XXVI SEMAT
XXVI Semana da Matemática
XVI SEMEST
XVI Semana da Estatística

20 a 24 de outubro de 2026

Crimes em Série: Modelos INAR para Prever Ocorrências em Uberlândia

Isabella Victoria Theodoro da Silva, Matheus Bartolo Guerrero (orientador)
Universidade Federal de Uberlândia, Instituto de Matemática e Estatística
PET Estatística

1 Introdução

Registros criminais mensais são **contagens**: valores inteiros e não negativos, com meses sem ocorrência, variância diferente da média e associação entre meses consecutivos.

O ARIMA gaussiano [3] tem distribuição preditiva contínua, que pode atribuir probabilidade a valores negativos ou não inteiros; os processos autorregressivos para valores inteiros, **INAR** [4, 5], respeitam esse suporte.

Como a dinâmica difere entre os 853 municípios de Minas Gerais [1], o foco é **Uberlândia**, cujos registros também podem ser consultados no painel interativo do PET Estatística da UFU [2].

Objetivo. Comparar ARIMA, PINAR(1), NB-INAR(1) e Geo-INAR(1) quanto à adequação distributiva e ao desempenho preditivo de um mês à frente em quatro séries mensais de crimes consumados em Uberlândia.

2 Dados

Banco de Crimes Violentos do Portal de Dados Abertos de Minas Gerais (SEJUSP) [7]: ocorrências **consumadas** em Uberlândia, agregadas por crime e mês.

São registros administrativos, não necessariamente a totalidade dos crimes ocorridos.

4	89	83	6
séries mensais, sem lacunas	meses: jan/2019 a mai/2026	meses de treinamento	meses de teste externo

Séries: **estupro**, **estupro de vulnerável**, **roubo** e **sequestro e cárcere privado**.

3 Modelos

No INAR(1) com desbaste binomial, a contagem do mês t depende diretamente apenas da contagem do mês anterior:

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t, \qquad \alpha \circ X_{t-1} = \sum_{j=1}^{X_{t-1}} B_{t,j}, \qquad B_{t,j} \overset{\text{iid}}{\sim} \text{Bernoulli}(\alpha),$$

com $0 \leq \alpha < 1$ e inovação ε_t inteira e não negativa.

Como $\mathrm{E}(X_t \mid X_{t-1} = x) = \alpha x + \mathrm{E}(\varepsilon_t)$, α mede a **persistência do nível** da série.

Crimes são eventos, não um estoque: o desbaste induz a associação entre meses consecutivos, sem significar que ocorrências passem de um mês para o outro.

ARIMA Referência contínua; normal preditiva discretizada nos inteiros não negativos.

PINAR(1) Desbaste binomial e inovações Poisson.

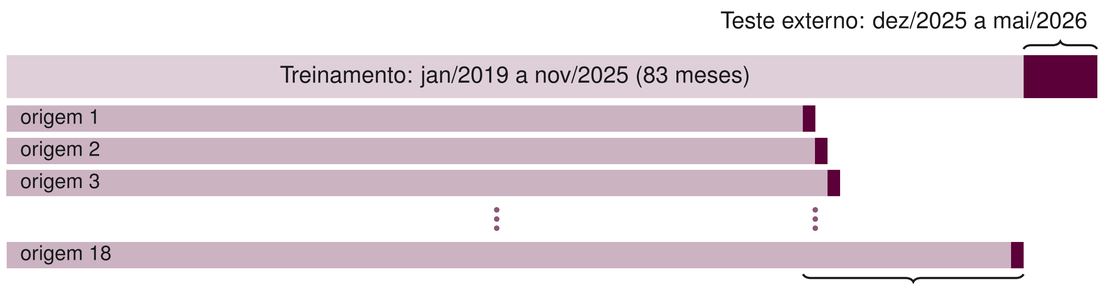
NB-INAR(1) Desbaste binomial e inovações binomiais negativas.

Geo-INAR(1) Marginal e inovações geométricas, com operador próprio [6].

Pela propriedade de Markov, a previsão usa $P(X_t = y \mid X_{t-1} = x)$, $y = 0, 1, 2, \ldots$; a previsão pontual é a **mediana** preditiva e os intervalos vêm dos quantis.

4 Avaliação e seleção

Validação temporal de **origem expansiva** [8]: a cada origem, o modelo é reajustado com os dados disponíveis até ali e prevê o mês seguinte.



Critério principal: **escore logarítmico negativo médio** (menor é melhor),

$$-\frac{1}{m}\sum_{t=1}^{m} \log P\left(X_t = x_t \mid \mathcal{H}_{t-1}\right),$$

em que $\mathcal{H}_{t-1}$ é a informação disponível antes de X_t.

Diferenças menores que um erro-padrão pareado foram tratadas como **empate prático**, decidido por resíduos, calibração, natureza discreta e parcimônia.

A calibração foi avaliada por histogramas **PIT** [9] (82 distribuições, 8 classes); a dependência remanescente, por resíduos de Pearson, autocorrelação residual e teste de Ljung–Box.

5 Resultados

Tabela 1: Características das séries e desempenho dos modelos selecionados.

Série	Média	I	Modelo	$\hat{\alpha}$	Validação	PIT L_1	Teste
Estupro	3,09	1,14	PINAR(1)	0,137	2,143	0,074	2,489
Estupro de vulnerável	7,74	1,36	PINAR(1)	$<$0,001	2,446	0,185	2,984
Roubo	116,18	29,83	Geo-INAR(1)	0,964	4,390	0,142	3,956
Sequestro e cárcere privado	0,85	1,03	PINAR(1)	0,022	1,143	0,073	1,357

I: índice de dispersão. Validação e Teste: escore logarítmico negativo médio.

Estupro — PINAR(1)
Empate com o NB-INAR(1) (diferença 0,002; erro-padrão 0,008); pesaram a parcimônia e o PIT ligeiramente melhor.

Estupro de vulnerável — PINAR(1)
Empate com o NB-INAR(1) (0,030; erro-padrão 0,037). Com $\hat{\alpha} < 0{,}001$, aproxima-se de uma sequência Poisson.

Roubo — Geo-INAR(1)
Superou o ARIMA (diferença 0,162; erro-padrão 0,079), com forte persistência de curto prazo e o menor escore no teste.

Sequestro e cárcere privado — PINAR(1)
Empate: ARIMA com 1,126 e PINAR(1) com 1,143 (erro-padrão 0,044). O PINAR(1) preserva o suporte e tem PIT mais uniforme.

6 Calibração e resíduos

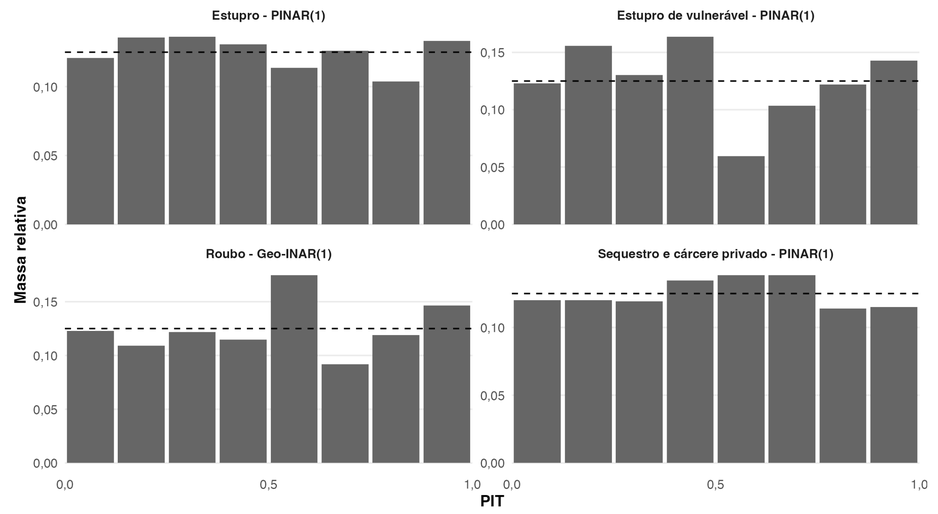


Figura: Histogramas PIT não aleatorizados (treinamento); tracejado: massa sob uniformidade.

Calibração próxima da uniformidade e nenhuma evidência de dependência residual pelo Ljung–Box (valores-p de 0,449 a 0,795). Em roubo, a autocorrelação residual no lag 1 ($-0{,}235$) superou levemente o limite aproximado de 0,216, mas sem significância global (valor-p 0,686).

7 Previsões no teste externo

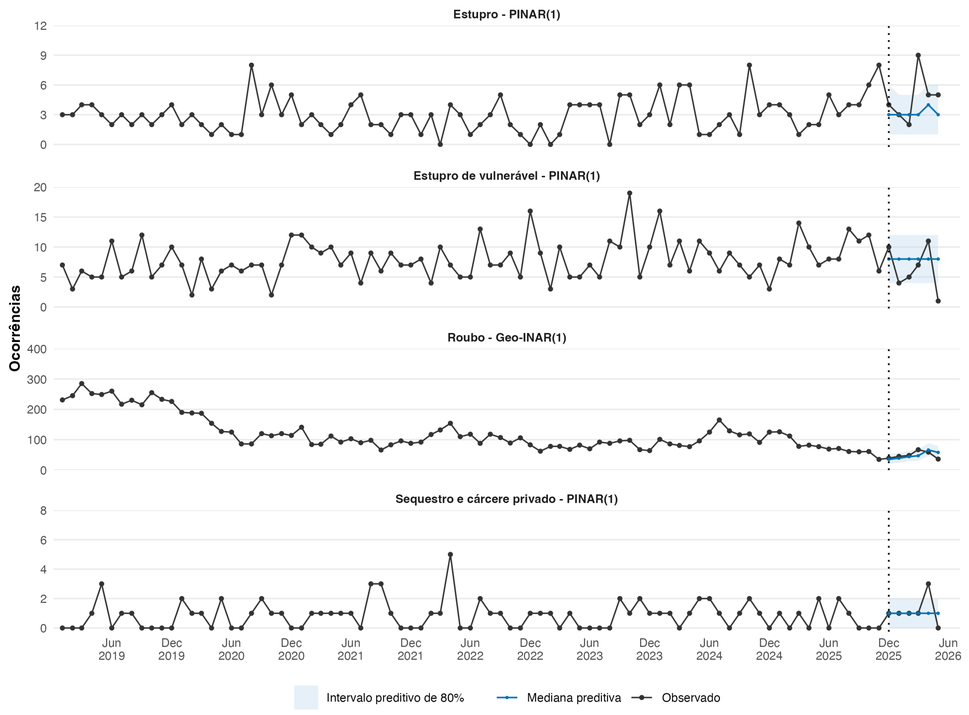


Figura: Observado, mediana preditiva e intervalo preditivo de 80%; a linha vertical marca o início do teste.

Os intervalos de 80% cobriram **5 das 6 observações** em todas as séries.

8 Considerações finais

3 de 4	0,964	5 de 6
séries com o PINAR(1) selecionado	$\hat{\alpha}$ do Geo-INAR(1) em roubo	observações no IP de 80%, em cada série

O melhor modelo depende da dispersão e da persistência de cada série. As previsões de um mês à frente são coerentes com o suporte das contagens; como roubo muda de nível, o Geo-INAR(1) deve ser reestimado à medida que novos dados chegam.

Referências Bibliográficas

[1] IBGE, *Censo Demográfico 2022: população e domicílios: primeiros resultados*, Rio de Janeiro, 2023.
[2] PET Estatística UFU, *Painel interativo sobre a criminalidade violenta em Minas Gerais*, 2024. https://pet-estatistica-ufu.shinyapps.io/crimes-mg/.
[3] P. A. Morettin e C. M. C. Toloi, *Análise de séries temporais: modelos lineares univariados*, 3. ed., Blucher, 2018.
[4] M. A. Al-Osh e A. A. Alzaid, *First-order integer-valued autoregressive (INAR(1)) process*, J. Time Series Analysis, v. 8, n. 3, p. 261–275, 1987.
[5] C. H. Weiß, *An Introduction to Discrete-Valued Time Series*, Wiley, 2018.
[6] M. B. Guerrero, W. Barreto-Souza e H. Ombao, *Integer-valued autoregressive processes with prespecified marginal and innovation distributions: a novel perspective*, Stochastic Models, v. 38, n. 1, p. 70–90, 2022.
[7] SEJUSP-MG, *Crimes Violentos*, Portal de Dados Abertos de MG, 2026. https://www.dados.mg.gov.br/dataset/crimes-violentos.
[8] R. J. Hyndman e G. Athanasopoulos, *Forecasting: Principles and Practice*, 3. ed., OTexts, 2021.
[9] C. Czado, T. Gneiting e L. Held, *Predictive model assessment for count data*, Biometrics, v. 65, n. 4, p. 1254–1261, 2009.

isabella.theodoro@ufu.br · matheusguerrero@ufu.br