

Indicador Geral de Criminalidade do Estado do Rio de Janeiro - IGCcrime RJ

NOTA METODOLÓGICA

Desenvolvido por:

Nicole Peçanha do Rego Barros
Cristiane Ramos Justen
Fellipe Figueiredo Silva
Mônica Licht Rocha
José Francisco Moreira Pessanha

Coordenadoria de Análises, Diagnósticos e Geoprocessamento – CADG / MPRJ em Mapas

Setembro, 2017

1. Metodologia de Construção do IGCrim RJ

A construção dos índices tem por objetivo subsidiar e auxiliar na tomada de decisão no que tange à atuação dos entes públicos de apoio à segurança do Estado do Rio de Janeiro. Dessa forma, tendo conhecimento do nível de criminalidade geral das regiões que compõe o Estado, tanto os responsáveis pela segurança pública quanto a população poderão ter uma atuação mais preditiva e até repressiva, por parte da sociedade.

Para a realização deste estudo foram consideradas, as informações de delitos, disponibilizados pelo Instituto de Segurança Pública do Estado do Rio de Janeiro (ISP)¹, informações populacionais oriundas do Instituto Brasileiro de Geografia e Estatística (IBGE) e dados de frota disponibilizados pelo Departamento Nacional de Trânsito (DENATRAN)². A análise se deu no período de 2012 a 2016.

Ressalta-se que as referidas variáveis foram obtidas para cada uma das 81 comarcas³ do Estado do Rio de Janeiro, conforme disposto no ANEXO I. A escolha por comarcas deve-se ao fato dos dados do ISP serem informados por Departamento de Polícia (DP), haja vista que nem todos os municípios são contemplados com instalação de DP.

Tendo em vista a temporalidade dos dados analisados, foram utilizados os dados de população projetados, disponibilizados pelo Departamento de Informática do SUS (DATASUS)⁴, de modo a capturar a proporcionalidade populacional dos municípios no decorrer dos anos abordados.

No que tange a variável frota, observou-se que havia três diferentes categorias: veículo, carga e coletivo. Para a categoria de veículos foram considerados automóveis, motoneta e motocicletas. A categoria de carga foi considerada caminhões, caminhonete, camioneta. Já para coletivos foi considerado micro-ônibus e ônibus. Todas as informações foram extraídas do DENATRAN.

Em relação aos dados de segurança pública, o ISP disponibiliza diversos tipos de delitos e variantes, como por exemplo, registro de ocorrência, apreensão de adolescente infrator e apreensão de adolescente por medida sócia educativa. No entanto, alguns destes não representam um delito propriamente dito e sim como uma medida de cumprimento, não fazendo, portanto, coerência considerá-los na construção do índice. A Tabela 1 apresenta as variáveis que foram consideradas neste estudo.

Tabela 1 – Variáveis selecionadas

Variáveis	
Homicídio doloso	Roubo a coletivo
Lesão corporal seguida de morte	Roubo aparelho celular
Latrocínio	Furto de veículos
Tentativa de homicídio	Extorsão
Lesão corporal dolosa	Estelionato
Homicídio culposos	Ameaça
Lesão corporal culposa	Estupro

¹ <http://www.isp.rj.gov.br/dadosoficiais.asp>;

² <http://www.denatran.gov.br/>

³ Divisão territorial do estado em municípios e áreas judiciárias e organização administrativa em núcleos regionais.

⁴ <http://datasus.saude.gov.br/informacoes-de-saude/tabnet>

Roubo de veículos	Apreensão de drogas
Roubo de carga	Apreensão de armas
Roubo a transeunte	

Um ponto muito relevante neste estudo trata da divisão geográfica dos batalhões de polícia nos municípios. Segundo o ISP, as Áreas Integradas de Segurança Pública (AISP) foram criadas como parte de uma política de segurança pública que tinha por objetivo estreitar a ligação entre as Polícias Civil e Militar, e estas com as comunidades abrangidas pelas AISP.

O contorno geográfico de cada AISP foi estabelecido a partir da área de atuação de um batalhão de Polícia Militar (BPM) e as circunscrições das delegacias de Polícia Civil (DP) contidas na área de cada batalhão. A atual divisão territorial do Estado do Rio de Janeiro, segundo o critério de Áreas Integradas de Segurança Pública, contempla um total de 39 AISP, que visou adequar os limites geográficos de atuação das unidades da Polícia Civil e Militar, de forma a torná-las compatíveis aos objetivos da gestão territorial da segurança pública segundo Regiões Integradas de Segurança Pública (RISP), Áreas Integradas de Segurança Pública (AISP) e Circunscrições Integradas de Segurança Pública (CISP).

A metodologia deste trabalho envolveu a criação de um indicador geral de criminalidade com base em 19 tipos de variáveis associadas aos crimes disponíveis na Secretaria de Segurança Pública do Estado Rio do Rio de Janeiro, abarcando os 92 municípios contemplados em 81 comarcas do Estado.

Para a construção do índice, foram consideradas as informações anuais de delitos e suas *proxys*⁵, disponibilizados pelo Instituto de Segurança Pública do Estado do Rio de Janeiro (ISP)⁶, além das informações populacionais oriundas do Instituto Brasileiro de Geografia e Estatística (IBGE) e dos dados de frotas disponibilizados pelo Departamento Nacional de Trânsito (DENATRAN)⁷, com periodicidade anual contemplando os anos de 2012 a 2016.

Com a informação da população de cada comarca (i) e com a incidência de cada tipo de crime (K), calculou-se a quantidade de crimes para cada 10.000 habitantes (C_{ki}). Para os crimes de Roubo de Veículo, Furto de Veículo, Roubo de Carga e Roubo a Coletivo utilizou-se a frota total de veículos por 10.000 veículos correspondentes a cada comarca. A principal finalidade da taxa por 10 mil é permitir a comparação entre locais com diferentes tamanhos de população e neutralizar o crescimento populacional, permitindo a comparação a médio e longos prazos.

Em seguida, foram obtidas as taxas mínimas (C_{kmin}) e máximas (C_{kmax}) em relação ao total registrado para cada um dos delitos. Estas foram definidas observando-se a menor e a maior taxa de ocorrência por 10 mil de cada tipologia criminal durante o período analisado. Estas taxas serão parâmetros entre os limites inferiores e superiores de referência para as amplitudes mínimas e máximas que poderão apresentar. Subtrai-se o crime ocorrido no município do crime mínimo entre todos os municípios. Posteriormente, se divide o resultado pela diferença entre os parâmetros máximo e mínimo. De forma algébrica tem-se:

$$TC_{ki} = \frac{(C_{ki} - C_{kmin})}{(C_{kmax} - C_{kmin})}$$

⁵ Variáveis de que expressam um fato não mensurável.

⁶ <http://www.isp.rj.gov.br/dadosoficiais.asp>;

⁷ <http://www.denatran.gov.br/>

Onde:

TCKi é a taxa de ocorrência do crime K na comarca i ponderado pelas amplitudes de máximo e mínimo de cada crime;

Cki é o crime ocorrido na cidade i para cada 10.000 habitantes/frota;

Ckmin é o menor taxa de crime observado em relação às 81 comarcas;

Ckmax é a maior taxa de crime observada em relação às 81 comarcas.

Na forma como o resultado desta fase se encontrava quanto mais próximo de zero fosse o valor, melhor o resultado, ou seja, menor a ocorrência de criminalidade. A seguir buscou-se estabelecer um critério de ponderação para a agregação de todas os 19 tipos de crimes em um único indicador. Diferente de Monteiro (2009) e Soares, Zabet e Ribeiro (2011) o caminho escolhido não foi análise fatorial. Buscou-se utilizar na metodologia deste trabalho um critério que correspondesse, pelo menos em parte, a forma como a sociedade brasileira classifica os crimes quanto a sua nocividade. Deste modo, o critério utilizado foi o ponto médio das penas dos crimes correspondentes ao Código Penal Brasileiro. Assim, há uma endogeneidade do parâmetro ponderador do peso que cada crime tem no IGCcrime definido pela própria sociedade que é fragilizada pela criminalidade.

A tabela a seguir mostra as penas mínimas e máximas praticadas no Brasil segundo o Código Penal na sua forma mais ampla, para as 19 variáveis de crimes da base de dados utilizada. Como exemplo, pode-se notar que o indivíduo que cometer um latrocínio pode ser sentenciado a um período de 20 a 30 anos, enquanto que aquele que pratica um estupro a pena é de 6 a 10 anos.

Tabela 2 - Penas por tipos de crimes praticados no Brasil segundo o Código Penal Brasileiro

Delitos	Penas em anos		Ponto Médio	Peso do crime no Índice (%)
	Mínimo	Máximo		
Homicídio Doloso	6	20	17	0,142
	12	30		
Lesão Corporal seguida por morte	4	12	8	0,067
Latrocínio	20	30	25	0,209
Tentativa de Homicídio	2	6,67	4,33	0,036
Lesão Corporal dolosa	0,25	1	0,63	0,005
Homicídio Culposos	1	3	2	0,017
Lesão Corporal Culposa	0,25	1	0,625	0,005
Roubo de veículos	4	10	7	0,58
Roubo de carga				
Roubo a transeunte				
Roubo a coletivo				
Roubo aparelho celular				
Furto de veículos	1	4	2,5	0,021
Extorsão	4	10	7	0,058
Estelionato	1	5	3	0,025
Ameaça	0,08	0,25	0,17	0,001
Estupro	6	10	8	0,067
Apreensão de drogas	5	15	10	0,084
Apreensão de armas	3	6	4,5	0,038
Total			119,75	

Fonte: Elaborado pelos autores com base no Código Penal Brasileiro.

Com definição dos pesos atribuídos a cada crime que compõe o índice, o passo seguinte foi tratar o componente populacional dos municípios do Rio de Janeiro. Devido a grande heterogeneidade na

distribuição populacional entre os municípios observou-se que grandes cidades concentravam uma parcela significativa do total dos crimes praticados no estado, a ocorrência de cada evento impacta de forma diferente sobre os índices de criminalidade de cidades por porte diferentes. A proporção representada pela ocorrência de um crime por 10 mil habitantes é substancialmente diferente entre os municípios de grandes e pequenos. Neste sentido, a ocorrência aleatória de um crime em um município de pequena população tem grande impacto, podendo gerar a informação de que o índice de criminalidade no município é elevado em determinado período, justificando a aplicação de recursos para o aumento do nível de segurança pública na área. No entanto, se o evento não se repete ao longo do tempo, não haveria necessidade de mobilizar tais recursos, uma vez que o evento é aleatório.

Com o intuito de se retirar a hipótese de aleatoriedade sobre as estimativas de crimes aplicou-se uma taxa bayesiana para as 81 comarcas do Estado do Rio de Janeiro. Buscou-se, assim, incorporar os riscos contidos em outras áreas, neste caso cidades, para estimar o risco de uma cidade específica. O método proposto por Marshall (1991) foi utilizado aqui e consiste em calcular uma taxa de risco ponderada por dois elementos, um contendo o evento ocorrido ponderado por uma constante c entre zero e um e, somado a outro elemento cuja constante é o complementar de c multiplicado pela taxa média dos eventos ocorridos em cidades de mesmo porte. Em termos de equação a taxa de risco de uma cidade segue a seguinte regra:

$$q = ct + (1 - c)m$$

Onde:

q = estimativa da taxa corrigida do crime;

c = parâmetro que amortece o valor do crime ocorrido no município quando a população é pequena. Para os menores municípios da amostra, a constante c tende a zero e para os maiores, o c tende a 1;

t = taxa bruta do crime ocorrido no município;

m = taxa de ocorrência média dos crimes obtida entre os municípios de população semelhantes.

Para a estimação da constante c , é necessária uma definição de classes populacional em que se dividirão as 81 comarcas. Para tanto, foi utilizado uma análise estatística de agrupamentos, *K-means*⁸, vide em anexo, onde foram indicados 5 classes de divisão populacional. A primeira classe apresentou 48 municípios com o c igual a 0,0121, as classes subsequentes receberam valores conforme a distribuição média da população contida em cada uma das classes. A última classe, recebeu o valor de c igual a 1 contemplando 4 municípios, cujos apresentaram população entre 867.067 a 6.390.290 sendo que esta última correspondia a população da capital do Estado do Rio de Janeiro, conforme exposto na Tabela 3.

Tabela 3 – Caracterização do valor atribuído ao grau de aleatoriedade através do coeficiente c e distribuição da população dos municípios

Classes K	Faixa de População		Quantidade Municípios	População Média por Faixa (A)	Participação de A/B
Classe 1	7.424	56.436	48	28.032	0,0121
Classe 2	72.679	157.986	15	107.438	0,0465
Classe 3	167.622	297.192	10	213.070	0,0922
Classe 4	460.062	491.807	4	474.691	0,2054
Classe 5	867.067	6.390.290	4	B = 2.311.192	1,0000

⁸ K-means é uma metodologia de análise multivariada

Fonte: Elaborado pelos autores

Conforme apresentado na tabela acima, a distribuição populacional dos municípios do Rio de Janeiro apresenta grande variabilidade, sendo assim, o ajuste bayesiano pode minimizar esta aleatoriedade principalmente quando de um impacto pontual de crimes. Logo, se os crimes ocorridos no município em análise corresponderem à média de crimes em cidades de mesmo porte, então a ocorrência neste município está seguindo um padrão. Do contrário, o critério bayesiano realiza o ajuste de acordo com o peso preestabelecido para o porte do município.

De acordo com a equação de Marshall (1991) quando c é igual a um implica que se está na classe de municípios de maior população e cuja taxa média à classe que pertencem não é aplicada a eles ($c = 1$). Já no outro extremo têm-se os municípios de menor população e para os quais, portanto, utiliza-se o valor de c próximo à zero. Isto significa que se utiliza a taxa média dos eventos ocorridos na classe do município para gerar a taxa de risco deste evento no município em análise. O IGC_{Crime} para cada comarca pode ser descrito através da seguinte expressão.

$$IGCrime_i = \frac{\sum[\gamma_i] \times [IGC_j]}{\max(IGCrime_i)}$$

Onde:

IGCrime_i é o índice de criminalidade geral para a comarca i resultante do peso da pena γ_i aplicado a cada uma das j tipologias de crimes;

γ_i é o peso da pena aplicado ao respectivo crime conforme a Tabela 1, matriz transposta 1x19;

IGC_j é o índice de criminalidade calculado para cada tipologia, matriz coluna 1x19.

Desta forma obteve-se o IGC_{Crime} de cada uma das 19 tipologias criminais para cada comarca.

2. Anexos – Análise de Agrupamentos – *K-means*

O método k-means é sistematizado por meio de um algoritmo matemático. Cada elemento amostral é colocado no cluster cujo centróide é o mais próximo dele. Originalmente, este método é composto por quatro passos. Primeiro são escolhidos k centróides (sementes) para iniciar o processo, seguidamente, cada elemento do conjunto de dados é comparado com cada centróide inicial através de uma medida de distância (em geral euclidiana). O elemento é alocado ao grupo cuja distância é menor. Depois de aplicar o primeiro e o segundo passo a cada um dos elementos do conjunto de dados, recalculam-se os valores dos centróides para cada grupo novo formado. Esse processo é repetido até que nenhuma realocação seja necessária.

Para a realização da primeira etapa do processo, faz-se necessário a escolha das sementes, conforme dito anteriormente. Para isso, existem alguns procedimentos que podem ajudar nesta decisão (MINGOTI, pag. 192). Tais como:

- Uso das técnicas hierárquicas aglomerativas – Utiliza-se algum dos métodos hierárquicos para a obtenção de k grupos iniciais. Calcula-se a média de cada grupo formado. Essas médias serão as sementes iniciais;
- Escolha aleatória – As k sementes são escolhidas aleatoriamente dentro do conjunto de dados a ser analisado;
- Uso de uma variável aleatória – Escolhe-se a variável aleatória de maior variância entre os p componentes de um vetor aleatório X. Divide-se o domínio da variável em k intervalos e a semente inicial será o centróide de cada intervalo;
- Uso de valores discrepantes – Através de uma análise estatística, selecionam-se k elementos discrepantes no conjunto de dados, que serão as k sementes;
- Escolha prefixada – Sementes escolhidas arbitrariamente pelo pesquisador (quando já se conhece o problema ou esteja buscando validar uma solução já existente);
- Escolha dos k primeiros valores do banco de dados – A maioria dos softwares usa como default as k primeiras observações do banco de dados, se as sementes não forem especificadas.

A seleção de entre método hierárquicos e métodos não hierárquicos é baseada em algumas considerações importantes. As soluções de Cluster Hierárquico são preferidas quando são analisadas alternativas de agrupamento e também quando o tamanho da amostra é moderado, em torno de 300 a 1000 observações ou uma sub-amostra de um grande banco de dados. Já os métodos não hierárquicos são preferíveis quando o número de grupos já é conhecido e as sementes podem ser especificadas por conhecimentos práticos, além disso, quando se trabalha com grande base de dados.

No entanto, também é muito recomendada uma combinação entre métodos hierárquicos e métodos não hierárquicos, na qual um procedimento hierárquico pode ser utilizado para selecionar o número de grupos e um vetor formado pelos centros destes grupos gerados pode ser usado como sementes para o procedimento não hierárquico. Sendo assim, o método não hierárquico seria utilizado para agrupar todas as observações através das sementes encontradas anteriormente, assim produziria grupos com membros mais preciso.

Por fim, vale ressaltar que não existem, em nenhuma literatura, informações ou regras de um procedimento simples e objetivo para determinar o número correto de grupos. Preferivelmente, opta-se por avaliar diversas soluções alternativas de grupos, na busca da quantidade de grupos ideal.