



VIABILIDADE DA CORREÇÃO BAYSEANA PARA TAXAS DE CRIMES VIOLENTOS.

Gilmar Rosa ¹

Maressa de Assis Silva Martins, Cap PM²

Leonardo Martins Miranda, 2^o Sgt PM³

Micael Henrique Silva, Cel PM⁴

Miller França Michalick, Ten-Cel PM⁵

Renato Pinheiro Batista, Ten-Cel PM⁶

Resumo

O presente artigo têm por objetivo mostrar a viabilidade do uso de taxas bayseanas corrigidas, com o propósito de minimizar a variabilidade do ajuste de taxas de crimes violentos, para localidades onde a população é pequena. Tal estimativa bayseana pode ser interpretada como o risco sob o qual aquela população está em relação à vitimização por crimes violentos. Em localidades a população é grande, verificamos efeito desprezível em relação ao valor da taxa bruta, comparada com a taxa bayseana corrigida. O estudo abrange o Estado de Minas Gerais, com nível de desagregação dos dados por de Batalhões e Cias Independentes, uma vez que tal hierarquia é composta por conjunto de municípios.

Palavras-chave: Taxas bayseanas, Crimes violentos e Estimador bayseano.

1 Introdução

Os cálculos que se baseiam na taxa bruta de crimes são delineados como os mais simples para o estudo da ocorrência de um evento, no rol das alternativas conhecidas. A taxa bruta é uma razão entre o número de eventos ocorridos e o número de pessoas expostas à ocorrência desse evento em uma determinada localidade (SILVA et al., 2011).

¹Doutor pelo Curso de Engenharia de Produção pela Universidade Federal de Minas Gerais - UFMG, glmrrs.dep@gmail.com;

²Especialista pelo Curso em Segurança Pública (CESP) pela Academia de Polícia Militar de Minas Gerais- maressa.asm@gmail.com

³Graduado pelo Curso de Matemática pela UNI/BH, 1488766@pmmg.mg.gov.br

⁴Especialista em Gestão Estratégica em Segurança Pública (CEGESP) pela Fundação João Pinheiro (FJP)

⁵Especialista em Gestão Estratégica em Segurança Pública (CEGESP) pela Academia da Polícia Militar de Minas Gerais, APM

⁶Especialista em Segurança Pública (CESP) pela Fundação João Pinheiro (FJP)



A taxa bruta é o estimador mais simples para o risco de ocorrência de um evento e deriva-se da razão entre o número de eventos ocorridos em uma área e o número de pessoas expostas à ocorrência desse evento.

Porém, há um relevante problema relacionado à utilização de taxas brutas: elas apresentam alta instabilidade para expressarem o risco de um determinado evento, quando este é raro ou quando a população da localidade onde ele ocorre é pequena (pequeno denominador) (SANTOS et al., 2005; SILVA et al., 2011).

A exemplo, cita-se a ocorrência de um ou dois casos do evento, a mais ou a menos, numa localidade, que resulta em um impacto grande no cálculo das taxas brutas, se a população de interesse for pequena, efeito que não se verifica em localidades de população de tamanho grande (SANTOS et al., 2005; SILVA et al., 2011).

Para Santos et al. (2005, p. 301)[1]:

As variações bruscas que ocorrem com estas taxas podem não dizer respeito ao fenômeno e sim com uma variabilidade associada às observações. Flutuações aleatórias casuais, como a ocorrência de um ou dois casos do evento, a mais ou a menos, numa localidade, causam variações substanciais nas taxas brutas se a sua população for pequena, o que não foi verificado em localidades de população grande. Ademais, quando não ocorrem casos do evento em algumas regiões, a taxa bruta estima o risco de ocorrência do evento como zero, algo irreal quando se trata de dados como doenças ou crimes. Mapas de eventos, baseados diretamente nas estimativas brutas, tornam-se de difícil interpretação e, frequentemente, resultam em falsas conclusões.

De acordo com [2], cerca de 25% dos municípios brasileiros possuem menos de 5.000 habitantes, logo, com frequência, podem ocorrer falsas interpretações e conclusões quando da análise de mapas de taxas brutas.

Frente à relevância da área, a utilização de taxas brutas por vezes, pode ser considerada pouco indicada devido à sua alta variabilidade aleatória, principalmente quando parte das unidades geográficas estudadas possui populações pequenas. Para a correção dessa instabilidade e variações, apresentam-se como solução diversas técnicas de suavização, como os métodos bayesianos, que estimam taxas corrigidas a partir dos valores observados, utilizando-se de conceitos de inferência bayesiana. Tais métodos suavizam os valores extremos e garantem melhor percepção do fenômeno estudado.



2 Estimador Bayseano

O método bayesiano empírico foi proposto por Roger J. Marshall [3]. Para Marshall o mapeamento de mortalidade e taxas de doença para mostrar a variabilidade geográfica da doença é uma ferramenta epidemiológica cada vez mais comum. A ideia não é nova, Howe [4], detalhou o seu desenvolvimento até o século XVIII, porém o advento dos computadores e da computação gráfica, tornaram a construção de mapas consideravelmente mais fácil, ainda assim o problema estatístico de derivar estimativas das quantidades a serem mapeadas, só recentemente tiveram maior atenção.

Os desenvolvimentos centraram-se na ideia de reunir informações entre áreas, através do uso de métodos empíricos de Bayes, para reduzir o erro quadrático médio total. Efron e Morris [5] parecem ser os primeiros a usar essa abordagem para taxas de doença toxoplasmose em El Salvador, através dos estimadores de James-Stein [6]. Clayton e Kaldor [7] propuseram procedimentos de estimação empíricos de Bayes usando uma probabilidade de Poisson e distribuição gama a priori, uma estrutura também usada por Stone [8], para ajustar níveis de significância para riscos locais. Tsutakawa et al. [9] havia proposto anteriormente uma abordagem bayseana empírica semelhante ao método de Leonard [10] para estimar proporções binomiais.

Clayton e Kaldor [7], também sugeriram estimar uma distribuição a priori de forma não paramétrica, usando um método proposto por Laird [11], e um método para modelagem a priori dos parâmetros por um esquema espacial autogressivo. Miyawaki e Chen [12] também propuseram o que é essencialmente um procedimento empírico de Bayes para estimar uma densidade a priori para uma taxa de mortalidade padronizada.

Marshall [3], expõe que o trabalho por ele realizado foi motivado por um estudo dos padrões espaciais de mortalidade na Região de Auckland da Nova Zelândia. O autor explica que um estimador bayesiano empírico é derivado dos resultados de Efron e Morris [5] para a construção de estimadores de James-Stein. É um procedimento não interativo. Trabalhos semelhantes de Clayton e Kaldor [7] também foram por ele discutidos. Contudo, nenhum desses procedimentos estudados tratou explicitamente do aspecto espacial dos dados, os quais eles se mostraram invariáveis à configuração espacial. O caráter espacial do problema é, portanto o que o torna interessante e ignorá-lo pareceu injustificável [3].

Marshall [3] propõe que seja considerada uma região dividida em N áreas i , onde ($i = 1; \dots; N$) e que eventos (mortes ou doenças) são nela registrados durante um período de tempo t , onde $t = (\text{ano}, \text{dia}, \text{mês}, \text{etc.})$.

Dado θ_i a taxa de determinado evento em um período de tempo, na área i e



seja r_i o total de eventos acumulados no respectivo tempo t , sendo uma variável aleatória com distribuição de *Poisson* com média condicional $E(r_i|\theta_i) = \theta_i/n_i$, onde n_i , é número de pessoas sob risco em determinado período de tempo t na área i . O estimador de máxima verossimilhança (ML) de θ_i é $x_i = r_i/n_i$. Marshal (1990), mostrou que x_i é um estimador mais grosseiro ou menos preciso para θ_i . Lógicamente essa simplificação proposta por Marshal têm implicação na distribuição condicional de θ_i e na variância condicional $Var(x_i|\theta_i) = \theta_i/n_i$. Na prática θ_i , geralmente requer estimativa por interpolação dos dados coletados, mas o erro intrínseco ao processo de interpolação é ignorado. Assumindo que a distribuição condicional de θ_i , os r_i são independentes, logo as questões relativas à dependência espacial e os r_i observados, não apareceram ou podem ser ignoradas.

Adotando a abordagem bayseana, suponha que θ_i , tenha distribuição a priori com média $m_i = E_\theta(\theta_i)$ e $A_i = Var_\theta(\theta_i)$ respectivamente. A média não condicional de x_i é $E_x(x_i) = E_\theta\{E(x_i|\theta_i)\} = E_\theta(\theta_i) = m_i$ e a variância não condicional é dada por:

$$Var_x(x_i) = Var_\theta\{E(x_i|\theta_i)\} + E_\theta\{Var(x_i|\theta_i)\}$$

$$Var_x(x_i) = Var_\theta(\theta_i) + E_\theta\{Var(x_i|\theta_i)\}$$

$$Var_x(x_i) = A_i + m_i/n_i \quad (1)$$

Dado m_i e A_i , são os melhores estimadores lineares de bayes, em termos da função perda quadrática [5], logo um estimador reduzido ou simplificado é dado por:

$$\theta_i = m_i + C_i(x_i - m_i) \quad (2)$$

Onde:

$$C_i = \frac{Var_\theta(\theta_i)}{Var_x(x_i)} = \frac{A_i}{\left(A_i + \frac{m_i}{n_i}\right)} \quad (3)$$

Os chamados estimadores de James-Stein (*James-Stein Estimator*) [5], de θ_i , são construídos assumindo x_i como uma amostra da densidade marginal de



x_i , para estimar m_i e C_i na Equação (2).

Uma possibilidade de simplificação do modelo Equação (3) é estabelecer $C_i = C$, $A_i = A$, Considere $m_i = m$ como a taxa global da região. Para encontrar uma estimativa para A podemos considerar a variância ponderada.

$$S^2 = \sum_i n_i \frac{(x_i - \tilde{m})^2}{n} \quad (4)$$

Assumindo que $\tilde{m} = \frac{\sum_i r_i}{n}$, onde r_i é a quantidade de eventos na área i e $n = \sum_i n_i$ é o total de pessoas sob risco no conjunto de todas as áreas que compõem estudo. Por exemplo n_i total de pessoas por bairro i , e n total de pessoas que compõem a região urbana da cidade sob estudo.

$$E_x(x_i - m)^2 = Var_x(x_i) = A + m/n_i$$

Logo:

$$E_x(S^2) = \sum_i n_i \frac{(x_i - m)^2}{n} = A + m/\bar{n}$$

Ignorando o erro de usar \tilde{m} como um estimador de m temos:

$$S^2 = A + \tilde{m}/\bar{n}$$

O estimador para A é dado por:

$$\tilde{A} = S^2 + \tilde{m}/\bar{n} \quad (5)$$

O estimador para a Equação (2) é:

$$\hat{\theta}_i = \tilde{m} + \tilde{C}_i(x_i - \tilde{m}) \quad (6)$$

O estimador da Equação (3) é:

$$\tilde{C}_i = \frac{S^2 - \tilde{m}/\bar{n}}{S^2 - \tilde{m}/\bar{n} + \tilde{m}/n_i} \quad (7)$$

Logo, esse método se utiliza dos próprios dados para construir a priori. Por esse motivo a taxa corrigida de Marshall [3] é denominada *empírica*.

Fazendo algumas manipulações na Equação (6), podemos reescrevê-la como:

$$\hat{\theta}_i = \tilde{C}_i x_i + (1 - \tilde{C}_i) \tilde{m} \quad (8)$$



Quanto ao fator de ponderação \tilde{C}_i , da Equação (8), ele será tanto menor quanto menor for a população da i -ésima área sob estudo, e reflete o grau de confiança a respeito de cada taxa. A taxa corrigida de áreas com populações suficientemente grandes, \tilde{C}_i se aproxima de 1 e, conseqüentemente, o peso $(1 - \tilde{C}_i)$ da média dos vizinhos, se aproxima de zero.

Quando as populações são reduzidas (pequenas), a confiança na taxa observada diminui e a estimativa da taxa bruta é influenciado pelo modelo a priori (ou seja, da taxa global \tilde{m}). Portanto, regiões com populações muito baixas sofrerão uma correção maior, ao passo que as regiões população maiores terão pouca alteração nas respectivas taxas. Para melhor entendimento do leitor apresentamos a Tabela 1, de forma a facilitar a implementação do estimador de bayseano empírico. À conclusão, Marshall [3], argumenta que a escolha de

Tabela 1: Descrição dos símbolos das Equações (7) e (8)

Símbolo	Fórmula	Descrição i	Observação
$\hat{\theta}_i$		Valor corrigido da taxa x_i	Ver Equação (8)
r_i		Total de eventos na área i	Exemplo: Crimes, incidência de doenças, etc.
n_i		Número de pessoas na área i	$\sum_i n_i = n$
x_i	$x_i = r_i/n_i$	Taxa de eventos em toda área i	
\tilde{m}	$\tilde{m} = \sum_i r_i / \sum_i n_i$	Taxa global dos eventos i	
\bar{n}	$\bar{n} = \sum_i n_i / N$ $\bar{n} = n/N$	Número médio de pessoas em risco	N é o número total de áreas sob estudo, por exemplo total de municípios
S^2	$Var(x_i)$	Variância das taxas x_i	

um estimador apropriado dependerá do verdadeiro, mas desconhecido, padrão de doença. Idealmente, exige-se um método que forneça boas estimativas em uma variedade de situações. No entanto, nenhum dos estimadores considerados cumpriu este requisito inteiramente. O estimador local θ_i emergiu, entretanto, como o menos sensível à raridade da doença e para o tipo de padrão, tendo sido também conceitualmente atraente na medida em que a configuração espacial é empregada, em que pese seu desempenho não ser tão bom para doenças que são raras.

Dois métodos então se destacaram: um global que torna uma região inteira a vizinhança, ao invés de cada unidade espacial ter sua própria e única vizinhança, e um “meio-termo”, caracterizado pelo agrupamento de áreas contíguas em bairros, ou zonas, se houver uma justificativa a priori para o zoneamento, por exemplo, em termos de características geográficas ou demográficas homogêneas.

Na literatura existem vários estudos que retratam a necessidade de suavização de taxas brutas e se que utilizam dos métodos bayesianos, principalmente no campo da Saúde Pública[13]. Em um deles, Assunção et al. [14], ao tratarem de taxas epidemiológicas, citam que Choynowski [15], em um artigo clássico e pioneiro. Nesse estudo, ao analisar a ocorrência de tumores cerebrais



em condados poloneses, o autor supracitado detectou que os dois casos observados na cidade de Lesko (Polônia), cuja população era de 17.000 habitantes, resultaram em uma taxa relativamente extrema, de 11,8 por 100 mil. Ao mesmo tempo, o registro de apenas um caso, ao invés dos dois observados, resultaria em uma taxa de 5,9, um valor considerado consistente com as taxas dos outros condados.

Sob essa ótica, Assunção et al. [14] afirmam que o mesmo problema foi detectado no estudo que conduziram, vez que observaram 15 municípios com nenhum óbito registrado e com número de nascidos vivos inferior a 30, implicando uma SMR (*standardized mortality ratio*), razão da taxa de mortalidade padronizada, igual a zero, porém se uma única morte houvesse sido registrada nesses municípios, as razões de mortalidade padronizada passariam de zero para valores entre 116 e 1048, sendo que este último valor corresponderia a um município com risco relativo estimado 10 vezes maior que a média do estado e passaria, então, a ser o mais alto, substituindo o valor extremo de fato observado, que foi 609. Os autores concluem que teriam, assim, os aspectos mais salientes do mapa resultante dos estudos, causados pela sensibilidade extrema da SMR quando das pequenas mudanças no numerador, possivelmente em virtude das flutuações aleatórias não associadas ao verdadeiro risco subjacente.

Segundo Silva et al [16], destaca que entre as abordagens bayesianas empregadas para a suavização de taxas, destacam-se a Bayesiana Empírica e a Bayesiana Padrão. A ascensão do uso dessas técnicas está atrelada ao grande avanço das tecnologias disponíveis e, também, à necessidade de se obter estimativas mais suavizadas, que traduzam adequadamente a característica subjacente do evento de interesse, sem a influência da alta variabilidade identificada nos dados brutos. Este trabalho se limitará à abordagem bayesiana empírica. Asseveram que se comparando as duas abordagens, os métodos Bayesianos Empíricos e Bayesianos Padrão apresentam resultados muito semelhantes, sendo que o primeiro se destaca pela facilidade na integração aos ambientes de geoprocessamento.

Marshall [3] propõe um método extremamente simples de ser implementado para o cálculo das estimativas bayesianas empíricas. A estimativa proposta por ele diminui a variabilidade das estimativas, ao restringir a flutuação aleatória ocasionada pelos eventos raros, e corresponde a uma média ponderada entre a taxa bruta do país (ou do Estado, no caso deste estudo) e a do próprio município. Pode-se afirmar que é adicionada uma suavidade espacial, condizente com a realidade da maior parte dos eventos, cuja ocorrência tem forte influência de fatores ambientais. Trata-se de uma técnica de aprimoramento de dados de taxas, Carvalho et al.[17].

A taxa bayesiana empírica local de um município é o resultado da soma pon-



derada entre a sua taxa bruta e a taxa média global de todo o mapa, segundo Marshall [3]. O autor delinea que essa soma possui fator de ponderação inversamente proporcional à população sob risco. Logo, quanto maior a população de um município, mais consistente é a sua estimativa via taxa bruta e, conseqüentemente, a taxa bayesiana empírica será aproximadamente igual a taxa bruta. Quando a população for pequena, porém, o valor estimado será ponderada pela tendência da taxa global, ao que se denomina taxa bayesiana empírica global. Outros trabalhos têm sido publicados em área diferente da saúde, na qual a técnica, foi originalmente proposta, especialmente em dados de crimes: [18], [19], [20], [21], [22]

3 Análise de Dados

3.1 Taxa de Crimes Violentos

A Taxa de Crimes Violentos (TVC) é um indicador calculado por meio da divisão do número total de crimes violentos registrados, pela população, multiplicado por 100.000.

Tal indicador sofreu, ao longo de sua existência, algumas alterações em relação às naturezas criminais contabilizadas, trazendo um acréscimo no número de crimes a serem avaliados, de modo que, atualmente, abrange os crimes de homicídio, estupro, estupro de vulnerável, roubo, sequestro e cárcere privado, extorsão e extorsão mediante sequestro, todos em suas modalidades consumadas e tentadas.

As naturezas são diferenciadas dependendo da Unidade da Federação, uma vez que na literatura não se encontra quais são as naturezas delituosas compõem tal indicador. De maneira geral, crime violento é aquele que por sua natureza têm o poder de trazer pânico à comunidade que o inclui, desta forma cada agência de segurança pública define o conjunto de naturezas que o compõem.

3.2 Estimativas Bayseanas das Taxas de Crimes Violentos

As taxas brutas, foram importadas para software R por meio de um arquivo .csv. Este arquivo contém quatro colunas, sendo Município, Código do Município, conforme IBGE, População por Município e Quantidade de Crimes Violentos em 2019 (valores absolutos), conforme Tabela 2.

Estes dados foram fornecidos pela Polícia Militar de Minas Gerais (PMMG), através do seu eficiente sistema de coleta de dados criminais. Maiores detalhes pode ser obtido em [23].

A rotina em R, **Listing 1**, a seguir, aplicada aos dados da Tabela 2, retorna os valores corrigidos, por município.



Tabela 2: Estrutura do arquivo .csv com valores brutos

ORD	Nome do Município	Código Município	População/2019	Crimes Violentos/2019
1	ABADIA DOS DOURADOS	310010	6.972	###
...
853	WENCESLAU BRAZ	317220	2.552	###

Através de pequenas modificações na rotina em R, apresentada a seguir, é possível obter correções bayesianas por Regiões e Batalhões/Companhias Independentes da Polícia Militar de Minas Gerais, uma vez um conjunto de Municípios distintos formam um Batalhão/Companhia. E a agregação de Batalhões/Companhias Independentes distintas formam uma Região da Polícia Militar.

Este estudo está restrito ao nível hierárquico de Regiões e Batalhões e/ou Companhias Independentes.

```
1 prgbayes<-function() {
2   banco <-read.csv2("E:/bancodados.csv", header=T, sep=";")
3   tx <- c <- theta<- V <- seq(0, 852)
4   for(i in 1:853){
5     tx[i]<-(c(banco[i,4])/c(banco[i,3]))
6   }
7   m<-(sum(c(banco[,4]))/sum(c(banco[,3])))
8   nb <- mean(c(banco[,3]))
9   for(i in 1:853){
10    V[i]<- ((banco[i,3])*(tx[i]-m)^2)/sum(c(banco[,3]))
11  }
12  S2 <- sum(V)
13  Num<-(S2-m/nb)
14  for(i in 1:853){
15    Den<-(S2-m/nb)+(m/c(banco[i,3]))
16    c[i]<-Num/Den
17  }
18  for(i in 1:853){
19    theta[i]=c[i]*tx[i]+(1-c[i])*(sum(c(banco[,4]))/sum(c(banco
20    [,3])))
21  }
22  saida_ind<-data.frame(theta, tx)
23  dados_map <-data.frame(banco, saida_ind*100000)
24  dados_map <-(cbind(theta*100000,tx*100000))
25  write.csv2(dados_map, file="E:/SaidaCorrigida.csv", row.names=
26  TRUE)
  head(dados_map)
}
```

Listing 1: Código fonte em R



3.3 Estimativas Bayseanas Por Regões da Polícia Militar de Minas Gerais

No que tange às RPM, 6 (seis) tiveram os resultados corrigidos com redução dos resultados do indicador, ou seja, os valores com correção Bayesiana ficaram menores que os valores brutos obtidos em 2019, à exemplo da 1ª RPM, que passou de 788,39 crimes violentos por 100 mil habitantes para 772,64, a partir da correção bayesiana, implicando em uma variação de -2,00%.

Outrossim, 13 (treze) RPM tiveram aumento dos resultados do indicador, em relação às taxas brutas de 2019. Cita-se por exemplo, a 13ª RPM que passou de uma TCV de 105,79 para 117,76 crimes violentos por 100 mil habitantes, o que equivale a uma variação percentual de 11,31%. conforme Tabela 3. Estas variações podem ser interpretadas, como o valor esperado, ou risco, nas regiões sob estudo.

No caso da 13ª RPM, seria esperado um risco maior, do que o resultado obtido/apurado. No entanto não foi possível no escopo deste estudo estabelecer as causas do resultado inesperado

A taxa estadual, somatório das taxas das Unidades teve uma variação pouco expressiva ao ser corrigida ou suavizada pelo método bayseano empírico, variou em -0,08%, tendo em vista que o Estado alcançou em 2019 a TCV de 328,17, valor que corrigido atingiu a taxa de 327,91 crimes violentos por 100 mil habitantes, conforme Tabela 3.

Os mapas constantes das Figuras 1 e 2, mostram as Taxas brutas e taxa bayseanas de criminalidade violenta por Batalhões da Polícia Militar de Minas Gerais.

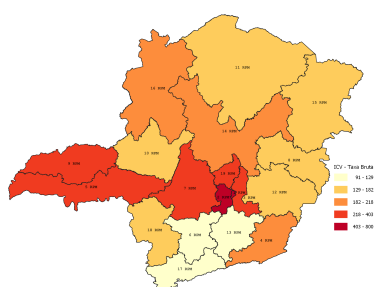


Figura 1: Taxa de Crimes Violentos

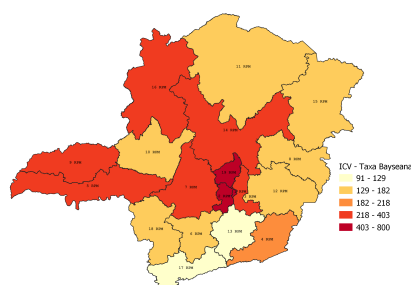


Figura 2: Taxa Corrigida



Tabela 3: Comparativo da Taxas Brutas de Criminalidade Violenta e Taxas Bayseanas de Crimes Violentos

RPM	TCV Apurado 2019	TCV Bayseano 2019	Var(%)	População 2019*
1ª RPM	788,39	772,64	-2,00	2.512.070
2ª RPM	733,9	718,8	-2,06	1.953.270
3ª RPM	316,83	316,74	-0,03	969.297
4ª RPM	199,02	200	0,49	1.664.130
5ª RPM	323,7	309,75	-4,31	828.846
6ª RPM	128,57	135,47	5,37	901.470
7ª RPM	336,36	339,51	0,94	1.182.350
8ª RPM	181,53	179,6	-1,06	773.976
9ª RPM	392,18	387,32	-1,24	1.084.714
10ª RPM	175,89	180,84	2,81	523.061
11ª RPM	161,64	168,64	4,33	1.508.947
12ª RPM	163,68	173,73	6,14	1.806.567
13ª RPM	105,79	117,76	11,31	879.131
14ª RPM	215,32	219,11	1,76	716.602
15ª RPM	153,66	160,23	4,28	890.267
16ª RPM	218,13	225,5	3,38	382.349
17ª RPM	90,55	99,3	9,66	1.009.403
18ª RPM	155,17	160,78	3,62	1.134.887
19ª RPM	402,72	407,37	1,15	447.454
MG	328,17	327,91	-0,08	21.168.791

* Nota: Estimativa Populacional-IBGE para o ano de 2019

3.4 Estimativas Bayseanas Por Batalhões da Polícia Militar

Quanto às UEOp, 33 (trinta e três) Unidades tiveram os resultados corrigidos com redução dos resultados do indicador, sendo os maiores percentuais de redução da TCV constatados no 6º BPM, com redução de 289,39 para 265,26 crimes violentos por 100 mil habitantes (-8,34%). Ao passo que 54 (cinquenta e quatro) Unidades tiveram as taxas aumentadas a partir da correção, destacando-se o 57º BPM com aumento de 43,00 para 53,85 crimes violentos (25,22%), conforme Tabela 4.

O mapa constante da Figura 3 mostra a suavização do indicador de Taxas de Criminalidade Bruta, Figura 4.

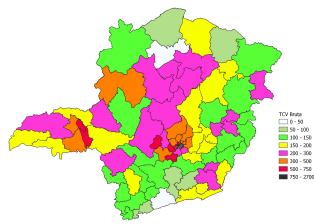


Figura 3: Taxa Bruta de Crimes

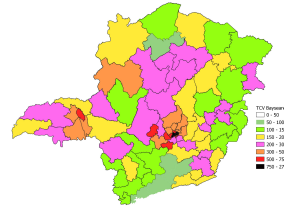


Figura 4: Taxa Bruta Bayseana

O método proposto por Marshall (1991) diminui a variabilidade das estimativas ao restringir a flutuação aleatória ocasionada pelos eventos raros, adicionando-se uma suavidade espacial, condizente com a realidade da maior parte dos eventos. Trata-se de uma técnica de aprimoramento de dados de taxas, que elimina parcialmente a variabilidade presente nas taxas brutas estimadas, as quais possuem alta instabilidade para expressarem o risco de um determinado evento quando este é raro ou quando a população da localidade onde ele ocorre é pequena (pequeno denominador).



Tabela 4: Taxas Brutas e Taxas Bayseanas de Crimes Violentos por UEOP/Mi-
nas Gerais/2019

RPM	UEOp	TCV APURADO 2019	TCV CORREÇÃO BAYESIANA 2019	Variação(%)
1ª RPM	1ª BPM	2.627,27	2566,78	-2,30
1ª RPM	13ª BPM	903,16	882,76	-2,26
1ª RPM	16ª BPM	772,14	758,84	-1,72
1ª RPM	22ª BPM	556,37	544,92	-2,06
1ª RPM	34ª BPM	817,63	802,19	-1,89
1ª RPM	41ª BPM	573,95	562,26	-2,04
1ª RPM	49ª BPM	814,03	799,08	-1,84
1ª RPM	5ª BPM	485,04	474,53	-2,17
2ª RPM	18ª BPM	498,03	477,58	-2,05
2ª RPM	33ª BPM	617,8	605,51	-1,99
2ª RPM	39ª BPM	1.137,24	1114,31	-2,02
2ª RPM	40ª BPM	597,87	592,49	-0,90
2ª RPM	48ª BPM	270,04	263,53	-2,41
2ª RPM	6ª CIA PM IND	405,37	393,55	-2,92
2ª RPM	66ª BPM	859,05	832,61	-3,08
2ª RPM	7ª CIA PM IND	517,62	509,19	-1,63
3ª RPM	1ª CIA PM IND	247,62	247,94	0,13
3ª RPM	35ª BPM	387,43	385,51	-0,50
3ª RPM	36ª BPM	396,39	393,54	-0,72
3ª RPM	52ª BPM	192,99	200,01	3,64
3ª RPM	61ª BPM	342,65	340,98	-0,49
3ª RPM	8ª CIA PM IND	312,45	309,36	-0,99
4ª RPM	10ª CIA PM IND	269,04	269,93	0,33
4ª RPM	2ª BPM	184,62	184,24	-0,21
4ª RPM	21ª BPM	207,96	205,2	-1,33
4ª RPM	27ª BPM	184,32	188,1	2,05
4ª RPM	47ª BPM	185,01	190,13	2,77
5ª RPM	3ª CIA PM IND	168,16	176,28	4,83
5ª RPM	37ª BPM	184,58	185,94	0,74
5ª RPM	4ª BPM	517,56	480,06	-7,25
5ª RPM	4ª CIA PM IND	195,06	201,58	3,34
5ª RPM	67ª BPM	438,14	408,04	-6,87
6ª RPM	16ª CIA PM IND	132,83	145,58	9,60
6ª RPM	24ª BPM	139,71	139,8	0,06
6ª RPM	8ª BPM	118,15	128,08	8,40
7ª RPM	19ª CIA PM IND	217,67	217,96	0,13
7ª RPM	23ª BPM	417,88	415,94	-0,46
7ª RPM	60ª BPM	696,02	689,79	-0,90
7ª RPM	63ª BPM	121,03	128,14	5,87
7ª RPM	7ª BPM	223,94	238,63	6,56
8ª RPM	15ª CIA PM IND	103,16	116,3	12,74
8ª RPM	18ª CIA PM IND	101,41	111,57	9,82
8ª RPM	5ª CIA PM IND	207,69	201,56	-2,95
8ª RPM	6ª BPM	289,39	265,26	-8,34
8ª RPM	65ª BPM	118,28	130,65	10,46
9ª RPM	17ª BPM	536,69	527,84	-1,65
9ª RPM	32ª BPM	448,07	440,03	-1,79
9ª RPM	53ª BPM	187,21	186,23	-0,52
9ª RPM	54ª BPM	246,2	249,38	1,29
10ª RPM	15ª BPM	136,69	139,63	2,15
10ª RPM	46ª BPM	234,7	242,65	3,39
11ª RPM	10ª BPM	222,78	225,89	1,40
11ª RPM	13ª CIA PM IND	47,79	59,56	24,63
11ª RPM	2ª CIA PM IND	95,67	107,2	12,05
11ª RPM	30ª BPM	95,52	104,74	9,65
11ª RPM	50ª BPM	235,33	243,35	3,41
11ª RPM	51ª BPM	180,68	183,48	1,55
12ª RPM	11ª BPM	122,82	135,97	10,71
12ª RPM	14ª BPM	205,7	205,27	-0,21
12ª RPM	17ª CIA PM IND	165,66	176,94	6,81
12ª RPM	21ª CIA PM IND	131,02	153,93	17,49
12ª RPM	26ª BPM	187,33	199,92	6,72
12ª RPM	58ª BPM	199,36	210,73	5,70
12ª RPM	62ª BPM	130,84	140,3	7,23
13ª RPM	31ª BPM	144,16	150,7	4,54
13ª RPM	38ª BPM	83,52	97,52	16,76
13ª RPM	9ª BPM	72,71	90,36	24,27
14ª RPM	23ª CIA PM IND	172,16	177,8	3,28
14ª RPM	3ª BPM	185,78	190,54	2,61
14ª RPM	42ª BPM	268,35	267,36	-0,37
14ª RPM	55ª BPM	243,83	247,86	1,65
15ª RPM	14ª CIA PM IND	191,93	194,23	1,20
15ª RPM	19ª BPM	139,73	146,08	4,54
15ª RPM	24ª CIA PM IND	210,93	220,91	4,73
15ª RPM	44ª BPM	136,25	143,07	5,01
16ª RPM	28ª BPM	141,9	151,34	6,65
16ª RPM	45ª BPM	300,02	305,18	1,72
17ª RPM	20ª BPM	113,24	122,29	7,99
17ª RPM	56ª BPM	86,55	94,09	8,71
17ª RPM	57ª BPM	43,00	53,85	25,23
17ª RPM	59ª BPM	100,65	106,65	5,96
18ª RPM	12ª BPM	202,43	208,91	3,20
18ª RPM	29ª BPM	128,85	134,76	4,59
18ª RPM	43ª BPM	130,94	136,92	4,57
18ª RPM	64ª BPM	160,78	164,81	2,51
19ª RPM	11ª CIA PM IND	413,66	415,75	0,51
19ª RPM	25ª BPM	398,6	404,2	1,40

4 Considerações finais

A aplicação de correção bayseana para taxas de crimes, seja de crimes violentos ou não, se mostrou bastante viável no contexto de incidência criminal em regiões ou localidades com pequena população, onde a incidência ou ocorrência de um evento criminal, traz impacto significativo na taxa calculada.



Neste contexto é viável interpretar a taxa bayseana como o risco que aquela população está submetida. Dessa forma é possível planejar ações com base no risco, uma vez que flutuações grande no valor da taxa bruta de crimes violentos pode conduzir a tomada de decisões desnecessárias ou desproporcional.

Agradecimentos

Os autores agradecem à Polícia Militar de Minas Gerais(PMMG), através do Centro Integrado de Defesa Social (CINDS), pelo apoio e cessão dos dados para realização deste estudo.

Referências

- [1] A. E. dos Santos, A. L. Rodrigues, and D. L. Lopes, “Aplicações de estimadores bayesianos empíricos para análise espacial de taxas de mortalidade.,” in *GeoInfo*, pp. 300–309, 2005.
- [2] A. L. Andrade, A. M. V. Monteiro, C. Barcellos, E. Lisboa, L. M. W. Acosta, M. C. d. M. Almeida, M. R. V. Brito, M. S. Carvalho, M. A. d. Santos, O. Cruz, *et al.*, “Introdução à estatística espacial para a saúde pública,” 2007.
- [3] R. J. Marshall, “Mapping disease and mortality rates using empirical bayes estimators,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 40, no. 2, pp. 283–294, 1991.
- [4] G. M. Howe *et al.*, “National atlas of disease mortality in the united kingdom,” *National Atlas of disease mortality in the United Kingdom. Revised and enlarged edition.*, 1970.
- [5] B. Efron and C. Morris, “Stein’s estimation rule and its competitors—an empirical bayes approach,” *Journal of the American Statistical Association*, vol. 68, no. 341, pp. 117–130, 1973.
- [6] G. Casella and J. T. Hwang, “Limit expressions for the risk of james-stein estimators,” *Canadian Journal of Statistics*, vol. 10, no. 4, pp. 305–309, 1982.
- [7] D. Clayton and J. Kaldor, “Empirical bayes estimates of age-standardized relative risks for use in disease mapping,” *Biometrics*, pp. 671–681, 1987.



- [8] R. A. Stone, “Investigations of excess environmental risks around putative sources: statistical problems and a proposed test,” *Statistics in medicine*, vol. 7, no. 6, pp. 649–660, 1988.
- [9] R. K. Tsutakawa, G. L. Shoop, and C. J. Marienfeld, “Empirical bayes estimation of cancer mortality rates,” *Statistics in medicine*, vol. 4, no. 2, pp. 201–212, 1985.
- [10] T. Leonard, “Bayesian methods for binomial data,” *Biometrika*, vol. 59, no. 3, pp. 581–589, 1972.
- [11] N. Laird, “Nonparametric maximum likelihood estimation of a mixing distribution,” *Journal of the American Statistical Association*, vol. 73, no. 364, pp. 805–811, 1978.
- [12] N. Miyawaki and S.-C. Chen, “A statistical consideration on the mapping of mortality,” *Social Science & Medicine. Part D: Medical Geography*, vol. 15, no. 1, pp. 93–101, 1981.
- [13] C. R. Thiersch, M. G. d. Andrade, M. F. B. Moreira, and S. Loibel, “Estimativa da relação hipsométrica em clones de eucalyptus sp. com o modelo de curtis ajustado por métodos bayesianos empíricos,” *Revista árvore*, vol. 37, no. 1, pp. 01–08, 2013.
- [14] R. M. Assunção, S. M. Barreto, H. L. Guerra, and E. Sakurai, “Mapas de taxas epidemiológicas: uma abordagem bayesiana,” *Cadernos de Saúde Pública*, vol. 14, no. 4, pp. 713–723, 1998.
- [15] M. Choynowski, “Maps based on probabilities,” *Journal of the American Statistical Association*, vol. 54, no. 286, pp. 385–388, 1959.
- [16] S. L. C. d. Silva, J. M. G. Fachel, S. K. Kato, and S. L. Bassanesi, “Visualização dos padrões de variação da taxa de mortalidade infantil no rio grande do sul, brasil: comparação entre as abordagens bayesiana empírica e totalmente bayesiana,” *Cadernos de Saúde Pública*, vol. 27, pp. 1423–1432, 2011.
- [17] A. X. Y. de Carvalho, G. D. M. da Silva, G. R. de Almeida Júnior, P. H. de Melo Albuquerque, *et al.*, “Mapeamento de taxas bayesianas, com aplicação ao mapeamento de homicídios,” tech. rep., 2011.
- [18] C. C. Beato Filho, R. M. Assunção, B. F. A. d. Silva, F. C. Marinho, I. A. Reis, and M. C. d. M. Almeida, “Conglomerados de homicídios e o tráfico de drogas em belo horizonte, minas gerais, brasil, de 1995 a 1999,” *Cadernos de Saúde Pública*, vol. 17, pp. 1163–1171, 2001.



- [19] T. A. de Freitas, A. F. Cadaval, and G. A. Gonçalves, “A estimação de um índice geral de criminalidade para os municípios do rio grande do sul-igricime rs,” *Ensaio FEE*, vol. 38, no. 3, pp. 499–520, 2017.
- [20] J. K. S. d. Santos, “Índice geral de criminalidade bayesiano para o estado de sergipe,” 2019.
- [21] P. R. Scalco, A. L. Amorim, and A. P. Gomes, “Eficiência técnica da polícia militar em minas gerais,” *Nova Economia*, vol. 22, no. 1, pp. 165–190, 2012.
- [22] A. M. L. de Barros, C. Zanotelli, E. G. Oliveira, E. C. Raizer, M. de Castro, and P. S. Lira, “Atlas da criminalidade violenta da grande vitória-es,”
- [23] P. T. G. de Oliveira and G. Rosa, “Preservação da memória documental: a criação do reds como instrumento de defesa social e integração institucional em minas gerais,” *REVISTA BRASILEIRA DE SEGURANÇA PÚBLICA-REBESP*, vol. 7, no. 2, 2014.