografia não registra estudos desse tipo para os estados brasileiros e, conforme será visto adiante, o teste da teoria leva a reflexões acerca da estrutura urbana gaúcha e sua evolução. A contribuição desse trabalho é, portanto, empírica.

O trabalho se inicia com uma apresentação mais cuidadosa da lei de Zipf. A seção apresenta a literatura sobre as suas causas. Os resultados para as cidades gaúchas constam da terceira seção. Considerações finais encerram o trabalho.

1 A LEI DE ZIPF

1.1 Apresentação

Por que em qualquer área o número de cidades grandes é baixo em relação ao número de cidades pequenas? Esse fenômeno aparentemente trivial chamou a atenção dos precursores da Ciência Regional.

A explicação de Zipf para a sua lei empírica não foi formalizada. Ele afirma que existem forças de diversificação e unificação do espaço. A primeira seria decorrência da

¹ Como costuma acontecer, a paternidade da idéia foi questionada por diversos autores.

² O valor estimado do expoente a para o Brasil em 2000 foi de -1,13. Apesar desses resultados, Soo (2002) tem uma apreciação bastante crítica da validade da Lei de Zipf.

localização da população nas proximidades das áreas fornecedoras de matérias-primas como forma de minimização de custos. Já as forças de unificação derivam da minimização dos custos de transporte das mercadorias finais para os mercados consumidores. A hierarquia urbana resultante seria ótima em relação a essas duas forças contrapostas (BERRY e GARRISON, 1958). As forças de diversificação levam a um número crescente de idades com população decrescente, e as de unificação para um número menor de cidades com população crescente. Para Zipf, esse fato faz com que se tenha uma relação linear quando a relação entre o *rank* das cidades e o seu tamanho é plotada em uma escala logarítmica dupla.

Em termos formais:

$$R_i = M x_i^{-\alpha} \quad (1)$$

Onde R_i = *rank* da cidade i ;

x = população da cidade i ;

M = tamanho da maior área metropolitana;

Aplicando-se logaritmo natural em ambos os lados da equação, tem-se:

$$\ln R_i = \ln M - \alpha \log x_i \quad (2)$$

Naquela que ficou conhecida como lei de Zipf há a exigência de que α seja próximo de -1. Assim, um gráfico que apresente as observações referentes aos logaritmos do *rank* nas ordenadas e da população nas abscissas será uma linha reta e com inclinação negativa e unitária.³

1.2 Uma panorâmica das Explicações para a Lei de Zipf

Brackman et alli (1999 apud REED, 2001) escreveu que a lei de Zipf ainda é uma “*empirical regularity in search of a theory*”. A apresentação completa das explicações possíveis para o fenômeno escapa, evidentemente, aos limites desse trabalho. Aqui será feita uma breve panorâmica das tentativas de se chegar a uma teoria que explique adequadamente o fenômeno.

Segundo Gabaix (1999) existem duas linhagens de explicações sobre a lei de Zipf: a econômica e a aleatória. Nesta, como a proposta por Simon (1955), existe um processo estocástico com características específicas que rege a forma com que novos contingentes populacionais seriam agregados às cidades existentes ou levariam à formação de novas cidades. Simon (1955) estudou teórica e empiricamente um tipo de função que retrata

³ Com $\alpha = -1$ tem-se a relação $x_i = M/R_i$ referida na introdução do trabalho.

desde a distribuição de palavras pela frequência de aparecimento em um texto; cientistas por número de trabalhos publicados; e, o que é de interesse aqui, a distribuição das cidades de acordo com o seu tamanho. A função é:

$$f(i) = a/i^k \quad (3)$$

Em que a e k são constantes e $1 < k < 2$. Ela indica que o número de fenômenos que ocorreram exatamente i vezes é igual a/i^k . A função (3) tem um formato da letra j , ou seja, poucos autores terão publicado muitos *papers*, enquanto muitos outros verão seu nome apenas uma vez. Conforme já se disse, esse é o mesmo perfil da lei de Zipf: poucas grandes cidades e muitas pequenas. Simon (1955) mostra que o número de palavras que aparecem i vezes no romance *Ulisses*, de James Joyce, é aproximadamente igual a a/i^k . No tocante às cidades, a questão está em buscar que processo que gera uma distribuição semelhante a $f(i) = a/i^k$.

Os processo estocástico desenvolvido por Simon se baseia em duas hipóteses: a) a probabilidade de uma novo contingente populacional se incorporar a uma cidade de tamanho i é proporcional à população total das cidades preexistentes com a mesma população i ; b) existe uma probabilidade constante de que o novo contingente populacional forme uma cidade nova. O resultado final desse processo, obtido formalmente ou via simulação, é uma distribuição exatamente igual a prevista por Zipf, inclusive com $\alpha = -1$ (ver FUJITA et al., 2002, p. 243).

A principal crítica que se tem em relação às explicações baseadas em processos puramente aleatórios é de ordem metodológica. Apenas porque conseguiu-se criar um processo aleatório capaz de gerar uma distribuição semelhante à observada, não significa que as razões subjacentes sejam aquelas. Ter-se-ia, portanto, não uma explicação, no sentido de uma cadeia de nexos causais, e sim uma coincidência. Simon diz que o paralelo entre o número de coices sofridos pelos soldados alemães e a distribuição de células sanguíneas nas lâminas de microscópio é que o mesmo processo aleatório pode representar de modo satisfatório o mesmo fenômeno (SIMON, 1955, p. 425). Contudo, aplicar a mesma argumentação para processos semelhantes ao crescimento das cidades não parece ser a melhor forma de se chegar a um entendimento da lei de Zipf.

As explicações econômicas buscam os microfundamentos que geram a distribuição de Zipf. Conforme argumentaram Berry e Garrison (1958), a lei de Zipf está relacionada com o modelo de Christaller. A área de mercado da cidade dominante abrange toda a região e disponibiliza todo o espectro de produtos. A cada descida de nível na hierarquia urbana, se encontra um número maior de cidades cada vez menores, com áreas de mercado menores e oferecendo um gama menor de bens. O resultado, em um espaço homogêneo, são as grandes áreas de mercado dos lugares centrais de formato hexagonal, que contêm os hexágonos menores aninhados referentes às cidades de posição mais baixa

na hierarquia. A estrutura urbana gerada por este também inclui uma tendência ao aumento do número de cidades a cada decréscimo na hierarquia das cidades de forma semelhante à lei de Zipf. Mais importante, com condições adicionais, tem-se a mesma relação fixa entre a população de cada cidade e do nível imediatamente superior.

Já o modelo de Henderson (1974) parte de uma assimetria entre economias e deseconomias externas nas cidades. As economias externas são específicas para cada indústria; enquanto as deseconomias atingem a todos os setores. A vantagem de se ter diversas firmas têxteis, por exemplo, em um mesmo lugar é que atingem apenas tal setor. Enquanto as deseconomias de uma cidade grande demais atingem todos os setores. Logo, esses modelos preveem uma diversidade de tamanhos de cidades, de acordo com o tipo de atividades nela localizadas. O problema com o modelo de Henderson é supor que a alocação dos setores econômicos pelas cidades seria exatamente aquele que gera uma relação como a de Zipf (FUJITA et al, 2002, p. 241).

Recentemente, o trabalho de Gabaix (1999) foi considerado por Duranton (1999) um marco. Ele propôs um modelo que gera a lei de Zipf com microfundamentos. Entretanto, Duranton sustentou que os pressupostos subjacentes ao modelo estão em desacordo com a evidência empírica. Modelos ainda mais sofisticados como o Fujita et al (1999)⁴ são capazes de gerar distribuições semelhantes as da lei de Zipf. Apesar desses avanços, os próprios autores de referência na Nova Geografia Econômica admitem que o problema de se encontrar uma explicação adequada para a lei de Zipf ainda não está solucionado. Nas suas palavras (FUJITA et al, 2002, p.237):

O problema que enfrentamos é que os dados oferecem uma imagem surpreendentemente nítida, difícil de reproduzir em qualquer modelo teórico plausível (ou mesmo implausível)

2 RESULTADOS PARA O RIO GRANDE DO SUL

Os dados desse trabalho se baseiam no CD-ROM "Um século de População do Rio Grande do Sul" o qual, por sua vez, tem nos censos do IBGE a fonte das informações. Como os testes necessitam a especificação da população urbana por município, a análise se restringiu aos anos de 1940, 1950, 1960, 1970, 1980, 1991 e 2000. Os dados utilizados referem-se à população urbana total, ou seja, incluindo os moradores urbanos fora do distrito sede do município.

O gráfico 1 mostra a relação entre os logaritmos naturais do *rank* das cidades gaúchas e sua população em 1940. Como a inspeção visual sugere, é uma relação linear

⁴ Ver McCann (2001) para uma apresentação acessível.

com inclinação próxima da unidade (em valores absolutos). Os gráficos referentes aos outros anos acham-se no Anexo do presente trabalho.

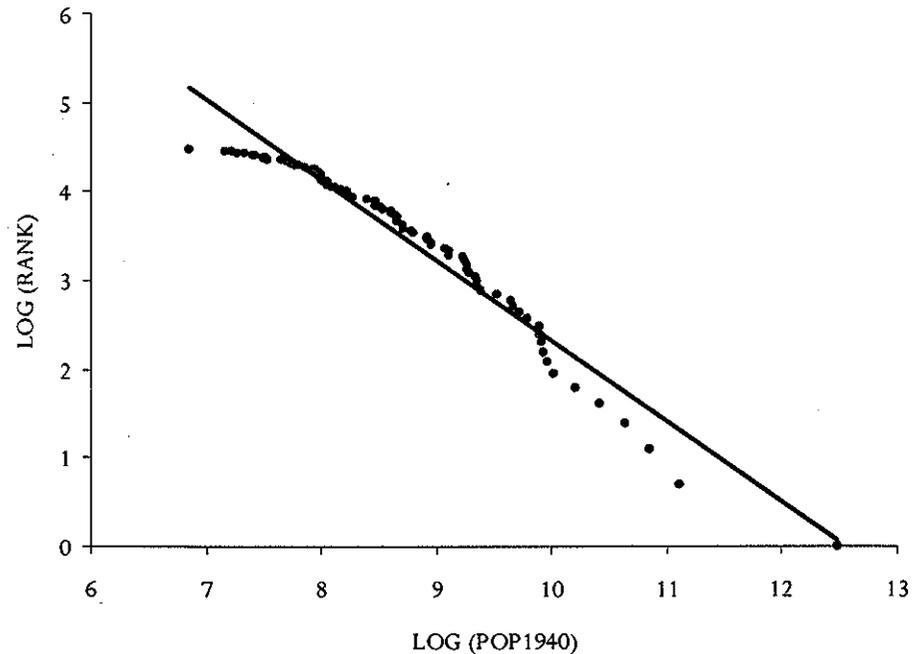


Gráfico 1 - Lei de Zipf para os municípios gaúchos de 1940.

Para melhor avaliar a adequação da lei de Zipf, fez-se as regressões simples baseadas na equação abaixo para cada ano censitário. Os resultados constam da Tabela 1.

$$\text{LN}(\text{Rank}) = \text{Constante} + \alpha \text{LN}(\text{População Urbana Total do Município}) + \text{erro} \quad (4)$$

Tabela 1 - Teste da lei de Zipf com todas observações

	1940	1950	1960	1970	1980	1991	2000
Constante	1.137.067	1.152.849	1.089.717	9.778.820	1.018.089	9.891.665	9.879.531
α	-0.905401	-0.893184	-0.782794	-0.636727	-0.648643	-0.59495	-0.578403
R ²	0.953644	0.947028	0.933216	0.915944	0.91603	0.9121	0.926213
Número de observações	88	92	150	232	232	333	466

Fonte: Cálculos do autor

O software utilizado foi o E-views 3.1 e o método, mínimos quadrados. Em todos os momentos os coeficientes foram negativos e significativos, tal como esperado. Contudo, como se pode ver o valor do estimador α é diferente de -1 e se distancia com o passar do tempo. Ou seja, existe uma relação tal como a apontada na equação 1, mas que, aparentemente, viola a lei de Zipf.

O que fez com que essa tendência surgisse? O que fez com que a distribuição na população do Rio Grande do Sul, ao contrário de outras partes do mundo, não estivesse de acordo com tal lei?

Uma das preocupações dos testes econométricos da Lei de Zipf decorre na dificuldade de se obter dados que representem adequadamente as áreas urbanas. Para evitar os problemas de lidar com as pequenas aglomerações, boa parte dos testes empíricos trata apenas das partes superiores da hierarquia urbana. Rosen e Resnick (1980), por exemplo, em seu estudo sobre 44 países examinaram apenas as 50 maiores cidades (exceto para seis países nos quais havia mais de 50 cidades com população maior do que 100 mil habitantes).

No caso presente existe um motivo adicional para se considerar a parte superior da distribuição: os critérios brasileiros de classificação de urbano e rural. Conforme mostrou Veiga (2002), foi durante o Estado Novo que foi instituída a norma de considerar como urbana toda a população da sede dos municípios, não importando se possuíam ou não as características estruturais ou funcionais de cidades. Mesmo após as mudanças na classificação do IBGE, os problemas, segundo o mesmo autor, continuaram.

A avassaladora criação de municípios no Brasil durante a segunda metade do século, que também ocorreu no Rio Grande do Sul, fez com que a população das sedes dessas novas unidades fosse considerada como urbana. Isso viesou o número de cidades pequenas, com uma aglomeração de observações com população pequena e próxima entre si. Isto fez com que surgisse um excessivo número de ditas "cidades" com população pequena e semelhante.

No caso em questão, optou-se por refazer as regressões com o mesmo número de municípios de 1940, ou seja, apenas os 88 mais populosos. À primeira vista isso pode parecer um limite alto, mas esse valor implica que as menores cidades consideradas vão desde 1,9 mil habitantes em 1950 até apenas 18 mil, em 2000. Apesar de algo arbitrário, a manutenção do mesmo número de observações torna mais fácil a comparação entre as regressões que se seguem:

Tabela II - Teste da lei de Zipf para as 88 maiores cidades

	1950	1960	1970	1980	1991	2000
Constante	1.200.435	1.318.336	1.332.714	1.373.362	1.406.875	1.434.675
α	-0.94338	-1.016.547	-0,994869	-0.992607	-0,998700	-1.009.374
R^2	0.967232	0.986271	0,997473	0.9668	0.973322	0.9756713
Teste de Wald para $\alpha = -1$	Não	Sim	Sim	Sim	Sim	Sim

Fonte: Cálculos do autor

Com essa amostra reduzida, todos os coeficientes alfa continuaram significativos (com mais de 99% de confiança) e com sinal esperado. Testando-se a restrição $\alpha = 1$, tem-se que de 1960 a 2000 a Lei de Zipf foi não falseada. O teste de Wald mostrou que se a hipótese nula de $\alpha = -1$ pode ser aceita. Em 1940 e 1950, os valores obtidos (-0,91 e -0,94) apesar de intuitivamente próximos de -1, desse valor são estatisticamente diferentes.⁵

Existe outra anomalia no caso do Rio Grande do Sul: a ausência de cidades intermediárias entre as duas maiores cidades, entre Porto Alegre e Caxias do Sul. Uma tem 1,3 milhão de habitantes e a outra pouco mais de 300 mil. Ao invés da relação de 2:1, prevista por Zipf, tem-se uma de 4:1. A partir daí, o erro é cada vez menor. Segundo McCann (1999, p. 80), essa tendência é freqüente em países recém-industrializados. Bangkok, por exemplo, é várias dezenas de vezes maior do que as cidades de segunda ordem da Tailândia.

Uma ressalva tem que ser feita. Utilizaram-se aqui os dados referentes apenas à população dos municípios e não das zonas metropolitanas. Caso se considerasse a Grande Porto Alegre, com seus mais de 3 milhões de habitantes, a anomalia supra citada seria ainda maior. Contudo, tal opção se justifica pela dificuldade de delimitação das áreas urbanas, especialmente quando se trata de dados de seis décadas atrás. Para evitar qualquer agregação arbitrária, preferiu-se preservar a divisão municipal como critério.⁶

3 CONSIDERAÇÕES FINAIS

Esse breve exercício empírico indicou que a lei de Zipf representa o perfil da hierarquia urbana do Rio Grande do Sul no período (1940-2000), de maneira razoavelmente adequada e estável. A revisão da literatura mostrou as dificuldades teóricas em se buscar uma explicação geral para o fenômeno.

⁵ A exclusão das observações referentes aos municípios menores da amostra leva à aceitação da lei de Zipf inclusive para 1940 e 1950. Tais regressões não foram incluídas, uma vez que não se deseja forçar a adequação dos dados à teoria.

⁶ Não só Porto Alegre deveria ter agregada a sua população a dos outros municípios da sua zona metropolitana. Por exemplo, hoje, a zona urbana de Pelotas deveria incluir, talvez, a do Capão do Leão.

Ainda há muito espaço para o avanço na pesquisa: seria interessante ter uma delimitação precisa e retrospectiva das zonas urbanas no estado. Já em relação à teoria, é difícil ser otimista acerca da possibilidade de uma explicação econômica parcimoniosa e geral da lei em questão. Os modelos feitos nas últimas cinco décadas foram importantes para a evolução da Ciência Regional, mas ainda é incerto se um dia serão capazes de explicar a ocorrência da lei de Zipf em tantos lugares e tempos distintos. Mesmo assim, vale atentar o recente trabalho de Durantton (2002) que tenta resolver os problemas das explicações anteriores mediante novos caminhos.

Ressalta-se que a lei de Zipf é uma regularidade empírica. Enquanto não houver boa fundamentação teórica é impróprio utilizá-la para prever, por exemplo, que Caxias do Sul crescerá até o tamanho de metade da Zona Metropolitana de Porto Alegre. Em outras palavras, é imprudente afirmar algo sobre o equilíbrio de longo prazo das cidades gaúchas com base na lei de Zipf.

O esforço de replicação desse estudo para outras regiões brasileiras pode ser interessante. A verificação das razões e condições nas quais a Lei de Zipf se aplica, e aquelas nas quais há uma violação, permitirá uma melhor compreensão dos mecanismos envolvidos na formação das hierarquias urbanas no Brasil.

REFERÊNCIAS

- BERRY, B. J. L.; GARRISON, W. J. *Alternate explanations of urban rank-size relationship*. Annals of the Association of American Geographers, v.48, n. 1, p. 83-91, 1958.
- DURANTTON, Gilles. *City size distributions as a consequence of the growth process*. London, 2002.
- FUJITA, Masashisa et al. *Economia espacial*. São Paulo: Futura, 2002.
- GABAIX, Xavier. Zip's Law and the growth of cities. *The American Economic Review*, v.89, n.2, p.129-132, 1999.
- HENDERSON, Vernon. The sizes and types of cities. *American Economic Review*, v.64, n.4, p.40-656, 1974.
- ISARD, Walter. *Location and space Economy*. John Wiley and Sons, New York, 1956.
- JARDIM, Maria de L. T. *Um século de população no Rio Grande do Sul*. Porto Alegre: Fundação de Economia e Estatística Siegfried Emanuel Heuser, 2001.
- McCANN, Philip. *Urban and regional Economics*. Oxford: Oxford University Press, 2001.
- REED, William. *On the rank-size distribution for human settlements*. Canada: Victoria, 2001.
- ROSEN, K., RESNICK, M. *The size distribution of cities: an examination fo the Pareto Law and primacy*. Journal of Urban Economics v.8, p.165-186, 1980.
- SIMON, Herbert. On a class of skew distributions functions. *Biometrika*, v.42, n.3/4, p.425-440, 1955.

SOO, Kwok T. *Zipf's Law for cities: a cross country investigation*. London, 2002.

VEIGA, José E. *Cidades imaginárias: o Brasil é menos urbano do que se calcula*. Autores Associados, São Paulo, 2002.

ZIPF, George. *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley, 1949.

Recebido para publicação em 22/01/04

Aceito para publicação em 15/10/04

ANEXO - LEI DE ZIPF (1950-2000)

