

de Carvalho, Alexandre Xavier Ywata; da Silva, Gabriela Drummond Marques; de Almeida Júnior, Gilberto Rezende; Albuquerque, Pedro Henrique de Melo

Working Paper

Mapeamento de taxas bayesianas, com aplicação ao mapeamento do homicídios nos municípios brasileiros

Texto para Discussão, No. 1662

Provided in Cooperation with:

Institute of Applied Economic Research (ipea), Brasília

Suggested Citation: de Carvalho, Alexandre Xavier Ywata; da Silva, Gabriela Drummond Marques; de Almeida Júnior, Gilberto Rezende; Albuquerque, Pedro Henrique de Melo (2011) : Mapeamento de taxas bayesianas, com aplicação ao mapeamento do homicídios nos municípios brasileiros, Texto para Discussão, No. 1662, Instituto de Pesquisa Econômica Aplicada (IPEA), Brasília

This Version is available at:

<https://hdl.handle.net/10419/91085>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

1662

TEXTO PARA DISCUSSÃO

MAPEAMENTO DE TAXAS BAYESIANAS, COM APLICAÇÃO AO MAPEAMENTO DE HOMICÍDIOS

Alexandre Xavier Ywata de Carvalho
Gabriela Drummond Marques da Silva
Gilberto Rezende de Almeida Júnior
Pedro Henrique de Melo Albuquerque

Instituto de Pesquisa
Econômica Aplicada

MAPEAMENTO DE TAXAS BAYESIANAS, COM APLICAÇÃO AO MAPEAMENTO DE HOMICÍDIOS NOS MUNICÍPIOS BRASILEIROS

Alexandre Xavier Ywata de Carvalho*
Gabriela Drummond Marques da Silva**
Gilberto Rezende de Almeida Júnior**
Pedro Henrique de Melo Albuquerque***

* Coordenador da Assessoria de Métodos Quantitativos da Diretoria de Estudos e Políticas Regionais, Urbanas e Ambientais (Dirur) do Ipea.

** Técnico de Planejamento e Pesquisa da Dirur/Ipea.

*** Pesquisador visitante da Dirur/Ipea e professor do Departamento de Administração da Universidade de Brasília (UnB).

Governo Federal

Secretaria de Assuntos Estratégicos da
Presidência da República

Ministro Wellington Moreira Franco

ipea Instituto de Pesquisa
Econômica Aplicada

Fundação pública vinculada à Secretaria de Assuntos Estratégicos da Presidência da República, o Ipea fornece suporte técnico e institucional às ações governamentais – possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiro – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidente

Marcio Pochmann

Diretor de Desenvolvimento Institucional

Geová Parente Farias

Diretor de Estudos e Relações Econômicas e Políticas Internacionais, Substituto

Marcos Antonio Macedo Cintra

Diretor de Estudos e Políticas do Estado, das Instituições e da Democracia

Alexandre de Ávila Gomide

Diretora de Estudos e Políticas Macroeconômicas

Vanessa Petrelli de Correa

Diretor de Estudos e Políticas Regionais, Urbanas e Ambientais

Francisco de Assis Costa

Diretor de Estudos e Políticas Setoriais, de Inovação, Regulação e Infraestrutura, Substituto

Carlos Eduardo Fernandez da Silveira

Diretor de Estudos e Políticas Sociais

Jorge Abrahão de Castro

Chefe de Gabinete

Fabio de Sá e Silva

Assessor-chefe de Imprensa e Comunicação

Daniel Castro

URL: <http://www.ipea.gov.br>

Ouvidoria: <http://www.ipea.gov.br/ouvidoria>

Texto para Discussão

Publicação cujo objetivo é divulgar resultados de estudos direta ou indiretamente desenvolvidos pelo Ipea, os quais, por sua relevância, levam informações para profissionais especializados e estabelecem um espaço para sugestões.

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade do(s) autor(es), não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou da Secretaria de Assuntos Estratégicos da Presidência da República.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

ISSN 1415-4765

JEL: H30, H50, H63

SUMÁRIO

SINOPSE

ABSTRACT

| | |
|---------------------------------|----|
| 1 INTRODUÇÃO | 7 |
| 2 INFERÊNCIA CLÁSSICA | 9 |
| 3 MATRIZ DE VIZINHANÇA | 12 |
| 4 INFERÊNCIA BAYESIANA..... | 13 |
| 5 TAXA BAYESIANA EMPÍRICA..... | 14 |
| 6 RISCO RELATIVO BAYESIANO..... | 16 |
| 7 TAXA ESPACIAL (TE)..... | 18 |
| 8 APLICAÇÃO | 19 |
| 9 CONCLUSÃO | 32 |
| REFERÊNCIAS | 33 |

SINOPSE

O mapeamento da taxa de ocorrência de eventos é uma importante ferramenta para o direcionamento de políticas públicas. Apesar disso, o tema não tem recebido a devida atenção por parte da maioria dos pesquisadores. Em muitos trabalhos, os índices utilizados apresentam sérios problemas quando as unidades geográficas possuem uma pequena população sob risco e quando o número de casos observados é muito baixo, como visto em Pringle (1996). Este trabalho pretende fazer uma revisão sobre o uso da estatística bayesiana como possibilidade para minimização da variabilidade dos dados. A estatística bayesiana discutida é chamada empírica, pois utiliza informação amostral para composição da distribuição *a priori*. As taxas estudadas serão exemplificadas por meio da análise de ocorrência de homicídios por município, em todo o Brasil, no ano de 2008.

ABSTRACTⁱ

The mapping of frequency rates of events is an important tool for guiding public policy. Nevertheless, this problem has not received adequate attention by most researchers. In many studies, the used ratios have serious problems when the geographic units have a small population at risk and when the observed number of cases is very low, as seen in Pringle (1996). This paper intends to review the use of Bayesian statistics in order to minimize data variability. The Bayesian statistics discussed is called empirical because it uses information for the composition of the sample prior distribution. The studied rates will be exemplified by analyzing the occurrence of homicides by throughout Brazil in 2008.

ⁱ. The versions in English of the abstracts of this series have not been edited by Ipea's editorial department.
As versões em língua inglesa das sinopses (abstracts) desta coleção não são objeto de revisão pelo Editorial do Ipea.

1 INTRODUÇÃO

A análise de taxas espacialmente referenciadas é importante em diversas áreas do conhecimento. Na epidemiologia, por exemplo, tem-se o interesse de representar mapas de mortalidade por doenças que sejam, tanto quanto possível, fidedignos com a situação real enfrentada pela população. Em Castro, Vieira e Assunção (2004), por exemplo, o estudo da distribuição espacial de taxas é utilizado com o objetivo de direcionar políticas públicas voltadas para o combate ao câncer de pulmão. Para isso, os dados de quatro estados brasileiros (Rio Grande do Sul, Santa Catarina, Paraná e São Paulo) foram analisados no período de 1996 a 2000. Os resultados obtidos nesse trabalho mostraram que as maiores razões de mortalidade localizam-se aos extremos norte e sul da região estudada. Já para as mulheres, o sul do Paraná e do Rio Grande do Sul compõem as áreas com maior risco.

Nas análises criminais, a intervenção preventiva tem sido o foco do enfrentamento da violência. Essa abordagem envolve, portanto, a análise investigativa propiciada pelos mapas de ocorrência de delitos. Para isso, trabalhos como o de Beato Filho *et al.* (2001) foram desenvolvidos. Nesse artigo, os autores utilizam uma análise espacial para verificar a distribuição dos homicídios ocorridos no município de Belo Horizonte e registrados pela Polícia Militar (PM) de Minas Gerais durante o período de 1995 até 1999. Considerando todas as regiões da cidade de Belo Horizonte, apenas dez apresentaram risco maior de homicídios – quase todas essas regiões mais perigosas estavam concentradas em favelas.

Assim como nos trabalhos já mencionados, na maior parte dos estudos de mapeamento, as unidades geográficas utilizadas apresentam diversos tamanhos, o que pode ocasionar algumas dificuldades de análise. Quando os municípios são muito pequenos, por exemplo, a estimação de taxas brutas (TBs) é pouco indicada. Resumidamente, é possível dizer que o pequeno número de observações leva a estimativas pouco representativas e, por isso, com alta variabilidade. Clayton e Kaldor (1987) solucionaram esse problema por meio de uma suavização bayesiana. Além desse trabalho, Marshall (1991) propôs uma estimativa de taxas de alta estabilidade, que inclui a localização geográfica do município como informação adicional. Tais metodologias são chamadas de empíricas, uma vez que, em ambos os casos, os dados auxiliares são obtidos da própria amostra. Pringle (1996) faz uma revisão de técnicas bayesianas empíricas tendo como base estudos de caso de câncer e mortalidade neonatal.

As melhorias nos estudos de mapeamento não se restringiram à estimação de valores absolutos das taxas. Catelan, Lagazio e Biggeri (2010), por exemplo, sugerem a utilização de uma abordagem bayesiana hierárquica para a execução de múltiplos testes com os dados de riscos de doenças em uma determinada área. As técnicas bayesianas hierárquicas também podem ser uma alternativa para cálculo do intervalo de predição dos riscos relativos de municípios, como visto em Lee, Jang e Lee (2011). O conceito de risco relativo será discutido na seção 4.

Algumas áreas já vêm incorporando análises empíricas para aprimorar o mapeamento de taxas em seus estudos. Na Economia, um campo que faz grande uso de tais métodos é a Economia da saúde. Segundo Andrade *et al.* (2007) a Economia da saúde se preocupa com a “ampliação da eficiência do sistema de saúde (...) no que se refere à provisão de serviços (...), tanto pelo lado da demanda, por meio da identificação das necessidades da população quanto pelo lado da oferta (...)”.

Considerando esse tratamento econômico, Asiimwe *et al.* (2011) utilizam técnicas de mapeamento em pequenas áreas para direcionar intervenções na redução de mortalidade de crianças com idade abaixo de 5 anos em Uganda. Já Souza *et al.* (2001) se apropriaram dos métodos bayesianos empíricos na análise da distribuição espacial de taxas de hanseníase, com o objetivo de identificar áreas de possível sub-registro de casos, enquanto Cavalini e Leon (2007) fizeram uso da mesma técnica para correção de sub-registros de óbitos por causas mal definidas pelo Sistema de Informações Hospitalares (SIH) do Sistema Único de Saúde (SUS).

Para o cálculo de taxas mapeadas no espaço, diversos *softwares* estão atualmente disponíveis. O *software* estatístico *R* (www.r-project.org) possui uma grande variedade de funções já programadas, o que permite ao usuário o cálculo de diferentes estimados para taxas de algum evento estudado. Rotinas para o cálculo dos estimadores descritos neste artigo podem ser encontradas também no *software* GEODA (<http://geodacenter.asu.edu/>). O Instituto de Pesquisa Econômica Aplicada (Ipea) vem desenvolvendo uma nova versão do *software* IpeaGEO (www.ipea.gov.br/ipeageo), na qual serão incluídas rotinas para o cálculo de taxas bayesianas.

O objetivo deste trabalho é fornecer uma revisão da literatura de algumas das técnicas mais comuns para mapeamento de taxas e mostrar a aplicação dessa

metodologia em uma base de dados reais. Mais especificamente, apresentaremos uma aplicação dos métodos de estimação de taxas para calcularmos a taxa de homicídios nos municípios brasileiros. Na seção 2 apresentaremos os estimadores clássicos de taxas mais usualmente utilizadas. Para embasar a discussão nas seções seguintes, a seção 3 traz uma revisão sumarizada da matriz de vizinhança espacial, enquanto a seção 4 apresenta uma discussão sucinta da ideia dos estimadores bayesianos. Na seção 5 discutimos a taxa bayesiana empírica, utilizada para corrigir algumas das deficiências observadas nas estimativas clássicas. Na seção 6, apresenta-se o estimador proposto por Clayton e Kaldor (1987); nesse caso, o risco relativo bayesiano é discutido com especificação das *prioris* e *posteriors*. A seção 7 apresenta a estimação das taxas espaciais. A aplicação dos estimadores discutidos neste artigo ao problema de estimação das taxas de homicídios municipais é apresentada na seção 8. Finalmente, a seção 9 conclui o artigo.

2 INFERÊNCIA CLÁSSICA

A forma mais usual de mapeamento de taxas utiliza a chamada TB, dada pela expressão:

$$TB_i = \frac{Y_i}{P_i} \times 100.000, i = 1, \dots, n \quad (1)$$

sendo Y_i o número de casos observados, P_i o tamanho total da população sob risco e n o número de unidades geográficas (por exemplo, municípios) contidas no mapa. Por meio desse cálculo obtém-se o número esperado de eventos no município i ($i=1, \dots, n$) em uma população base de 100 mil pessoas (esse padrão é utilizado, por exemplo, para a apresentação de taxas de homicídios, conforme será apresentado no estudo de caso mais adiante, neste documento).

Além da TB, outra medida bastante utilizada é a razão de mortalidade padronizada – Standardized Mortality Ratio (SMR) –, expressa como:

$$SMR_i = \frac{Y_i}{E_i}, i = 1, \dots, n \quad (2)$$

considerando

$$E_i = \frac{\sum_{k=1}^n Y_k}{\sum_{k=1}^n P_k} P_i = m P_i \quad (3)$$

Sendo m a taxa global, o SMR constitui então uma comparação entre o número de ocorrências observado e o esperado sob a hipótese de que a taxa de ocorrência do evento seja constante ao longo do espaço. Assim, se o SMR de um município é 1,1, pode-se dizer que o risco de ocorrência do evento nessa localidade é 10% maior que o esperado, caso a informação espacial não fosse conhecida. Note que o SMR não se constitui em um estimador diferente na taxa de ocorrências de um determinado evento; trata-se simplesmente de uma forma relativizada de apresentação dos resultados.

Ao seguir a abordagem convencional do mapeamento de risco, é implícito dizer que o risco relativo da unidade geográfica i , θ_i , é um parâmetro desconhecido a ser estimado. Considerando as suposições em que Y_i segue um processo de contagem de Poisson, o estimador de máxima verossimilhança (EMV) de θ_i é o SMR, $\hat{\theta} = \frac{Y_i}{E_i}$. Entretanto, o conjunto de estimativas $\{\hat{\theta}_i, i = 1, \dots, N\}$ não é necessariamente o melhor estimador de θ_i , pois possui variabilidade alta e, com isso, baixa eficiência.

Com o tempo, o SMR foi aprimorado pela modificação do cálculo de E_i . Para isso, considerou-se que a taxa global é constante no espaço dado, entretanto, determinadas características intrínsecas à população, tais como gênero e idade, devem ser levadas em conta. Na razão de mortalidade, os valores esperados são redefinidos como:

$$E_i = \sum_{j=1}^z E_{ij} \quad (4)$$

sendo z o número total de categorias determinadas e

$$E_{ij} = \frac{\sum_{i=1}^n Y_{ij}}{\sum_{i=1}^n P_{ij}} P_{ij} \quad (5)$$

Essa medida fragmentada por grupos é muito utilizada na demografia. Ela é útil para corrigir medidas quando um município possui uma proporção maior de idosos, por exemplo, levando a taxas de mortalidade mais altas não necessariamente pelo efeito do evento de interesse.

Na maior parte das análises o interesse é que TB e SMR sejam utilizados simultaneamente. A TB é importante, pois permite estimar a taxa de ocorrência do evento em um município. No risco relativo, entretanto, o valor bruto se perde em benefício de uma medida que indica quanto o valor observado no município pode ser considerado baixo ou alto, de acordo com o cenário previsto para o mapa como um todo. Essa informação é de suma importância, embora sua interpretação não seja tão clara quanto a TB.

Para as situações em que o evento de interesse é raro, como número de homicídios (em países desenvolvidos) ou ocorrência de câncer, o mapeamento deve ser mais cuidadoso. Nesses casos, muitas vezes a taxa estimada pode ser igual a zero, embora seja razoável dizer que a probabilidade de que a taxa real seja exatamente igual a zero é muito baixa. A estimativa igual a zero pode ser resultado simplesmente de não termos uma janela temporal longa o suficiente, por exemplo.

Outro problema marcante nessas duas medidas ocorre quando a população sob risco, P_i , é muito pequena e, por isso, a variabilidade das estimativas se torna muito alta. Nesses casos, o SMR e a TB podem ser extremos no mapa e dominar o seu padrão. Como os valores obtidos são pouco confiáveis, técnicas de estimação em pequenas áreas (*small area estimation*) foram desenvolvidas. Um livro que aborda esse tema é o de Rao (2003).

Nas próximas seções veremos soluções providas da estatística bayesiana para enfrentar os desafios mencionados nos dois parágrafos anteriores. Conforme veremos, o fator espacial desempenha um papel importante na correção das estimativas. Por esse motivo, antes de prosseguir com o assunto sobre estimações bayesianas propriamente ditas, faremos uma breve discussão sobre o conceito de matriz de vizinhança, presente em diversos textos de análise de dados espaciais.

3 MATRIZ DE VIZINHANÇA

Em grande parte das técnicas de estatística espacial considera-se que as informações de um determinado polígono estão fortemente associadas aos dados de seus vizinhos. Para tanto, os vizinhos são estabelecidos como as regiões mais próximas ao polígono, de acordo com um determinado critério predefinido. Nesse caso, o critério geográfico é o mais comumente empregado. Para captar a informação da vizinhança de um território, pode-se utilizar a chamada matriz de vizinhança W . Considere que cada observação esteja associada a um polígono e a um sistema georreferenciado. O elemento W_j^i da matriz assume valor $W_j^i = 1$, caso os polígonos i e j sejam vizinhos, e $W_j^i = 0$, caso contrário. A diagonal principal de W possui todos os elementos iguais a zero, por definição.

Para identificar polígonos (municípios, setores censitários etc.) vizinhos, pode-se considerar uma vizinhança do tipo *queen*, quando os dois polígonos possuem pelo menos um vértice em comum, ou uma vizinhança do tipo *rook*, quando os polígonos possuem pelos menos um lado inteiro em comum. Note que a vizinhança do tipo *queen* é menos restritiva que a vizinhança do tipo *rook*. Além da vizinhança de primeira ordem, na qual se consideram vizinhos os municípios que compartilham limites geográficos, pode-se utilizar também vizinhanças de ordem maior. Na definição de vizinhança de segunda ordem, por exemplo, os polígonos i e j são vizinhos caso exista outro polígono k , para o qual i e k sejam vizinhos de primeira ordem, e j e k também sejam vizinhos de primeira ordem.

A matriz W , com elementos 0 ou 1, é conhecida como matriz de vizinhança binária não normalizada, em contraposição à matriz W^* normalizada. A matriz W^* normalizada é construída a partir da matriz W original (não normalizada), dividindo-se todos os elementos de cada linha de W pela soma da linha. Portanto, a matriz W^* possui todas as linhas com soma igual a 1. Por sua vez, a matriz W original é simétrica, o que não vale para a matriz W^* . A matriz de contiguidade neste trabalho foi referida simplesmente como W , independentemente de ser uma matriz normalizada ou não normalizada.

4 INFERÊNCIA BAYESIANA

Um teorema muito utilizado em estatística é o chamado teorema de Bayes que postula sobre probabilidades condicionais (BAYES, 1763). A expressão para o teorema de Bayes é:

$$P(B \setminus A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A \setminus B)P(B)}{P(A)} \quad (6)$$

Logo, a fórmula indica como a informação sobre a ocorrência de um determinado evento B muda quando dados sobre um evento A são fornecidos. Na expressão (6), $P(A)$ é a probabilidade de ocorrência de um determinado evento A ; $P(B)$, a probabilidade de ocorrência do evento B ; $P(A|B)$ a probabilidade condicional de o evento A ocorrer dado que B aconteça (similarmente para $P(B|A)$); e $P(A \cap B)$ é a probabilidade de ocorrência dos eventos A e B conjuntamente. O teorema de Bayes, aparentemente simples, é importante para a elaboração do conceito de estatística bayesiana.

A partir do teorema de Bayes, a abordagem na estatística bayesiana permite que os parâmetros do modelo sejam tratados não mais como valores fixos (e desconhecidos) da natureza. Considerados fixos na abordagem clássica, na abordagem bayesiana os parâmetros são considerados como variáveis aleatórias através da incorporação da incerteza. Um especialista, por exemplo, pode não conhecer o valor exato de um determinado parâmetro, mas pode ter alguma informação preliminar sobre ele. Com essa informação, pode-se construir a distribuição de probabilidade inicial que mostra em que valores há maior probabilidade de ocorrência do parâmetro.

Considere um experimento no qual uma amostra da variável X com função $f(x|\theta)$ é obtida. Chamamos distribuição *a priori* de θ a que descreve a nossa crença sobre o comportamento do parâmetro, antes de conhecermos os dados. Nesse caso, o teorema de Bayes é útil para utilizar a função de verossimilhança $f(x|\theta)$ para atualizar a informação sobre θ . Com base no teorema de Bayes, a informação atualizada sobre o parâmetro θ pode ser expressa como:¹

$$f(\theta \setminus x) = \frac{f(\theta)f(x \setminus \theta)}{\int f(\theta)f(x \setminus \theta)\partial x} \quad (7)$$

1. Para maiores detalhes sobre inferência bayesiana, o leitor pode recorrer a Gelman *et al.* (2004).

A função obtida ao final desse processo é a chamada distribuição *a posteriori* de θ . Logo, encontrar um estimador plausível do parâmetro, em estatística bayesiana, corresponde a fazer inferências da distribuição *a posteriori*. Um estimador pontual, construído a partir da distribuição *a posteriori*, muito utilizado, é a média *a posteriori*:

$$E(\theta | x) = \int \theta f(\theta | x) d\theta \quad (8)$$

Em algumas situações, entretanto, é importante fornecer a acurácia associada à estimativa pontual. Nesse caso, a melhor opção é utilizar o estimador intervalar, conhecido na estatística bayesiana como intervalo de credibilidade. Com ele é possível encontrar uma região $C_\alpha(x)$ com $100(1 - \alpha)\%$ de probabilidade de conter o verdadeiro valor do parâmetro. Para calculá-lo, basta encontrar os valores que satisfazem a equação:

$$\int_{C_\alpha(x)} f(\theta | x) d\theta = 1 - \alpha \quad (9)$$

5 TAXA BAYESIANA EMPÍRICA

A Taxa Bayesiana Empírica (proposta por MARSHALL, 1991) é construída considerando-se que o número de eventos observados no município i , Y_i segue uma distribuição de Poisson com média condicional:

$$E(Y_i | \Gamma_i) = P_i \Gamma_i \quad (10)$$

O EMV Γ_i é a taxa bruta TB_i que possui média Γ_i e variância não condicional Γ_i / P_i . Utilizando-se uma abordagem bayesiana, pode-se atribuir uma distribuição *a priori* para Γ_i com média $m_i = E_\Gamma(\Gamma_i)$ e variância $V_i = Var_\Gamma(\Gamma_i)$. Nesse caso, a média não condicional de TB_i será m_i e a variância não condicionada será igual a

$$Var_{TB}(TB_i) = V_i + \frac{m_i}{P_i} \quad (11)$$

Dados m_i e V_i , o melhor estimador linear para Γ_i considerando-se uma função de perda quadrática é o estimador

$$\hat{\Gamma}_i = w_i TB_i + (1 - w_i) m_i \quad (12)$$

$$w_i = \frac{V_i}{V_i + \frac{m_i}{P_i}} \quad (13)$$

Logo, w_i é a razão entre a variância *a priori* de Γ_i e a variância não condicional de TB_i . Uma possibilidade para a redução desse modelo é estabelecer $V = V_i$ e $m = m_i$. Considerando-se m como a taxa global da região, para encontrar uma estimativa para V podemos considerar a variância ponderada

$$S^2 = \frac{\sum_i P_i (TB_i - m)^2}{\sum_{j=1}^n P_j} \quad (14)$$

Sabe-se que $E_{TB} (TB_i - m)^2 = Var_{TB} (TB_i) = V + m/P_i$. Logo, $E_{TB} (S^2) = V + m/\bar{P}$ sendo que \bar{P} corresponde à média de populações em todas as unidades geográficas. Assim, o estimador de momentos de V , é $V = S^2 - m/\bar{P}$. Logo, esse método se utiliza dos próprios dados para construir *a priori*. Por esse motivo a taxa de Marshall (1991) é denominada empírica.

A Taxa Empírica Bayesiana de um município é então uma soma ponderada entre a sua TB e a taxa global da região em que o município está inserido. Tal soma possui fator de ponderação inversamente proporcional à população em risco P_i . Logo, quanto maior a população de um município, mais confiável é a sua estimativa via TB_i e por isso $\bar{\Gamma}_i$ será aproximadamente igual a TB_i . Quando, entretanto, a população é pequena, o valor estimado será aproximadamente igual à taxa global m .

Como pode ser visto, a estimativa proposta por Marshall (1991) diminui a variabilidade das estimativas ao restringir a flutuação aleatória ocasionada pelos eventos raros. Essa taxa, entretanto, pode ser ainda mais aprimorada com a inclusão de efeitos

espaciais de vizinhança em seus cálculos. Para que isso seja feito, um recurso possível é convergir os valores observados para uma média local, que será a taxa média dos vizinhos e do próprio município, m_i . Logo, ao modelo será adicionada uma suavidade espacial, condizente com a realidade da maior parte dos eventos cuja ocorrência tem forte influência de fatores ambientais. Em geral, a definição dos vizinhos se utiliza da matriz de vizinhança espacial, apresentada na seção 3.

6 RISCO RELATIVO BAYESIANO

A taxa empírica proposta por Clayton e Kaldor (1987) se destaca das demais, pois estabelece uma estrutura clara para a distribuição do risco relativo θ_i , utilizada como informação *a priori*. O resultado final é um modelo de misturas (*continuous mixtures*) para o qual se assume também que, condicional em θ_i , Y_i segue uma distribuição de Poisson com média θ_i , E_i . A partir dessas especificações é possível encontrar a distribuição marginal de Y_i que permite a estimação dos parâmetros de $f(\theta_i)$. Logo, esse modelo também é denominado empírico, pois os parâmetros da priori não são definidos pelo conhecimento anterior do pesquisador, e sim segundo a evidência dos próprios dados. A esperança *a posteriori* de θ_i dado Y_i é então a estimativa empírica bayesiana final. Em seu artigo, os autores definiram duas especificações diferentes para $f(\theta_i)$. Nas próximas seções, descreveremos a metodologia mais claramente, tendo cada uma dessas distribuições como ponto de partida.

6.1 O MODELO GAMA

Neste modelo, assume-se que as variáveis aleatórias θ_i são independentes e seguem uma distribuição gama com parâmetro de escala α e parâmetro de forma v ; ou seja, com média v/α e variância v/α^2 . Considerando a distribuição já mencionada de Y_i condicionada a θ_i , é fácil mostrar que a distribuição *a posteriori* é também uma distribuição gama com parâmetro de escala $(E_i + \alpha)$ e parâmetro de forma $(Y_i + v)$. Logo, a esperança *a posteriori* é:

$$E(\theta_i | Y_i, \alpha, v) = \frac{O_i + v}{E_i + \alpha} \quad (15)$$

Note que o valor encontrado irá variar entre o SMR (Y_i/E_i) e $\hat{v}/\hat{\alpha}$, a média estimada da distribuição de θ_i .

Resta ainda estimar os valores de v e α . Para isso, calcula-se a distribuição não condicional de Y_i , que é uma Binomial Negativa com:

$$E(Y_i) = \frac{E_i v}{\alpha} \quad (16)$$

$$Var(Y_i) = \frac{E_i v}{\alpha} + \frac{E_i^2 v}{\alpha^2} \quad (17)$$

A partir disso, os valores estimados correspondem àqueles encontrados por simulações que solucionam simultaneamente fórmulas vindas do EMV e o estimador de momentos. Juntas, essas fórmulas são utilizadas recursivamente, em um procedimento iterativo padrão, para cálculo das estimativas de v e α .

O modelo para a distribuição de θ_i pode ser estendido para a inclusão de covariáveis z_i por meio de um modelo log-linear. Essa extensão permite que diferentes parâmetros de escala α_i sejam obtidos de acordo com as características de cada unidade geográfica, representadas pelas variáveis explicativas. Nesse caso, a média *a priori* da distribuição de θ_i é

$$E(\theta_i) = \frac{v}{\alpha_i} = \exp(z_i' \Phi) \quad (18)$$

6.2 O MODELO LOG-NORMAL

Nesse modelo, considera-se que θ possui uma distribuição Log-Normal multivariada. Essa proposição é equivalente a supor que β , o vetor do logaritmo dos riscos relativos [$\theta_i = \exp(\beta_i)$], possui distribuição normal multivariada com média μ e matriz de covariância Σ . Com essa suposição a distribuição de θ_i condicional a Y_i não possui forma fechada. Por isso, utiliza-se uma aproximação quadrática para a verossimilhança de β , $\psi(\beta)$, obtendo-se então uma fórmula explícita para a estimativa empírica de θ_i . A

aproximação quadrática exige que $\psi(\beta)$ seja expandido em torno de um valor plausível β . O valor ideal nesse caso é o estimador não viciado do logaritmo do risco relativo:

$$\tilde{\beta}_i = \log\left(\frac{Y_i + 0.5}{E_i}\right) \quad (19)$$

No caso mais simples, em que os β_i 's são independentes, sua distribuição pode ser redefinida por dois parâmetros: $\phi = \mu_i$ e $\Sigma = \sigma^2 \mathbf{I}$. Com isso, a estimativa Bayesiana empírica é da forma:

$$b_i = \frac{\hat{\phi}' + (Y_i + 0.5)(\sigma^2)' \log\left[\frac{Y_i + 0.5}{E_i - (\sigma^2)' / 2}\right]}{1 + (Y_i + 0.5)(\sigma^2)'} \quad (20)$$

Os parâmetros ϕ' e $(\sigma^2)'$ por sua vez são encontrados com o uso do algoritmo EM (*Expectation-Maximization*) proposto por Dempster, Laird e Rubin (1977).

Pode-se observar que, assim como ocorreu com a estimativa do modelo Gama, b_i assume valores entre β_i , uma estimativa local, e ϕ , a média global da distribuição dos β_i 's. Além disso, como já visto em outras estimativas bayesianas, as medidas baseadas em grandes números são menos afetadas que aquelas baseadas em pequenos números. Na próxima seção, faremos uma discussão mais especificamente para medidas de risco utilizando-se informações das unidades geográficas na vizinhança.

7 TAXA ESPACIAL (TE)

Supondo que o espaço possui influência sobre o evento de interesse, espera-se que os valores dos riscos das unidades geográficas correspondam a uma superfície que varia suavemente ao longo do território. Na prática, entretanto, ruídos aleatórios dificultam a visualização dessa superfície. Esses ruídos aleatórios podem ser amenizados utilizando taxas que priorizem o efeito do espaço, de forma a obtermos estimativas mais acuradas das taxas implícitas a serem estimadas.

Para obter uma estimativa suavizada pode-se substituir o valor de cada localidade pela taxa encontrada para a região em que ela está localizada. Tal região é composta pelo próprio município e seus vizinhos. Considerando-se apenas os tipos de vizinhança do *queen* ou *rook* (conforme discutido na seção 3), todos os vizinhos possuem a mesma importância e, portanto, a observação estimada não precisa incluir o peso dos vizinhos.² Assim, a TE pode ser representada por:

$$TE_i = \sum_{j=1}^n \frac{w_{ij} Y_j}{w_{ij} P_j}, j = 1, \dots, n \quad (21)$$

sendo que a matriz W é uma modificação da matriz de vizinhança definida na seção 3, substituindo-se todos os elementos da diagonal principal por 1 (na versão original, W possui diagonal principal com elementos todos nulos). Além disso, quando existem taxas já calculadas, como as de Clayton e Kaldor (1987), pode-se substituir o valor obtido pela média da região em que o município se encontra:

$$be_i = \sum_{j=1}^n \frac{w_{ij} b_j}{n}, j = 1, \dots, n \quad (22)$$

8 APLICAÇÃO

Nesta seção, será analisada a distribuição dos homicídios nos 5.566 municípios brasileiros no ano de 2008. Todos os dados estão disponíveis na base de dados do Sistema de informação de mortalidade (SIM) do Datasus. Nesse sistema, optou-se por utilizar as siglas dos Códigos Internacionais de Doenças (CIDs) cujos óbitos foram motivados por causas que sugerem homicídios ou que foram comprovadamente ocasionadas por homicídios. Nesse grupo, enquadram-se as mortes por agressão, negligência e abandono, disparo de arma de fogo e contato com objeto cortante penetrante. Os CIDs relacionados são X85 a Y09 e Y22 a Y24 e Y28 e Y29 da CID-10, como utilizados em Castro, Assunção e Durante (2003). A seguir, é apresentada tabela com as medidas descritivas das taxas calculadas neste artigo, utilizando-se as diferentes metodologias discutidas acima.

2. Essa hipótese pode ser relaxada, através da consideração de médias ponderadas dos vizinhos, onde os pesos podem ser, por exemplo, inversamente proporcionais à distância d_{ij} entre o polígono i e os vizinhos j .

Pode-se concluir, por meio da tabela 1, que a média simples das TBs de homicídios no Brasil é de 15,70 óbitos por 100 mil habitantes. Comparado com outros países, essa é uma taxa elevada, uma vez que, nos Estados Unidos, o número de mortes por 100 mil habitantes foi de seis pessoas em 2006, enquanto na Argentina foram registrados apenas 5,8 homicídios em 2004, para uma população base de mesmo tamanho. A TB do Brasil, entretanto, pode estar sendo subestimada pelo grande número de municípios pequenos no país. Estas regiões podem não ter registrado nenhum homicídio, influenciando fortemente na média da TB e deslocando-a para valores mais baixos.

TABELA 1
Brasil: estatísticas descritivas – taxas de homicídio por 100 mil habitantes - 2008

| | Média | Mediana | Desvio-padrão | Mínimo | Máximo | CV* |
|----------------------------------|-------|---------|---------------|--------|--------|------|
| TB | 15,70 | 10,70 | 18,74 | 0,00 | 211,10 | 1,19 |
| Taxa Bayesiana Empírica | 21,47 | 18,80 | 11,54 | 2,50 | 125,90 | 0,54 |
| TE | 18,93 | 14,85 | 14,11 | 0,00 | 109,60 | 0,75 |
| Taxa Bayesiana Empírica Espacial | 17,10 | 13,70 | 16,37 | 0,00 | 146,80 | 0,96 |
| Taxa Bayesiana Clayton e Kaldor | 17,88 | 14,30 | 11,42 | 2,10 | 120,50 | 0,64 |

Fonte: Elaboração dos autores com dados do Datasus.

* CV = coeficiente de variação.

Obs.: Taxas por 100 mil habitantes.

A tabela 2 apresenta os menores municípios brasileiros e exemplifica o que foi comentado. Em grande parte dessas regiões, a TB de homicídios registrada foi zero. Esses valores, entretanto, precisam ser analisados com cautela, uma vez que as observações extremas podem ser fruto da variabilidade aleatória presente nos dados. Além disso, o sub-registro nas informações do SIM-Datasus, relatado em Waiselfisz (2010), é outro fator que deve ser levado em conta. Nesse trabalho, o autor destacou que em pequenos municípios, localizados em regiões onde é difícil o acesso aos serviços públicos, existem casos de sepultamentos feitos sem a identificação detalhada das causas do óbito.

TABELA 2
Brasil: indicadores dos 20 menores e dos 20 maiores municípios – 2008

| Município | População | Homicídios | TB | Taxa Bayesiana Empírica | TE | Taxa Bayesiana Empírica Espacial | Taxa Bayesiana Clayton e Kaldor |
|-----------------------------|------------|------------|------|-------------------------|------|----------------------------------|---------------------------------|
| Borá (SP) | 834 | 0 | 0 | 24,1 | 10,6 | 0 | 16,4 |
| Serra da Saudade (MG) | 889 | 0 | 0 | 24 | 12,8 | 0 | 16,3 |
| Anhanguera (GO) | 1.007 | 0 | 0 | 23,6 | 18,3 | 0 | 16,1 |
| Nova Castilho (SP) | 1.111 | 0 | 0 | 23,4 | 0 | 0 | 15,9 |
| Oliveira de Fátima (TO) | 1.114 | 0 | 0 | 23,4 | 16 | 0 | 15,9 |
| Araguainha (TO) | 1.138 | 0 | 0 | 23,3 | 21,2 | 0 | 15,9 |
| Miguel Leão (PI) | 1.228 | 0 | 0 | 23,1 | 2,5 | 2,1 | 15,8 |
| Cedro do Abaeté (MG) | 1.228 | 0 | 0 | 23,1 | 17,5 | 0 | 15,8 |
| André da Rocha (RS) | 1.262 | 1 | 79,2 | 34 | 9 | 15,8 | 25,4 |
| Chapada de Areia (TO) | 1.274 | 0 | 0 | 23 | 14,4 | 0 | 15,7 |
| Parari (PB) | 1.283 | 0 | 0 | 22,9 | 9,2 | 7,8 | 15,7 |
| Crixás do Tocantins (TO) | 1.299 | 0 | 0 | 22,9 | 25,6 | 0 | 15,6 |
| Lagoa Santa (GO) | 1.308 | 0 | 0 | 22,9 | 17 | 0 | 15,6 |
| Uru (SP) | 1.397 | 0 | 0 | 22,6 | 7,4 | 0 | 15,5 |
| Serra Nova Dourada (MT) | 1.407 | 0 | 0 | 22,6 | 20,3 | 14 | 15,5 |
| Nova Aliança do Ivaí (PR) | 1.426 | 0 | 0 | 22,6 | 16,6 | 0 | 15,5 |
| Santa Salete (SP) | 1.441 | 0 | 0 | 22,5 | 3,8 | 3,4 | 15,4 |
| Cachoeira de Goiás (GO) | 1.441 | 0 | 0 | 22,5 | 23,7 | 0 | 15,4 |
| Tupirama (TO) | 1.448 | 0 | 0 | 22,5 | 18,1 | 0 | 15,4 |
| São Félix do Tocantins (TO) | 1.451 | 0 | 0 | 22,5 | 4,7 | 0 | 15,4 |
| Natal | 798.065 | 324 | 40,6 | 40,5 | 39,6 | 40,6 | 40,3 |
| São Bernardo do Campo (SP) | 801.580 | 121 | 15,1 | 15,2 | 16,2 | 15,2 | 15,1 |
| Nova Iguaçu (RJ) | 855.500 | 302 | 35,3 | 35,2 | 34,2 | 35,2 | 35,1 |
| Duque de Caxias (RJ) | 864.392 | 485 | 56,1 | 55,8 | 33,9 | 55,2 | 55,7 |
| Maceió | 924.143 | 933 | 101 | 100,3 | 93,9 | 100,6 | 100,1 |
| São Gonçalo (RJ) | 982.832 | 367 | 37,3 | 37,3 | 35,5 | 37,2 | 37,2 |
| São Luís | 986.826 | 360 | 36,5 | 36,4 | 33,8 | 36,2 | 36,3 |
| Campinas (SP) | 1.056.644 | 156 | 14,8 | 14,9 | 14,2 | 14,7 | 14,8 |
| Goiânia | 1.265.394 | 458 | 36,2 | 36,1 | 35,2 | 36,1 | 36,1 |
| Guarulhos (SP) | 1.279.202 | 362 | 28,3 | 28,3 | 17,2 | 27,5 | 28,2 |
| Belém | 1.424.124 | 742 | 52,1 | 52 | 51,8 | 52,1 | 51,9 |
| Porto Alegre | 1.430.220 | 570 | 39,9 | 39,8 | 43 | 40 | 39,7 |
| Recife | 1.549.980 | 976 | 63 | 62,8 | 62,3 | 62,9 | 62,7 |
| Manaus | 1.709.010 | 648 | 37,9 | 37,9 | 35,4 | 37,8 | 37,8 |
| Curitiba | 1.828.092 | 774 | 42,3 | 42,3 | 48 | 42,4 | 42,2 |
| Belo Horizonte | 2.434.642 | 879 | 36,1 | 36,1 | 38 | 36,2 | 36 |
| Fortaleza | 2.473.614 | 874 | 35,3 | 35,3 | 35,6 | 35,4 | 35,3 |
| Brasília (DF) | 2.557.158 | 814 | 31,8 | 31,8 | 38,1 | 31,9 | 31,8 |
| Salvador | 2.948.733 | 1.878 | 63,7 | 63,6 | 65,2 | 63,7 | 63,5 |
| Rio de Janeiro | 6.161.047 | 1.821 | 29,6 | 29,6 | 34,2 | 29,6 | 29,5 |
| São Paulo | 10.990.249 | 1.701 | 15,5 | 15,5 | 18,7 | 15,5 | 15,5 |

Fonte: Elaboração dos autores com dados do Datasus.

Obs.: Taxas por 100 mil habitantes

Diante desses problemas, uma alternativa para tornar as estimativas mais apuradas é o aumento do número de dados. Assim, as taxas de um município são corrigidas por meio das informações obtidas no seu entorno, como feito nas estimativas bayesianas e espaciais. Com essas taxas, têm-se evidências de que a TB está subestimada, uma vez que, em todas as demais taxas, as médias encontradas foram superiores a 17 homicídios por 100 mil habitantes. Tendo a mediana como foco de análise, as diferenças nas estimativas tornam-se ainda mais expressivas. Nesse caso, pelo menos 50% dos municípios brasileiros possuem uma TB igual ou inferior a 10,70 homicídios por 100 mil habitantes. Considerando-se as outras taxas, entretanto, diz-se que ao menos metade dos municípios do país implicou um número de óbitos, por 100 mil habitantes, igual ou menor que 14.

Observando-se as medidas já comentadas, as taxas corrigidas produzem resultados muito parecidos, excluindo-se apenas a Taxa Bayesiana Empírica. A não concordância, nesse caso, indica que não basta simplesmente utilizar as suavizações sugeridas. Antes de tudo é necessário identificar qual das estimativas se adapta melhor ao cenário analisado. No caso do Brasil, por exemplo, a Taxa Bayesiana Empírica é pouco útil, pois utiliza a média geral dos dados. Como o país é extremamente diversificado, a correção das taxas por uma única média estática pode não ser o mais adequado.

Os problemas de estimação da TB não se restringem ao sub-registro. Em regiões com um número pequeno de dados, outro fenômeno observado é a superestimação. Como pode ser visto, a TB é a única que qualificou um dos municípios do país como uma região de extremo risco de homicídios, tendo uma taxa de 211,10 óbitos estimados para cada 100 mil habitantes. Esse valor, entretanto, não é intuitivamente aceitável e deve, portanto, ser objeto de um estudo mais amplo. Pode-se supor que, devido às condições sociais, econômicas e ambientais, um município com altos índices de violência esteja localizado em uma região igualmente crítica. Nesse caso, as taxas corrigidas também iriam acusar a alta periculosidade do município, o que não ocorreu. Em nenhuma das demais taxas o valor máximo superou o número de 150 mortes por 100 mil habitantes. O município dito o mais perigoso do país pela estimação obtida com a TB registrou, na verdade, cinco homicídios. Trata-se do distrito de Nova Marilândia, localizado em Mato Grosso, cuja população é de apenas 2.369 pessoas. Segundo a Taxa Clayton e Kaldor, entretanto, a estimativa dessa região é reduzida para 57,8 mortes por 100 mil habitantes, enquanto, de acordo com a TE, esse número é de 35,7 óbitos por 100 mil habitantes.

Logo, os dois problemas mencionados, de super e de subestimação, podem ser resumidos em único desafio, a baixa precisão da TB. Observando-se o desvio-padrão, já é possível obter indícios de que a variabilidade desse estimador é bem maior do que a variabilidade dos demais. Essa evidência é confirmada por meio do coeficiente de variação. A TB é a única estimativa em que o coeficiente de variação é maior do que 1,0. Como o coeficiente de variação é o desvio-padrão dividido pela média, esta é a única taxa em que a variabilidade é maior até do que a estimativa pontual.

Outra característica importante a ser avaliada é a simetria dos dados. Essa informação pode ser captada avaliando a posição da mediana em relação à média. Variáveis como as observadas, em que a mediana é inferior à média, indicam que as observações possuem simetria positiva. Isso sugere a presença de municípios com taxas discrepantes, nos quais os índices de homicídios são muito superiores às taxas dos demais. Logo, existem regiões no país que necessitam de atenção específica. Locais em que a intervenção pública deve ser executada como medida de emergência.

Após a análise geral, é importante entender o comportamento dos homicídios em um nível mais desagregado. Nas análises anteriores observou-se uma forte evidência de que o tamanho da população está relacionado ao valor das taxas. Nesse caso, é interessante averiguar as particularidades de cada uma das taxas nos maiores e menores municípios do país. Diante disso, a tabela 2 apresenta os indicadores utilizados nesse trabalho, calculados para os 20 menores e os 20 maiores municípios do Brasil, respectivamente.

Por meio da tabela 2, é possível observar que a maior parte dos pequenos municípios possui uma TB igual a zero. Seria errado, entretanto, inferir que estas são regiões completamente isentas de risco de homicídios. É razoável pensar que, mesmo que a chance seja pequena, ela não é exatamente igual a zero e, como as populações são menores, a estimação do verdadeiro valor é prejudicada. Assim, as correções bayesianas e espacial apresentam uma taxa não nula para quase todos os municípios do grupo. É importante destacar, contudo, que apenas a Taxa Bayesiana Empírica Espacial manteve taxas iguais a zero para alguns municípios. Isso ocorre porque a suavização por esse método é feita utilizando-se uma média local. Caso um município esteja localizado em uma região pouco populosa, sem ocorrência de homicídios, a média local pode ser igual a zero. Nessas situações, caso seja do interesse do pesquisador, é possível utilizar vizinhanças de ordem maior do que um, para as quais a suavização será mais acentuada.

A superestimação da TB pode ser vista por meio dos resultados para o município de André da Rocha. Esse município, que é o nono menor do país, registrou apenas um homicídio, mas, apesar disso, apresentou uma taxa de 80 homicídios por 100 mil habitantes. Trata-se, portanto, de uma região cujos indicadores precisam ser obtidos pelos métodos de correção. O número de homicídios encontrados utilizando-se, por exemplo, as taxas com suavização espacial, reduz-se para menos de um quarto da TB.

A menor redução observada com relação à TB ocorreu para a Taxa Bayesiana Empírica. Isso pode ser explicado, uma vez que essa taxa possui um método de suavização mais conservador. Como pode ser observado, a Taxa Bayesiana Empírica fez com que os indicadores dos menores municípios convergissem para um valor próximo à TB do Brasil, que é de 26,6 homicídios por 100 mil habitantes. Todos os resultados, entretanto, foram menores do que 26,6, pois a Taxa Bayesiana Empírica corresponde a uma média ponderada entre a TB do país e a do próprio município.

Para os municípios maiores, percebe-se que a TE é a única que apresenta valores destoantes das demais taxas. Isto pode ser visto, por exemplo, para municípios como Duque de Caxias, Guarulhos, Curitiba, Brasília e Rio de Janeiro. Como a TE é apenas uma média não ponderada do valor dos vizinhos e do próprio município, a taxa dos vizinhos possui uma influência maior do que nos demais métodos de suavização. Isto pode causar problemas caso um município com grande população seja vizinho de uma região pouco populosa. Nesse caso, se nos municípios vizinhos não houver ocorrência do evento em estudo, a taxa dos grandes municípios apresentará uma redução substancial. Esse fenômeno pode ser exemplificado com o caso de Guarulhos, em que a TE difere em quase 10 óbitos por 100 mil habitantes das demais, influenciada pelo município vizinho, São Paulo.

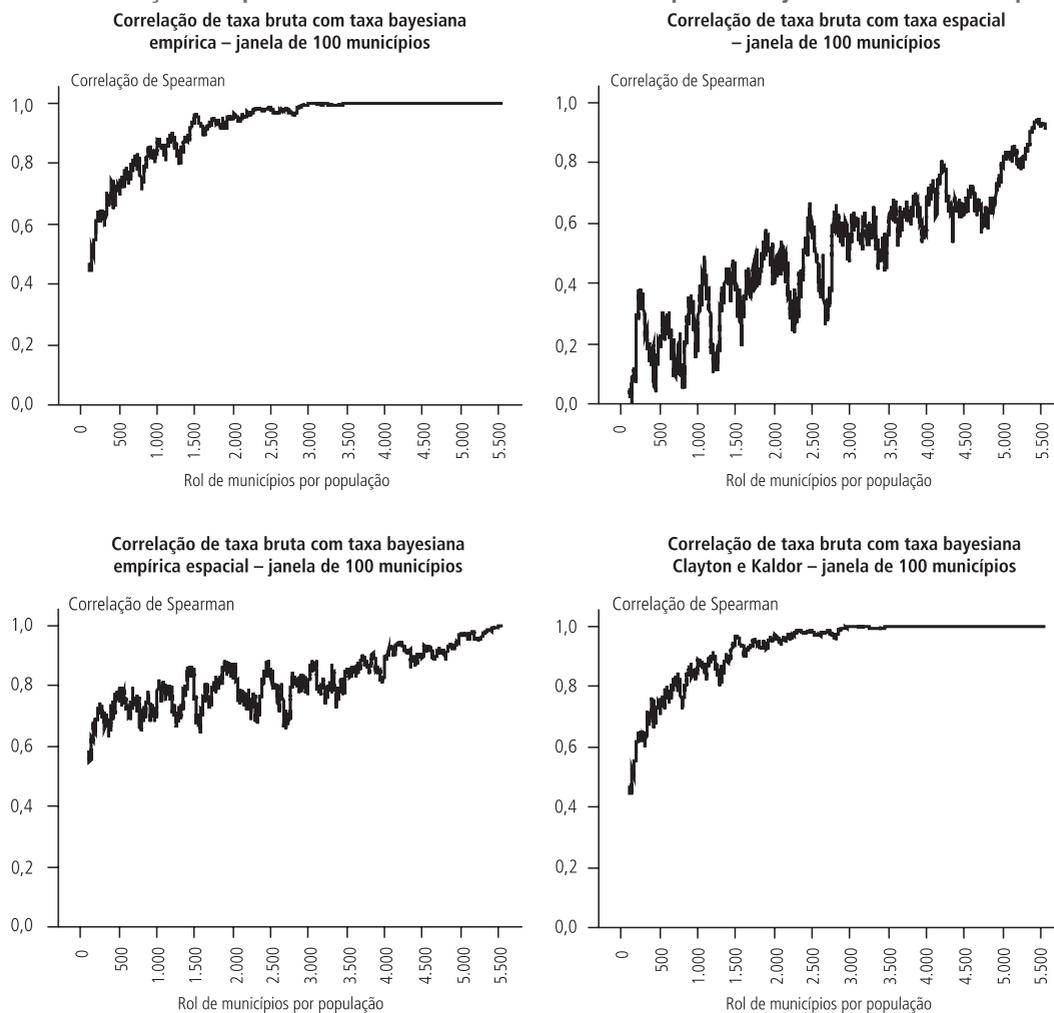
Dentre os 20 maiores municípios listados na tabela 2, mesmo após a utilização dos métodos de correção, quase todas as taxas encontradas obtiveram valor elevado. Destaque para o município de Maceió, cujas taxas suavizadas apresentaram valores superiores a 100 óbitos por 100 mil habitantes. Logo, o cenário de violência é realidade em muitas das grandes metrópoles brasileiras, que devem ser alvo da intervenção de políticas públicas.

É possível verificar que as TE e as bayesianas são importantes para o melhoramento das taxas. Esse melhoramento, entretanto, possui intensidade de suavização direcionada

pelo tamanho da população. É interessante então quantificar a relação entre o tamanho do município e a quantidade de correção da TB. Para isso, foi realizado um ensaio apresentado no gráfico 1. Cada gráfico representa a correlação de *Spearman* entre cada uma das taxas corrigidas e a TB, em uma janela de 100 municípios.

GRÁFICO 1

Correlação de Spearman entre a TB e as demais taxas para uma janela de 100 municípios



Fonte: Elaboração dos autores com base nos dados do Datasus.

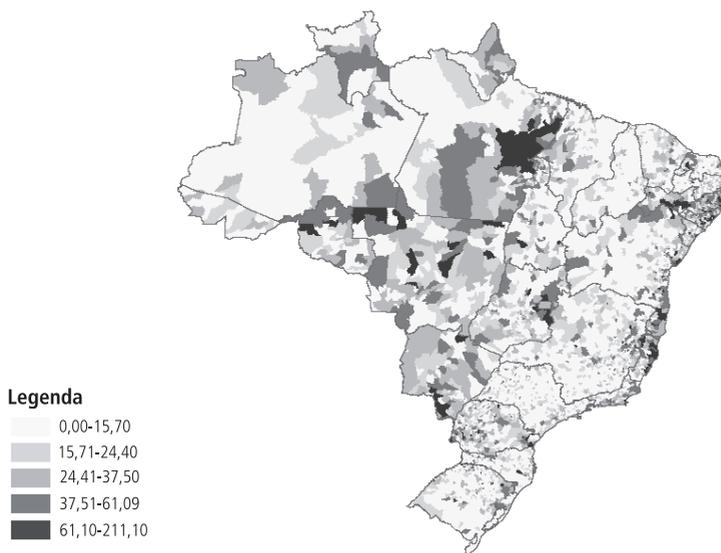
Segundo o gráfico 1, a atuação das correções na TB em municípios de população pequena é maior do que nos grandes municípios. Isso pode ser observado ao verificar o padrão crescente nos quatro gráficos. Embora a tendência de crescimento seja observada em todos os quatro casos, existem dois comportamentos diferentes quanto

à intensidade. A relação entre a TB e as taxas que possuem um componente espacial, apresenta um crescimento aproximadamente linear. A correlação entre a TB e as taxas não espaciais, por sua vez, possuem um ritmo de crescimento mais rápido, devido a um comportamento aparentemente log-linear. Isso significa que todas as taxas fazem correções maiores para os municípios pequenos. Como nos municípios pequenos os valores tendem a ser extremos, a suavização também será mais abrupta. Os indicadores não espaciais, entretanto, levam a uma menor suavização das taxas para cidades de porte médio do que os indicadores espaciais.

Quanto ao nível, pode-se dizer que quase todas as taxas possuem uma correlação alta com a TB, independentemente do tamanho da população. Isso não é verdade, entretanto, para a TE, cuja associação é pequena para tamanhos de amostras pequenos. Como a TE é uma média da taxa do município e de seus vizinhos, a TB do próprio município tende a possuir pouca influência no resultado final. Para as demais taxas, entretanto, a correlação alta com a TB é uma constatação positiva, pois indica que elas corrigem algumas distorções da TB, sem distorcê-la completamente. Já a TE parece ser uma boa estimativa apenas globalmente, pois faz com que as taxas apresentem apenas variações suaves no espaço. Localmente, entretanto, a TE pode apresentar resultados não condizentes com a realidade dos pequenos municípios.

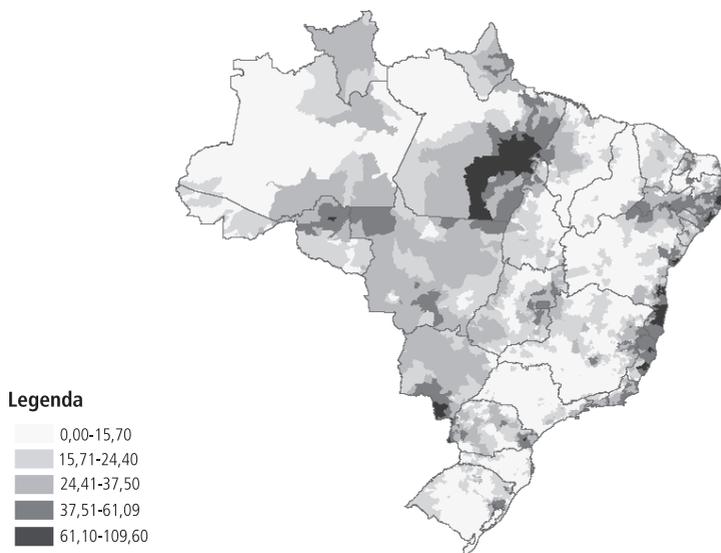
Uma vez que a o tamanho do município é componente do peso utilizado na Taxa Bayesiana Empírica e na Taxa Bayesiana de Clayton e Kaldor, a suavização, nesse caso, é função do tamanho do município. Com isso, segundo o gráfico 1, é possível observar que a variabilidade em torno da média é menor para as taxas não espaciais. Dentre as taxas espaciais, aquela em que a correlação apresentada possui maior variabilidade em torno da média é a TE não bayesiana. Logo, municípios de mesmo tamanho podem apresentar taxas com graus diferentes de suavização de acordo com as taxas espaciais, principalmente. Assim, caso o pesquisador escolha considerar a influência do espaço, o tamanho da população terá uma participação minimizada na correção da taxa. Caso o interesse, entretanto, seja obter taxas mais apuradas, considerando-se que o porte do município é o item mais importante para a qualidade da estimativa, em detrimento do fator espacial, é mais interessante utilizar as taxas não espaciais. Para uma visualização do fenômeno discutido, os mapas 1 a 5 apresentam a distribuição de cada uma das taxas para os municípios brasileiros em 2008.

MAPA 1
TB de homicídios nos municípios brasileiros em 2008



Fonte: Elaboração dos autores com base nos dados do Datasus.

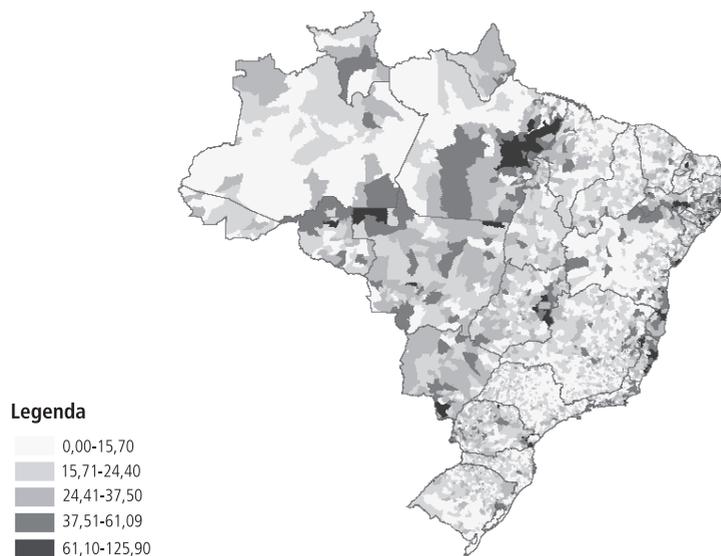
MAPA 2
TE de homicídios nos municípios brasileiros em 2008



Fonte: Elaboração dos autores com base nos dados do Datasus.

MAPA 3

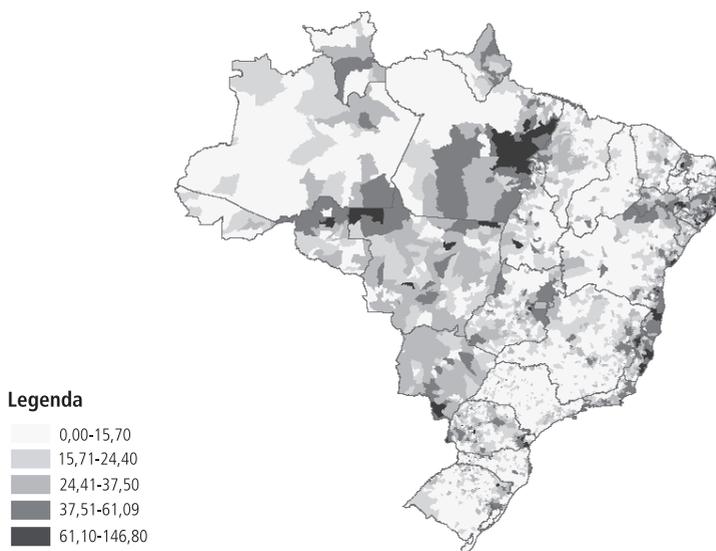
Taxa bayesiana empírica de homicídios nos municípios brasileiros em 2008



Fonte: Elaboração dos autores com base nos dados do Datasus.

MAPA 4

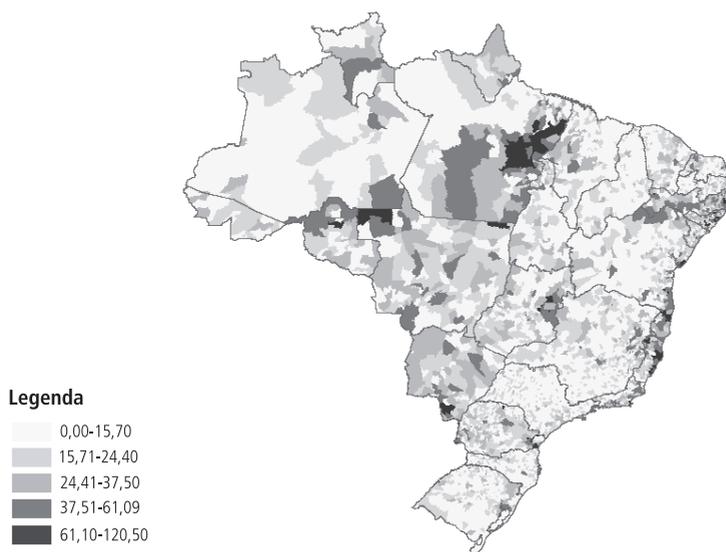
Taxa bayesiana empírica espacial de homicídios nos municípios brasileiros em 2008



Fonte: Elaboração dos autores com base nos dados do Datasus.

MAPA 5

Taxa bayesiana Clayton e Kaldor de homicídios nos municípios brasileiros em 2008



Fonte: Elaboração dos autores com base nos dados do Datasus.

Do mapa 1 até o mapa 5 (na sequência), é importante destacar que, em todos, as taxas estão divididas nas mesmas categorias, com o objetivo de possibilitar a comparação entre eles. De acordo com esses mapas, é possível observar, primeiramente, que o mapa da TE é o que possui maior diferença em relação aos demais padrões. Isso se deve pelo fato de que ele utiliza somente o fator espacial para a correção das taxas. Dessa forma, a visualização dos dados é priorizada, uma vez que as taxas se modificam suavemente no espaço. Como o Brasil é formado por 5.566 pequenos municípios, a variação é ainda mais branda. Pode-se observar também que, no litoral norte do Espírito Santo e no sul da Bahia, os municípios de Guarapari (ES), Linhares (ES) e Porto Seguro (BA) apresentam grande influência na TE, aumentando a taxa dos municípios vizinhos.

Outra característica importante é a presença de uma maior quantidade de valores extremos no mapa da TB, representados pela categoria mais escura. Embora também fosse esperada uma presença maior dos valores na classe inferior, isso não ocorre. Provavelmente os valores extremos inferiores permaneceram na classe branca após correção, uma vez que, as suavizações não foram muito bruscas em sua maioria. Neste mesmo mapa pode-se observar que, no Norte, a presença de grandes municípios, leva a leituras erradas da realidade na região. Este é o caso do município de Altamira (PA), considerado o maior município do mundo em extensão territorial (159.696 km²),

e como está na quarta classe, pode nos induzir a uma falsa impressão de maior periculosidade do estado como um todo.

O mapa da Taxa Empírica Bayesiana apresenta muitos municípios de valores entre 24,41 e 37,50, o que indica a influência da média geral que é de 26,70 homicídios por 100 mil pessoas. É possível notar ainda que o comportamento dos mapas baseados na Taxa Bayesiana Empírica e da Taxa Bayesiana Empírica Espacial é muito semelhante. O mapa com o componente espacial, entretanto, parece reproduzir o mapa da Taxa Bayesiana Empírica de forma suavizada. Em todas as situações pode-se perceber que a Taxa Bayesiana Empírica e a Taxa Bayesiana de Clayton e Kaldor são muito parecidas. Isso indica que a média global é próxima da esperança *a priori*.

Uma análise complementar à visualização do mapa de homicídios no Brasil é a listagem desse evento nas 27 capitais brasileiras. Por meio desse estudo, apresentado na tabela 3, será possível identificar os valores exatos dos índices em alguns pontos estratégicos espalhados por todo o país.

Na tabela 3 são apresentadas as taxas das capitais brasileiras, organizadas de forma ascendente, de acordo com a TB. Por meio dela, destaca-se, primeiramente, que as três capitais brasileiras mais perigosas, em termos de números de homicídios, fazem parte da região Nordeste. Esses foram os únicos municípios que registraram taxas superiores a 60 homicídios por 100 mil habitantes. Mesmo para as taxas corrigidas é possível observar tal realidade. Comparando-se com o mapa, essas regiões estão localizadas no extremo nordeste do país, na faixa mais escura do mapa que permeia os estados de Pernambuco e Maceió.

As demais regiões que possuem representantes entre as cinco capitais com maior taxa de homicídio são Norte e Sudeste. Entre as capitais do Norte, a única que possui um número de mortes maior que 50 por 100 mil habitantes é Belém. Já no Sudeste, a capital com maior grau de periculosidade é Vitória, que apresenta uma taxa de homicídios de quase 60 mortes para cada 100 mil pessoas. Em Vitória, a situação é ainda mais alarmante, pois há uma forte evidência de que as cidades no seu entorno possuem um grau de violência alto. Isso pode ser visto por meio da TE igual a 80,70, muito superior aos demais indicadores observados. Essa região pode ser visualizada no mapa como uma mancha mais escura localizada sobre o centro-norte do Espírito Santo e sul da Bahia.

TABELA 3
Indicadores das 27 capitais brasileiras – 2008

| Município | População | Homicídios | TB | Taxa Bayesiana Empírica | TE | Taxa Bayesiana Empírica Espacial | Taxa Bayesiana Clayton e Kaldor |
|----------------|------------|------------|--------|-------------------------|-------|----------------------------------|---------------------------------|
| Maceió | 924.143 | 933 | 101,00 | 100,30 | 93,90 | 100,60 | 100,14 |
| Salvador | 2.948.733 | 1.878 | 63,70 | 63,60 | 65,20 | 63,70 | 63,55 |
| Recife | 1.549.980 | 976 | 63,00 | 62,80 | 62,30 | 62,90 | 62,70 |
| Vitória | 317.817 | 186 | 58,50 | 57,80 | 80,70 | 59,80 | 57,39 |
| Belém | 1.424.124 | 742 | 52,10 | 52,00 | 51,80 | 52,10 | 51,88 |
| João Pessoa | 693.082 | 347 | 50,10 | 49,80 | 49,50 | 49,50 | 49,65 |
| Curitiba | 1.828.092 | 774 | 42,30 | 42,30 | 48,00 | 42,40 | 42,22 |
| Natal | 798.065 | 324 | 40,60 | 40,50 | 39,60 | 39,60 | 40,34 |
| Cuiabá | 544.737 | 220 | 40,40 | 40,20 | 42,40 | 40,80 | 40,02 |
| Porto Alegre | 1.430.220 | 570 | 39,90 | 39,80 | 43,00 | 40,00 | 39,71 |
| Macapá | 359.020 | 143 | 39,80 | 39,60 | 34,60 | 38,90 | 39,29 |
| Porto Velho | 379.186 | 147 | 38,80 | 38,50 | 36,90 | 38,70 | 38,28 |
| Manaus | 1.709.010 | 648 | 37,90 | 37,90 | 35,40 | 37,80 | 37,81 |
| São Luís | 986.826 | 360 | 36,50 | 36,40 | 33,80 | 36,20 | 36,31 |
| Goiânia | 1.265.394 | 458 | 36,20 | 36,10 | 35,20 | 36,10 | 36,06 |
| Belo Horizonte | 2.434.642 | 879 | 36,10 | 36,10 | 38,00 | 36,20 | 36,03 |
| Fortaleza | 2.473.614 | 874 | 35,30 | 35,30 | 35,60 | 35,40 | 35,27 |
| Brasília | 2.557.158 | 814 | 31,80 | 31,80 | 38,10 | 31,90 | 31,78 |
| Rio de Janeiro | 6.161.047 | 1.821 | 29,60 | 29,60 | 34,20 | 29,60 | 29,54 |
| Aracaju | 536.785 | 143 | 26,60 | 26,60 | 31,60 | 27,10 | 26,49 |
| Boa Vista | 260.930 | 69 | 26,40 | 26,50 | 26,00 | 26,00 | 26,15 |
| Rio Branco | 301.398 | 78 | 25,90 | 25,90 | 22,60 | 22,60 | 25,64 |
| Campo Grande | 747.189 | 183 | 24,50 | 24,50 | 24,80 | 24,80 | 24,41 |
| Teresina | 793.915 | 178 | 22,40 | 22,50 | 19,70 | 22,20 | 22,37 |
| Florianópolis | 402.346 | 89 | 22,10 | 22,20 | 20,10 | 20,50 | 22,03 |
| São Paulo | 10.990.249 | 1.701 | 15,50 | 15,50 | 18,70 | 15,50 | 15,48 |
| Palmas | 184.010 | 27 | 14,70 | 15,20 | 16,70 | 16,50 | 14,82 |

Fonte: Datasus.

Obs.: Taxas por 100 mil habitantes.

Entre as capitais com menores níveis de mortalidade por homicídios, destacam-se São Paulo e Palmas. Esses foram os únicos municípios, em meio aos listados, que apresentaram taxas de homicídios inferiores a 20 mortes por 100 mil habitantes. É interessante notar que as taxas corrigidas são concordantes com essa realidade. Nesse caso, São Paulo, que durante a década de 1980 foi considerada uma região com nível de violência crescente e preocupante, tem conseguido reverter esse cenário (PERES *et al.*,

2011). Embora em termos absolutos apresente um dos maiores registros de homicídios listados, percentualmente a capital se caracteriza como uma das mais seguras do país.

Em relação ao grau de suavização das taxas corrigidas, pode-se observar que a maior parte delas não possui mudanças substanciais com relação à TB. A diferença, nesse caso, não chega a ultrapassar uma unidade, na maior parte dos casos. Uma vez que todas as capitais possuem um número alto de habitantes, apenas a TE é destoante para algumas capitais. Essa é a situação, por exemplo, de Maceió, Vitória, Curitiba e Brasília, nas quais a alteração repercutiu em uma variação de pelo menos seis homicídios por 100 mil habitantes. Assim, a ordenação dos municípios não apresenta modificações bruscas, se arranjada de acordo com qualquer um dos indicadores utilizados.

9 CONCLUSÃO

Neste trabalho, foi feita uma revisão de algumas das principais técnicas de aprimoramento de dados de taxas. Mostrou-se a importância dos trabalhos de mapeamento, que são úteis para o direcionamento de políticas públicas relativas à segurança. Frente à relevância da área, a utilização de TBs foi considerada pouco indicada devido a sua alta variabilidade aleatória, principalmente quando parte das unidades geográficas estudadas possui populações muito baixas. Os métodos sugeridos, em contraposição, suavizaram os valores extremos por meio de técnicas bayesianas e espaciais.

Diante disso, um exemplo da funcionalidade das medidas empíricas foi feito com dados de homicídio no Brasil em 2008. Em um país com extensão continental como o Brasil, onde as diferenças regionais são bruscas, não se aconselha a aplicação da Taxa Bayesiana Empírica, pois a correção desta é feita com base na média do país. Uma melhor opção é a Taxa Bayesiana Empírica Espacial que atribui maior influência aos municípios geograficamente próximos, apresentando resultados mais coerentes regionalmente.

A TE é apenas a média dos vizinhos do município, incluindo o município, e é utilizada para obter uma superfície suavizada do evento, sendo por isso pouco útil para obter valores pontuais. No caso desta aplicação, as taxas que melhor se adaptaram foram as bayesianas Empírica, Empírica Espacial e Clayton e Kaldor. No caso da Taxa Bayesiana de Clayton e Kaldor, a distribuição *a priori* influencia muito as estimativas

para os municípios pequenos, e temos que verificar se a distribuição *a priori* utilizada faz sentido na aplicação.

Portanto, a escolha do indicador a ser utilizado dependerá do intuito a ser atingido com o trabalho. Caso o mapa utilizado apresente muitos polígonos e grandes diferenças regionais, como é o caso do Brasil, aconselha-se a utilização da Taxa Bayesiana Empírica Espacial ou a Taxa de Clayton e Kaldor. A TE é mais indicada se o objetivo for a análise visual dos dados. Se, entretanto, deseja-se obter estimativas mais apuradas, desconsiderando-se o efeito do espaço, deve-se utilizar a Taxa de Clayton e Kaldor, que possui menor variabilidade.

REFERÊNCIAS

- ANDRADE, E. I. G. *et al.* Pesquisa e produção científica em economia da saúde no Brasil. **Revista de Administração Pública**, v. 41, n. 2, p. 211-235, 2007.
- ASIIMWE, J. B. *et al.* Examining small area estimation techniques for public health intervention: lessons from application to under-5 mortality data in Uganda. **Journal of Public Health Policy**, v. 32, p. 1-14, 2011.
- BAYES, T. An essay towards solving a problem in the doctrine of chances. **Philosophical Transactions of the Royal Society**, v. 53, p. 370-418, 1763.
- BEATO FILHO, C. C. *et al.* Conglomerados de homicídios e o tráfico de drogas em Belo Horizonte, Minas Gerais, Brasil, de 1995 a 1999. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 17, n. 5, p. 1163-1171, Fiocruz, 2001.
- CASTRO, M. S. M. de; ASSUNÇÃO, R. M.; DURANTE, M. O. Comparação de dados sobre homicídios entre dois sistemas de informação, Minas Gerais. **Revista de Saúde Pública**, 37, n. 2, p. 168-176, 2003.
- _____.; VIEIRA, V. A.; ASSUNÇÃO, R. Padrões espaço-temporais da mortalidade por câncer de pulmão no Sul do Brasil. **Rev. Bras. Epidemiol**, v. 7, n. 2, p. 1-13, 2004.
- CATELAN, D.; LAGAZIO, C.; BIGGERI, A. A hierarchical Bayesian approach to multiple testing in disease mapping. **Biometrical Journal**, v. 52, p. 784-797, 2010.
- CLAYTON, D.; KALDOR, J. Empirical bayes estimates of age-standardized relative risks for use in disease mapping. **Biometrics**, v. 43, n. 3, p. 671-681, 1987.
- CAVALINI, L. T.; LEON, A. C. M. P. Correção de sub-registros de óbitos e proporção de internações por causas mal definidas. **Revista de Saúde Pública**, v. 41, n. 1, p. 85-93, 2007.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood for incomplete data via the EM algorithm (with discussion). **Journal of the Royal Statistical Society**, v. 39, n. 1, p. 1-38, 1977 (Series B).

GELMAN, A. *et al.* **Bayesian Data Analysis**. 2. ed. Londres: Chapman & Hall/CRC, 2004.

LEE, Y.; JANG, M.; LEE, W. Prediction interval for disease mapping using hierarchical likelihood. **Computational Statistics**, v. 26, p. 159-179, 2011.

MARSHALL, R. J. Mapping disease and mortality rates using empirical bayes estimators. **Journal of the Royal Statistical Society**, v. 40, p. 2, p. 283-294, 1991 (Series C - Applied Statistics).

PERES, M. F. T. *et al.* Queda dos homicídios em São Paulo, Brasil: uma análise descritiva. **Rev. Panam. Salud Publica**, v. 29, n. 1, p. 17-26, 2011.

PRINGLE, D. G. Mapping disease risk estimates based on small numbers: An assessment of empirical bayes techniques. **Economic and Social Review**, v. 27, p. 341-363, 1996.

RAO, J. N. K. **Small area estimation**. New Jersey: Wiley, 2003.

SOUZA, W. V. *et al.* Aplicação de modelo Bayesiano empírico na análise espacial da ocorrência de hanseníase. **Revista de Saúde Pública**, v. 35, n. 5, p. 474-480, 2001.

WAISELFISZ, J. J. **Mapa da Violência 2010** - Anatomia dos Homicídios no Brasil. São Paulo: Instituto Sangari, 2010.

EDITORIAL

Coordenação

Cláudio Passos de Oliveira

Supervisão

Andrea Bossle de Abreu

Revisão

Eliezer Moreira

Elisabete de Carvalho Soares

Fabiana da Silva Matos

Lucia Duarte Moreira

Luciana Nogueira Duarte

Míriam Nunes da Fonseca

Editoração

Roberto das Chagas Campos

Aeromilson Mesquita

Aline Cristine Torres da Silva Martins

Carlos Henrique Santos Vianna

Maria Hosana Carneiro Cunha

Paula Mascarenhas Rodrigues de Almeida (estagiária)

Capa

Luís Cláudio Cardoso da Silva

Projeto Gráfico

Renato Rodrigues Bueno

Livraria do Ipea

SBS – Quadra 1 – Bloco J – Ed. BNDES, Térreo.

70076-900 – Brasília – DF

Fone: (61) 3315-5336

Correio eletrônico: livraria@ipea.gov.br

Tiragem: 500 exemplares

Missão

Produzir, articular e disseminar conhecimento para aperfeiçoar as políticas públicas e contribuir para o planejamento do desenvolvimento brasileiro.



Ipea – Instituto de Pesquisa
Econômica Aplicada



SECRETARIA DE
ASSUNTOS ESTRATÉGICOS
DA PRESIDÊNCIA DA REPÚBLICA

