

**Modelo generalizado autorregressivo e de
média móveis Bernoulli-Geométrico: um
novo modelo e seu gráfico de controle
Shewhart modificado**

Lucas de Oliveira Ferreira de Sales

TESE APRESENTADA AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA UNIVERSIDADE DE SÃO PAULO
PARA OBTENÇÃO DO TÍTULO DE
DOUTOR EM CIÊNCIAS

Programa: Estatística

Orientadora: Prof^a. Dr^a. Airlane Pereira Alencar

Coorientadora: Prof^a. Dr^a. Linda Lee Ho

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento
de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001

São Paulo
Março de 2023

**Modelo generalizado autorregressivo e de
média móveis Bernoulli-Geométrico: um
novo modelo e seu gráfico de controle
Shewhart modificado**

Lucas de Oliveira Ferreira de Sales

Esta versão da tese contém as correções e
alterações sugeridas pela Comissão Julgadora
durante a defesa da versão original do
trabalho, realizada em 31 de Março de 2023.

Uma cópia da versão original está
disponível no Instituto de Matemática e
Estatística da Universidade de São Paulo.

Comissão julgadora:

Prof^a. Dr^a. Airlane Pereira Alencar (orientadora) – IME - USP

Prof. Dr. André Luís Santos de Pinho – DEST- UFRN

Prof^a. Dr^a Chang Chiann – IME - USP

Prof^a. Dr^a Glaura da Conceição Franco – DEST - UFMG

Prof. Dr. Marcelo Pereira Bourguignon – DEST - UFRN

Ficha catalográfica elaborada com dados inseridos pelo(a) autor(a)
Biblioteca Carlos Benjamin de Lyra
Instituto de Matemática e Estatística
Universidade de São Paulo

Oliveira Ferreira de Sales, Lucas de
Modelo generalizado autorregressivo e de média móveis
Bernoulli-Geométrico: um novo modelo e seu gráfico de
controle Shewhart modificado / Lucas de Oliveira Ferreira de
Sales; orientadora, Airlane Pereira Alencar; coorientadora,
Linda Lee Ho. - São Paulo, 2023.

Tese (Doutorado) - Programa de Pós-Graduação em
Estatística / Instituto de Matemática e Estatística
/ Universidade de São Paulo.

Bibliografia
Versão simplificada

1. Gráficos de controle modificados. 2. modelo BerG-
GARMA. 3. séries temporais de contagem. 4. superdispersão. 5.
subdispersão. I. Pereira Alencar, Airlane. II. Título.

Bibliotecárias do Serviço de Informação e Biblioteca
Carlos Benjamin de Lyra do IME-USP, responsáveis pela
estrutura de catalogação da publicação de acordo com a AACR2:
Maria Lúcia Ribeiro CRB-8/2766; Stela do Nascimento Madruga CRB 8/7534.

À minha mãe, por todo seu esforço, dedicação e amor.

Agradecimentos

Sou grato ao meu bom Deus, por todas as dádivas em minha vida, na vida dos meus familiares e amigos.

À minha família, por seu incondicional apoio e amor. Em especial, à minha mãe Fransuilma, que demonstra a cada segundo de sua existência a pessoa espetacular que és. Ao meu padrasto Percy, por todo carinho ao longo dos anos. E ao meu pai Mozart, pelo suporte e confiança.

Agradeço as minhas orientadoras, Airlane Alencar (Lane) e Linda Ho, pela orientação acadêmica, profissional e de vida. Me sinto extremamente honrado por tê-las comigo neste percurso.

Aos meus amigos do coração, por sempre demonstrarem cuidado, atenção e apoio nos diversos momentos da vida. Dentre eles, cito: Brendda, Carla, Felipe, Giovana, Mattheus, Rayland e Rodrigo. Agradeço a Deus por me presentear com pessoas como vocês em minha vida.

Um agradecimento especial, àqueles que extrapolaram os vínculos de amizade: Erika, Félix e Heverton John. Por toda à fraternidade, momentos de descontração e companheirismo durante nossa jornada em uma nova cidade morando juntos.

Aos professores do Departamento de Estatística do IME-USP, por toda dedicação, esforço e novos ensinamentos dados. Especialmente à Airlane Alencar, Pedro Morettin, Gilberto de Paula e Silvia Ferrari.

Aos professores e amigos do Departamento de Estatística da UFRN, em especial aos professores André Pinho, Carla Vivacqua, Marcelo Bourguignon, Bruno Monte, Damião Nóbrega, Luz Milena, Dione, Pledson e Moisés Medeiros. Me orgulho em poder falar que fui aluno de vocês.

Aos funcionários e servidores do IME-USP, que com seus esforços e dedicação superaram o contexto da pandemia e auxiliaram do melhor modo possível os alunos.

À CAPES pelo apoio financeiro.

Resumo

Lucas de Oliveira Ferreira de Sales. **Modelo generalizado autorregressivo e de média móveis Bernoulli-Geométrico: um novo modelo e seu gráfico de controle Shewhart modificado.** Tese (Doutorado). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2023.

Ao longo dos anos, devido aos avanços tecnológicos, houve um crescimento substancial no armazenamento de diferentes tipos de dados. Em particular, os dados de contagem coletados ao longo de um determinado período de tempo ganharam cada vez mais importância e relevância em diversas áreas da ciência como economia, saúde, políticas públicas, etc. Deste modo, surge a necessidade de modelos estatísticos mais flexíveis (em termos de dispersão) para modelar diferentes tipos de dados de contagem. Neste contexto, este trabalho propõe um novo modelo generalizado autorregressivo e de média móvel (GARMA) que utiliza a distribuição Bernoulli-Geométrica (BerG) para modelar a média condicional de séries temporais de contagem. Adicionalmente, na segunda etapa do trabalho, construímos e apresentamos um gráfico de controle de Shewhart modificado para monitorar a média de dados que se ajustem ao modelo BerG proposto. As principais contribuições desse trabalho são: propor um modelo GARMA com a variável resposta seguindo uma distribuição BerG, a qual comporta inflação (ou deflação) de zeros. Além disso, o modelo proposto combina a flexibilidade de dispersão da distribuição BerG, com a inclusão de covariáveis e termos defasados para modelar a média condicional, induzindo assim uma estrutura de autocorrelação (relevante na análise de séries temporais). Apresentamos também as expressões para a estimação dos parâmetros do modelo, construção de teste de hipótese, uma forma simples de análise de diagnóstico e um procedimento para previsões. Com relação ao monitoramento estatístico, um gráfico de controle proposto foi construído utilizando método de Monte Carlo e um bootstrap paramétrico. Avaliamos essa ferramenta considerando o efeito da estimação dos parâmetros sob diferentes cenários e sua performance foi medida através do número de amostras até detectar um alarme (verdadeiro ou falso). Em ambos os trabalhos, proposição do modelo BerG-GARMA e monitoramento via gráficos de controle, ilustramos a aplicabilidade dos procedimentos utilizando dois conjuntos de dados reais, um apresentando superdispersão e outro com subdispersão.

Palavras-chave: Gráficos de controle modificados. modelo BerG-GARMA. séries temporais de contagem. superdispersão. subdispersão.

Abstract

Lucas de Oliveira Ferreira de Sales. **Generalized autoregressive and moving average Bernoulli-Geometric: a new model and its modified Shewhart control chart.**

Thesis (Doctorate). Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2023.

Over the years, due the technological advances, we have had substantial growth in the storage and collection of different types of data. More specifically, the use of count data over time has become more important in different fields of science such as economics, biological/ health, and public policy, to name a few. Thus, a necessity for new flexible statistical models (in terms of dispersion) to model different types of count data arises. Based on this, here we propose a new generalized autoregressive moving average (GARMA) model, based on the Bernoulli-Geometric (BerG) distribution, to model the conditional mean of counting time series. In addition, in the second part of this work, we built a modified Shewhart control chart to monitor the mean of data fitted by the proposed BerG model. The main contributions of this work are: to propose a GARMA model with the response variable following a BerG distribution, which includes zero inflation (or deflation). Furthermore, the proposed model combines the dispersion flexibility of the BerG distribution, with the inclusion of covariates and lagged terms to model the conditional mean, inducing an autocorrelation structure (relevant in time series analysis). We also present expressions for estimating model parameters, building hypothesis testing, a simple form of diagnostic analysis, and a procedure for predictions. Regarding the statistical monitoring, a proposed control chart was built using the Monte Carlo method and a parametric bootstrap. We evaluated the control chart considering the effect of parameter estimation in different scenarios and its performance was measured through percentiles and average of the in-control and out-of-control run lengths. In both parts of this work, we illustrate the applicability of the procedures using two real data sets, one showing overdispersion and the other one with underdispersion.

Keywords: BerG-GARMA. count time series. modified control chart. overdispersion. underdispersion.