

ANÁLISE BAYESIANA DE CLASSES LATENTES APLICADA À CRIMINALIDADE NO ESTADO DO RIO DE JANEIRO

Matheus Jun Onishi da Silva¹

Jony Arrais Pinto Junior²

Resumo

A criminalidade no estado do Rio de Janeiro se caracteriza por uma diversidade de delitos distribuídos de forma desigual entre suas regiões. Este estudo utilizou a Análise de Classes Latentes (LCA), sob a perspectiva bayesiana, para analisar os dados de criminalidade deste estado entre 2020 e 2023, com o objetivo de identificar perfis ocultos de incidência criminal e agrupar as Circunscrições Integradas de Segurança Pública (CISPs) com base em padrões semelhantes. A segmentação revelou duas classes distintas: uma com alta e outra com baixa criminalidade, com maior concentração de delitos na Região Metropolitana e em parte do interior. A análise espacial evidenciou que áreas com baixa incidência formam clusters mais coesos, enquanto regiões de alta criminalidade apresentaram maior dispersão geográfica. Os resultados mostram que a LCA é uma ferramenta eficaz para subsidiar o diagnóstico e o planejamento de políticas de segurança pública diante da complexidade dos padrões criminais. **Palavras-chave:** Análise de Classes Latentes, Inferência Bayesiana, Criminalidade.

Abstract

Crime in the state of Rio de Janeiro is characterized by a variety of offenses unevenly distributed across regions. This study applied Latent Class Analysis (LCA), from a Bayesian perspective, to crime data from 2020 to 2023 with the aim of identifying hidden profiles of criminal incidence and grouping the Integrated Public Security Districts (CISPs) based on similar patterns. The segmentation revealed two distinct classes: one with high and the other with low crime rates, with offenses concentrated mainly in the Metropolitan Region and parts of the interior. Spatial analysis showed that low-incidence areas tend to form more cohesive clusters, while high-crime regions exhibited greater geographic dispersion. The findings indicate that LCA is an effective tool to support the diagnosis and strategic planning of public security policies in the face of complex criminal patterns. **Keywords:** Latent Class Analysis; Bayesian Inference; Crime.

Introdução

A criminalidade no Rio de Janeiro abrange uma diversidade de delitos, desde crimes violentos até furtos e crimes patrimoniais, distribuídos de forma heterogênea entre as regiões.

¹Universidade Federal Fluminense (UFF), majun@id.uff.br

²Universidade Federal Fluminense (UFF), jarrais@id.uff.br

O Instituto de Segurança Pública (ISP-RJ) registra essas ocorrências por tipologias, mas a análise isolada dessas categorias pode ocultar padrões complexos, em que diferentes crimes coexistem, formando padrões criminais específicos.

Nesse contexto, a **Análise de Classes Latentes** (LCA) surge como uma ferramenta eficaz para identificar grupos ocultos com base em múltiplas tipologias criminais, permitindo agrupar regiões com perfis semelhantes de incidência de diversos tipos de crimes. A aplicação da LCA sob a ótica bayesiana, por meio do amostrador de Gibbs, reforça essa capacidade analítica ao incorporar a incerteza dos dados e fornecer estimativas probabilísticas mais robustas, facilitando o diagnóstico e o monitoramento das dinâmicas criminais no estado. Parte superior do formulário

Objetivo

Este estudo tem por objetivo identificar classes de criminalidade, formadas por múltiplos tipos de delitos, por meio da Análise de Classes Latentes (LCA) sob a ótica bayesiana, aplicada aos dados do estado do Rio de Janeiro entre 2020 e 2023. A proposta é revelar perfis ocultos de incidência criminal e subsidiar o planejamento de políticas públicas. A metodologia adotada possibilita a segmentação das Circunscrições Integradas de Segurança Pública (CISPs) em grupos de alta e baixa criminalidade, oferecendo uma visão integrada das dinâmicas criminais no estado.

Material e Método

A área de estudo compreende o estado do Rio de Janeiro, dividido em 137 Circunscrições Integradas de Segurança Pública (CISPs), no período de 2020 a 2023. Para cada CISP, calculou-se o total acumulado de crimes violentos, crimes de trânsito, roubos, furtos, crimes contra o patrimônio, atividades policiais e outros delitos. As ocorrências foram posteriormente binarizadas, classificando-se como de alta incidência ou de baixa incidência.

Para identificar possíveis classes de criminalidade e estimar a probabilidade de cada CISP pertencer a essas classes, foi utilizado o modelo de Análise de Classes Latentes. Nesse modelo, uma variável latente C representa a heterogeneidade não observável entre subgrupos populacionais, sendo amplamente adotado para variáveis categóricas (LI et al., 2018). Cada observação $\mathbf{Y}_i$, de dimensão $K \times 1$, corresponde ao vetor de respostas da unidade de estudo (CISPs) i a K variáveis observáveis. O modelo LCA assume independência condicional entre essas variáveis, dado $C = j$.

Nesse estudo, todas as variáveis Y_{ik} foram tratadas como dicotômicas, uma escolha frequente em aplicações práticas que visa facilitar a implementação computacional dos modelos bayesianos. A distribuição condicional de Y_{ik} , dado que o indivíduo pertence à classe j , é dada por:

$$P(Y_{ik} = y_{ik} | C = j) = \theta_{jk}^{y_{ik}} (1 - \theta_{jk})^{1-y_{ik}},$$

em que θ_{jk} representa a probabilidade de uma CISP ter alta ocorrência na criminalidade k , condicional à classe j . A contribuição da observação i para a função de verossimilhança do modelo LCA com G classes latentes é dada por:

$$P(\pi, \theta \vee Y_i = y_i) = \sum_{j=1}^G \pi_j \prod_{k=1}^K \theta_{jk}^{y_{ik}} (1 - \theta_{jk})^{1-y_{ik}},$$

em que π_j é a probabilidade da variável latente C assumir a classe j , $P(C = j)$. O objetivo da modelagem com LCA é estimar o vetor de probabilidades de pertencimento às classes latentes $\pi = (\pi_1, \dots, \pi_G)$ e as probabilidades condicionais $\theta_j = (\theta_{j1}, \dots, \theta_{jK})$.

Sob a ótica bayesiana, considera-se que a variável latente C seja representada por um vetor one-hot, $C_i = (c_{i1}, c_{i2}, \dots, c_{iG})$, tal que c_{ij} indica se a CISP i pertence à classe j . A estimação dos parâmetros foi realizada via amostrador de Gibbs utilizado no software R (COLLINS; LANZA, 2010), adotando distribuições a priori: $\pi \sim \text{Dirichlet}(\gamma)$ e $\theta_{jk} \sim \text{Beta}(\alpha_{jk}, \beta_{jk})$, o que facilita a obtenção das distribuições condicionais completas.

$$\pi \vee \theta, y, c \sim \text{Dirichlet}\left(\sum_{i=1}^N c_{i1} + \gamma_1, \sum_{i=1}^N c_{i2} + \gamma_2; \dots; \sum_{i=1}^N c_{iG} + \gamma_G\right);$$

$$\theta_{jk} \vee \pi, y, c \sim \text{Beta}\left(\alpha_{jk} + \sum_{i=1}^N y_{ik} c_{ij}; \beta_{jk} + \sum_{i=1}^N (1 - y_{ik}) c_{ij}\right);$$

$$C_i \vee \cdot \sim \text{Multinomial}\left(1; \frac{\pi_1 \prod_{k=1}^K \theta_{1k}^{y_{ik}} (1 - \theta_{1k})^{1-y_{ik}}}{\sum_{j=1}^G \pi_j \prod_{k=1}^K \theta_{jk}^{y_{ik}} (1 - \theta_{jk})^{1-y_{ik}}}, \dots, \frac{\pi_G \prod_{k=1}^K \theta_{Gk}^{y_{ik}} (1 - \theta_{Gk})^{1-y_{ik}}}{\sum_{j=1}^G \pi_j \prod_{k=1}^K \theta_{jk}^{y_{ik}} (1 - \theta_{jk})^{1-y_{ik}}}\right).$$

Resultados e Discussão

Os 7 diferentes tipos de crimes foram dicotomizados com base no primeiro quartil de suas distribuições: uma CISP foi classificada como de alta incidência para um determinado crime se o total de registros estivesse entre os 75% maiores valores observados para aquele tipo de ocorrência. A aplicação do modelo LCA bayesiano resultou na identificação de duas classes latentes bem definidas. A Classe 1, que abrange a maioria das CISPs (72,62%), apresenta altas probabilidades de ocorrência em todas as categorias de crime, sendo, portanto, associada a um perfil de alta criminalidade. Por outro lado, a Classe 2 concentra regiões com baixas probabilidades de ocorrência, indicando menor vulnerabilidade criminal. Essa segmentação evidencia a existência de padrões criminais distintos entre as regiões, fornecendo subsídios relevantes para o planejamento e a priorização de políticas públicas de segurança.

Tabela 1: Probabilidades das Classes Latentes e Condicionais com Intervalos de Credibilidade (95%)

Parâmetros	Classe 1	Classe 2	IC classe 1	IC classe 2
Probabilidade da Classe (π_j)	0,73	0,27	[0,65;0,80]	[0,19;0,34]
Crimes Violentos (θ_{j1})	0,97	0,13	[0,93;0,99]	[0,02;0,23]
Crimes de Trânsito (θ_{j2})	0,97	0,14	[0,93;0,99]	[0,03;0,24]
Roubos (θ_{j3})	0,94	0,21	[0,89;0,98]	[0,08;0,33]
Furtos (θ_{j4})	0,97	0,13	[0,93;0,99]	[0,03;0,24]
Crimes contra o Patrimônio (θ_{j5})	0,98	0,11	[0,95;0,99]	[0,02;0,20]
Atividade Policial (θ_{j6})	0,91	0,28	[0,85;0,96]	[0,14;0,42]
Outros Registros (θ_{j7})	0,97	0,13	[0,93;0,99]	[0,03;0,23]

Fonte: Elaborado pelo autor

Com base nas estimativas apresentadas na Tabela 1, foram calculadas as probabilidades a posteriori de pertencimento, $C \vee Y, \pi, \theta$, de cada CISP pertencer à Classe 1, caracterizada pela alta incidência conjunta de diferentes tipos de delitos. A Figura 1 apresenta a distribuição espacial dessas probabilidades, permitindo visualizar o agrupamento das CISPs com perfis criminais semelhantes ao longo do estado do Rio de Janeiro.

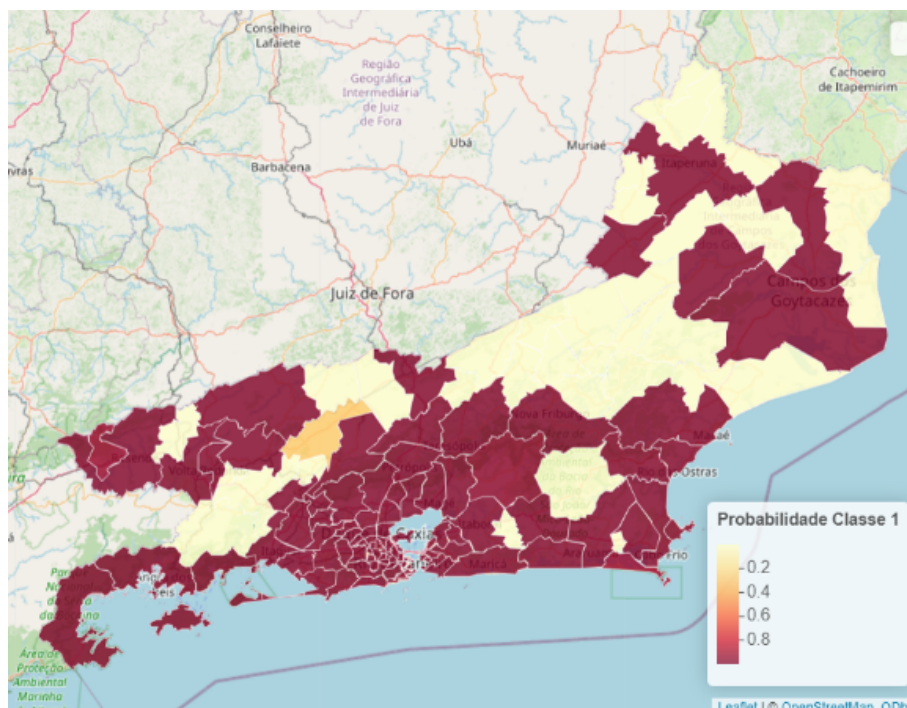


Figura 1: Distribuição espacial das probabilidades a posteriori de pertencimento à Classe 1, associada a alta incidência de criminalidade entre 2020 à 2023 nas CISPs do estado do Rio de Janeiro.

Fonte: Elaborado pelo autor

A Figura 1 revela que a maioria das CISPs apresenta probabilidades a posteriori ex-

tremas, indicando uma segmentação clara entre regiões de alta e baixa criminalidade, sem zonas de transição nítidas entre os perfis latentes. Destacam-se como áreas de baixa incidência as regiões do norte fluminense e da Serra, enquanto a Região Metropolitana — especialmente a capital e a Baixada Fluminense — concentra os maiores níveis de criminalidade.

Conclusão

Diante da complexidade dos padrões criminais no estado do Rio de Janeiro, a Análise de Classes Latentes (LCA) mostrou-se uma abordagem adequada para identificar perfis multivariados de criminalidade, permitindo segmentar as CISP em grupos de alta e baixa incidência de delitos. Embora o modelo utilizado não incorpore diretamente a dependência espacial entre regiões, a análise das probabilidades a posteriori ofereceu uma visão inicial sobre a distribuição geográfica dos perfis criminais. Os resultados reforçam o potencial da LCA como ferramenta exploratória, ao mesmo tempo em que sinalizam a importância da incorporação explícita da estrutura espacial em estudos futuros, visando aprimorar a robustez e a aplicabilidade dos achados. Pesquisas subsequentes podem explorar extensões espaciais da LCA, como os modelos de Classes Latentes Espaciais, para capturar padrões de contiguidade e dependência geográfica, contribuindo de forma mais eficaz para o desenvolvimento de políticas públicas de segurança com base em evidências.

Referências

COLLINS, Linda M.; LANZA, Stephanie T. Latent class and latent transition analysis: with applications in the social, behavioral, and health sciences. Hoboken: Wiley, 2010.

LI, Y.; LORD-BESSEN, J.; SHIYKO, M.; LOEB, R. Bayesian latent class analysis tutorial. *Multivariate Behavioral Research*, v. 53, n. 3, p. 430–451, 2018.

WHITE, Arthur; MURPHY, Thomas Brendan. BayesLCA: an R package for Bayesian latent class analysis. *Journal of Statistical Software*, v. 61, n. 13, p. 1–28, 2014. URL: <http://www.jstatsoft.org/v61/i13/>.

Anexo

<https://github.com/MaJOS0707/Poster-SER-Matheus-Onishi->