
UNIVERSIDADE ESTADUAL DE MARINGÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

LUIZ GUSTAVO DE ANDRADE ALVES

PADRÕES ESTATÍSTICOS E DINÂMICOS EM HOMICÍDIOS
E MÉTRICAS URBANAS

Maringá

2014

UNIVERSIDADE ESTADUAL DE MARINGÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

LUIZ GUSTAVO DE ANDRADE ALVES

PADRÕES ESTATÍSTICOS E DINÂMICOS EM HOMICÍDIOS
E MÉTRICAS URBANAS

Dissertação de mestrado apresentada como parte dos requisitos para obtenção do título de mestre em Física, do Programa de Pós-Graduação em Física, da Universidade Estadual de Maringá.

Orientador: Prof. Dr. Renio dos Santos Mendes
Co-orientador: Dr. Haroldo Valentin Ribeiro

Maringá

2014

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Central - UEM, Maringá – PR., Brasil)

A474p Alves, Luiz Gustavo de Andrade
 Padrões estatísticos e dinâmicos em homicídios e
 métricas urbanas / Luiz Gustavo de Andrade Alves
 Augusto. -- Maringá, 2014.
 58 f. : il. col., figs., tabs.

 Orientador: Prof. Dr. Renio dos Santos Mendes.
 Coorientador: Prof. Dr. Haroldo Valentin Ribeiro.
 Dissertação (mestrado) - Universidade Estadual de
 Maringá, Programa de Pós-Graduação em Física, 2014.

 1. Física estatística - Sistemas complexos. 2.
 Leis de escala. 3. Análise de dados - Sistemas
 sociais - Homicídios - Métricas urbanas. I. Mendes,
 Renio dos Santos, orient. II. Ribeiro, Haroldo
 Valentin, coorient. III. Universidade Estadual de
 Maringá. Programa de Pós-Graduação em Física. IV
 Título.

CDD 21.ed. 530.13

ECSL-001287

Agradecimentos

Gostaria de agradecer ao Prof. Dr. Renio dos Santos Mendes pela orientação nesta dissertação e, também, por ter me orientado durante os últimos anos na pesquisa científica. Agradeço ao Dr. Haroldo Valentin Ribeiro pela co-orientação nesta dissertação e pela assistência/colaboração na elaboração dos trabalhos aqui apresentados. Também, gostaria de agradecer ao Prof. Dr. Ervin Kaminski Lenzi pelas discussões e conselhos que muito contribuíram para minha formação. Agradeço aos integrantes do grupo de Física Estatística e Sistemas Complexos e aos meus colegas pós-graduandos pelos incontáveis cafés e conversas que tivemos. Por fim, agradeço a meus pais, Luiz Carlos Alves e Mirian Cristina de Andrade Alves, e a meu irmão, Cássio Roberto de Andrade Alves, que sempre apoiaram minhas escolhas.

Resumo

A Física Estatística tem provado ser frutífera para descrever fenômenos fora do seu domínio tradicional. Aqui, consideramos esse cenário na tentativa de entender a dinâmica e padrões estatísticos de homicídios e métricas urbanas. No capítulo 1, caracterizamos e analisamos aspectos quantitativos de assassinatos no período de 1980 a 2009 em cidades brasileiras. Encontramos que a distribuição das taxas de crescimento logarítmicas anuais, bianuais e tri- anuais exibem a mesma forma funcional para escalas distintas, isto é, um comportamento invariante de escala. Também, identificamos relações de decaimentos assintóticos tipo lei de potência entre os desvios padrão dessas três taxas de crescimento e o valor inicial. Por fim, discutimos similaridades com organizações complexas. No capítulo 2, reportamos uma análise quantitativa das relações entre o número de homicídios, o tamanho da população e outras dez métricas urbanas. Usando dados de cidades brasileiras, mostramos que leis de escala médias bem definidas com o tamanho da população emergem quando investigamos relações entre a população e o número de homicídios assim como entre a população e as métricas urbanas. Ainda, mostramos que as flutuações em torno da lei de escala são distribuídas log-normalmente, o que permite modelá-las por uma equação estocástica, na qual o ruído é multiplicativo e log-normalmente distribuído. Por causa das leis de escala, argumentamos que é melhor empregar logaritmos para descrever o número de homicídios em função de métricas urbanas via análise de regressão. Além da análise de regressão, propomos uma abordagem para correlacionar o crime e métricas urbanas via cálculo das distâncias entre o valor do número de homicídios (assim como o valor das métricas urbanas) e o valor que é esperado pela lei de escala com o tamanho da população. No capítulo 3, relatamos sobre a conexão existente entre distribuições lei de potência e alometrias. Apresentamos uma caracterização extensa dessa conexão quando todos os pares possíveis de relações de 12 indicadores urbanos das cidades brasileiras são considerados. Nossas análises revelam que todos os doze indicadores urbanos são assintoticamente distribuídos como leis de potência e que, quando um par de indicadores apresenta uma alometria estatisticamente significativa, o expoente alométrico se torna completamente definido em termos dos expoentes das distribuições lei de potência. Além disso, também descobrimos que as flutuações residuais ao redor de todas as possíveis alometrias apresentam uma variância quase constante e distribuições log-normais.

Palavras-chave: Sistemas complexos. Física Estatística. Leis de Escala. Homicídios. Sistemas Sociais. Métricas Urbanas.

Abstract

Statistical physics has proven to be fruitful to describe phenomena very far from their traditional domain. Here, we consider this framework in an attempt to understand the dynamics and statistical patterns of crime and urban metrics. In chapter 1, we characterize and analyze quantitative aspects of murders in the period from 1980 to 2009 in Brazilian cities. We find that the distribution of the annual, biannual and triannual logarithmic homicide growth rates exhibit the same functional form for distinct scales, that is, a scale invariant behavior. We also identify asymptotic power-law decay relations between the standard deviations of these three growth rates and the initial size. Further, we discuss similarities with complex organizations. In chapter 2, we report on a quantitative analysis of relationships between the number of homicides, population size and other ten urban metrics. By using data from Brazilian cities, we show that well defined average scaling laws with the population size emerge when investigating the relations between population and number of homicides as well as population and urban metrics. We also show that the fluctuations around the scaling laws are log-normally distributed, which enabled us to model these scaling laws by a stochastic-like equation driven by a multiplicative and log-normally distributed noise. Because of the scaling laws, we argue that it is better to employ logarithms in order to describe the number of homicides in function of the urban metrics via regression analysis. In addition to the regression analysis, we propose an approach to correlate crime and urban metrics via the evaluation of the distance between the actual value of the number of homicides (as well as the value of the urban metrics) and the value that is expected by the scaling law with the population size. In chapter 3, we report on the existing connection between power-law distributions and allometries. We present an extensive characterization of this connection when considering all possible pairs of relationships from twelve urban indicators of Brazilian cities. Our analyzes reveal that all twelve urban indicators are asymptotically distributed as power-laws and that, when a pair of indicators present a statistically significant allometry, the allometric exponent becomes completely defined in terms of the exponents of the power-law distributions. In addition, we have also found that the residuals fluctuations surrounding all possible allometries present an almost constant variance and log-normal distributions.

Key-words: Complex Systems. Statistical Physics. Scaling Law. Homicides. Social Systems. Urban Metrics.

Sumário

Introdução	1
1 Leis de escala na dinâmica das taxas de crescimento do crime	4
1.1 Apresentação dos dados	4
1.2 Resultados	7
1.3 Discussão	11
1.4 Conclusões e perspectivas	13
2 Distância às leis de escala: uma abordagem útil para revelar relações entre crimes e métricas urbanas	15
2.1 Apresentação dos dados	16
2.2 Leis de escala entre crime, métricas urbanas e população	16
2.3 Modelo de regressão: homicídios versus métricas urbanas	20
2.4 Uma métrica relativa: distância às leis de escala	22
2.5 Conclusões e perspectivas	26
3 Análise empírica da conexão entre distribuições lei de potência e alometrias para indicadores urbanos	28
3.1 Apresentação dos dados	29
3.2 Métodos e resultados	30
3.3 Conclusões e perspectivas	38
Visão geral e perspectivas	40

A	Métodos e técnicas estatísticas	42
A.1	Método <i>bootstrapping</i>	42
A.2	Teste de hipótese de Kolmogorov-Smirnov	43
A.3	Teste de hipótese de Cramér-von Mises	44
A.4	Coeficiente de correlação de Pearson	45
A.5	Método de Clauset para o ajuste de distribuições lei de potência	45
A.6	Cálculo de Gomez-Lievano <i>et al.</i>	46
B	Definição das métricas urbanas, fontes dos dados e comentários adicionais	48
	Referências Bibliográficas	50

Introdução

A Física Estatística tem provado ser frutífera para descrever fenômenos fora do seu domínio tradicional [1–3]. O estudo de sistemas sociais complexos tem sido foco de pesquisa intensa nas últimas décadas [2–5]. Ao lado disso, o crescimento da integração tecnológica em nossas vidas tem criado um enorme volume de dados relacionados ao comportamento cotidiano da sociedade. Tais dados abrem novas e excitantes oportunidades para um estudo quantitativo de sistemas sociais via métodos computacionais e estatísticos [5]. Eleições [6–9], crescimento populacional [10–13], economia [14, 15] e linguagem [16–18] são apenas alguns exemplos de atividades sociais que foram investigadas recentemente. Tais investigações devem proporcionar uma melhor compreensão de como nossa sociedade é organizada e, também, apontar melhores estratégias para gestão de recursos, alocação de serviços e estratégias políticas. No contexto social, crime é uma das atividades mais preocupantes para nossa sociedade, e entender e prevenir atos criminosos é um grande desafio [19–21].

Particularmente, foi observado que sistemas sociais podem exibir propriedades universais similarmente a sistemas termodinâmicos na criticalidade. Por exemplo, leis de escala têm sido reportadas na pesquisa científica [25, 26], sistemas biológicos [27, 28], atividades econômicas [29–31] e religiosas [32, 33]. Além disso, há algumas evidências sobre a relação entre métricas urbanas e o tamanho da população, em que não linearidades são manifestadas explicitamente por leis de potência [34–38]. Mais recentemente, transições de fase foram encontradas em modelos de criminalidade [20, 21].

Apesar de não haver uma definição bem estabelecida sobre o termo sistema complexo, sistemas que são compostos por muitas partes interagindo entre si e que apresentam um comportamento emergente são, em geral, chamados de sistemas complexos. Exemplos de sis-

temas complexos incluem eleições [6–9], crescimento populacional [10–13], linguagem [16–18], terremotos [22, 23], atividades de raios [24]. A existência de propriedades emergentes é uma das características mais distinguíveis de sistemas complexos, além da interdisciplinariedade intrínseca a esses sistemas [3]. A criminalidade se enquadra nesse último contexto, sendo um assunto multidisciplinar, considerado em diversas áreas do conhecimento, desde as ciências humanas [39–51] às ciências exatas [20, 21, 34–38, 52–58].

A criminalidade é, provavelmente, uma das mais antigas atividades humanas [20] e tem sido uma preocupação constante para a sociedade ao longo de toda a história. Desde as primeiras organizações sociais até os dias de hoje, tem sido um desafio entender e prevenir a atividade criminosa [19–21]. Estamos em um período em que a maioria das pessoas vive em cidades [59] e, portanto, há um grande interesse em entender características do crime em conexão com sistemas urbanos. Muitas razões para cometer-se um crime foram apontadas e uma boa parte da literatura considera os criminosos como indivíduos desviantes [20]. Apesar do enorme interesse em entender o crime, apenas em 1968 foi apresentada uma análise econômica dos custos e benefícios do crime [39]. Usando um ponto de vista similar, outras teorias econômicas foram desenvolvidas na tentativa de explicar a participação em atividades ilegítimas [42, 60]. Além disso, mostrou-se que as interações sociais podem explicar a grande variação do crime em cidades com diferentes concentrações de pessoas [41, 51].

Nas últimas décadas, diferentes abordagens foram usadas no estudo da violência em cidades [61]. Uma grande parte da literatura investigou dados empíricos dos Estados Unidos. Por exemplo, foram investigados, por meio de regressões lineares, crimes violentos em conexão com desigualdade social, pobreza e armas de fogo [44, 62–64]. O ciclo eleitoral de contratações de policiais também foi investigado para estimar o efeito da polícia no crime [43]. Considerando dados de séries temporais, buscou-se a conexão entre desemprego e crime [48]. Ainda, foi sugerido que ambientes degradados podem conduzir a um comportamento de vandalismo/degradação [49]. Dados empíricos de crimes em cidades da América Latina também têm sido foco da investigação científica. Por exemplo, a violência e o desenvolvimento econômico nas cidades da Colômbia foram investigados via análise de regressão [61]. Violência urbana e relações com governabilidade democrática na Argentina também foram estudadas [47]. O papel da desigualdade, do desemprego e das mudanças estruturais foram consideradas com o objetivo de explicar o crime na Argentina [65, 66]. Mais recentemente, desigualdade e crime foram relacionados por meio de dados de cidades brasileiras [50] e análises estatísticas também foram empregadas na caracterização do crime em cidades brasileiras, colombianas e mexicanas [38].

O homicídio, como a expressão máxima da violência na sociedade humana, é um aspecto largamente investigado [38]. Muitas razões são apontadas para escolher esse tipo de dado, como a sua grande disponibilidade em nível municipal e o fato de que a informação é, em geral, mais confiável devido a gravidade do crime. Nesta dissertação, investigamos uma base de dados de homicídios das cidades brasileiras na tentativa de encontrar relações entre crime, tamanho das cidades e métricas urbanas. No capítulo 1, estudamos a dinâmica das taxas de crescimento de crimes, assim como a relação com o tamanho das cidades [67]. Também, no capítulo 2, analisamos correlações entre o número de crimes e métricas urbanas, assim como o surgimento de leis de escala no contexto de sistemas urbanos [68]. Por fim, no capítulo 3, estudamos um cenário estatístico que conecta as leis de escala alométricas entre métricas urbanas com distribuições tipo lei de potência [69]. Para que a leitura se torne mais agradável, apresentamos as técnicas e os métodos estatísticos nos apêndices. Por fim, apresentamos uma visão geral dos problemas estudados aqui, assim como perspectivas para trabalhos futuros no estudo e na modelagem da criminalidade e outros sistemas sociais.

Leis de escala na dinâmica das taxas de crescimento do crime

Apesar das amplas investigações e da relevância social dos estudos sobre criminalidade em conexão com a urbanização, nossa compreensão da universalidade e da abrangência de padrões nesse assunto permanece limitada. Neste capítulo, mostraremos uma abordagem estatística para investigar padrões e características do crescimento de homicídios. De forma quantitativa, nosso estudo fornece ideias sobre como o crescimento do crime e o tamanho das cidades estão relacionados. Mais precisamente, estudamos as taxas de crescimento de homicídios nas cidades brasileiras com base em dados de 1980 a 2009 (seção 1.1). Investigamos as taxas de crescimento de homicídios nas cidades com o objetivo de compreender os mecanismos que regem a criminalidade (seção 1.2). Mostramos que nossos resultados têm uma estrutura formal semelhante àquelas encontradas em outros sistemas complexos tais como pesquisa científica, sistemas biológicos, atividades econômicas e religiosas, fato este que coloca a universalidade dos nossos resultados em evidência (seção 1.3). Motivados por essa universalidade, indicamos um cenário para nossos resultados em termos de modelos estocásticos propostos no âmbito das organizações complexas (seção 1.3).

1.1 Apresentação dos dados

Para o estudo de propriedades estatísticas do crime, consideramos uma base de dados contendo o número anual de homicídios em todas as cidades brasileiras, abrangendo os anos de 1980 a 2009. Os dados foram obtidos a partir da base de dados de estatísticas vitais do

DATASUS [70]. O número total de habitantes das cidades foram obtidos da base de dados demográfica e socioeconômica do DATASUS [70], cujos anos 1980, 1991, 1996 e 2000 reportam o número de indivíduos obtido pelo censo populacional realizado pelo IBGE [71], enquanto os demais anos registram valores aproximados da população estimados pela agência IBGE. Devido a isso, nossa análise será focada principalmente na base de dados de homicídios.

Selecionamos as 528 cidades que apresentam um número significativo de homicídios (ao menos um por ano) no período de 1980 a 2009. Elas são cerca de 10% das cidades brasileiras, mas representam aproximadamente 80% do número total de homicídios no Brasil em todo o período considerado. Além disso, a porcentagem média da população do país vivendo nessas cidades durante esse período é de cerca de 60%. Uma ilustração de nossa base de dados é apresentada na Fig. 1.1. Nela, uma lei de escala típica pode ser observada se considerarmos apenas cidades grandes (população maior que 53.465). Encontramos um expoente bem próximo àqueles encontrados em outros estudos sobre métricas urbanas e crimes [34–38].

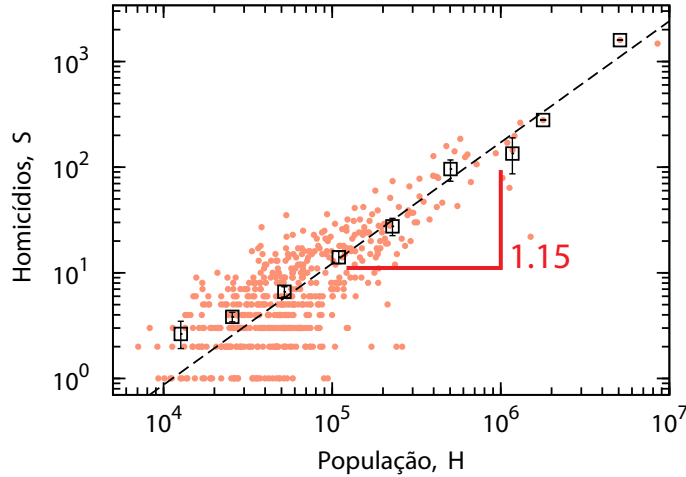


Figura 1.1: Os pontos vermelhos são os dados do número de homicídios versus a população das 528 cidades selecionadas no ano de 1980 (em escala log-log). Os quadrados pretos são os valores médios dos dados divididos em janelas. A linha tracejada em preto é um ajuste linear da tendência média dos dados em escala log-log, onde encontramos um expoente próximo a 1.15. As barras de erros representam o intervalo de confiança de 95% calculado via *bootstrapping* [73] (apêndice A.1).

Em termos do número total de homicídios, $S(t)$, no ano t , em uma dada cidade, a taxa de crescimento logarítmica anual, bianual e trianual são definidas como

$$R_n(t) = \ln \left[\frac{S(t+n)}{S(t)} \right], \quad (1.1)$$

com $n = 1, 2$ e 3 , respectivamente. Para simplificar a notação, omitiremos o subíndice n

quando referirmos à taxa de crescimento anual, isto é, empregaremos $R(t)$ para representar $R_1(t)$. Exemplos da evolução temporal de $R(t)$ para algumas cidades são mostrados na Fig. 1.2. Essa figura ilustra a presença de flutuações nas taxas de crescimento do homicídio $R(t)$. Ela também mostra que as flutuações em $R(t)$ são geralmente maiores em cidades pequenas que em cidades grandes. Isso é um resultado esperado que motiva nossa análise nos desvios padrão em função do tamanho da cidade, visto que as flutuações são maiores em sistemas pequenos que em sistemas grandes, como tem sido verificado em diferentes contextos, tais como pesquisas científicas, sistemas biológicos, atividades econômicas e religiosas [25–31, 33].

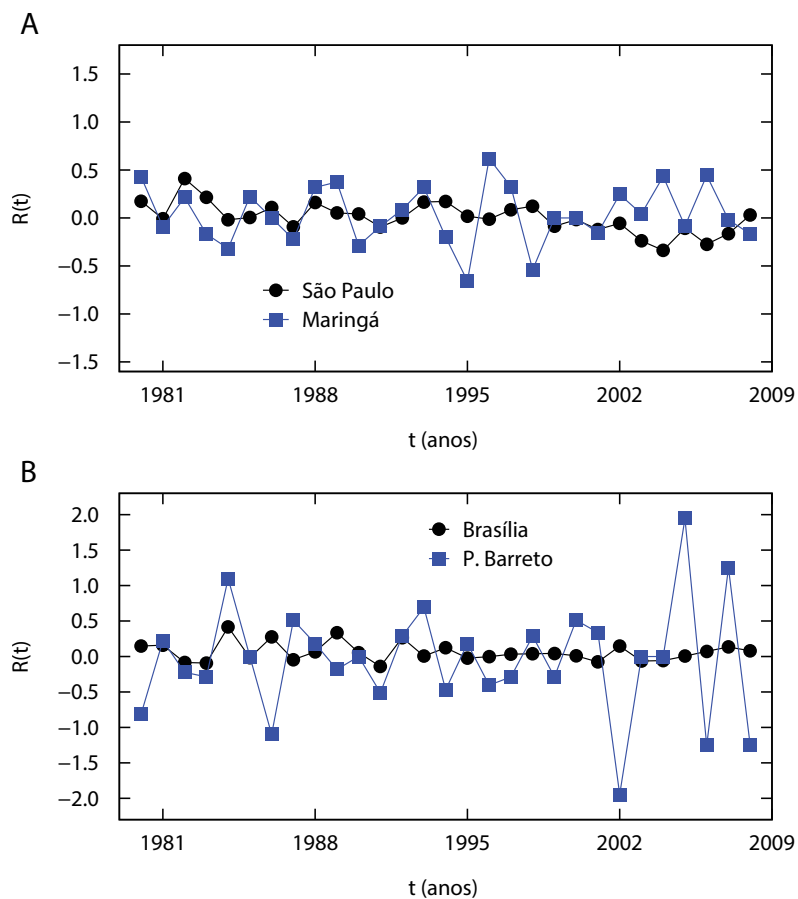


Figura 1.2: Exemplos de $R(t)$ dado pela equação (1.1) para algumas cidades no período de 1980 a 2009. (a) São Paulo ($S = 1480$) e Maringá ($S = 15$). (b) Brasília ($S = 145$) e Pereira Barreto ($S = 9$). Observe que as flutuações em $R(t)$ são geralmente maiores em cidades com um número pequeno de homicídios S no ano de 1980.

1.2 Resultados

Mesmo sem uma descrição detalhada do sistema, os estudos sobre sistemas complexos têm mostrado que as propriedades de escala das flutuações das quantidades macroscópicas podem levar a uma melhor compreensão de algumas características dessas variáveis. Nesse sentido, uma variedade de organizações complexas tem sido investigadas e foi demonstrado que, apesar das diferenças, todas elas exibem distribuições não-gaussianas em forma de tenda em suas taxas de crescimento anual, com um comportamento de lei de potência média em seus desvios padrão [25–31,33]. Esses estudos sugerem que existe algum mecanismo universal, independente dos detalhes particulares, que regula as flutuações das taxas de crescimento de organizações complexas. Esses fatos nos motivam a investigar aspectos do crescimento da criminalidade com foco nessas propriedades estatísticas.

Primeiramente, consideramos a distribuição condicional de $R(t)$ levando em conta alguns intervalos de valores iniciais para $S(t)$. Com a intenção de reduzir os efeitos de flutuações estatísticas, empregamos distribuições cumulativas (CDFs). Para usar esse método, separamos as taxas de crescimento $R(t)$ em valores positivos e negativos. Por exemplo, para $R(t)$ positivo, temos a CDF

$$p_c(R|S) = \int_R^\infty p(R'|S) dR', \quad (1.2)$$

em que $p(R'|S)$ é a função densidade de probabilidade condicional. A Fig. 1.3 mostra exemplos de CDFs empíricas considerando três intervalos de S (aqui usamos S para representar $S(t = 1980)$) entre o conjunto de valores possíveis, nos quais encontramos uma boa concordância entre os dados empíricos e a família de distribuições de Laplace

$$p(R|S) = \frac{1}{\sqrt{2} \sigma} \exp\left(-\frac{\sqrt{2} |R - \mu|}{\sigma}\right), \quad (1.3)$$

com μ sendo o valor médio e σ o desvio padrão de R dentro de cada intervalo de S selecionado.

O comportamento de escala pode ser graficamente representado considerando uma mudança apropriada de escala das variáveis independentes e dependentes. Se a escala existe, os dados para vários valores iniciais, S , devem colapsar em uma única curva. Por exemplo, notamos na equação (1.3) que

$$r(t) = \frac{R(t) - \mu}{\sigma} \quad (1.4)$$

é a mudança de variável que realiza o colapso das distribuições. Essa mudança de variável leva a

$$p(r|S) = \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2} |r|\right), \quad (1.5)$$

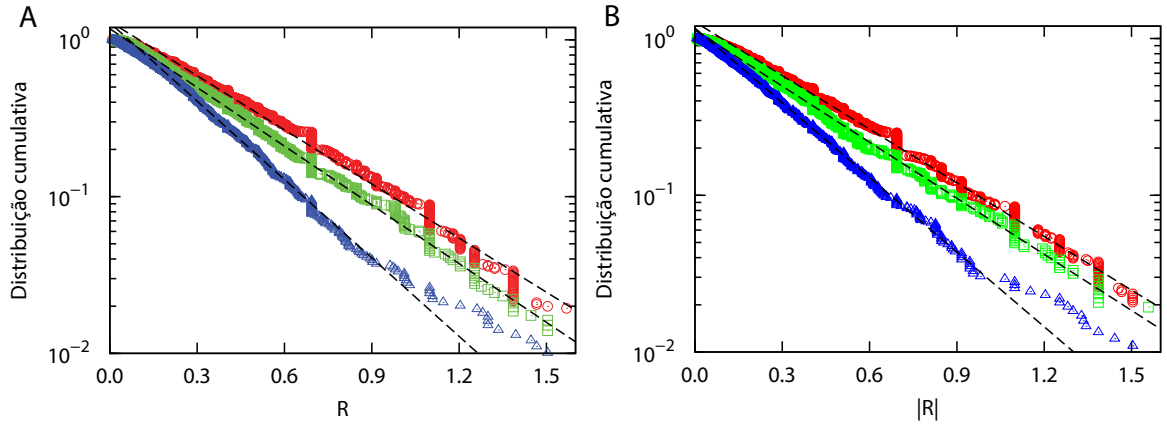


Figura 1.3: Distribuições cumulativas $p_c(R|S)$ para alguns intervalos do número inicial de homicídios S : círculos vermelhos ($5 \leq S \leq 8$), quadrados verdes ($9 \leq S \leq 13$) e triângulos azuis ($14 \leq S \leq 23$). Observe que os gráficos estão em escala mono-log. As linhas tracejadas são ajustes da distribuição de Laplace. (a) Valores positivos de R e (b) módulo dos valores negativos de R .

em que $p(r|S)$ é a distribuição de probabilidade da variável escalada r , que não possui parâmetros.

Dividimos os dados em intervalos log-espaçados de S . Então, usando a variável escalada r , verificamos que todas as curvas, para diferentes intervalos de S , colapsam aproximadamente em uma única CDF (veja a Fig. 1.4), indicando que a distribuição das taxas de crescimento exhibe a mesma forma funcional para escalas de tamanho distintos. A partir dessa figura, notamos, também, que a distribuição de Laplace apresenta desvios pequenos (mas sistemáticos) dos dados empíricos.

Na tentativa de contornar esses desvios, ajustamos os dados usando uma distribuição com um parâmetro

$$p(r|S) = \frac{c}{2} \sqrt{\frac{\Gamma(3/c)}{\Gamma(1/c)^3}} \exp \left[- \left(\frac{\Gamma(3/c)}{\Gamma(1/c)} \right)^{c/2} |r|^c \right]. \quad (1.6)$$

Essa distribuição é a gaussiana alongada (ou *stretched Gaussian*) [74, 75] que recupera a gaussiana padrão para $c = 2$ e a distribuição de Laplace para $c = 1$. Em nosso caso, encontramos que $c \approx 1.2$ fornece uma descrição mais precisa para as CDFs empíricas (veja a Fig. 1.4). Note que, se considerarmos diretamente $p(r|S)$ no lugar de sua CDF, obtemos a distribuição em forma de tenda mostrada na Fig. 1.5.

Retornando à Fig. 1.3, notamos que, para valores pequenos de S , as distribuições possuem caudas longas, que se tornam curtas com o crescimento de S . Esse resultado e a Fig. 1.2 indicam que flutuações grandes nas taxas de homicídios são observadas em pequenas cidades e que o desvio padrão σ tende a decrescer com o crescimento de S . Nesse ponto, estamos

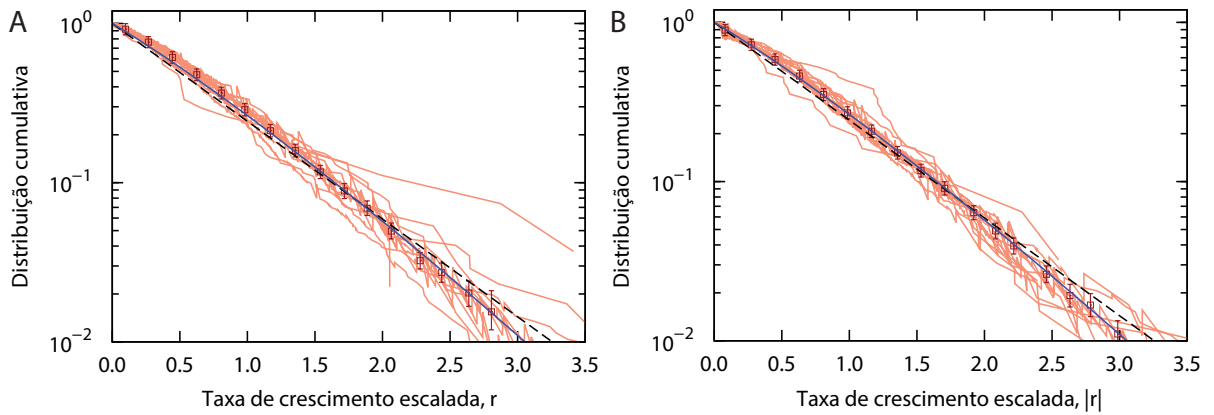


Figura 1.4: Distribuição cumulativa normalizada em escala mono-log. As linhas contínuas em vermelho são as distribuições cumulativas da taxa de crescimento normalizada r , para cada intervalo log-espaçado de valores iniciais S . Os quadrados são valores médios das CDFs por janela na taxa de crescimento escalada r . As linhas tracejadas azuis mostram a distribuição de Laplace com desvio padrão unitário e média zero (Eq. 1.5). As linhas contínuas se referem à distribuição gaussiana alongada (Eq. 1.6), com $c = 1.2$. (a) Valores positivos de r e (b) módulo dos valores negativos de r . As barras de erros representam o intervalo de confiança de 95% calculado via *bootstrapping* [73].

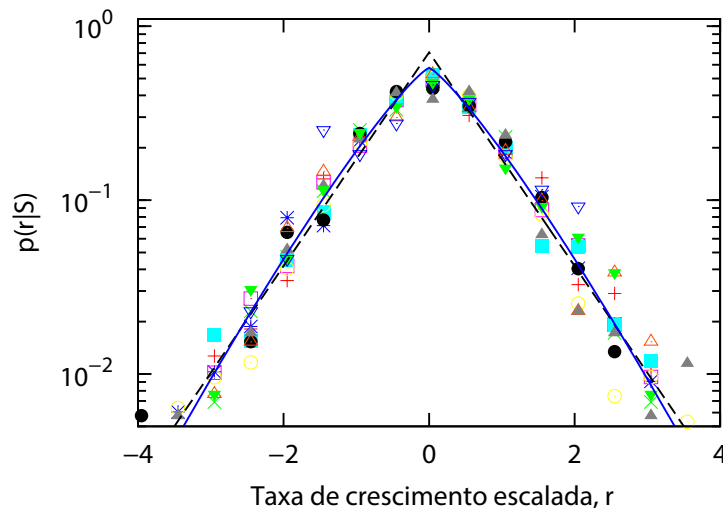


Figura 1.5: Função densidade de probabilidade condicional $p(r|S)$ das taxas de crescimento anual r em escala mono-log. Os diferentes símbolos representam cada intervalo de S empregado na análise da distribuição. Aqui, os intervalos de S são os mesmos usados na Fig. 1.4. A linha tracejada em preto é a distribuição de Laplace (1.5) com desvio padrão unitário e média zero. A linha contínua em azul é a distribuição gaussiana alongada (1.6) com $c = 1.2$.

considerando o número de homicídios como uma medida do tamanho da cidade, uma vez que, em princípio, o número médio de crimes aumenta com o número de habitantes (veja a

Fig. 1.1 e [38]). Em nossa análise, encontramos que a principal tendência entre σ e S é bem ajustada pela lei de potência

$$\sigma \sim S^{-\beta_S}, \quad (1.7)$$

com $\beta_S \approx 0.36$ para $S > 8$ (veja Fig. 1.6a). Entretanto, a relação entre o número de crimes e a população é não linear [38] e, ainda, não entendida por completo quando consideramos todos os tamanhos de cidades, isto é, todas as cidades brasileiras. Por essa razão, investigamos diretamente σ como função da população total H de cada cidade no ano de 1980. Como resultado, verificamos que $\sigma \sim H^{-\beta_H}$ com $\beta_H \approx \beta_S$, como mostrado na Fig. 1.6b. Esse resultado está em concordância com a Fig. 1.1, visto que, quando levamos em conta as 528 cidades, a relação entre o número de homicídios e a população é aproximadamente linear, conforme discutido na seção 1.1.

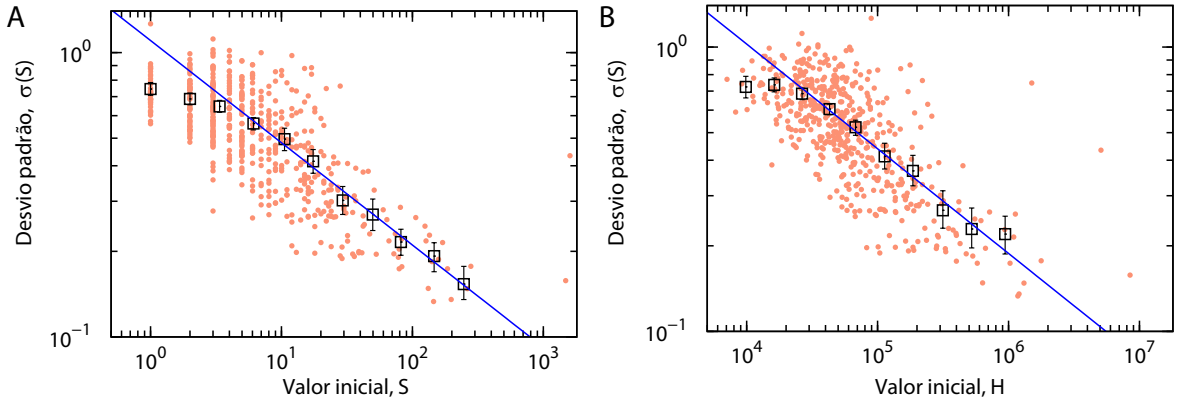


Figura 1.6: Desvio padrão versus o tamanho da cidade em escala log-log. (a) Os pontos vermelhos são os desvios padrão das taxas de crescimento anual observadas versus o valor inicial de S da série temporal dos homicídios das 528 cidades examinadas. Os quadrados em preto são médias por janela do desvio padrão σ das taxas de crescimento R em função do valor médio de S . (b) Aqui, substituímos o valor inicial de homicídios S pela população H . Os quadrados em preto são médias por janela do desvio padrão σ das taxas de crescimento R em função do valor médio de H . Para ambos, as linhas contínuas são ajustes lineares da principal tendência do dado em escala log-log (equação (1.7)), resultando em um expoente $\beta_H \approx \beta_S \approx 0.36$. As barras de erros representam o intervalo de confiança de 95% calculado via *bootstrapping* [73] (apêndice A.1).

Além de $R(t)$, também investigamos as taxas de crescimento bianual, $R_2(t)$, e trianual, $R_3(t)$, de homicídios. Nesses últimos dois casos, temos o mesmo cenário das taxas de crescimento anual. Contudo, os valores numéricos dos parâmetros relacionados às taxas de crescimento anual, bianual e trianual são, em geral, diferentes, como mostrado na Tabela 1.1. Note que cada valor de β_S e o correspondente β_H são bem próximos um do outro e que

eles decaem com o crescimento do intervalo de tempo. Se considerarmos σ como função de $S(t)$ ou $H(t)$ em outros anos no lugar de $t = 1980$, também obtemos um comportamento assintótico lei de potência. Entretanto, o expoente β decresce com o crescimento de t . Esse fato é mais aparente quando consideramos σ como função de $S(t)$ do que $H(t)$. No caso de $H(t)$, os expoentes são aproximadamente constantes. Como uma última preocupação, consideramos taxas de crescimento anual $R(t)$ das cidades com, ao menos, um homicídio por ano em períodos menores, isto é, começando em anos subsequentes a 1980 e terminando em 2009. Naturalmente, esse tipo de análise contém condições iniciais diferentes. Porém, os valores de β_S e β_H flutuam em torno de 0.33 e 0.36, respectivamente. Todos esses resultados, assim como os anteriores, são robustos perante a mudanças no tamanho das janelas utilizadas.

Tabela 1.1: **Resumo dos resultados**

Taxas de crescimento	Forma da distribuição	c (SG)	β_S	β_H
Anual	<i>Tent-shaped</i>	1.2	0.36	0.36
Bianual	<i>Tent-shaped</i>	1.2	0.29	0.28
Triannual	<i>Tent-shaped</i>	1.2	0.23	0.21

Um resumo dos nossos resultados sobre a taxa logarítmica de crescimento dos homicídios anual, bianual e triannual (veja a equação (1.1)) para as cidades brasileiras é apresentado considerando os dados de 1980 a 2009. As segunda e terceira colunas especificam a forma das distribuições colapsadas, em que o parâmetro c se refere a gaussiana alongada (SG) e indica o desvio da distribuição de Laplace ($c = 1$) (veja a equação (1.6)). Os parâmetros β_S e β_H são expoentes assintóticos da tendência principal do desvio padrão σ das taxas de crescimento logarítmicas (veja equação (1.7)).

1.3 Discussão

Uma grande parte da literatura sobre criminalidade tenta relacionar índices de criminalidade com possíveis variáveis explicativas como renda média, desemprego, desigualdade social, gênero, idade, escolaridade, etnia e outras relacionadas à delinquência [56]. Certamente, cidades com o mesmo tamanho podem apresentar números diferentes de crime (como mostrado na Fig. 1.1) devido, usualmente, a aspectos socioeconômicos. Contudo, em nossa análise, não especificamos tais dependências, visto que nosso resultado surge de um contexto estatístico mais geral. Esse fato pode ser visto como uma reminiscência dos conceitos de universalidade encontrados em Física Estatística, em que diferentes sistemas podem ser caracterizados por uma mesma lei fundamental, independente dos detalhes microscópicos [29]. Sistemas com os

mesmos valores de expoente de ponto crítico e as mesmas funções de escala pertencem a uma mesma classe de universalidade [77]. Encontrar um expoente para a relação entre o desvio padrão e o tamanho da cidade é, de alguma forma, indicar a classe a qual o sistema pertence. De fato, em outros contextos [25–31, 33], esses expoentes geralmente assumem valores diferentes.

Encontramos que as distribuições da taxa de crescimento dos homicídios exibem uma estrutura de escala invariante, considerando cidades de tamanhos distintos, como mostrado na Fig. 1.4. Nossos resultados mostram que as distribuições das taxas de crescimento de homicídios anual, bianual e trianual são próximas da distribuição de Laplace. Também, observamos que o desvio padrão das taxas de crescimento tem uma tendência de decrescer com o crescimento do número inicial de crimes (Fig. 1.2 e Fig. 1.3). Além disso, o valor médio dos desvios padrão é bem ajustado por leis de potência (Fig. 1.6). Resultados similares em taxas de crescimento anual e leis de potência também foram reportados para outras organizações complexas, mas, usualmente, com valores diferentes para o expoente β [25–31, 33]. Esse paralelo entre criminalidade e organizações complexas sugere que há similaridades em suas dinâmicas de crescimento.

Como visto, há uma universalidade no modo com o qual as flutuações na taxa de crescimento são distribuídas, mesmo que elas sejam diferentes em muitos aspectos, incluindo os detalhes sobre educação das pessoas, saúde pública, desenvolvimento cultural e etnia. Nossos resultados também indicam que a dependência tipo lei de potência nesse sistema social poderia surgir devido a algum cenário emergente, tal como ocorre na dinâmica das taxas de crescimento de sistemas físicos e biológicos, nas publicações científicas e nas atividades econômicas e religiosas. Aqui, tentando abranger as características universais acima e, em uma perspectiva diferente de vários trabalhos que aplicaram uma variedade de técnicas e modelos na busca de uma caracterização do comportamento de crime [56], apresentamos motivações para modelos estocásticos propostos no contexto de organizações complexas. Embora a modelagem não seja o principal objetivo de nossas investigações empíricas, discutiremos alguns aspectos de nossos resultados que indicam uma conexão entre a criminalidade e o quadro de organizações complexas.

Podemos considerar que as organizações complexas têm uma estrutura interna que envolve subunidades [25–31, 33]. Nesse contexto, o modelo de Gibrat, inicialmente proposto para descrever o crescimento de firmas [76], faz o papel de uma primeira tentativa de descrever organizações complexas. No nosso contexto de crimes, ele pode ser escrito como

$$S(t+1) - S(t) = A \epsilon(t) S(t), \quad (1.8)$$

em que $S(t)$ é o número de homicídios em uma cidade no ano t , $\epsilon(t)$ é um número aleatório que segue uma distribuição gaussiana com média zero e desvio padrão unitário e A é uma constante positiva. Como podemos notar, esse modelo assume que o crescimento do homicídio não é correlacionado ao tempo. Infelizmente, essa abordagem simples não é suficiente para descrever nossos resultados empíricos tal como as distribuições em forma de tenda nas taxas de crescimento e a lei de potência assintótica nos desvios padrão. Contudo, sua simplicidade pode ser vista como uma primeira ideia para a compreensão do mecanismo que governa as taxas de crescimento de homicídios. Generalizações desse modelo podem incluir interações entre as subunidades (por exemplo, a não linearidade) e a pertinência do número prévio de homicídios (memória). Como uma última observação nessa breve discussão sobre modelagem, ilustramos que, se a não linearidade e a memória são incorporadas ao modelo de Gibrat, uma melhoria substancial é obtida. Para essa ilustração, seguimos Ref. [26], isto é, substituímos $A\epsilon(t)S(t)$ por $\lambda(t)S^k(t)$, em que $\lambda(t) = [A + B\lambda(t-1)]\epsilon(t)$ e A , B e k são constantes positivas ($k = 1$ e $B = 0$ recupera o modelo de Gibrat). O valor do parâmetro de memória B basicamente dita o grau em que a distribuição de probabilidade das taxas de crescimento do homicídio se aproxima da forma de tenda. Por outro lado, o parâmetro k essencialmente fixa o valor de β . Assim, esse modelo é suficiente para reproduzir dois aspectos qualitativos básicos dos resultados empíricos: a forma de tenda da distribuição escalada e o comportamento lei de potência para os desvios padrão.

1.4 Conclusões e perspectivas

Nosso estudo sobre as taxas de crescimento de homicídios fornece uma nova caracterização quantitativa da criminalidade. Usando um quadro de organizações complexas, investigamos a relação entre o número de homicídios e as correspondentes taxas logarítmicas de crescimento anual, bianual e trianual para as cidades brasileiras com pelo menos um homicídio por ano no período de 1980 a 2009. Obtivemos dois principais resultados empíricos com base nessas três taxas de crescimento. Um deles é que os desvios padrão das taxas de crescimento em função do número de assassinatos iniciais seguem um decaimento assintótico tipo lei de potência. Além disso, se o número de habitantes é considerado no lugar do valor inicial de crimes, a lei de potência nos desvios padrão permanece essencialmente inalterada. Outro resultado empírico importante é que a distribuição das taxas de crescimento é não-gaussiana para escalas de tamanhos distintos, porém apresenta uma mesma forma, indicando a presença de leis de escala na dinâmica da taxa de crescimento do crime. Essa distribuição não-gaussiana é

bem descrita pela distribuição de Laplace e pequenas correções podem ser feitas empregando-se uma gaussiana alongada. De um modo quantitativo, todos esses aspectos são resumidos na Tabela 1.1.

Relatamos que comportamentos empíricos semelhantes também foram observados em uma variedade de organizações complexas, sugerindo uma universalidade nos padrões de sua dinâmica de crescimento. Uma vez que esse comportamento universal não é dependente de vários detalhes do sistema, modelos de organizações complexas devem ser bons candidatos quando se investiga a dinâmica do crime. A partir da literatura dessas organizações, consideramos um modelo que reproduz dois aspectos qualitativos básicos de nossos dados: as distribuições não-gaussianas e o expoente da lei de potência dos desvios padrão.

Distância às leis de escala: uma abordagem útil para revelar relações entre crimes e métricas urbanas

Vários trabalhos apontam a existência de relações entre o número de atos criminosos e indicadores urbanos tais como renda, desemprego e desigualdade social [39–42, 51, 60]. A maioria desses trabalhos emprega análise de regressão, em que a variável dependente é o indicador de crime (usualmente o número de um ato criminoso particular) e as variáveis independentes são indicadores urbanos [43–48, 50, 61–66]. Contudo, a maioria desses estudos não consideram a forma funcional da relação entre crime, indicadores urbanos e o tamanho da população; usualmente assumem que essas relações são lineares [56]. Por outro lado, muitos trabalhos mostram que crimes e indicadores urbanos obedecem a leis de escala com o tamanho da população das cidades e, também, entre si [34–38]. Por exemplo, o número de homicídios cresce super linearmente com a população [38, 67]. Não considerar essas leis de escala pode ser uma das razões de muitos trabalhos baseados em regressões lineares conduzirem a conclusões controversas [56]. Além disso, se assumirmos que essas leis de escala com o tamanho da população são, de alguma maneira, uma expressão natural de como as cidades são organizadas, considerar o fenômeno de escala também é importante para que se tenha uma comparação mais justa entre cidades com populações de tamanhos diferentes.

Nesse capítulo, investigamos um procedimento que pode ajudar a resolver esse problema. A abordagem consiste em definir uma “distância” entre o crime ou indicadores urbanos e a principal tendência esperada pela lei de escala com a população. Esse procedimento é baseado

na ideia de competitividade relativa proposta por Podobnik *et al.* [79] no contexto econômico. Os resultados estão organizados da seguinte forma. Começamos apresentando os dados de métricas urbanas e crimes das cidades brasileiras e também uma caracterização intensiva das leis de escala existentes entre esses indicadores e o tamanho da população. Também empregamos um modelo de regressão linear para explicar o número de atos criminosos (homicídios) em termos dos indicadores urbanos. A seguir, usamos a distância previamente discutida em uma tentativa de investigar relações/padrões entre crimes e métricas urbanas que não aparecem na análise de regressão. Finalmente, apresentamos um resumo de nossos resultados na seção 2.5.

2.1 Apresentação dos dados

Acessamos dados de cidades brasileiras no ano de 2000 disponíveis gratuitamente pelo sistema público de saúde do Brasil — DATASUS [70]. Aqui, apesar de existirem outras definições [80], consideramos que cidades são as menores unidades com um governo local. Os dados consistem no tamanho da população (N) e no número de homicídios (H), assim como outros dez indicadores urbanos (Y) em nível municipal: número de casos de trabalho infantil, tamanho da população idosa (idade maior que 60 anos), tamanho da população feminina, produto interno bruto (PIB), PIB per capita, número de analfabetos (idade maior que 15 anos), renda média familiar, tamanho da população masculina, número de residências com instalações sanitárias e número de desempregados (idade maior que 16 anos). Mais detalhes sobre esses indicadores podem ser encontrados na apêndice B. Observe que escolhemos o número de homicídios como sendo o nosso indicador de crime. Como já dissemos, essa é uma escolha vastamente usada [38] devido ao fato de que dados sobre homicídios são mais confiáveis, visto que esse tipo de violência é quase sempre reportada. Também, nossos dez indicadores urbanos são usualmente listados como relacionados ao crime [56]. Além disso, consideramos apenas as cidades com pelo menos um caso de homicídio para nossa análise.

2.2 Leis de escala entre crime, métricas urbanas e população

Começamos revisando a questão sobre a hipótese de homicídios e métricas urbanas apresentarem leis de escala com o tamanho da população (veja também as Refs. [34–38, 67]). Queremos, então, discutir se Y é uma lei de potência em função de N , isto é, $Y \sim N^\beta$, sendo

β o expoente da lei de potência. A Fig 2.1 mostra o gráfico de dispersão de $\log_{10} Y$ versus $\log_{10} N$ para todos os indicadores urbanos, começando pelos homicídios e passando por todas as dez métricas urbanas. Notamos que, apesar da presença de um ruído considerável em algumas relações, as leis de escalas com a população são perceptíveis. Com a intenção de contornar esse ruído e descobrir a principal tendência nessas relações, dividimos os dados em w janelas igualmente espaçadas em $\log_{10} N$ e calculamos o valor médio dos pontos dentro de cada janela. Os símbolos quadrados mostrados na Fig. 2.1 representam esses valores médios e as linhas tracejadas são ajustes lineares. Note que a função linear descreve razoavelmente bem todas as relações médias, ou seja, a equação

$$\langle \log_{10} Y \rangle_w = A + \beta \log_{10} N \quad (2.1)$$

vale para todos os indicadores urbanos. Aqui, $\langle \log_{10} Y \rangle_w$ é o valor médio de Y dentro de cada uma das w janelas, A é uma constante e β é o expoente da lei de potência (mostrado na Fig. 2.1) e a igualdade deve ser entendida como o melhor ajuste. Então, confirmamos que há leis de escala entre os valores médios dos indicadores urbanos Y e a população N . Vale a pena lembrar que essas relações médias são muito robustas quando variamos o número de janelas w (veja a Fig. 2.2).

Outra característica marcante da Fig. 2.1 é a flutuação em torno da lei de potência média. Temos observado que o desvio padrão

$$\sigma_w = \sqrt{\langle (\langle \log_{10} Y \rangle_w - \log_{10} Y)^2 \rangle_w} \quad (2.2)$$

dentro de cada janela praticamente não muda com o tamanho da população N para todos os indicadores urbanos (Fig. 2.3A). Também, verificamos que os resíduos normalizados em torno da lei de potência,

$$\xi = \frac{\log_{10} Y(N) - \langle \log_{10} Y \rangle_w}{\sigma_w}, \quad (2.3)$$

são distribuídos normalmente com média zero e desvio padrão unitário (Fig. 2.3B). Em particular, o teste de Kolmogorov-Smirnov [81] (apêndice A.2) não pode rejeitar a normalidade de ξ para todos os indicadores urbanos (os valores p são todos maiores que 0.51). Os resultados anteriores sobre a distribuição dos resíduos e o comportamento constante do desvio padrão também foram verificados no contexto de eleições [8].

Nossa análise anterior, então, permite uma formulação elegante para as leis de escala médias e, também, para os ruídos dessas tendências. Matematicamente, podemos escrever

$$Y = \mathcal{A} \eta(N) N^\beta \quad (2.4)$$

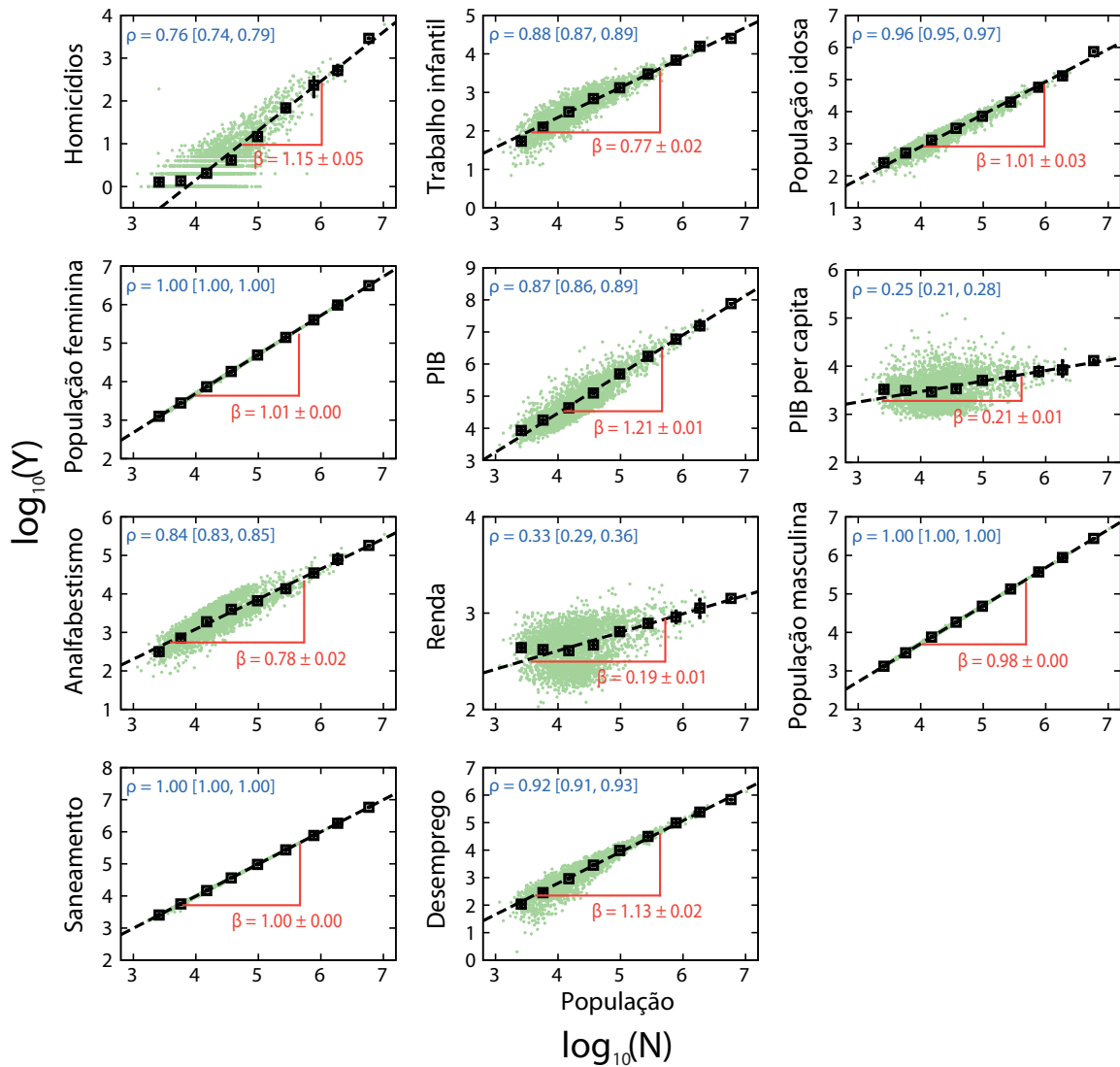


Figura 2.1: Leis de escala entre o tamanho da população e os indicadores urbanos. Em cada gráfico, os pontos verdes são os valores dos indicadores urbanos (Y) em escala logarítmica base-10 versus o tamanho da população (N) para cada cidade. Os quadrados pretos são valores médios dos dados divididos em 10 janelas igualmente espaçadas e as barras de erros são os intervalos de confiança de 95% para os valores médios obtidos via *bootstrapping* [73]. Os valores dos coeficientes de correlação de Pearson ρ (assim como o intervalo de confiança de 95%) dessas relações são mostrados em cada gráfico. As linhas tracejadas são ajustes lineares (via método dos mínimos quadrados) da relação média e as inclinações dessas retas são iguais ao expoente da lei de potência β (mostrado em cada gráfico).

ou, equivalentemente,

$$\log_{10} Y = \log_{10} \mathcal{A} + \beta \log_{10} N + \log_{10} \eta(N), \quad (2.5)$$

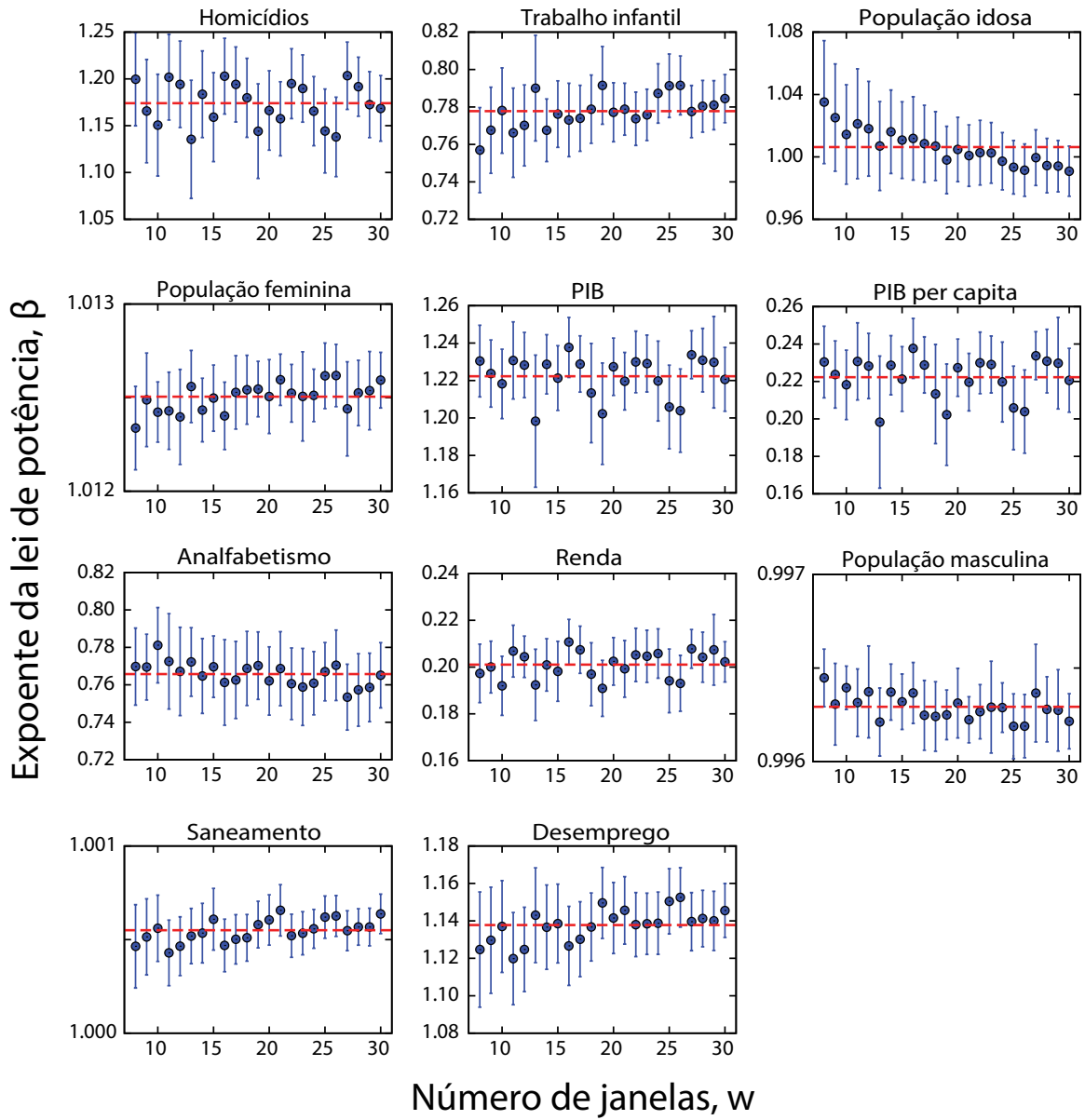


Figura 2.2: A robustez do expoente da lei de potência em relação ao número de janelas empregadas nas relações médias. O valor do expoente β da lei de potência versus o número de janelas w empregadas no cálculo da relação média entre $\log_{10} Y$ e $\log_{10} N$. As barras de erros são os intervalos de confiança de 95% para os valores de β e as retas horizontais em vermelho são os valores médios de β em relação a w . Notamos um comportamento quase constante de β em função de w .

em que, ao compararmos com (2.1) e (2.3), $\log_{10} \mathcal{A} = A$ e $\log_{10} \eta(N) = \sigma_w \xi(N)$. Note que, visto que $\xi(N)$ é normalmente distribuído, $\eta(N)$ deve ser distribuído de acordo com uma distribuição log-normal. Além de descrever as leis de escala média, a Eq. (2.4) representa um

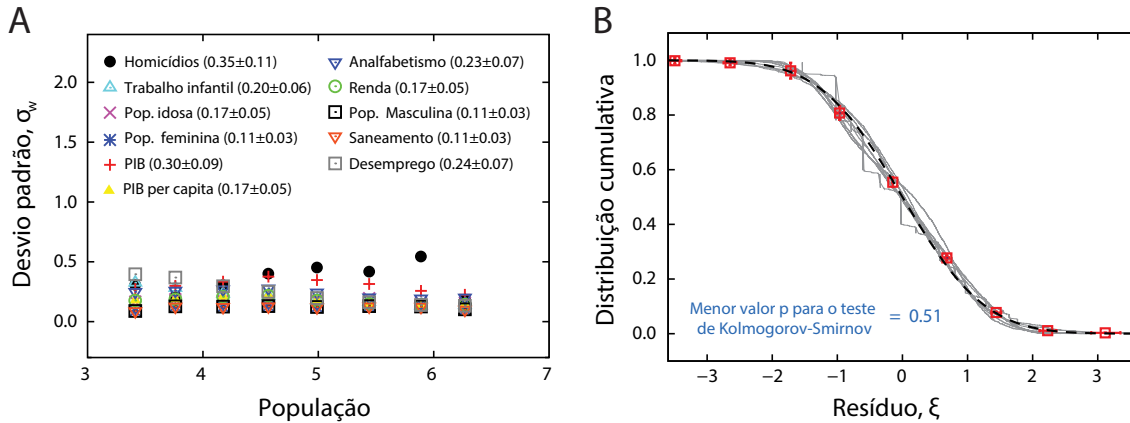


Figura 2.3: Flutuações em torno da lei de escala. (A) Desvio padrão σ_w das flutuações (em escala logarítmica base-10) em cada uma das $w = 10$ janelas igualmente espaçadas. Notamos que o desvio padrão é uma função quase constante para todos os indicadores urbanos. Os valores médios de σ_w em relação às janelas da população é mostrado na legenda do gráfico. (B) Distribuições cumulativas normalizadas de ξ em torno da lei de escala. Nesse gráfico, cada linha cinza é uma distribuição de um dado indicador, os quadrados são os valores médios dessas distribuições cumulativas e as barras de erros são os intervalos de confiança de 95% obtidos via *bootstrapping* [73]. Notamos que a distribuição gaussiana (linha tracejada) descreve razoavelmente bem essas distribuições. Em particular, o menor valor p do teste de Kolmogorov-Smirnov é 0.51, mostrando que não podemos rejeitar a normalidade das flutuações.

processo estocástico em que o indicador urbano Y segue uma relação lei de potência com a população N conduzida por um ruído multiplicativo distribuído log-normalmente.

2.3 Modelo de regressão: homicídios versus métricas urbanas

Como mencionamos anteriormente, uma parte considerável da literatura sobre criminalidade tenta correlacionar indicadores de crime com outras métricas urbanas. Usualmente, essas relações são obtidas a partir de modelos de regressão linear, apesar da não linearidade explícita presente nessas variáveis, tais como as leis de escalas previamente apresentadas. Nesse contexto, não é incomum observar análises baseadas em regressão linear que levam a resultados controversos [56]. Uma maneira alternativa que pode levar em conta essas não linearidades é empregar o logaritmo das variáveis, ou seja,

$$\log_{10} H(i) = C_0 + \sum_k C_k \log_{10} Y_k(i) + \epsilon(i). \quad (2.6)$$

Aqui, $H(i)$ é o número de homicídios na cidade i , $Y_k(i)$ é o k -ésimo ($k > 1$) indicador urbano da cidade i , C_0 é o coeficiente de intercepção, C_k ($k > 1$) é o coeficiente linear que quantifica o efeito explicativo de $\log_{10} Y_k(i)$ e $\epsilon(i)$ é um ruído que considera o efeito de outros fatores não mensuráveis.

Aplicamos o modelo anterior em nossos dados utilizando o método dos mínimos quadrados com correção para heteroscedasticidade¹ [86] e os resultados são resumidos na Tabela 2.3. Primeiramente, notamos que, exceto pelo saneamento e desemprego, todos os indicadores urbanos possuem potencial explicativo para descrever o número de homicídios. Além disso, o valor do R^2 ajustado aponta que o modelo descreve 62% da variância no número de homicídios. Quando analisamos os efeitos individuais dos indicadores urbanos, notamos que trabalho infantil, população idosa, população feminina, PIB per capita e população masculina são negativamente correlacionados com o número de homicídios (H decresce com o crescimento desses indicadores). Por outro lado, PIB, analfabetismo e renda são positivamente correlacionados com o número de homicídios (H cresce com o crescimento desses indicadores). Apesar da falta de uma comparação mais adequada com nossos dados, o resultado de nossa regressão concorda e também discorda com outros resultados empíricos da literatura de criminologia. Por exemplo, encontramos que a correlação entre homicídios e desemprego não é estatisticamente significativa, enquanto há uma correlação positiva e estatisticamente relevante entre homicídios e analfabetismo. Contudo, esses indicadores estão entre aqueles em que há mais conclusões controversas, como apontado por Gordon [56].

Naturalmente, nosso modelo de regressão é muito simples e várias melhorias poderiam ser adicionadas. Por exemplo, algumas dessas métricas podem mostrar correlações e, consequentemente, uma métrica pode afetar a previsibilidade de outra, um fenômeno conhecido como mediação [87]. Uma possível maneira de reduzir esse efeito é combinar algumas métricas e aplicar modelos de regressão diferentes. Outra possibilidade é empregar a análise de componente principal (PCA) para reduzir a redundância entre as métricas urbanas. Contudo, outros problemas como imprecisões na seleção das métricas urbanas e dificuldades em extrair conclusões qualitativas em termos dos eixos da PCA ainda estariam presentes. Aqui, em vez de discutirmos as possíveis controversas que a Tabela 2.1 pode exibir, assim como outras maneiras possíveis de melhorar nosso resultado de regressão, iremos comparar essa análise de regressão simples com nossa nova abordagem baseada em desvios da lei de escala.

¹O termo significa “variâncias diferentes” e vem do grego hetero (diferente) e skedasis (dispersão). Em estatística, uma coleção de variáveis aleatórias é heteroscedástica se há subpopulações que possuem variâncias diferentes.

Tabela 2.1: Coeficientes do modelo de regressão. Valores dos coeficientes lineares C_k obtidos via método dos mínimos quadrados com correção para a heteroscedasticidade [86]. Aqui, t é o valor da estatística t e p é o p -valor para o teste de hipótese de que o coeficiente C_k é diferente de zero.

k	Indicador Y_k	Coeficiente C_k Intervalo de confiança de 95%	Erro Padrão	t	$p > t $
0	Interceção	322.932 [156.944, 488.920]	84.653	3.81	0.000
1	Trabalho infantil	-0.146 [-0.216, -0.076]	0.035	-4.11	0.000
2	População idosa	-0.647 [-0.777, -0.518]	0.066	-9.81	0.000
3	População feminina	-56.644 [-87.015, -26.274]	15.488	-3.66	0.000
4	PIB	121.127 [59.605, 182.648]	31.375	3.86	0.000
5	PIB per capita	-120.987 [-182.509, -59.465]	31.375	-3.86	0.000
6	Analfabetismo	0.213 [0.111, 0.314]	0.051	4.11	0.000
7	Renda	0.223 [0.079, 0.367]	0.073	3.05	0.002
8	População masculina	-62.459 [-93.967, -30.952]	16.068	-3.89	0.000
9	Saneamento	-0.665 [-2.487, 1.156]	0.929	-0.72	0.474
10	Desemprego	-0.026 [-0.082, 0.028]	0.028	-0.94	0.347

R^2 ajustado = 0.62

2.4 Uma métrica relativa: distância às leis de escala

Além de empregar o logaritmo dos indicadores urbanos para contornar as não linearidades, também devemos considerar o comportamento de escala entre os indicadores urbanos, o homicídio e o tamanho da população (Fig 2.1) para uma comparação mais justa entre as cidades com populações de tamanhos diferentes. Então, propomos calcular a diferença entre o valor de um indicador urbano e o valor esperado pelo ajuste lei de potência, isto é,

$$\begin{aligned}
 D_Y &= \log_{10} Y - \langle \log_{10} Y \rangle_w \\
 &= \log_{10} Y - (A + \beta \log_{10} N).
 \end{aligned}
 \tag{2.7}$$

Note que D_Y indentifica se o indicador urbano para uma dada cidade está acima ($D_Y > 0$) ou abaixo ($D_Y < 0$) da lei de escala média, assim como quão longe ele está. Também calculamos

essa distância para o número de homicídios, ou seja, $D_H = \log_{10} H - \langle \log_{10} H \rangle_w$ (note que estamos cometendo um abuso de terminologia quando denotamos D como uma distância). Essa é a mesma ideia proposta por Podobnik *et al.* [79] para quantificar a competitividade entre países, comparando a taxa de crescimento do PIB com a distância do indicador de competitividade global (GCI) ao valor médio esperado pela lei de potência.

Aqui, estudamos relações entre a distância calculada para o indicador de homicídios (D_H) e as outras métricas urbanas (D_Y). A Fig. 2.4 mostra um gráfico de dispersão de D_Y versus D_H , na qual notamos que todas as distâncias das métricas urbanas (exceto pelo desemprego) possuem correlações estatisticamente significantes com a distância do homicídio (veja os valores de correlação de Pearson ρ nesses gráficos). Também observamos que o sinal do coeficiente de correlação ρ concorda com o valor do coeficiente linear C_k para os indicadores trabalho infantil, população idosa, população feminina, PIB, renda, saneamento e desemprego. Contudo, para os indicadores PIB per capita, analfabetismo e população masculina, o sinal de ρ é oposto ao sinal de C_k . Esse resultado significa, por exemplo, que enquanto a análise de regressão sugere que o crescimento da população masculina leva a um decréscimo do número de homicídios, o resultado considerando as distâncias relativas mostra que, quanto mais a população masculina está abaixo da tendência lei de potência, mais o número de homicídios está abaixo da tendência lei de potência. Conclusões controversas similares são obtidas para os indicadores PIB per capita e analfabetismo.

Em adição ao cálculo da correlação de Pearson ρ , os gráficos na Fig. 2.4 revelam outros padrões intrigantes. Notamos que a relação entre a distância dos homicídios e a dos indicadores PIB, PIB per capita, analfabetismo e renda são caracterizados por dois picos na densidades de pontos, enquanto que, para outros indicadores, a densidade de pontos mostra apenas um pico. Observamos, também, que ambos os picos dessas distribuições bimodais estão localizados em torno de $D_H \approx 0$. Esse resultado indica que, apesar dos valores positivos de ρ , há um número considerável de cidades que apresentam valores da distância D_Y abaixo e acima da tendência lei de potência com aproximadamente o mesmo valor para a distância D_H , sugerindo que tais indicadores podem não ser tão bons quanto os outros na descrição do número de homicídios.

Outra maneira de extrair informações significativas da Fig. 2.4 é calculando valores médios. Para fazer isso, agrupamos as cidades em dois conjuntos: aquelas com $D_H > 0$ (homicídios acima da lei de potência) e aquelas com $D_H < 0$ (homicídios abaixo da média). A seguir, calculamos o valor médio de D_Y para cada grupo e consideramos as cidades com um valor absoluto de D_H maior que um limiar Δ . A Fig. 2.5 mostra esses valores médios em

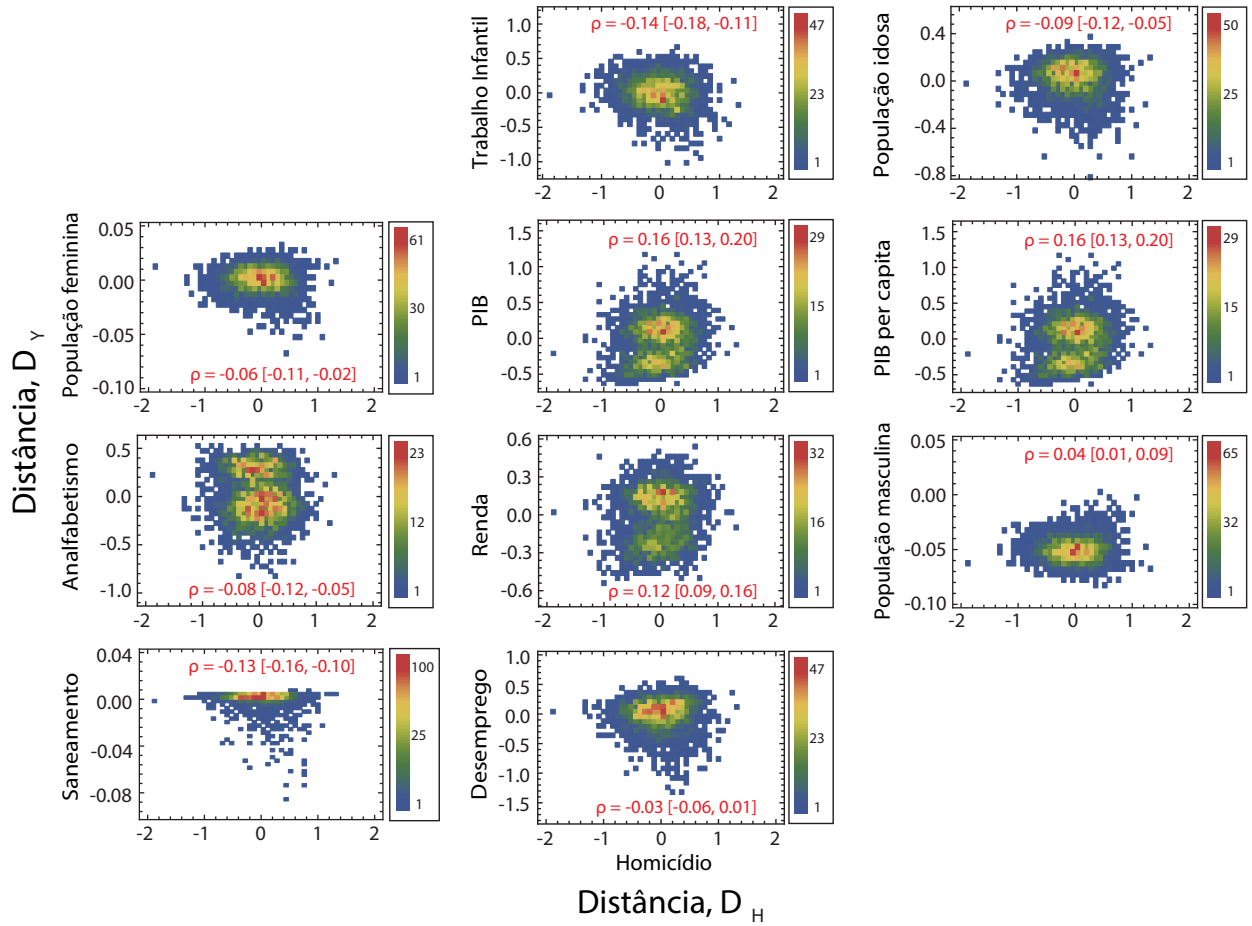


Figura 2.4: Distância à lei de escala avaliadas para os indicadores urbanos versus o número de homicídios. Gráfico da dispersão das distâncias à lei de escala calculada para os indicadores urbanos (D_Y) versus a distância calculada para o número de homicídios (D_H). O código de cores representa a densidade de pontos, indo do azul (densidade baixa) ao vermelho (densidade alta). Mostramos em cada gráfico o valor e o intervalo de confiança de 95% do coeficiente de correlação de Pearson ρ . Notamos que as distâncias D_Y calculadas para os indicadores PIB, PIB per capita, renda e população masculina são positivamente correlacionadas com D_H , enquanto que as D_Y relacionadas ao trabalho infantil, população idosa, população feminina, analfabetismo, saneamento e desemprego são negativamente correlacionados com D_H . Também, observamos distribuições bimodais na relações envolvendo PIB, PIB per capita, analfabetismo e renda.

função do limiar Δ .

Observamos que, para os indicadores trabalho infantil, analfabetismo e saneamento, os valores médios de D_Y são significativamente diferentes entre os dois grupos de cidades e também que a média de D_Y cresce com o crescimento de Δ para as cidades com $D_H < 0$

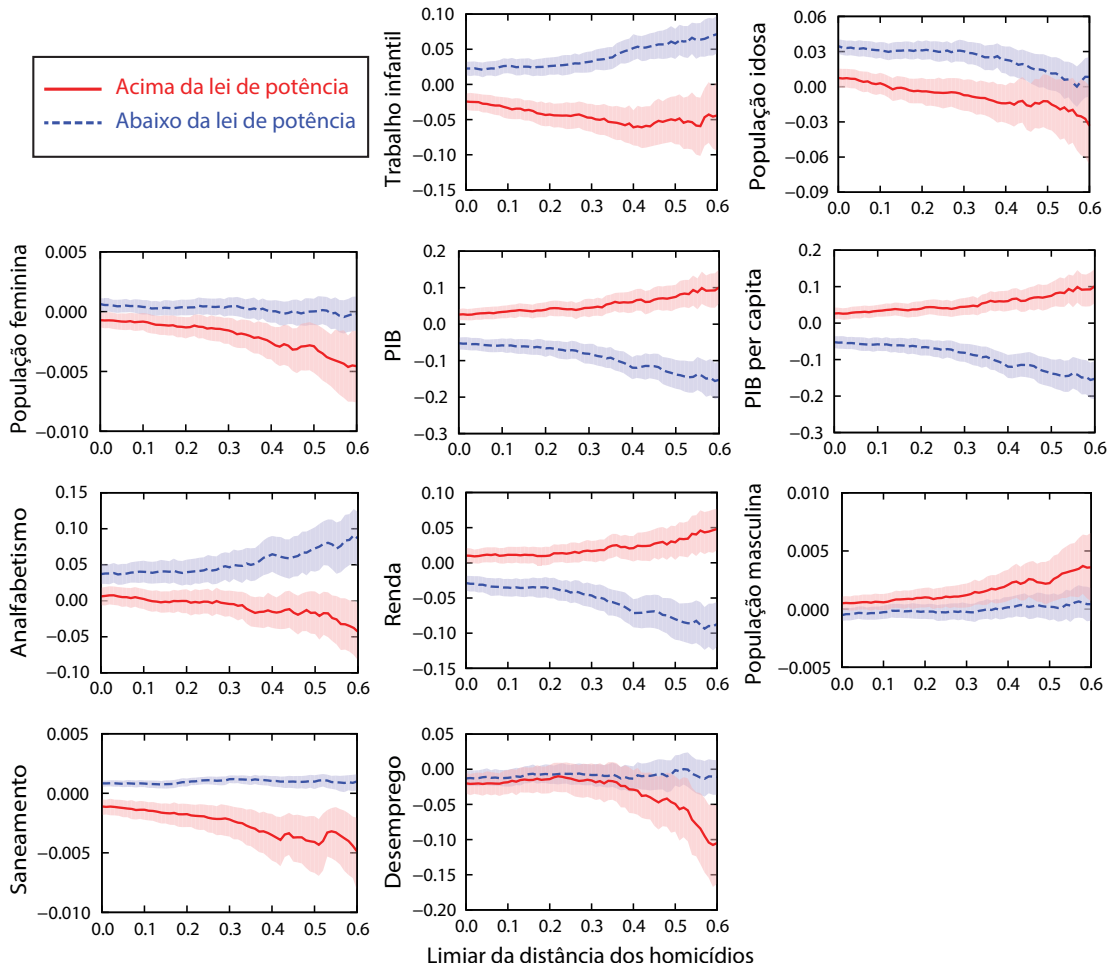


Figura 2.5: Valores médios das distâncias à lei de escala versus o limiar da distância dos homicídios. Os valores médios das distâncias calculados para cada indicador urbano em função do limiar da distância dos homicídios Δ , depois de reagrupar as cidades em acima (linha contínua vermelha) e abaixo (linha tracejada azul) da lei de escala com o tamanho da população. As regiões sobreadas são os intervalos de confiança de 95% para essas médias obtidos via *bootstrapping* [73] (apêndice A.1).

e decresce para aquelas com $D_H > 0$. O oposto ocorre para os indicadores PIB, PIB per capita e renda, ou seja, a média de D_Y decresce com o crescimento de Δ para as cidades com $D_H < 0$ e cresce para aquelas com $D_H > 0$. Curiosamente, para o indicador de população idosa, observamos que cidades com D_H abaixo da lei de potência apresentam uma média para D_Y maior que aquelas com D_H acima da lei de potência; contudo, essa diferença só é significativa para $\Delta \lesssim 0.45$. Esse resultado sugere que, para cidades com número muito grande ou muito pequeno de homicídios comparado ao esperado pela lei de potência, a população idosa

pode não ter potencial explicativo. Similarmente, para o indicador de desemprego, nenhuma diferença é observada entre os valores médios de D_Y abaixo e acima até $\Delta \gtrsim 0.56$. Para um valor um pouco menor de Δ , o valor médio de D_Y (para o desemprego) para as cidades acima da lei de potência começa a decrescer sistematicamente e, para $\Delta \approx 0.56$, uma diferença estatisticamente significativa é observada. Esse resultado, portanto, nos fornece uma pista para uma melhor compreensão do potencial explicativo do indicador de desemprego, apontando que (em nossos dados) seu efeito só é manifestado quando D_H está muito acima do valor médio esperado pela lei de potência.

A Fig. 2.5 também fornece indicativos para um efeito de gênero no número de homicídios. Para a população feminina, notamos que as cidades com o número de homicídios acima da lei de potência ($D_H > 0$) são caracterizadas por valores médios de $D_Y < 0$ que decrescem com o crescimento do valor de Δ . Também, observamos que os intervalos de confiança para os valores médios de D_Y acima e abaixo da lei de potência praticamente não se sobrepõem. Esses resultados, então, apontam que, em cidades onde o número de homicídios está acima do valor esperado, a população feminina é sistematicamente menor que o valor esperado pela lei de escala. Para a população masculina, apesar da sobreposição dos intervalos de confiança para a média de D_Y , observamos um comportamento oposto, ou seja, cidades com número de homicídios acima da lei de potência são também caracterizadas pela população masculina acima da lei de potência.

2.5 Conclusões e perspectivas

Caracterizamos extensivamente algumas relações entre crime e métricas urbanas. Inicialmente, mostramos que os indicadores urbanos obedecem a leis de escala médias bem definidas com o tamanho da população e, também, que a flutuação em torno dessas tendências são distribuídas log-normalmente. Usando esses resultados, mostramos que as leis de escalas podem ser representadas por uma equação multiplicativa estocástica (Eq. 2.4) conduzida por um ruído log-normal. A seguir, abordamos o problema de aplicar a análise de regressão para explicar o número de homicídios H em termos dos indicadores urbanos Y . Devido a não linearidade intrínseca, argumentamos que é melhor empregar o logaritmo dessas variáveis quando realizamos a análise de regressão linear (Eq. 2.4 e Tab. 2.1). Além disso, também discutimos que considerar o fenômeno de escala é importante para uma comparação mais justa entre cidades com diferentes tamanhos de população. Então, propomos o cálculo das distâncias entre o número de homicídios H (D_H), assim como o valor dos indicadores urbanos Y (D_Y) e o

esperado pela lei de escala média. Investigando a correlação de Pearson (ρ) das relações entre D_H e D_Y , encontramos que o valor de ρ tem o mesmo sinal do coeficiente linear C_k para os indicadores trabalho infantil, população idosa, população feminina, PIB, renda, saneamento e desemprego. Por outro lado, para PIB per capita, analfabetismo e população masculina os sinais de ρ e C_k são opostos. Além dos valores de ρ , analisamos os valores médios de D_Y depois de reagrupar as cidades em dois conjuntos: aquelas com o número de homicídios acima da lei de potência ($D_H > 0$) e aquelas abaixo da lei de potência ($D_H < 0$). Essa análise revelou padrões intrigantes que não foram mostrados pela regressão linear. Em particular, nossos resultados para as cidades brasileiras apontam que: *i*) a população idosa pode não ter potencial explicativo quando o número de homicídios está muito acima ou muito abaixo do valor esperado pela lei de potência; *ii*) que o efeito do desemprego no número de homicídios é observado apenas para cidades com D_H consideravelmente maior que o valor esperado pela lei de potência; *iii*) que existe uma diferença de gênero no número de homicídios, pois as cidades com população feminina abaixo do valor esperado são caracterizadas pelo número de homicídios acima da lei de potência, enquanto cidades com o número de homicídios abaixo do valor esperado são caracterizadas por uma população feminina acima da lei de potência. Acreditamos que a abordagem apresentada pode ser empregada em outras bases de dados a fim de produzir resultados mais robustos sobre a relação entre indicadores de crime e métricas urbanas.

Análise empírica da conexão entre distribuições lei de potência e alometrias para indicadores urbanos

Tem havido um interesse notável e crescente nas últimas décadas em investigar sistemas sociais através do quadro de Física Estatística [2, 5, 94–97]. A maioria desses estudos são, entretanto, focados em modelos e propriedades que frequentemente lembram aquelas do contexto de transição de fase [2]. Apesar de sua importância evidente, consideravelmente menos e desigual atenção tem sido dada para trabalhos em dados empíricos relacionados a sistemas sociais. Por exemplo, investigações empíricas sobre os mercados financeiros [14, 98] são muito mais abundantes que aquelas relacionadas a sistemas urbanos. Além disso, visto que mais da metade da população humana vive em áreas urbanas [59, 78], é crucial identificar e entender padrões em sistemas urbanos.

Entre várias propriedades de sistemas urbanos, há duas que são notáveis e onipresentes: a emergência de distribuições lei de potência assintóticas e alometrias para uma vasta gama de variáveis que, de alguma forma, caracterizam os sistemas urbanos (indicadores urbanos ou métricas). Por um lado, população [90, 99], tamanho de prédios [90], fortuna pessoal [100], tamanho de empresas [101], número de patentes [102] e outros indicadores foram verificados seguirem distribuições assintóticas tipo lei de potência. Por outro lado, alometrias com o tamanho da população têm sido reportadas para crime [37, 67, 68], várias métricas urbanas, incluindo patentes, postos de gasolina, produto interno bruto [34, 36, 91] e, também, para número de candidatos [8], de membros dos partidos [103], entre outros. As alometrias foram

recentemente modeladas por Bettencourt [104] via um conjunto pequeno de princípios simples e localmente baseados. Contudo, apenas em 2012 a conexão entre essas duas características foi elucidada por Gomez-Lievano *et al.* [38]. Em seu trabalho, eles mostraram, para relação entre homicídios e população, que, quando ambos indicadores urbanos são assintoticamente distribuídos como leis de potência, esses indicadores urbanos poderão apresentar uma relação alométrica particular. Além disso, o expoente da alometria pode ser completamente determinado a partir dos expoentes das distribuições lei de potência. Nesse capítulo, pretendemos estender esses resultados de forma empírica via uma caracterização extensiva dessa conexão considerando 12 indicadores urbanos e todas as possíveis relações de pares entre eles. Nossos resultados mostram que a conexão proposta por Gomez-Lievano *et al.* é satisfeita para nossos dados quando há correlação suficiente nas relações alométricas. Também discutimos que nem todos os expoentes são independentes e que, na verdade, as alometrias entre pares de indicadores urbanos podem ser entendidas como uma consequência da relação alométrica entre o indicador urbano e a população.

Esse capítulo está organizado como segue. Primeiramente, descrevemos nossa base de dados dos 12 indicadores urbanos das cidades brasileiras. Em seguida, investigamos a hipótese de que esses indicadores urbanos são assintoticamente distribuídos como leis de potência empregando o procedimento estatístico proposto por Clauset *et al.* [105]. Depois, caracterizamos todas as possíveis relações entre os pares de indicadores urbanos considerando o grau de correlação entre eles. Então, discutimos que, quando existe correlação suficiente na alometria, o expoente alométrico pode ser determinado a partir dos expoentes das distribuições lei de potência dos indicadores urbanos. Também, discutimos que o comportamento constante da variância das flutuações em torno das alometrias e a distribuição log-normal dos resíduos são condições necessárias para aplicabilidade da conexão entre leis de potência e alometria. Finalmente, apresentamos um resumo dos resultados e das perspectivas.

3.1 Apresentação dos dados

A base de dados usada nesse capítulo é a mesma apresentada na seção 2.1. Para essa análise, selecionamos 2862 cidades (cerca de 50% das cidades brasileiras), cujos indicadores urbanos estão disponíveis. Mais detalhes sobre os dados podem ser vistos no apêndice B.

3.2 Métodos e resultados

Começamos nossa investigação testando a hipótese de que os 12 indicadores urbanos são distribuídos de acordo com leis de potência assintóticas. A fim de fazer isso, primeiramente avaliamos a função distribuição cumulativa para todos os indicadores urbanos. Como mostrado na Fig. 3.1, a forma das distribuições se assemelha à lei de potência [$P_i(Y_i) \propto Y_i^{-\alpha_i}$], evidenciada aqui pelo comportamento linear (nos gráficos log-log) das distribuições para grandes valores dos indicadores urbanos. Para fortalecer esse resultado, empregamos o procedimento estatístico proposto por Clauset *et al.* [105] (ver apêndice A.5), que é uma maneira sistemática para testar estatisticamente distribuições empíricas tipo lei de potência. Devido à natureza discreta de nossos indicadores urbanos, consideramos a versão discreta do procedimento. Nesse caso, aplica-se o procedimento de ajuste de máxima-verossimilhança considerando a função distribuição de probabilidade

$$P_i(Y_i) = \frac{Y_i^{-\alpha_i}}{\zeta(\alpha_i, Y_{i,\min})}, \quad (3.1)$$

em que $\zeta(\alpha_i, Y_{i,\min}) = \sum_{n=0}^{\infty} (Y_{i,\min} + n)^{-\alpha_i}$ é a função zeta de Hurwitz (que garante a normalização de $P_i(Y_i)$ quando Y_i assume apenas valores discretos), $Y_{i,\min}$ é o parâmetro que representa o início do regime lei de potência e α_i é o expoente da lei de potência. Após encontrar os melhores parâmetros para o ajuste, avaliamos a sua significância via o teste de Cramér-von Mises (apêndice A.3).

Aplicamos o procedimento de Clauset *et al.* em nossos dados e obtivemos os valores de α_i e $Y_{i,\min}$, assim como o p -valor do teste de Cramér-von Mises cujos valores são mostrados na Fig. 3.1. Note que não podemos rejeitar a hipótese lei de potência com confiança de 99% em todos os indicadores urbanos. A única exceção é o número de homicídios, para o qual o p -valor é 0.002. Observamos, contudo, que a estatística de Cramér-von Mises é dominada pelos valores extremos do número de homicídios, na qual notamos um *cut-off* no comportamento. Nesse caso, depois de remover as cidades com os 10 maiores valores, a hipótese lei de potência também não pode ser rejeitada. Verificamos, então, que as distribuições dos 12 indicadores urbanos podem ser modeladas por leis de potência assintóticas, confirmando a hipótese inicial e, também, em concordância com trabalhos anteriores baseados em dados de outros países [34, 36, 91].

Agora, vamos focar na seguinte questão intrigante: se a distribuição lei de potência para indicadores urbanos também implica uma relação alométrica entre pares de indicadores. Como havíamos mencionado, essa questão foi posta no trabalho de Gomez-Lievano *et al.* [38] em que investigou-se a relação entre o tamanho da população e o número de homicídios. Os

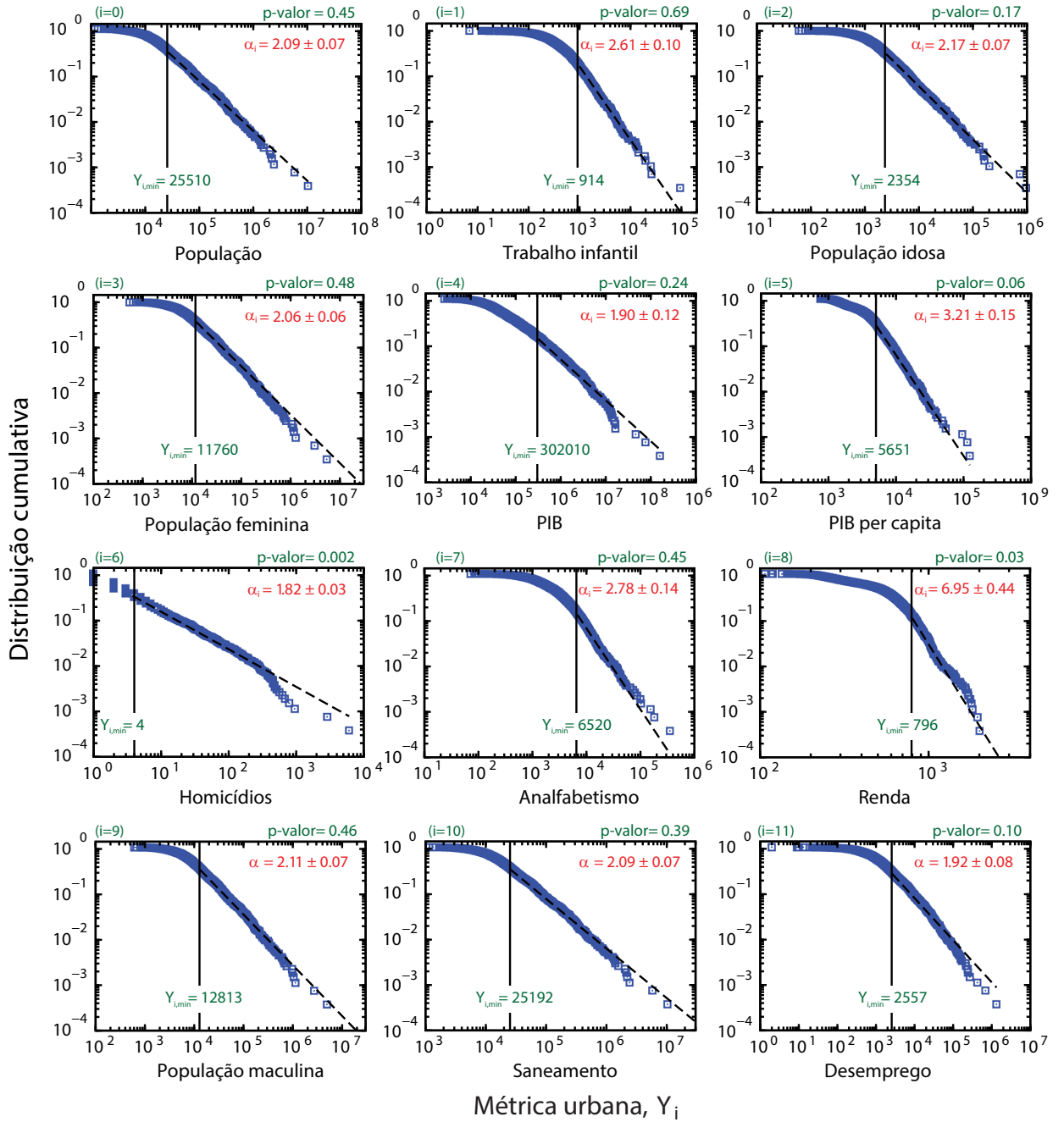


Figura 3.1: Distribuições cumulativas dos 12 indicadores urbanos das cidades brasileiras em 2000. Notamos que todas as distribuições mostram decaimentos assintóticos lei de potência que são bem descrito por funções lei de potência $P_i(Y_i) \sim Y_i^{-\alpha_i}$ para $Y_i \geq Y_{i,\min}$ (linhas tracejadas). Em cada caso, os parâmetros α_i e $Y_{i,\min}$ (mostrados nos gráficos) foram obtidos via máxima-verossimilhança. Os p -valores mostram que não podemos rejeitar a hipótese lei de potência com confiança de 99% para todos os indicadores, exceto para o número de homicídios (veja discussão no texto).

seguintes desenvolvimentos analíticos são apenas uma revisão dos resultados apresentados em seu artigo. Gomez-Lievano *et al.* proporam que, se a forma da alometria é bem estabelecida, podemos escrever

$$P_i(Y_i) = \sum_{Y_i=Y_i^*}^{\infty} P_{i,j}(Y_j | Y_i) P_j(Y_j), \quad (3.2)$$

em que $P_j(Y_j)$ é a distribuição de probabilidade do indicador urbano Y_i e $P_{i,j}(Y_j | Y_i)$ é a probabilidade condicional, representando as flutuações em torno da alometria, as mesmas que discutimos na seção 2.2. Já sabemos que $P_i(Y_i)$ tem a forma lei de potência da Eq. 3.1 e seguindo os achados empíricos, vamos assumir que a probabilidade condicional tem a forma log-normal

$$P_{i,j}(Y_i | Y_j) = \frac{1}{Y_i \sqrt{2\pi \sigma_{i,j}^2(Y_j)}} \exp \left\{ -\frac{[\ln Y_i - \mu_{i,j}(Y_j)]^2}{2\sigma_{i,j}^2(Y_j)} \right\}, \quad (3.3)$$

em que $\mu_{i,j}(Y_i)$ e $\sigma_{i,j}^2(Y_i)$ representam possíveis dependências dos parâmetros da log-normal na variável condicional Y_i . Em particular, devido também a estarmos assumindo que existe uma alometria entre Y_i e Y_j (isto é, $Y_i \propto Y_j^{\beta_{i,j}}$), a forma funcional de $\mu_{i,j}(Y_i)$ é

$$\mu_{i,j}(Y_i) = \ln[\mathcal{A} Y_i^{\beta_{i,j}}], \quad (3.4)$$

se considerarmos ainda que o desvio padrão não depende de Y_i (*i.e.*, $\sigma_{i,j}^2(Y_j) = \sigma_{i,j}^2$).

Após considerarmos todas as suposições anteriores, podemos resolver a Eq. 3.2 substituindo a soma por uma integral e considerando Y_j^* suficientemente pequeno. De fato, após alguns cálculos (apêndice A.6), temos

$$P_i(Y_i) \propto Y_i^{-\frac{(\alpha_j-1)}{\beta_{i,j}}-1}, \quad (3.5)$$

Agora, lembrando que $P_i(Y_i) \sim Y_i^{-\alpha_i}$, podemos escrever a relação entre o expoente alométrico ($\beta_{i,j}$) e as leis de potência (α_i e α_j):

$$\beta_{i,j} = \frac{\alpha_j - 1}{\alpha_i - 1}. \quad (3.6)$$

Vale a pena notar que a mesma relação poderia ter sido obtida usando o fato de que as densidades de probabilidades em áreas diferenciais são quantidades invariantes sob mudança de variáveis, ou seja, $P_i(Y_i)dY_i = P_j(Y_j)dY_j$ ($i \neq j$). Contudo, quando fazemos esse cálculo, também assumimos que a alometria entre Y_i e Y_j é uma expressão exata (não apresentando alguma aleatoriedade), o que não é o caso. Além disso, é igualmente intrigante que a log-normal parece ser a única distribuição para a qual podemos obter uma forma analítica para a relação entre $\beta_{i,j}$, α_i , e α_j (apêndice A.6).

Verificamos analiticamente que, quando os indicadores urbanos apresentam distribuições lei de potência, eles também poderão apresentar relações alométricas particulares. A fim de testar empiricamente esse resultado, investigamos todas as possíveis alometrias entre pares de indicadores urbanos de nossos dados. Especificamente, analisamos todas essas relações considerando o logaritmo dos indicadores urbanos e ajustando (via método dos mínimos quadrados) a função linear

$$\log_{10} Y_i = A_{i,j} + \beta_{i,j}^{(f)} \log_{10} Y_j \quad (3.7)$$

para todos eles. Note que uma análise análoga foi apresentada na seção 2.2 para leis de escala alométricas com a população. Aqui $A_{i,j} = \log_{10} \mathcal{A}_{i,j}$ é uma constante e $\beta_{i,j}^{(f)}$ é o valor empírico do expoente alométrico. A Fig. 3.2 mostra exemplos de alometrias entre os indicadores urbanos população, homicídios e trabalho infantil. Para encontrar o valor de $\beta_{i,j}^{(f)}$, aplicamos um corte no eixo das abscissas considerando apenas os valores de Y_j que são maiores que $Y_{j,\min}$ (o início do regime lei de potência na distribuição de Y_j , veja Fig. 3.1). Os pontos azuis na Fig. 3.2 são os que obedecem a essa condição e a linha tracejada representa a função linear ajustada aos dados. Repetimos esse procedimento para todas as possíveis alometrias e a Fig. 3.3 mostra todos os valores de $\beta_{i,j}^{(f)}$. Vale a pena notar que o expoente alométrico da relação entre Y_i e Y_j deve ser o inverso do obtido para a relação entre Y_j e Y_i . Para os valores empíricos, observamos pequenos desvios nesse comportamento devido aos diferentes cortes aplicados no eixo das abscissas. Contudo, quando consideramos os intervalos de confiança para os valores de $\beta_{i,j}^{(f)}$, verificamos que a relação entre os valores dos expoentes alométricos é satisfeita na maioria dos casos.

Caracterizamos, ainda, a qualidade dessas alometrias calculando o coeficiente de correlação de Pearson $\rho_{i,j}$, que é mostrado na Fig. 3.3 por meio do código de cores. Notamos que a maioria das relações apresentam valores de $\rho_{i,j}$ maiores que 0.5 ($\approx 70\%$). Contudo, também observamos duas exceções sistemáticas: as alometrias com os indicadores urbanos PIB per capita e renda, o que possivelmente se deve ao fato desses indicadores envolverem a população em suas definições. A Fig. 3.2, portanto, confirma que, devido às distribuições lei de potência, a maioria dos pares de indicadores também exibe relações alométricas bem definidas. Ainda, podemos comparar quantitativamente os valores de $\beta_{i,j}^{(f)}$ (estimados a partir das alometrias) com os valores de $\beta_{i,j}$ (obtidos analiticamente a partir da Eq. 3.6). A Fig. 3.4a mostra o gráfico de $\beta_{i,j}^{(f)}$ versus $\beta_{i,j}$ para as alometrias com correlação de Pearson $\rho_{i,j}$ maior que 0.5. Observamos que a função linear (sem constante aditiva) descreve muito bem a relação entre $\beta_{i,j}^{(f)}$ e $\beta_{i,j}$, especialmente se considerarmos os erros padrões dos valores de $\beta_{i,j}^{(f)}$ (área sombreada). Também, observamos que a qualidade dessa relação linear é dete-

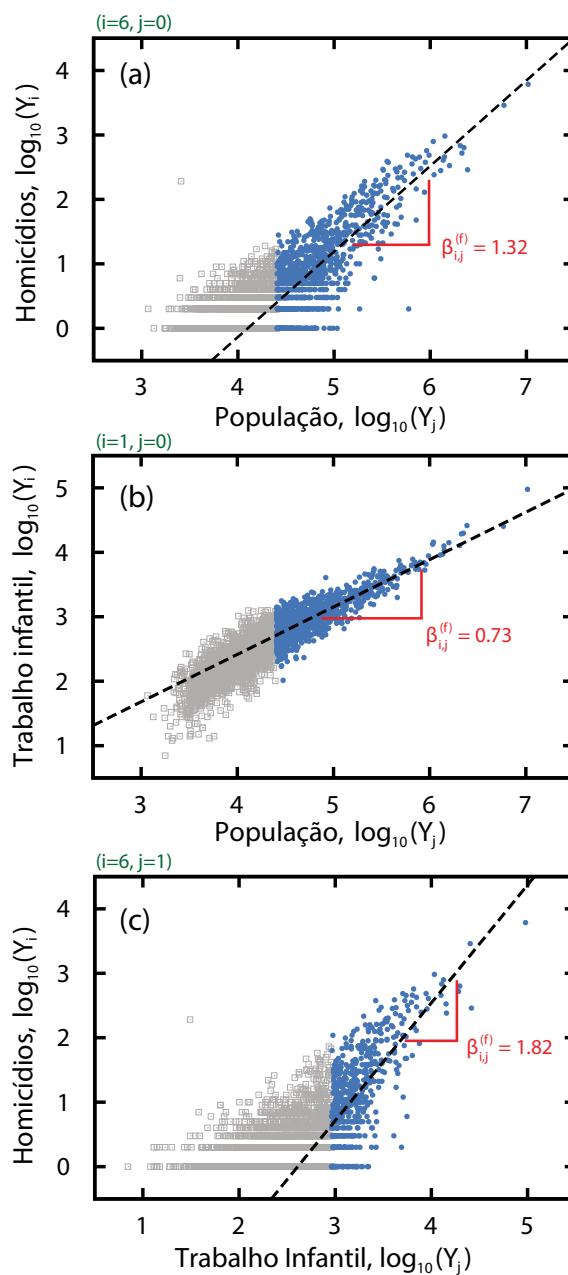


Figura 3.2: Exemplo de alometrias entre indicadores urbanos: (a) homicídios versus população, (b) trabalho infantil versus população e (c) homicídios versus trabalho infantil. Os pontos são os valores logarítmicos na base-10 dos indicadores Y_i versus Y_j para cada cidade. Os pontos azuis representam aquelas cidades que possuem $Y_j \geq Y_{j,\min}$ e os quadrados cinzas são aquelas que $Y_j < Y_{j,\min}$ ($Y_{j,\min}$ é o mesmo obtido na Fig. 3.1). A linha tracejada é um ajuste via método dos mínimos quadrados dos pontos obedecendo à condição $Y_j \geq Y_{j,\min}$ e a inclinação dessa linha é igual ao expoente alométrico $\beta_{i,j}^{(f)}$ (mostrado no gráfico).

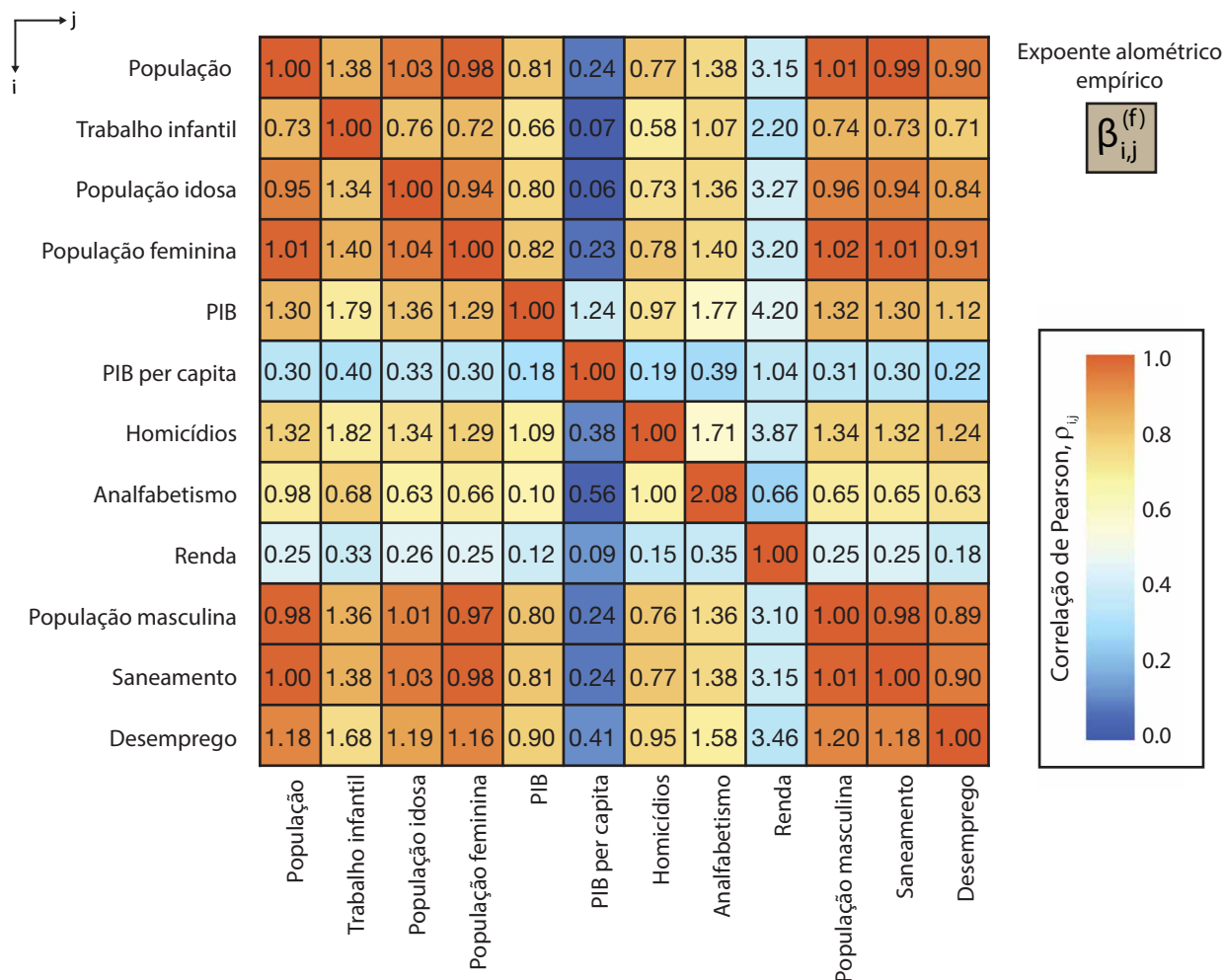


Figura 3.3: Valores empíricos dos expoentes alométricos $\beta_{i,j}^{(f)}$ obtidos para todas as combinações possíveis de relações entre pares dos 12 indicadores urbanos. Essa matriz mostra o valor de $\beta_{i,j}^{(f)}$, em que as colunas representam o caso no qual o indicador urbano é empregado como a variável independente (eixo x) e as linhas representam o caso em que o indicador urbano é empregado como variável dependente (eixo y). O código de cores mostra o valor do coeficiente de correlação de Pearson $\rho_{i,j}$ para cada alometria.

riorada quando começamos considerar alometrias com valores menores de $\rho_{i,j}$. Então, além de serem distribuídos como leis de potência, os indicadores Y_i e Y_j também devem apresentar uma alometria de boa qualidade, (valor grande para $\rho_{i,j}$) a fim de ser válida a relação da Eq. 3.6.

Vale a pena notar que apesar de considerarmos todos os expoentes das leis alométricas $\beta_{i,j}^{(f)}$ independentes, eles são, na verdade, dependentes entre si. Em primeiro lugar, o expoente alométrico da relação entre Y_i e Y_j é supostamente o inverso daquele obtido para relação entre

Y_j e Y_i . Para os valores empíricos, observamos pequenos desvios nessa relação devido aos diferentes cortes aplicados no eixo das abscissas. Contudo, quando levamos em conta os intervalos de confiança dos valores de $\beta_{i,j}^{(f)}$, verificamos que a relação inversa entre os valores dos expoentes alométricos é satisfeita para a maioria dos casos. Em segundo lugar, quando temos que $Y_i \sim Y_k^{\beta_{i,k}}$ e $Y_j \sim Y_k^{\beta_{j,k}}$ segue direto que $Y_i \sim Y_j^{\beta_{i,j}}$ com $\beta_{i,j} = \beta_{i,k}/\beta_{j,k}$ (para alometrias sem ruídos). Esse resultado implica que as alometrias entre todos os pares de indicadores Y_i ($i \neq 0$) podem ser entendidas como uma consequência da relação alométrica entre o indicador Y_i e o tamanho da população. Contudo, o fato de que as relações alométricas possuem ruídos e que $P_i(Y_i)$ não é uma lei de potência perfeita pode ter um papel não trivial nos valores empíricos de $\beta_{i,j}$. Na Fig. 3.4(b), testamos empiricamente a relação $\beta_{i,j}^{(f)} = \beta_{i,0}^{(f)}/\beta_{j,0}^{(f)}$ e o resultado mostra que a relação é válida para as alometrias com $\rho_{i,j}$ maior que 0.5. Também, nesse caso, a qualidade da relação linear entre $\beta_{i,j}^{(f)}$ e $\beta_{i,0}^{(f)}/\beta_{j,0}^{(f)}$ é deteriorada a medida que os valores de $\rho_{i,j}$ diminuem.

Agora abordaremos a questão sobre as flutuações em torno das alometrias para verificar as duas hipóteses restantes dos cálculos de Gomez-Lievano *et al.*: a distribuição log-normal das flutuações e o comportamento constante do parâmetro log-normal $\sigma_{i,j}^2(Y_i)$. Para ter acesso a essas hipóteses, dividimos os valores das relações de alometria linearizadas ($\log_{10} Y_i$ versus $\log_{10} Y_j$ com $Y_j > Y_{j,\min}$) em w janelas igualmente espaçadas e, para cada uma, avaliamos a variância σ_{i,j_w}^2 . Note que, nessa escala log-log, o valor da variância σ_{i,j_w}^2 é aproximadamente igual ao valor do parâmetro da log-normal $\sigma_{i,j}^2(Y_i)$; portanto, verificar que σ_{i,j_w}^2 não depende de $\log_{10} Y_i$ é equivalente a mostrar que $\sigma_{i,j}^2(Y_i)$ é constante. A Fig. 3.5a mostra o comportamento de σ_{i,j_w}^2 pelo valor médio por janela de $\log_{10} Y_i$. Observamos que não há uma dependência clara nessas relações e o comportamento pode ser aproximado por uma função constante, em que o platô de valores muda de indicador para indicador. Finalmente, para estudar a hipótese log-normal, avaliamos as flutuações normalizadas em torno das relações alométricas linearizadas, isto é,

$$\xi_{i,j} = \frac{\log_{10} Y_j - \langle \log_{10} Y_j \rangle_w}{\sigma_{i,j_w}}, \quad (3.8)$$

em que $\langle \log_{10} Y_j \rangle_w$ representa o valor médio por janela de $\log_{10} Y_j$. A Fig. 3.5b mostra a distribuição cumulativa de $\xi_{i,j}$ para todos os nossos indicadores urbanos. Observamos que essas distribuições estão em boa concordância com a gaussiana padrão. Portanto, se $\xi_{i,j}$ é normalmente distribuído, $P_{i,j}(Y_i | Y_j)$ (que representa as flutuações na escala usual) deve seguir uma distribuição log-normal, confirmando a hipótese inicial e, também, conectando as flutuações em torno da alometria a um processo multiplicativo, assim como foi mostrado na seção 2.2 para alometrias com a população. Analogamente aos expoentes alométricos,

nem todas essas distribuições são independentes, visto que se conhecermos $P_{i,0}(Y_i | Y_0)$ e $P_{j,0}(Y_j | Y_0)$ podemos calcular $P_{i,j}(Y_i | Y_j)$.

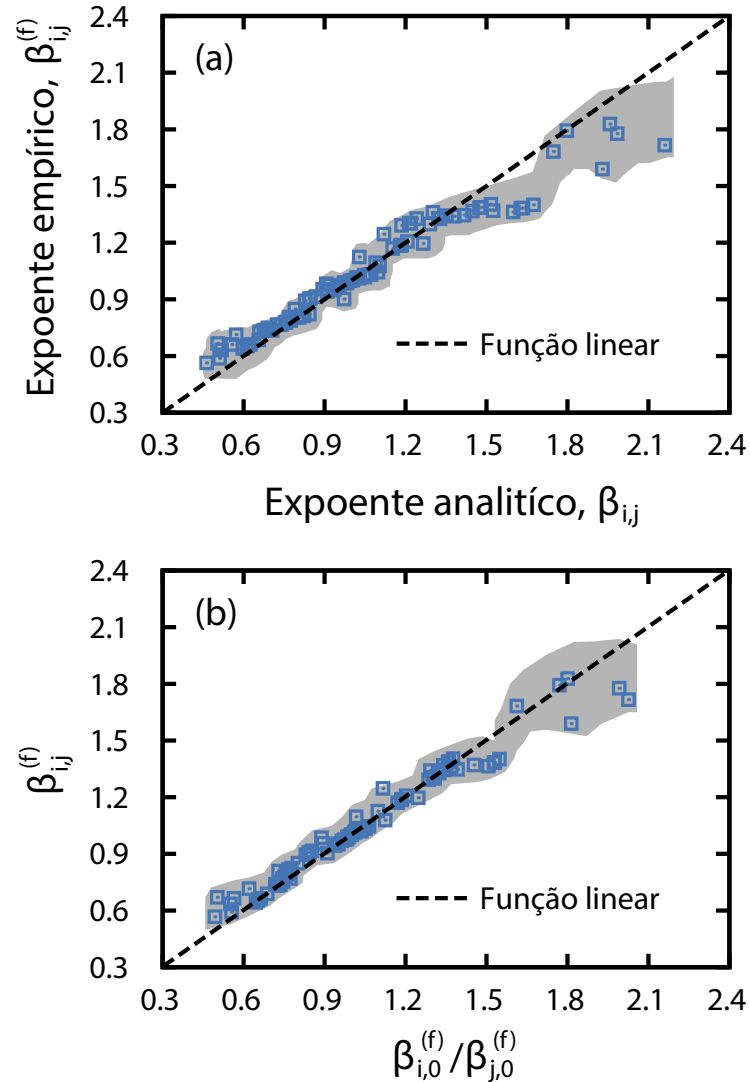


Figura 3.4: (a) Comparação entre os valores dos expoentes alométricos empíricos $\beta_{i,j}^{(f)}$ (obtidos via ajustes lineares) versus os valores analíticos de $\beta_{i,j}$ (obtidos a partir da Eq. 3.6). Os quadrados são os valores de $\beta_{i,j}^{(f)}$ versus $\beta_{i,j}$, em que o valor do coeficiente de correlação de Pearson $\rho_{i,j}$ caracterizando a alometria é maior que 0.5. (b) Verificação empírica da relação linear entre $\beta_{i,j}^{(f)}$ e $\beta_{i,0}^{(f)}/\beta_{j,0}^{(f)}$ (quadrados) quando consideramos $\rho_{i,j} > 0.5$. Em ambas figuras, a área sombreada representa o erro padrão de $\beta_{i,j}^{(f)}$ após suavizar os dados com um filtro de média móvel com janela de tamanho 5.

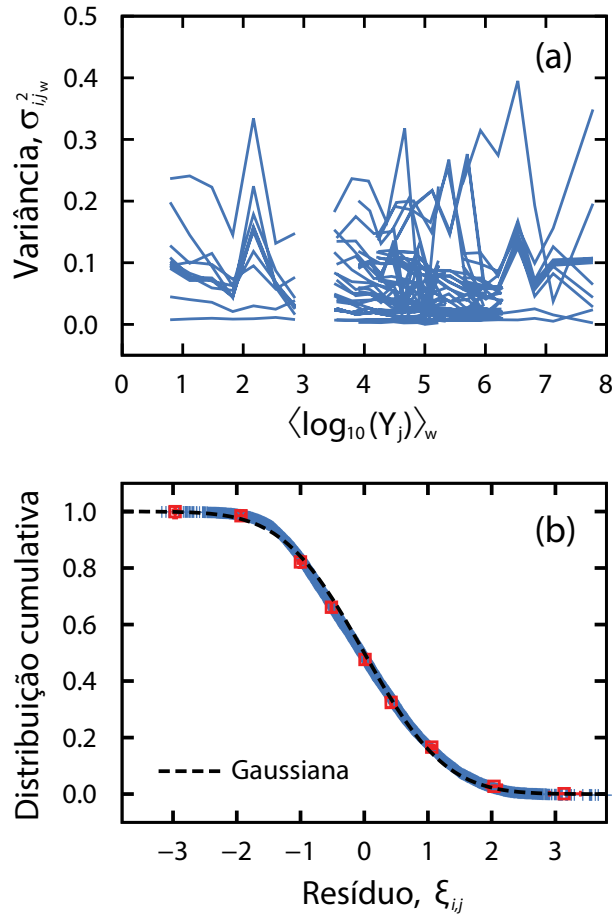


Figura 3.5: (a) Variâncias σ_{i,j_w}^2 das flutuações em torno das alometrias (em escala logarítmica base-10) para cada indicador urbano. Vale a pena notar que σ_{i,j_w}^2 não apresenta uma forma particular e pode ser aproximado por uma função constante. (b) Distribuição cumulativa das flutuações ξ em torno das alometrias normalizadas (cruzes). Os quadrados são os valores médios sobre todas as distribuições, as barras de erros representam o intervalo de confiança de 95% obtidos via *bootstrapping* [73] e a linha tracejada é uma distribuição gaussiana padrão (média zero e variância unitária). Em ambos gráficos, consideramos apenas as relações alométricas em que a correlação de Pearson $\rho_{i,j}$ é maior que 0.5.

3.3 Conclusões e perspectivas

Apresentamos uma caracterização extensiva da conexão entre distribuições lei de potência e alometrias considerando 12 indicadores urbanos de cidades brasileiras. Inicialmente, verificamos que esses 12 indicadores estão distribuídos como leis de potência via um método estatístico rigoroso. Os resultados mostram que a hipótese de lei de potência não pode ser rejeitada para todos os casos. Depois, revisamos os cálculos de Gomez-Lievano *et al.* [38] para

mostrar que, sob certas hipóteses, as distribuições lei de potência dos indicadores urbanos podem ser relacionadas às alometrias entre eles. Verificamos empiricamente as predições desses cálculos e, também, se as hipóteses necessárias são satisfeitas. Em particular, verificamos que a relação entre os expoentes das leis de potência e os expoentes das alometrias é satisfeita para a maioria dos casos e que uma condição adicional é a qualidade da alometria, medida aqui pelo coeficiente de Pearson. Também argumentamos que os expoentes alométricos não são independentes entre si, e que, na verdade, as alometrias entre os pares de indicadores (Y_i e Y_j com $i, j \neq 0$) podem ser entendidas como uma consequência da relação alométricas de cada indicador com a população (Y_0). Finalmente, confirmamos as duas hipóteses subjacentes nos cálculos de Gomez-Lievano *et al.*, isto é, o comportamento constante da variância e as distribuições log-normais das flutuações em torno da alometria. Portanto, acreditamos que nossas investigações empíricas contribuem para a consolidação da conexão entre distribuições lei de potência e alometrias e, também, revelam condições adicionais subjacentes a essas relações.

Visão geral

Nessa dissertação, analisamos dinâmica e padrões estatísticos em homicídios e métricas urbanas utilizando ferramentas de Física Estatística e de Sistemas Complexos. O estudo da criminalidade e métricas urbanas é estimulante por ser um assunto multidisciplinar e, também, por ser um conteúdo de grande interesse para sociedade contemporânea, que tem buscado, cada vez mais, soluções para problemas relacionados a criminalidade e a urbanização. Por essa razão, dedicamos ao entendimento e caracterização de padrões relacionados ao crime no sentido de tentar dar novas ideias sobre como é a dinâmica do crescimento de crimes e como os indicadores urbanos estão relacionados ao número de homicídios das cidades brasileiras.

Os principais resultados de nossas investigações podem ser resumidos como segue. No capítulo 1, encontramos que a distribuição das taxas de crescimento logarítmicas anual, bi-anual e trianual de homicídios exibe a mesma forma funcional. Verificamos um decaimento tipo lei de potência na relação entre o desvio padrão das taxas de crescimento logarítmica e o valor inicial (número) de atos criminosos. Também, argumentamos uma possível conexão com o quadro de organizações complexas. No capítulo 2, caracterizamos leis de escalas de crimes e métricas urbanas com o tamanho da população e mostramos que elas podem ser descritas por equações estocásticas. Também, argumentamos que é melhor empregar variáveis logarítmicas quando se utiliza análise de regressão linear. Além disso, propusemos uma nova abordagem para correlacionar o crime e métricas urbanas via o cálculo da distância entre o valor do número de homicídios (assim como o valor das métrica urbanas) e o valor que é esperado pela lei de escala com a população. Mostramos que esse método é robusto, revelando padrões que não haviam sido detectados na análise de regressão linear. No capítulo 3, caracterizamos extensivamente esses 12 indicadores urbanos das cidades brasileiras. Verificamos

que esses indicadores são assintoticamente distribuídos como leis de potência, que existem leis alométricas entre qualquer dois indicadores com alto grau de correlação e que os expoentes das alometrias são determinado a partir dos expoentes das distribuições lei de potência. Também, mostramos que as flutuações em torno das alometrias podem estar relacionadas a um processo multiplicativo subjacente, guiado por um ruído log-normal.

Esperamos que, com os resultados e as técnicas utilizados aqui, novos problemas possam ser abordados, tanto em criminalidade como em outros temas em sistemas complexos. Por exemplo, poderíamos considerar correlações espaciais nas séries temporais utilizadas no capítulo 1. A abordagem empregada no capítulo 2 poderia ser aplicada para estudar outros indicadores urbanos. Outra possibilidade seria empregar essa mesma técnica para o estudo de base de dados de outros países. Uma questão que ainda deveria ser verificada é se outras leis alométricas, tais como leis biológicas de alometria, apresentam um comportamento similar, isto é, se é possível relacionar a distribuição das medidas de tamanho com o surgimento de leis alométricas.

A.1 Método *bootstrapping*

Bootstrapping é um método estatístico de reamostragem proposto por Bradley Efron em 1979 [72] que pode ser usado para construir intervalos de confiança [73]. Suponha uma dada amostra aleatória $\mathcal{A} = \{x_1, x_2, x_3, \dots, x_n\}$, da qual se deseja obter alguma medida estatística $\hat{\theta}$, tal como a média ou a variância de \mathcal{A} . O procedimento para obter o intervalo de confiança da medida $\hat{\theta}$ pode ser feito, de acordo com [73], como mostrado a seguir.

Geramos subconjuntos da amostra \mathcal{A} em que cada elemento dessas subamostras são escolhidos aleatoriamente da série original, isto é, geramos uma série

$$\mathcal{B} = \{y_1, y_2, y_3, \dots, y_n\}, \quad (\text{A.1})$$

em que y_i é um elemento escolhido aleatoriamente a partir da amostra \mathcal{A} . Então, calculamos a medida estatística $\hat{\theta}$ desejada para cada subamostra resultante, ou seja, obtemos o seguinte conjunto de valores:

$$\Theta = \{\hat{\theta}[\mathcal{B}_1], \hat{\theta}[\mathcal{B}_2], \hat{\theta}[\mathcal{B}_3], \dots, \hat{\theta}[\mathcal{B}_n]\} \quad (\text{A.2})$$

do qual podemos obter o intervalo de confiança de \mathcal{A} , $(\hat{\theta}[\mathcal{A}]_{inferior})$ e $(\hat{\theta}[\mathcal{A}]_{superior})$, com nível de precisão desejado α , definidos como

$$(\hat{\theta}[\mathcal{A}]_{inferior}) = \hat{Q}_{\alpha/2}[\Theta], \quad (\text{A.3})$$

$$(\hat{\theta}[\mathcal{A}]_{superior}) = \hat{Q}_{1-\alpha/2}[\Theta], \quad (\text{A.4})$$

em que o \hat{Q}_β é o β -quantil da distribuição de probabilidade dos elementos de Θ . \hat{Q}_β retorna o valor de x tal que a probabilidade de encontrar um evento menor que x seja β . Por exemplo, $\hat{Q}_{1/2}[\mathcal{A}]$ retorna a mediana do conjunto \mathcal{A} .

A.2 Teste de hipótese de Kolmogorov-Smirnov

O teste de Kolmogorov-Smirnov (K-S teste) é um teste não paramétrico para igualdade de distribuições de probabilidades contínuas em que uma amostra $\mathcal{A} = \{x_1, x_2, x_3, \dots, x_n\}$ é comparada com uma distribuição de probabilidade referência $P_r(x)$ (por exemplo, a distribuição gaussiana) [81, 82]. A estatística de Kolmogorov-Smirnov quantifica a distância entre a função distribuição acumulada empírica e a função distribuição acumulada de referência $P_r(x)$. Em linguagem estatística, temos duas hipóteses:

$$H_0 : P(x) = P_r(x), \quad (\text{A.5})$$

$$H_1 : P(x) \neq P_r(x), \quad (\text{A.6})$$

em que H_0 é a hipótese nula e H_1 a hipótese alternativa.

Então, temos que definir a função de probabilidade empírica acumulada,

$$F_n(x) = P_n(X \leq x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x), \quad (\text{A.7})$$

em que a função I retorna 1 se $X_i \leq x$ e 0 caso contrário. Assim, o somatório representa o número de elementos da amostra \mathcal{A} que são menores que x . Já a função distribuição acumulada de referência é definida por,

$$F_r(x) = \int_{-\infty}^x P_r(y) dy, \quad (\text{A.8})$$

em que $P_r(y)$ é a função distribuição de referência. Devemos, então, calcular a distância entre $F_n(x)$ e $F_r(x)$,

$$D_n = \sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F_r(x)|, \quad (\text{A.9})$$

em que $\sup_{x \in \mathbb{R}}$ é o supremo do conjunto das distâncias. Então, temos a seguinte decisão a ser feita:

Não podemos rejeitar H_0 se $D_n \leq c_\alpha$;

Não podemos rejeitar H_1 se $D_n > c_\alpha$.

O valor de c_α depende da significância estatística α e pode ser encontrado em livros e tabelas estatísticas [83]. Os valores dessas tabelas estatísticas são encontrados resolvendo-se a equação

$$\alpha = 1 - K(c_\alpha), \quad (\text{A.10})$$

em que K é a distribuição acumulada de Kolmogorov definida por [84] como

$$\lim_{n \rightarrow \infty} P(\sqrt{n} D_n \leq c_\alpha) = K(c_\alpha) = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1} \exp(-2i^2 c_\alpha^2). \quad (\text{A.11})$$

Para termos 95% de probabilidade de que a hipótese esteja correta ($\alpha = 0.05$), por exemplo, teríamos

$$0.05 = 1 - K(c_{0.05}) \rightarrow c_{0.05} \approx 1.36. \quad (\text{A.12})$$

Também, é comum calcular o chamado p_{valor} (valor p ou p -valor), que é a probabilidade de uma distância estatística menor que a D_n ser observada, supondo que a hipótese nula seja verdadeira. O valor p pode ser calculado a partir da distribuição K (equação A.11) para o limite em que $n \rightarrow \infty$ [84]. Assim, podemos aproximar o valor p por $p_{\text{valor}} \approx K(D_n)$. O teste de Kolmogorov-Smirnov é considerado estatisticamente significativo para $p_{\text{valor}} > \alpha$.

A.3 Teste de hipótese de Cramér-von Mises

O teste de Cramér-von Mises (C-M) é um teste utilizado para julgar a qualidade do ajuste de uma distribuição de probabilidade de referência $P_r(x)$ em comparação a uma determinada distribuição empírica $P(x)$. A estatística de Cramér-von Mises quantifica a distância entre a função distribuição acumulada de probabilidade empírica e a função distribuição acumulada de referência $P_r(x)$. Em linguagem estatística, temos duas hipóteses

$$H_0 : P(x) = P_r(x), \quad (\text{A.13})$$

$$H_1 : P(x) \neq P_r(x), \quad (\text{A.14})$$

sendo H_0 a hipótese nula e H_1 a hipótese alternativa. Então, a partir da distribuição acumulada ($F_r(x) = \int_{-\infty}^x P_r(x) dx$), o teste de Cramér-von Mises verifica se uma dada amostra $\mathcal{A} = \{x_1, x_2, x_3, \dots, x_n\}$ segue uma distribuição de referência $P_r(x)$, baseando-se nos valores esperados para $[(F(x) - F_r(x))^2]$ em que $F(x)$ é a distribuição acumulada empírica e $F_r(x)$ é a distribuição acumulada de referência [106]. Para dados univariados ($\mathcal{A} = \{x_1, x_2, x_3, \dots, x_n\}$), o teste estatístico é dado por

$$T = \frac{1}{12n} + \sum_{i=1}^n \left(\frac{2i-1}{2n} - F(x_i) \right)^2. \quad (\text{A.15})$$

Se este valor T for maior que o valor tabelado [83], a hipótese de que os dados são distribuídos de acordo com $F_r(x)$ pode ser rejeitada.

A.4 Coeficiente de correlação de Pearson

Medidas de correlação são bastante comuns em estatística e uma das mais populares é o coeficiente de correlação de Pearson [85]. O coeficiente de correlação de Pearson r (ou r de Pearson) é uma medida do grau de linearidade entre duas variáveis quantitativas. O cálculo do coeficiente de correlação de Pearson entre duas variáveis é definido como a covariância das duas variáveis dividido pelo produto de seus desvio padrões. Matematicamente, sejam dois conjuntos de variáveis x_i e y_i , com $i \in \{1, 2, 3, \dots, n\}$, o r de Pearson fica

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (\text{A.16})$$

em que $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ e $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. Esse coeficiente varia entre os valores -1 e 1 . O valor 0 (zero) significa que não há relação linear entre as duas variáveis, o valor 1 indica uma relação linear perfeita e o valor -1 também indica uma relação linear perfeita, mas inversa, ou seja, quando uma das variáveis aumenta a outra diminui. Quanto mais próximo estiver de 1 ou -1 , maior é a associação linear entre as duas variáveis.

A.5 Método de Clauset para o ajuste de distribuições lei de potência

Clauset *et al.* [93] desenvolveram uma metodologia para estimar os parâmetros de ajustes da curva lei de potência. Uma função lei de potência pura pode ser escrita como

$$P(x) = C \left(\frac{x}{x_{min}} \right)^{-\alpha}, \quad (\text{A.17})$$

em que C é uma constante de normalização. Em linhas gerais, as regras propostas para a análise de dados empíricos que seguem uma distribuição lei de potência podem ser resumidas como segue: *i*) estima-se os parâmetros x_{min} e α via máxima-verossimilhança (*maximum-likelihood estimation*); *ii*) calcula-se a precisão do ajuste da lei de potência em descrever os dados empíricos via teste de Kolmogorov-Smirnov ou Cramér-von Mises.

Como no caso discreto, x assume apenas um conjunto de valores. Então, a distribuição deve divergir perto de zero, o que implica que deve existir um valor mínimo x_{min} em que se inicia o comportamento lei de potência. Assim, calculando a constante de normalização, a função de distribuição de probabilidade fica,

$$P(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{min})}, \quad (\text{A.18})$$

em que

$$\zeta(\alpha, x_{min}) = \sum_{n=0}^{\infty} (n + x_{min})^{-\alpha} \quad (\text{A.19})$$

é a função zeta generalizada ou função de Hurwitz.

Se tomarmos x_i , $i = 1, 2, \dots, n$, sendo o conjunto no qual se deseja saber a distribuição, por exemplo, o número de homicídios anuais de cada cidade. Então, podemos escrever a log-máxima-verossimilhança associada a distribuição A.18 como

$$\mathcal{L}(\alpha) = \sum_{i=1}^n \log \left(\frac{x_i^{-\alpha}}{\zeta(\alpha, x_{min})} \right) \quad (\text{A.20})$$

ou, equivalentemente,

$$\mathcal{L}(\alpha) = -n \log [\zeta(\alpha, x_{min})] - \alpha \sum_{i=1}^n \log [x_i]. \quad (\text{A.21})$$

A estimativa numérica de α é feita maximizando a função de máxima-verossimilhança para cada escolha de x_{min} , ou seja, resolvendo a equação em que $\partial \mathcal{L} / \partial \alpha = 0$. Feito isso, devemos usar o teste K-S ou C-M para verificar se a hipótese de que a lei de potência ajusta os dados empíricos é ou não rejeitada.

A.6 Cálculo de Gomez-Lievano *et al.*

Aqui, fornecemos detalhes adicionais do cálculo que resulta na equação 3.6. Podemos escrever $P_j(Y_j)$ em termos de $P_i(Y_i)$ e $P_{i,j}(Y_i|Y_j)$,

$$P_i(Y_i) = \sum_{Y_j=Y_j^*}^{\infty} P_{i,j}(Y_i | Y_j) P_j(Y_j) \quad (\text{A.22})$$

que, com a substituição das formas explícitas de $P_j(Y_j)$ e $P_{i,j}(Y_i|Y_j)$, fica

$$P_i(Y_i) = \sum_{Y_j=Y_j^*}^{\infty} \frac{1}{Y_i \sqrt{2\pi \sigma_{i,j}^2(Y_j)}} \exp \left\{ \frac{[\ln Y_i - \mu_{i,j}(Y_j)]^2}{2 \sigma_{i,j}^2(Y_j)} \right\} \frac{(Y_j)^{-\alpha_j}}{\zeta(\alpha_j, Y_{j,min})}. \quad (\text{A.23})$$

Substituindo a soma por uma integral e assumindo que Y_j^* é suficientemente pequeno tal que permita integrar sobre todo o intervalo dos números não negativos, obtemos

$$P_i(Y_i) \propto \int_0^\infty \frac{1}{Y_i} \exp \left[-\frac{1}{2\sigma_{i,j}^2} \left(\ln Y_i - \ln A_{i,j} Y_j^{\beta_{i,j}} \right) - (\alpha_j - 1) \ln Y_j \right] dY_j \quad (\text{A.24})$$

$$\propto \frac{1}{Y_i} \int_0^\infty \frac{1}{Y_j} \exp \left[-\frac{\beta_{i,j}^2}{2\sigma_{i,j}^2} \left(\ln^2 Y_j - 2 \ln Y_j \left(\frac{1}{\beta_{i,j}} \ln \frac{Y_i}{A_{i,j}} - \frac{\sigma_{i,j}^2}{\beta_{i,j}^2} (\alpha_j - 1) \right) + \frac{1}{\beta_{i,j}^2} \ln^2 \frac{Y_i}{A} \right) \right] dY_j. \quad (\text{A.25})$$

Completando o quadrado e rearranjando os termos, temos

$$P_i(Y_i) \propto \frac{\exp \left(-(\alpha_j - 1) \ln \left(\frac{Y_i}{A_{i,j}} \right)^{1/\beta_{i,j}} + \frac{\sigma_{i,j}^2}{2\beta_{i,j}^2} (\alpha_j - 1)^2 \right)}{Y_i} \times \int_0^\infty \frac{1}{Y_j} \exp \left[-\frac{\beta_{i,j}^2}{2\sigma_{i,j}^2} (\ln Y_j - f(Y_i; \theta))^2 \right] dY_j. \quad (\text{A.26})$$

Agora, notamos que o termo dentro da integral é uma distribuição log-normal, integrada sobre todo o seu domínio. Conseqüentemente, a integral é constante, independentemente da forma de $f(Y_i; \theta)$, em que θ representa os parâmetros $A_{i,j}$, $\beta_{i,j}$, $\sigma_{i,j}$ e α_j . Mantendo apenas os termos em Y_i , obtemos

$$P_i(Y_i) \propto \frac{\exp \left(-(\alpha_j - 1) \ln Y_i^{1/\beta_{i,j}} \right)}{Y_j} \propto Y_i^{-\frac{(\alpha_j - 1)}{\beta_{i,j}} - 1}, \quad (\text{A.27})$$

ou seja,

$$P_i(Y_i) \propto Y_i^{-\alpha_i}, \quad (\text{A.28})$$

em que $\alpha_i = \frac{(\alpha_j - 1)}{\beta_{i,j}} + 1$ ou

$$\beta_{i,j} = \frac{\alpha_j - 1}{\alpha_i - 1}. \quad (\text{A.29})$$

Definição das métricas urbanas, fontes dos dados e comentários adicionais

Obtivemos dados das cidades brasileiras no ano de 2000, disponibilizados gratuitamente pelo sistema público de saúde do Brasil — DATASUS [70]. Abaixo, descrevemos os indicadores e fornecemos alguns detalhes sobre os dados.

Homicídios: causado por lesões infligidas por outra pessoa com a intenção de ferir ou matar, por qualquer meio [88]. Esse indicador nos dá o número de mortes causadas por agressões. Seleccionamos as mortes no website do DATASUS cuja causa está inclusa nos códigos X85-Y09 da Classificação Internacional de Doenças (CID-10) [88].

População: esse indicador é derivado do censo populacional de 2000 realizado pelo IBGE (Instituto Brasileiro de Geografia e Estatística) [71] e reporta o número total de habitantes de cada cidade. Essa base de dados também contém informações sobre faixa etária e gênero.

Analfabetismo: fornece o número de habitantes da população total de uma dada área geográfica, no ano corrente, com 15 anos ou mais, que não podem ler ou escrever ao menos um simples bilhete no idioma que sabem.

Renda: esse indicador fornece a renda média domiciliar per capita dos moradores de uma determinada área geográfica no ano corrente. Foi considerado como renda domiciliar per capita a soma da renda mensal do domicílio, em reais, dividido pelo número de residentes.

Desemprego: fornece o número da população economicamente ativa com 16 anos ou mais que está a procura de trabalho durante a semana de referência, em uma dada área

geográfica no ano corrente. É definido como População Economicamente Ativa (PEA) o número de pessoas com 10 anos ou mais que está trabalhando ou procurando trabalho. Para esse indicador, foi considerado apenas a população com 16 anos ou mais.

Trabalho infantil: proporção da população de 10 a 15 anos que está trabalhando ou procurando por trabalho durante a semana de referência, em uma dada área geográfica no ano corrente.

PIB per capita: O indicador de Produto Interno Bruto (PIB) per capita é o valor do PIB per capita municipal, sendo calculado como o PIB municipal do ano dividido pela população no mesmo ano. Os valores são apresentados em moeda corrente (reais), não sendo aplicado deflator ou nenhum fator de correção.

PIB: fornece o valor do PIB municipal. Os valores são dados em milhares em moeda real corrente, não sendo aplicado deflator ou nenhum fator de correção.

População idosa: número de habitantes de uma dada cidade com 60 anos ou mais.

Saneamento: fornece o número de habitantes que possui acesso a instalações sanitárias, coleta de lixo e fornecimento de água.

Referências Bibliográficas

- [1] R. J. Gaylord, P. R. Wellin, *Computer Simulations with Mathematica: Explorations in Complex Physical and Biological Systems* (Springer, New York, 1994).
- [2] C. Castellano, S. Fortunato, V. Loreto, *Statistical physics of social dynamics*, *Reviews of Modern Physics* **81**, 591 (2009).
- [3] N. Boccaro, *Modeling Complex Systems* (Springer, New York, 2010).
- [4] L. A. N. Amaral, J. M. Ottino, *Augmenting the framework for the study of complex systems*, *The European Physical Journal B* **38**, 147 (2004).
- [5] R. Conte, N. Gilbert, G. Bonelli, C. Cioffi-Revilla, G. Deffuant, *et al. Manifesto of computational social science*, *The European Physical Journal Special Topics* **214**, 325 (2012).
- [6] M. C. González, A. O. Sousa, H. J. Herrmann, *Opinion formation on a deterministic pseudo-fractal network*, *International Journal of Modern Physics C* **15**, 45 (2004).
- [7] L. E. Araripe, R. N. Costa Filho, H. J. Herrmann, J. S. Andrade Jr., *Plurality voting: the statistical laws of democracy in Brazil*, *International Journal of Modern Physics C* **17**, 1809 (2006).
- [8] M. C. Mantovani, H. V. Ribeiro, M. V. Moro, S. Picoli Jr., R. S. Mendes, *Scaling laws and universality in the choice of election candidates*, *Europhysics Letters* **96**, 48001 (2011).

- [9] A. Chatterjee, M. Mitrović, S. Fortunato, *Universality in voting behavior: an empirical analysis*, Scientific Reports **3**, 1049 (2013).
- [10] H. D. Rozenfeld, D. Rybski, J. S. Andrade Jr., M. Batty, H. E. Stanley, H. A. Makse, *Laws of population growth*, Proceedings of the National Academy of Sciences U. S. A. **105**, 18702 (2008).
- [11] D. Rybski, S. V. Buldyrev, S. Havlin, F. Liljeros, H. A. Makse, *Scaling laws of human interaction activity*, Proceedings of the National Academy of Sciences U. S. A. **106**, 12640 (2009).
- [12] A. Hernando, R. Hernando, A. Plastino, A. R. Plastino, *The workings of the maximum entropy principle in collective human behaviour*, Journal of The Royal Society Interface **10**, 20120758 (2013).
- [13] A. Hernando, R. Hernando, A. Plastino, *Space-time correlations in urban sprawl*, Journal of The Royal Society Interface **11**, 20130930 (2014).
- [14] R. N. Mantegna, H. E. Stanley, *Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge, Cambridge University Press, 2007).
- [15] T. K. D. Peron, L. da F. Costa, F. A. Rodrigues, *The structure and resilience of financial market networks*, Chaos **22**, 013117 (2012).
- [16] S. Wichmann, *The emerging field of language dynamics*, Language and Linguistics Compass **2**, 442-455 (2008).
- [17] A. M. Petersen, J. Tenenbaum, S. Havlin, H. E. Stanley, *Statistical laws governing fluctuations in word use from word birth to word death*, Scientific Reports **2**, 313 (2012).
- [18] D. R. Amancio, O. N. Oliveira Jr., L da F. Costa, *Unveiling the relationship between complex networks metrics and word senses*, Europhysics Letters **98**, 18002 (2012).
- [19] R. W. Kates , T. M. Parris, *Long-term trends and a sustainability transition*, Proceedings of the National Academy of Sciences U. S. A. **100**, 8062 (2003).
- [20] M. B. Gordon, J. R. Iglesias, V. Semeshenko, J. P. Nadal, *Crime and punishment: the economic burden of impunity*, The European Physical Journal B **68**, 133 (2009).
- [21] J. R. Iglesias, V. Semeshenko, E. M. Scheneider, M. B. Gordon, *Crime and punishment: does it pay to punish?* Physica A **391**, 3942 (2012).

- [22] R. S. Mendes, L. C. Malacarne, R. P. B. Santos, H. V. Ribeiro, S. Picoli Jr., *Earthquake-like patterns of acoustic emission in crumpled plastic sheets*, EPL, 92, 29001 (2010).
- [23] H. Kawamura, T. Hatano, N. Kato, S. Biswas, B. K. Chakrabarti, *Statistical physics of fracture, friction, and earthquakes*, Reviews of Modern Physics, 84 (2012) 839.
- [24] H. V. Ribeiro, F. J. Antonio, L. G. A. Alves, E. K. Lenzi, R. S. Mendes, *Long-range spatial correlations and fluctuation statistics of lightning activity rates in Brazil*, EPL 104, 69001 (2013).
- [25] F. Havemann, M. Heinz, R. Wagner-Dobler, *Firm-like behaviour of journals? Scaling properties of their output and impact growth dynamics*, Journal of the American Society for Information Science and Technology **56**, 3 (2005).
- [26] S. Picoli Jr., R. S. Mendes, L. C. Malacarne, E. K. Lenzi, *Scaling behaviour in the dynamics of citations to scientific journals*, Europhysics Letters **75**, 673 (2006).
- [27] T. H. Keitt, L. A. N. Amaral, S. V. Buldyrev, H. E. Stanley, *Scaling in the growth of geographically subdivided populations: invariant patterns from a continent-wide biological survey*, Philosophical Transactions of the Royal Society of London B **357**, 627 (2002).
- [28] F. A. Labra, P. A. Marquet, F. Bozinovic, *Scaling metabolic rate fluctuations*, Proceedings of the National Academy of Sciences U. S. A. **104**, 10900 (2007).
- [29] M. H. R. Stanley, L. A. N. Amaral, S. V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger, H. E. Stanley, *Scaling behaviour in the growth of companies*, Nature **379**, 804 (1996).
- [30] L. A. N. Amaral, S. V. Buldyrev, S. Havlin, M. A. Salinger, H. E. Stanley, *Power law scaling for a system of interacting units with complex internal structure*, Physical Review Letters **80**, 1385 (1998).
- [31] D. Fu, F. Pammolli, S.V. Buldyrev, M. Riccaboni, K. Matia, K. Yamasaki, H. E. Stanley, *The growth of business firms: theoretical framework and empirical evidence*, Proceedings of the National Academy of Sciences U. S. A. **102**, 18801 (2005).
- [32] M. Ausloos, F. Petroni, *Statistical dynamics of religions and adherents*, Europhysics Letters **77**, 38002 (2007).

- [33] S. Picoli Jr., R. S. Mendes, *Universal features in the growth dynamics of religious activities*, Physical Review E **77**, 036105 (2008).
- [34] L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kuhnert, G. B. West, *Growth, innovation, scaling, and the pace of life in cities*, Proceedings of the National Academy of Sciences U. S. A. **104**, 7301 (2007).
- [35] L. M. A. Bettencourt, J. Lobo, G. B. West, *Why are large cities faster? Universal scaling and self-similarity in urban organizations and dynamics*, The European Physical Journal B **63**, 285 (2008).
- [36] L. M. A. Bettencourt, G. B. West, *A unified theory of urban living*, Nature **467**, 912 (2010).
- [37] L. M. A. Bettencourt, J. Lobo, D. Strumsky, G. B. West, *Urban scaling and its deviations: revealing the structure of wealth, innovation and crime across cities*, PLoS ONE **5**, e13541 (2010).
- [38] A. Gomez-Lievano, H. Youn, L. M. A. Bettencourt, *The statistics of urban scaling and their connections to Zipf's law*, PLoS ONE **7**, e40393 (2012).
- [39] G. S. Becker, *Crime and punishment: an economic approach*, Journal of Political Economy **76**, 169 (1968).
- [40] J. Crane, *The epidemic theory of ghettos and neighbourhood on dropping out and teenage childbearing*, The American Journal of Sociology **96**, 1226 (1991).
- [41] E. L. Glaeser, B. Sacerdote, J. A. Scheinkman, *Crime and social interactions*, Quarterly Journal of Economics **111**, 507 (1996).
- [42] I. Ehrlich, *Crime, punishment, and the market for offenses*, Journal of Economic Perspectives **10**, 43 (1996).
- [43] S. D. Levitt, *Using electoral cycles in police hiring to estimate the effect of police on crime*, The American Economic Review **87**, 270 (1997).
- [44] B. P. Kennedy, I. Kawachi, D. Prothrow-Stith, K. Lochner, V. Gupta, *Social capital, income inequality, and firearm violent crime*, Social Science & Medicine **47**, 7 (1998).

- [45] S. F. Messner, L. Anselin, R. D. Baller, D. F. Hawkins, G. Deane, S. E. Tolnay, *The spatial patterning of county homicide rates: an application of exploratory spatial data analysis*, Journal of Quantitative Criminology **15**, 423 (1999).
- [46] A. Blumstein, F. P. Rivara, R. Rosenfeld, *The rise and decline of homicide-and why*, Annual Review of Public Health **21**, 505 (2000).
- [47] L. Tedesco, *La ñata contra el vidrio: urban violence and democratic governability in Argentina*, Bulletin of Latin American Research **19**, 527 (2000).
- [48] S. D. Levitt, *Alternative strategies for identifying the link between unemployment and crime*, Journal of Quantitative Criminology **17**, 377 (2001).
- [49] K. Keizer, S. Lindenberg, L. Steg, *Study shows how degraded surrounding can degrade behaviour*, Science **322**, 1681 (2008).
- [50] A. Sachsida, M. J. C. Mendonça, P. R. A. Loureiro, M. B. S Gutierrez, *Inequality and criminality revisited: further evidence from Brazil*, Empirical Economics **39**, 93 (2010).
- [51] E. L. Glaeser, B. Sacerdote, *Why is there more crime in cities?*, Journal of Political Economy **107**, S225 (1999).
- [52] J. Crane, N. Boccara, K. Higdon, *The dynamics of street gang growth and policy response*, Journal of Policy Modeling **22**, 1 (2000).
- [53] W. Cook, P. Ormerod, E. Cooper, *Scaling behaviour in the number of criminal acts committed by individuals*, Journal of Statistical Mechanics: Theory and Experiment, P07003 (2004).
- [54] J. C. Nuno, M. A. Herrero, M. Primicerio, *A triangle model of criminality*, Physica A **387**, 2926 (2008).
- [55] M. B. Short, M. R. D'orsogna, V. B. Pasour, G. E. Tita, P. J. Brantingham, A. L. Bertozzi, L. B. Chayes, *A statistical model of criminal behaviour*, Mathematical Models and Methods in Applied Sciences **18**, 1249 (2008).
- [56] M. B. Gordon, *A random walk in the literature on criminality: A partial and critical view on some statistical analyses and modelling approaches*, European Journal of Applied Mathematics **21**, 283 (2010).

- [57] M. B. Short, A. L. Bertozzi, P. J. Brantingham, *Nonlinear patterns in urban crime: hotspots, bifurcations, and suppression*, The SIAM Journal on Applied Dynamical Systems **9**, 462 (2010).
- [58] M. B. Short, P. J. Brantingham, A. L. Bertozzi, G. E. Tita, *Dissipation and displacement of hotspots in reaction-diffusion models of crime*, Proceedings of the National Academy of Sciences U. S. A. **107**, 3961 (2010).
- [59] P. Crane, A. Kinzig, *Nature in the metropolis*, Science **308**, 1225 (2005).
- [60] I. Ehrlich, *The deterrent effect of capital punishment: a question of life and death*, The American Economic Review **65**, 397 (1975).
- [61] A. C. Poveda, *Violence and economic development in Colombian cities: a dynamic panel data analysis*, Journal of International Development **24**, 809 (2012).
- [62] J. R. Blau, P. M. Blau, *The cost of inequality: metropolitan structure and violent crime*, American Sociological Review **47**, 114 (1982).
- [63] W. Bailey, *Poverty, inequality and homicide rates*, Criminology **22**, 531 (1984).
- [64] M. Kelly, *Inequality and crime*, The Review of Economics and Statistics **82**, 530 (2000).
- [65] D. E. Hojman, *Explaining crime in Buenos Aires: the roles of inequality, unemployment, and structural change*, Bulletin of Latin American Research **21**, 121 (2002).
- [66] D. E. Hojman, *Inequality, unemployment and crime in Latin American cities*, Crime, Law & Social Change **41**, 33 (2004).
- [67] L. G. A. Alves, H. V. Ribeiro, R. S. Mendes, *Scaling laws in the dynamics of crime growth rate*, Physica A **392**, 2672 (2013).
- [68] L. G. A. Alves, H. V. Ribeiro, E. K. Lenzi, R. S. Mendes, *Distance to the scaling law: a useful approach for unveiling relationships between crime and urban metrics*, PLoS ONE **8**, e69580 (2013).
- [69] L. G. A. Alves, H. V. Ribeiro, E. K. Lenzi, R. S. Mendes, *Empirical analysis on the connection between power-law distributions and allometries for urban indicators* (Em preparação, 2014).

- [70] Sistema público de saúde do Brasil (SUS), Departameto de processamento de dados (DATASUS). Disponível em: <<http://www.datasus.gov.br/>>. Acesso em: abr. 2012, fev. e out. 2013.
- [71] Instituto Brasileiro de Geografia e Estatística. Disponível em: <<http://www.ibge.gov.br/>>. Acesso em: abr. 2012 e fev. 2013.
- [72] B. Efron, *Bootstrap methods: another look at the jackknife*, The annals of Statistics **7**, 1 (1979).
- [73] B. Efron, R. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993).
- [74] L. F. Richardson, *Atmospheric diffusion shown on a distance-neighbour graph*, The Proceedings of the Royal Society A **110**, 709 (1926).
- [75] R. S. Mendes, H. V. Ribeiro, F. C. M. Freire, A. A. Tateishi, E. K. Lenzi, *Universal patterns in sound amplitudes of songs and music genres*, Physical Review E **83**, 017101 (2011).
- [76] R. Gibrat, *Les Inégalités Economiques* (Sirey, Paris, 1931).
- [77] H. E. Stanley, *Scaling, universality, and renormalization: Three pillars of modern critical phenomena*, Reviews of Modern Physics **71**, S358 (1999).
- [78] World Resources Institute. Disponível em: <<http://www.wri.org>>. Acesso em: fev. e out. 2013.
- [79] B. Podobnik, D. Horvatic, D. Y. Kenett, H. E. Stanley, *The competitiveness versus the wealth of a country*, Scientific Reports **2**, 678 (2012).
- [80] S. Angel, C. S. Sheppard, D. L. Civco, P. Buckley, A. Chabaeva, L. Gitlin, A. Kralej, J. Parent, M. Perlin, *The Dynamics of Global Urban Expansion* (Washington DC, World Bank, 2005).
- [81] G. W. Corder, D. I. Foreman, *Nonparametric Statistics for non-Statisticians: a Step-by-Step Approach* (New Jersey, Wiley, 2009).
- [82] W. T. Eadie, D. Drijard, F. E. James, M. Roos, B. Sadoulet, *Statistical Methods in Experimental Physics* (North-Holland Publishing, Amsterdam, 1971).

- [83] E. S. Pearson, H. O. Hartley, *Biometrika Tables for Statisticians* (Cambridge University Press, Cambridge, 1972).
- [84] G. Marsaglia, W. W. Tsang, J. Wang, *Evaluating Kolmogorov's distribution*, Journal of Statistical Software **8**, 1 (2003).
- [85] R. V. Hogg, A. Craig, *Introduction to Mathematical Statistics* (Prentice Hall, New York, 1995).
- [86] R. Davidson, J. G. MacKinnon, *Estimation and Inference in Econometrics* (New York, Oxford University Press, 1993).
- [87] K. J. Preacher, A. F. Hayes, *Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models*, Behavior Research Methods **40**, 879 (2008).
- [88] World Health Organization. Disponível em: <<http://apps.who.int/classifications/icd10/>>. Acesso em: fev. 2013.
- [89] S. Fortunato, C. Castellano, *Word of mouth and universal voting behaviour in proportional elections*, Physical Review Letters. **99**, 138701 (2007)
- [90] M. Batty, *The size, scale, and shape of cities*, Science **319**, 769 (2008).
- [91] S. Arbesman, J. M. Kleinberg, W. H. Strogatz, *Superlinear scaling for innovation in cities*, Physical Review E **79**, 016115 (2009)
- [92] S. Arbesman, N. A. Christakis, *Scaling of prosocial behavior in cities*, Physica A **390**, 2155 (2011).
- [93] A. Clauset, C. R. Shalizi, M. E. J. Newman, *Power-law distributions in empirical data*, SIAM Review **51**, 661 (2009).
- [94] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D. U. Hwang, *Complex networks: Structure and dynamics*, Physics Reports **424**, 175 (2006).
- [95] S. Galam, *Sociophysics: A Physicist's Modeling of Psycho-political Phenomena* (Springer, Heidelberg, 2012).
- [96] A. Vespignani, *Modelling dynamical processes in complex socio-technical systems*, Nature Physics **8**, 32 (2012).

- [97] T. Yokoyama, T. Takahashi, *Mathematical neurolaw of crime and punishment: the q-exponential punishment function*, Applied Mathematics **4**, 1371 (2013).
- [98] Y. Malevergne, D. Sornette, *Extreme financial risks: from dependence to risk management* (Springer, Heidelberg, 2006).
- [99] M. Marsili, Y-C. Zhang, *Interacting individuals leading to Zipf's law*, Physical Review Letters **80**, 2741 (1998).
- [100] M. E. J. Newman, *Power laws, Pareto distributions and Zipf's law*, Contemporary Physics **46**, 323 (2005).
- [101] R. L. Axtell, *Zipf distribution of U.S. firm sizes*, Science **293**, 1818 (2001).
- [102] D. R. J. O'Neale, S. C. Hendy, *Power law distributions of patents as indicators of innovation*, PLoS ONE **7**, e49501 (2012).
- [103] M. C. Mantovani, H. V. Ribeiro, E. K. Lenzi, S. Picoli Jr., R. S. Mendes, *Engagement in the electoral processes: Scaling laws and the role of political positions*, Physical Review E **88**, 024802 (2013).
- [104] L. M. A. Bettencourt, *The origins of scaling in cities*, Science **340**, 1438 (2013).
- [105] A. Clauset, C. R. Shalizi, M. E. J. Newman, *Power-law distributions in empirical data*, SIAM Review **51**, 661 (2009).
- [106] T. W. Anderson, *On the distribution of the two-sample Cramér-von Mises criterion*, Annals of Mathematical Statistics **33**, 1148(1962).